

# Verteilte Anerkennung

## Wie künstliche Intelligenz die Theorie der Anerkennung verändert

---

Natalia Juchniewicz

**Abstract:** *The article argues that, based on contributions from social philosophy, the relationships between humans and AI and the effects of AI on social life can also be conceptualized with the necessary differentiation. This is done in continuation of discussion approaches that refer to A. Honneth's theory of recognition. It is shown how this theory makes it possible to include non-human persons as ›partial persons‹ and in emotionally important relationships; and why a (classical) argument from reciprocity and consciousness is not a necessary element of the recognition relationship itself. Recognition theory, for its part, must be expanded to include collective intelligence and distributed action. However, this also leads to a new challenge for ethics and social research: that of distributed responsibility.*

**Keywords:** *artificial intelligence (AI); recognition; partial persons; distributed responsibility; distributed recognition*

Bei den Blickwinkeln, unter denen soziale Beziehungen im 21. Jahrhundert analysiert werden, kommt der Digitalisierung der Technologien, die wir nutzen, und den entsprechenden sozialen Praktiken zweifellos eine zentrale Rolle zu. Der zunehmenden Bedeutung künstlicher Intelligenz (KI), die häufig Bedenken in normativ sensiblen Bereichen, im Zusammenhang mit Entscheidungsprozessen und vorgeschlagenen Lösungen hervorruft, ist dabei ein besonderes Gewicht zuzumessen. »Almost every day, the news media report on achievements of AI helping to overcome a great variety of real-world problems.« (Peeters et al. 2021: 217) Sosehr KI es beschleunigt, verschiedene Probleme zu diagnostizieren, und sosehr sie schnellere Entscheidungsfindung durch die Analyse von Big Data und *machine computing* ermöglicht, gilt es doch zugleich sich dessen bewusst zu sein, dass die Bestimmung dessen, was gerecht, akzeptabel oder förderlich für das Wohlergehen der Menschen

ist, nicht allein auf der Bewertung von Daten beruhen sollte. Dies darf schon allein deswegen nicht außer Acht kommen, weil zahlreiche Aspekte des menschlichen Daseins über die messbaren Indikatoren hinausgehen, die KI zur Verfügung stehen. Auch im Hinblick auf das soziale Leben kann KI, wenn sie ihrer Technik gemäß unsere Identitäten auf der Grundlage *vergänger* Aktivitäten bestimmt und definiert, uns nur auf Vorbestimmtes beschränken und ließe wenig Raum für menschliche Veränderungen oder die Begegnung mit Neuem aus dem Internet. Was KI nur vermag, ist, aufgrund früherer Nutzeraktivitäten etwas zu projizieren, und basiert dabei auf der Voraussetzung, dass die menschliche Identität fest (van Dijck 2013) und vorhersehbar ist.

»With their predictive capabilities and relentless nudging, ubiquitous but imperceptible, AI systems can shape our choices and actions easily and quietly. This is not necessarily detrimental. [...] It may foster social interaction and cooperation. [But it] may also exert its influencing power beyond our wishes or understanding, undermining our control on the environment, societies, and ultimately our choices, projects, identities, and lives.« (Taddeo/Floridi 2018: 752)

Mein Fokus in diesem Artikel liegt darauf, welche Antworten aus der Sozialphilosophie abzuleiten sind, um den Veränderungen und Herausforderungen Rechnung zu tragen, die in unseren intersubjektiven Beziehungen aufgrund der zunehmenden Präsenz künstlicher Intelligenz in unseren alltäglichen Aktivitäten sich ergeben. Im Letzten zielen die vorgestellten Überlegungen darauf ab, die umfassend gewordenen Auswirkungen des zunehmenden Einflusses und der Präsenz von KI in unserer gesellschaftlichen Welt zu erforschen. Dazu suche ich zu erläutern, warum das Konzept der *Anerkennung* eine entscheidende Rolle auch bei der Reflexion der Beziehung zwischen Mensch und KI spielen sollte und wie gegenwärtige Theorien der Anerkennung nicht-menschliche Entitäten wie ›KI‹ in ihren Rahmen einbeziehen könnten.

Unzweifelhaft scheint, dass die zunehmende Präsenz von KI verschiedenste Bedenken und Hoffnungen hervorruft, Verschiedenes, das mit unserer Privatsphäre, unseren Beschäftigungen, Entscheidungsprozessen, der Erkenntnis von Strukturen (›Muster‹) und der Vorausschau zukünftiger Ereignisse zusammenhängt. All dies beeinflusst in der Tat, wie wir uns als Individuen wahrnehmen, und es wirkt sich nicht weniger auf die Anerkennung aus, die wir in sozialen Beziehungen erfahren. Unsere alltäglichen Interaktionen mit künstlicher Intelligenz, die durch verschiedene Technologien wie Browser, soziale Medien und Chatbots an uns kommt, prägen denn viele normative Konsequenzen innerhalb unseres Lebens. Von einfachen Algorithmen bis hin zum *machine learning* und *deep learning*, diese Technologien formen zahlreiche Gegebenheiten, mit denen wir alltäglich zu tun haben. So sind wir, und das ist ein entscheidender Punkt, bei der Konstruktion unseres Selbstgefühls nun heute erheblich auf die Vermittlung durch Elemente von KI angewiesen. Dies führt

zu einem nicht zu unterschätzenden Aspekt des sozialen Problems der Anerkennung – dass unsere erfahrene Anerkennung nun wesentlich durch die Beziehungen zwischen Menschen und KI geprägt werden könnte. Diese Veränderungen auf jeden Fall mitreflektieren zu müssen, gilt umso mehr, als Individuen nicht unbedingt bemerken, dass KI keine menschliche Person ist, da sie menschenähnliches Verhalten zeigen kann (nicht nur den Turing-Test besteht, sondern auch in verschiedenen sozialen Situationen zufrieden stellende Interaktionen liefert; s. Borenstein/Pearson 2010).

Der nachfolgende Artikel ist in drei Abschnitte gegliedert. Im ersten werde ich zunächst allgemein die soziale Bedeutung von KI erklären und dabei den aktuellen Ansatz, wie die aus den Beziehungen zwischen Mensch und KI sich ergebenden sozialen Probleme bestimmt werden können, vorblickend unter Verwendung der Sprache der Anerkennungstheorie erläutern. Im zweiten Abschnitt werde ich die Theorie der Anerkennung einführen und mich insbesondere auf ihre ausgereifteste Form fokussieren, wie sie sich von Axel Honneth entwickelt findet. Näheres zum Potenzial dieses Ansatzes lässt sich sodann durch Analyse der Kommentare von Arto Laitinen und Heikki Ikäheimo gewinnen; beider Einsichten heben wichtige Probleme und Begriffe hervor, die in Bezug auf Erörterungen von KI und deren Rolle in den menschlichen Praktiken verwendet werden können. Laitinens Perspektive ermöglicht eine Diskussion über verschiedene Formen der Anerkennung, einschließlich eines weitreichenden Verständnisses, das einbefassend verschiedenartige Objekte veranschlagen kann. Ikäheimos Sichtweise hilft zu erkennen, dass auch schon generell in verschiedenen Dimensionen der Anerkennung eine Wechselseitigkeit der Anerkennung nicht immer als Voraussetzung steht. Davon ausgehend kann ein Ansatz dahin entwickelt werden, bei dem statt des einen menschlichen Parts in Anerkennungsrelationen nun eine Instanz von KI stünde. Konkret lässt sich so in einem nächsten Schritt darlegen, wie das auch in KI-Interaktionen zentral bestehende Problem, welches Bewusstsein die Subjekte der Anerkennung haben oder entwickeln, angegangen werden kann. Dazu werde ich dem Vorschlag von Nolen Gertz über ›den unerlaubten Schritt‹ (›illicit move‹) in der Hegelschen Tradition der Anerkennung folgen. Im dritten und letzten Abschnitt schließlich werde ich einen Ansatz vorschlagen, der auf Studien zu dem Neuen in dem Verhältnis von Mensch und KI angewandt werden kann. Dabei stütze ich mich das Konzept der »partial persons« (Hirvonen 2017), was auf KI Anwendung finden kann, das generative Modell der Anerkennung (Laitinen) – Anerkennung als performative Formung bzw. Veränderung von Normativem und Sozialem – und werde mich der Betonung der Notwendigkeit anschließen, einen neuen Ansatz zur Beschreibung von verteilter Agency und von Verantwortung in den Relationen zwischen Mensch und KI zu entwickeln (Peeters et al. 2021; Taddeo/Floridi 2018). Im Ganzen werde ich dafür argumentieren, dass die Entwicklung verantwortlicher Haltungen zu den Wirklichkeiten wie Möglichkeiten von KI und ihren sich entwickelnden Formen nicht ausschließlich auf KI selbst fo-

kussieren sollte – auf ihre Gestaltung, die Algorithmen oder die Nutzung –, sondern auch auf die Menschen. Sind doch die Menschen nicht nur Nutzer/Anwender der Technologie, sondern Individuen, die ihr Empfinden von Wert, sozialer Anerkennung und Respekt in ihren Beziehungen damit gestalten.

Wenn es um die Anerkennung zwischen Menschen und KI geht, gibt es zwei mögliche Perspektiven zu bedenken. Die erste konzentriert sich auf die Analyse von Anerkennungstheorien und Argumente, die es möglich machen, KI und die Beziehung dazu in ihren Rahmen einzubeziehen. Die zweite Perspektive beinhaltet die Analyse von KI selbst und ihrer Einwirkungen im Sozialen, um zu bestimmen, wann und warum ein Problem von Anerkennung entsteht (s. Jacobs 2024). Die erste Perspektive lässt sich theoretisch und die zweite empirisch nennen. Im vorliegenden Artikel wird mein Schwerpunkt hauptsächlich auf der theoretischen Perspektive liegen. Ich werde versuchen zu zeigen, dass innerhalb der bestehenden Anerkennungstheorien ein Raum vorhanden ist, der es uns ermöglicht, auch über KI zu diskutieren und neue Aspekte der Anerkennung zu beleuchten, die berücksichtigt werden müssen.

## 1. KI, Gesellschaft und Gesichtserkennungstechnologien

Im Jahr 2005 prognostizierte Ray Kurzweil, dass an einem bestimmten, nicht allzu fernen Zeitpunkt Maschinen leistungsfähiger sein würden als menschliche Intelligenz und an dem wir, um effektiver denken zu können, mit den Maschinen verschmelzen müssen. Bekannt wurde dies unter dem zum Schlagwort gewordenen »Singularität«, mit dem der Einschnitt in allem bisherigen menschheitlichen Geschehen, von da an das Eintreten in ein fundamental neues Seinsstadium gekennzeichnet sein sollte. (Kurzweil 2005) Dass diese Vision einer allgemeinen künstlichen Intelligenz verwirklicht ist, ist immer noch eher spekulativ als auf echten Technologien basierend. Doch ist offenkundig, dass die zeitgenössische KI bereits eine eminente Rolle spielt in verschiedenen Bereichen menschlicher Entscheidungspraxis, wie zum Beispiel in allem, was mit Logistik zu tun hat, bei Versicherungen, im Gesundheitswesen und auch in der Bildung (Coeckelbergh 2020: 3). Zusätzlich hat sich der Gedanke, künstliche Intelligenz in eine spezifische Technologie zu integrieren, wie etwa in eine Robotermaschine oder maschinelle Assistenz, um menschliche Tätigkeiten und Verhalten zu verbessern, zum Konzept des Internets der Dinge (*Internet of Things*) entwickelt. Dieser Wandel zielt auf die Schaffung eines Netzwerks von jeweiligen, miteinander verbundenen technologischen Umgebungsfeldern, die Verbindung über das Internet haben können. Diese Perspektive hat den Einsatz von KI erweitert, die nun in allen digitalen Technologien implementiert werden kann, genauer gesagt, in allen Geräten mit Internetzugang (und dies nimmt inzwischen schon lange den bei weitem höchsten Anteil des Datenauf-

kommens im Internet ein). So nimmt künstliche Intelligenz gegenwärtig zahlreiche Formen an, und unsere Interaktionen und Beziehungen zu ihr sind, auch wo im Einzelnen nicht erkannt, zu integralen Bestandteilen unseres täglichen Lebens geworden.

In der vorhandenen Literatur wurde der soziale Einfluss von KI hauptsächlich aus einer normativen Perspektive analysiert. Der Fokus liegt dabei auf den möglichen Folgen. Es wurden sowohl die verbundenen Hoffnungen als auch die im Allgemeinen erkennbaren Risiken dessen, dass KI in unseren Entscheidungsprozessen mit enthalten ist, herausgehoben, unter Letzterem die Beschränkung der persönlichen Autonomie, die Reproduktion von Ungleichheit, Diskriminierung, Rassismus, die Destabilisierung der Demokratie, die Förderung von Polarisierung und nicht zuletzt die ökologischen Umweltfolgen usw. (Avnoon et al. 2023: 2). In KI stecken Potenziale, soziale Ungerechtigkeiten gerade zu perpetuieren, indem sie Stereotypen und bestehende Diskriminierung verstärkt und gleichzeitig neue Formen epistemischer Ungerechtigkeit hervorruft. (Rafanelli 2022) Was dabei jedoch in der Forschung bislang weniger häufig untersucht wird, sind die gewissen zur Theorie der Anerkennung gehörenden Aspekte, mit denen diese Bewertungsdimensionen offenkundig in engem Zusammenhang stehen. Dazu zunächst ein erster Aufriss.

So enormen Einfluss künstliche Intelligenz auf die Gesellschaft hat, die Bewertung ihrer Wirkung ist ambivalent. KI könnte uns helfen zu verstehen, wer wir sind, aber auch menschliche Fähigkeiten abwerten und die autonome Selbstverwirklichung verringern; sie kann das Spektrum der Dinge, die wir tun können, erweitern und menschliches Handeln ausdehnen, aber auch dazu führen, dass menschliche Verantwortung dabei wegfällt; KI kann uns dabei helfen, mehr zu erreichen (individuell und kollektiv), aber auch die menschliche Kontrolle über verschiedene Prozesse reduzieren; und schließlich kann sie Einfluss nehmen auf die gesellschaftliche Zusammengehörigkeit, indem sie menschliche Interaktionen mit anderen Menschen und der Welt weiterentwickelt, aber sie kann auch die Selbstbestimmung des Menschen untergraben. (Vgl. Floridi et al. 2018) Für die Ziele meiner Erörterung ist es interessant, dass bei der Suche nach einer Sprache, um in der Konstatierung dieser ambivalenten Bewertungen unsere sozialen Beziehungen zu KI zu beschreiben, auf Lösungen aus der Ethik und Bioethik zurückgegriffen wird (vgl. ebd.) und höchstens in geringerem Maße auf etwas aus der Sozialphilosophie. Wie im Folgenden gezeigt, sollte jedoch die Theorie der Anerkennung eine bedeutende Rolle in unserer aktuellen Diskussion über KI spielen. Denn sie ist ein weitreichendes Konzept, das es ermöglicht, über Phänomene wie das Selbstverständnis in vermittelten Beziehungen (mit verschiedenen Menschen, aber auch Technologien) und Selbstachtung im Zeitalter der zunehmenden Rolle von KI im Rechtlichen und der Rechtsmacht von Verhältnissen und in ökonomischen Markt-Beziehungen zu sprechen. Integral einbefasst ist darin, dass die Präsenz von KI in unserem täglichen Leben sowohl bewusst sein kann, wie wenn wir Entwicklungen auf dem Markt be-

obachten, die von aufstrebenden Technologien beeinflusst werden, als auch unbewusst, wenn wir unsere Webbrowser nutzen und unbewusst unser Selbstwertgefühl durch KI-Algorithmen gestalten, d.h. gestalten lassen (zum Beispiel, wenn jemand unter Fettleibigkeit leidet und nach einem geeigneten Arzt online sucht, wird dieser Mensch dabei oder danach aller Wahrscheinlichkeit nach Benachrichtigungen über neue Diäten und gesunde Lebensstile erhalten und der private Nachrichten-Feed wird sich mit Bildern von trainierten Athleten füllen – was potenziell die Bemühungen zur Selbstfürsorge gerade untergraben wird).

Ein aktuelles Problemfeld mag die Tragweite, die die Perspektive auf Anerkennung haben könnte, illustrieren. Es bringt die konkrete Frage einer ›Politik der Anerkennung‹ ein. Rosalie A. Waelen schlägt vor, dass die sozialen und politischen Auswirkungen von KI im Licht der Anerkennungstheorie von Ch. Taylor verstanden werden können, da diese die universellen Prinzipien der Gleichheit und Würde für alle Bürger ansetzt und doch zugleich die Bedeutung dessen betont, deren Besonderheit in Unterschieden zwischen Individuen und Gruppen Rechnung zu tragen. (Waelen 2022: 217) Dies spielt eine entscheidende Rolle etwa auch bei der Bewertung der sozialen und normativen Konsequenzen von Technologien, die für Waelen das empirische Beispiel sind, das von ihr dazu untersucht wurde – die Technologien, die für die Gesichtserkennung entwickelt wurden. Da Anerkennung (qua Erkennen) in ihrem unmittelbarsten Sinne »Identifizieren« oder »Kategorisieren« bedeutet, vollziehen Technologien, die auf der Identifikation oder Kategorisierung menschlicher Gesichter basieren, buchstäblich das (An)Erkennen von Menschen. Selbst wenn KI lediglich die Gesichtsmerkmale einer Person mit bestimmten Mustern abgleicht, sind die sozialen und normativen Auswirkungen bedeutend.

Waelen unterscheidet dabei drei Formen der Fehleinschätzung (»misrecognition«), die auftreten können, wenn Gesichtserkennungstechnologien eingesetzt werden, um Menschen zu identifizieren. Die erste ist die umfassende Fehlidentifikation. Sie tritt auf, wenn eine Person durch KI aufgrund von ungenauen Daten mit etwas oder jemandem fälschlich identifiziert wird. Ein Beispiel dafür könnte eine falsche Identifizierung einer Person als Verdächtiger durch die Strafverfolgungsbehörden sein, was strukturell in besonderem Maße dunkelhäutige Menschen betrifft, da die KI dunkelhäutige Menschen nicht richtig erkennt. Dies ist nicht nur in Bezug auf eine falsche Darstellung von Menschen mit dunkler Haut beleidigend. Sondern es hat auch praktische Auswirkungen, da für sie das Risiko steigt, fälschlicherweise als jemand identifiziert zu werden, der ein Verbrechen begangen hat (Waelen 2022: 218). Die zweite Form ist die Fehlkategorisierung im Einzelnen. Sie tritt auf, wenn die KI verschiedene Aspekte menschlicher Gesichter als konkrete Verhaltensweisen oder Emotionen interpretiert. Eine solche (An)Erkennung impliziert, dass menschliche Gesichter Emotionen grundsätzlich in ähnlicher Weise zeigen, sozusagen anthropologisch universell, was es der KI ermöglicht, Individuen basierend auf ihren Gesichtsausdrücken als ›traurig‹, ›glücklich‹ oder ›zwinkernd‹ zu kategorisie-

ren. Das Letzte, das ›Zwinkern‹, kann als überzeugende Illustration dafür stehen, wie eine KI Asiaten möglicherweise fälschlich als ›zwinkernd‹ kategorisiert, weil ihre Gesichtsmerkmale von denen abweichen, die weiße westliche Personen gemeinhin haben. »Miscategorization implies that people have elements of their identity misunderstood or ignored, or worse, that they are treated wrongly because of certain elements of their identity.« (Waelen 2022: 219) Die dritte Form der Fehleinschätzung liegt in der Unfähigkeit, die subjektive Identität zu erkennen. Ihre Bedeutung ergibt sich aus der Tatsache, dass in KI-Gesichtserkennungstechnologien eine Person ausschließlich basierend auf ihrem Gesicht interpretiert wird, das in einer stark polaren Matrix beurteilt wird. Unsere Gesichter müssen vorgegebenen Kategorisierungen entsprechen. Das Problem entsteht besonders für Personen, deren äußeres Erscheinungsbild nicht-binär ist, für die, die sich als nicht-binär identifizieren, oder für solche mit gemischtethnischem Hintergrund (Waelen 2022: 220). In all solchen Fällen können potenzielle Fehlklassifizierungen von Individuen erhebliche schädliche Auswirkungen haben und ihr Selbstwertgefühl, das sie aus dem entwickeln, was die Gesellschaft ihnen entgegenbringt – in Gestalt ihrer eingesetzten KI –, negativ beeinflussen. Dies unterstreicht, welche Bedeutung eine nuanciertere Analyse von Anerkennungstheorien bei Untersuchung der Beziehungen zwischen Mensch und KI und ihrer sozialen Folgen haben wird. (S. auch Waelen/Wieczorek 2022)

Die Theorie der Anerkennung kann, wie ich zeigen werde, dazu beitragen, die Fragen zu verstehen, die sorgfältig berücksichtigt werden müssen, um eine verantwortungsbewusste Implementierung von KI in die Gesellschaft zu ermöglichen, da KI die Anerkennung und ebenso Fehlbeurteilung von Individuen beeinflusst. Dies beinhaltet ein ganzes Untersuchungsprogramm. Im Folgenden werde ich dazu zunächst die Theorie von Axel Honneth heranziehen und erläutern, wie verschiedene Aspekte dieser Theorie zur Erörterung der Beziehungen zwischen Mensch und KI beitragen können. Anschließend werde ich die Theorie von Laitinen vorstellen, die Konzepte liefert, die helfen können, die Anwendung der Anerkennungstheorie in diesem spezifischen Kontext zu strukturieren. In Hinsicht auf die beiden Schlüsselprobleme im Zusammenhang mit der Erweiterung der Anerkennungstheorie auf KI, nämlich Gegenseitigkeit und Bewusstsein, wird sich sodann eine Perspektive von Ikäheimo zur Theorie der Anerkennung als fruchtbar erweisen, die es in Anschlag zu bringen erlaubt, dass Gegenseitigkeit nicht immer in allen Formen der Anerkennung die Voraussetzung ist. Schließlich wird ein letzter Unterteil das Problem des Bewusstseins anhand von Gertz' Interpretation dieser Frage thematisieren.

## 2. Anerkennungstheorie

Anerkennung besteht in der Fähigkeit, sich selbst als Subjekt nur durch Interaktion mit einem anderen Subjekt zu verstehen, d.h. durch Intersubjektivität. (Haber-



mas 1968) In diesem Sinne liefert Anerkennung eine starke Grundlage für Identität und ist psychologisch entscheidend. (Taylor 1992; Honneth 1994) Der Prozess der Anerkennung basiert auf einer Identifizierung, worin bestimmte Merkmale durch eine\*n Anerkenner\*in qualifiziert werden, die das anerkannte Individuum zu einem *Subjekt normativer Überlegungen* machen. Sobald wir bestimmte Merkmale, die ein Wesen beispielsweise als bewusst definieren, identifizieren, ist es unvermeidlich, die Implikationen dieses Merkmals auch auf praktische Aspekte zu erstrecken, wie freien Willen und Autonomie. Daher ist Anerkennung auch aus normativer Sicht wichtig, was sich in Ethik, Sozial- und politischer Theorie widerspiegelt (Kloc-Konkołowicz 2015). Anerkennung kann auch durch ihre gegenteilige Position, die fehlende oder versagte Anerkennung, diagnostiziert werden. Diese tritt auf, wenn ein Individuum bedeutsame Merkmale besitzt, die es beispielsweise zu einem autonomen Wesen machen, diese Merkmale jedoch von anderen nicht anerkannt werden. Fehlende oder versagte Anerkennung führt oft zu einem Kampf um Anerkennung und kann Grundlage für soziale und politische Bewegungen sein (Taylor 1992; Butler 1987).

Ein wichtiger Aspekt der Anerkennung, für einige Forscher sogar der entscheidende (Brandom 2007), ist die Annahme der Gegenseitigkeit. Anerkannt zu werden bedeutet – dies auf den ersten Blick –, eine Verständigung zu gewinnen über beiderseitig verschiedene normative Wesentlichkeiten oder Bedürfnisse, dabei unter der Voraussetzung, dass wir diese auf Gegenseitigkeit gewährleisten können. Was darin jedoch nicht so klar ist, ist das, ob Gegenseitigkeit notwendiges Element jedweder Formen der Anerkennung ist, oder ob dies nur der Anerkennung in ihrem Ideal entspricht. Wie in den folgenden Teilen erläutert, gibt es im Konkreten menschlicher Sozialexistenz verschiedene Formen der Anerkennung, von denen durchaus nicht alle auf Gegenseitigkeit beruhen. Zum Beispiel Eltern, die den Bedürfnissen ihrer Kinder gerecht zu werden suchen, erwarten von diesen nicht unbedingt Gegenseitigkeit, jedenfalls nicht in der idealen Form elterlicher Liebe; sie anerkennen ihre Kinder allem voran als Wesen, die normativ gerechtfertigt eine Fürsorge erwarten (dürfen). Ein weiteres Problem, das mit dem Punkt der Gegenseitigkeit aufgeworfen ist, betrifft das, was auch in den Beziehungen zwischen Mensch und KI oft als entscheidend betrachtet wird, nämlich das Bewusstsein beim Anerkennenden wie Anerkannten. Kann jemand (oder allgemeiner: eine Subjekt-Verkörperung) Anerkanntsein haben, ohne sich dessen bewusst zu sein? Ist es dann eine echte Anerkennung? Ich werde argumentieren, dass Anerkennungstheorien in diesem Aspekt von Gegenseitigkeit und Bewusstsein nicht streng festgelegt sind – und dadurch die Möglichkeit eröffnen, den Bereich der anerkannten Objekte auch auf künstliche Intelligenz auszudehnen.



## 2.1 Die Theorie von Axel Honneth im Kontext von KI

Eine der einflussreichsten Interpretationen der Anerkennungstheorie wurde von Axel Honneth vorgelegt. Im Anschluss an Hegel und G. H. Mead kennzeichnet Honneth, dass es drei Formen der Anerkennung gibt: Liebe, Recht und Solidarität (verstanden als Sittlichkeit oder Wertschätzung; Honneth 1994). Liebe ist notwendig, um elementar Selbstvertrauen aufzubauen, und sollte in einer lebensübergreifenden Verbundenheit von ›Familie‹ vermittelt werden, in der die Bedürfnisse der Menschen, insbesondere der Kinder, Gegenstand der Fürsorge der Eltern sind. Durch Liebe kann eine Person das grundlegende Gefühl für sich selbst und ihre körperliche Integrität aufbauen. Recht ist ein Ausdruck des kognitiven Respekts der Menschen vor einem menschlichen Wesen, das in der Lage ist, sich moralisch verantwortungsbewusst zu verhalten. Zu wissen, dass jemand als eine Person nicht nur von der eigenen Familie anerkannt wird, sondern auch ein Subjekt des Rechts ist, stärkt unter allen Beteiligten das Selbstwertgefühl und das Gefühl der Autonomie, was Grundlage für die Rechte im Rechtssystem bildet. Solidarität schließlich wird mit anderen Menschen geteilt, wenn wir ihren Beitrag zu einer Gesellschaft, der auf ihren Werten, ihrem Beruf, ihrem Engagement usw. basiert, hochschätzen. Dies bedeutet, dass eine Person aufgrund ihrer Bedeutung für die Gesellschaft soziale Anerkennung erreichen kann. – Nach Honneth können all diese Formen der Anerkennung, insofern sie fundamental sind, auch ihre Gegenpositionen haben, wie Missbrauch oder Vergewaltigung, Verweigerung von Rechten, oder in der Gesellschaftsgemeinschaft Ausschluss und Herabwürdigung oder Beleidigung. Dieses Negative sollte ebenfalls berücksichtigt werden, wenn es die soziale Bedeutung der Anerkennung zu analysieren gilt.

Alle diese drei Formen der Anerkennung: Liebe, Recht und Solidarität, waren in Honneths Theorie auf empirische Forschung gegründet (hauptsächlich Psychologie/ Psychoanalyse und Sozialwissenschaften). Das hat die Diskussion über Anerkennung in der Bedeutung von Anerkennungsprozessen für Menschen gestärkt, aber auch eingeschränkt. Doch ist offenkundig, dass es, wie in der Forschung über verschiedene Formen menschlicher Interaktion und menschlichen Tuns sich erwiesen hat, unterschiedliche Objekte menschlicher Fürsorge, Achtung oder Wertschätzung gibt, die nicht alle zwangsläufig menschenförmig sein müssen. Das lässt sich unschwer illustrieren.

Die Erfahrung zu lieben, wohl eine der komplexesten Emotionen der menschlichen Natur, kann sich auf Gott, die Natur, Tiere oder sogar auf verschiedene Tätigkeiten richten (z.B. wenn wir Liebe zu unserer Arbeit ausdrücken). Wenn Liebesempfindungen etwas sind, das bei der Entstehung von Selbstvertrauen, größerer Selbstbestimmung oder einem Gefühl der Sicherheit aufkommt, können sie auch durch ein Technisches zustande kommen. – Pflegeroboter dienen als bemerkenswertes Beispiel für solches Technische. Entwickelt wurden sie, um

Einzelpersonen in Krankenhäusern oder Pflegeheimen zu unterstützen, Medikamente zu verabreichen, bei der Mobilität zu helfen oder Gespräche zu führen, wenn menschliche Pflegekräfte beschäftigt sind. Forschungsergebnisse zeigen, dass solche Beziehungen zu Robotern genauso befriedigend sein können wie diejenigen, die mit einem menschlichen Gegenüber entstehen, und starke emotionale Bindungen wachsen lassen. Dies kann gleichwohl nicht bedeuten, dass dies nicht noch ethisch überdacht werden müsste. (Vallor 2011) Gerade weil Menschen starke Beziehungen zu künstlichen Intelligenzen aufbauen können, die ihr eigenes Verhalten beeinflussen, ist es so wichtig, die verschiedenen Ebenen der Anerkennung und des Mangels an Anerkennung in solchen Beziehungen zu berücksichtigen (Cappuccio et al. 2019). Solchen emotionalen Aspekten bei künstlicher Intelligenz wird heute auch für die Gestaltung von sozialen Robotern stark Rechnung getragen. »This places strong demands on social robotics to build robots that understand others' actions, intentions, and emotions and show emotions themselves; that know when to listen to the human or act on its own preferences; that develop social competence, can keep up a normal conversation, form social relationships, learn from experience, and perhaps have a personality.« (Brinck/Balkenius 2020: 54; s. auch Breazeal 2002; Dautenhahn 2007) Diese Fähigkeit der künstlichen Intelligenz, menschliches Verhalten zu imitieren, wirft die Frage auf, ob die Nutzer von z.B. sozialen Robotern von den Entwicklern dieser Technologien subjektiv und mit Respekt vor ihrer Würde behandelt werden oder ob sie einer emotionalen Manipulation erliegen, die ethisch fragwürdig ist.

Ebenfalls in der Literatur diskutiert in Bezug auf Anerkennung wurde das Problem der Rechte, und es gibt viele Lösungsvorschläge, die auf KI angewendet werden können, um sie zu einem Rechtssubjekt zu machen. (Rodrigues 2020) Rechte auf KI auszuweiten, kann durch das, wie die Theorie der Anerkennung für Gruppen und Gruppenidentitäten angewandt wurde, gerechtfertigt werden. O. Hirvonen stellt hierzu fest: »In our everyday social practices we grant the status of personhood to any agent that is [capable of an addressive performance towards us]« (Hirvonen 2017: 147), was bedeutet, dass wir eine Zurechenbarkeit und Verantwortlichkeit zuschreiben und dies auf den kommunizierenden Agenten übertragen können. Personenhaftigkeit kann dabei nach Hirvonen in der Linie der Theorie der Anerkennung von Hegel und Honneth gefasst sein, und dementsprechend muss, um respektvoll behandelt zu werden, das gegeben sein, dass eine Entität ihrerseits in der Lage ist, abstraktes, moralisches Denken zu entwickeln, dass sie über sprachliche Fähigkeiten verfügt, dass sie selbstbewusst und dass sie rational ist. Dies könnte in gewisser Weise auf solche Formen von KI wie Chat-GPT oder fortgeschrittenere Roboter zutreffen, die normativen Erwartungen ähnlich folgen wie Menschen dem Gesetz. Hirvonen schlägt in seiner Argumentation für die Personenhaftigkeit von Gruppen vor, dass wir dort von »partial persons« sprechen können – Entitäten, die nicht alle

Kriterien erfüllen, die wir auf Menschen anwenden, aber dennoch in einigen Dimensionen der Anerkennung als Personen betrachtet werden können.

Das Beispiel von Gruppen, wie von Hirvonen dargestellt, verdeutlicht allerdings zugleich den Unterschied in den Dimensionen von Liebe und Recht in Bezug auf Prozesse von Anerkennung. Wir lieben keine Gruppen als abstrakte Entitäten, sondern lieben vielmehr bestimmte Individuen innerhalb dieser Gruppen (ansonsten die Gruppe vielleicht in einem übertragenen Sinne, wenn sie die Verkörperung einer von uns geliebten Idee ist). Dennoch sollten Gruppen, auch ohne Liebe, als Entitäten mit Rechten anerkannt werden, und dass man ebenso in Verhältnissen der Solidarität mit ihnen stehen kann. Ähnlich dann stellt sich das Problem aber auch bei KI (s. Gunkel 2014). Kann KI im Sinne des Vorschlags von Hirvonen »teilweise« die Kriterien der Personenhaftigkeit erfüllen? Die Antwort könnte bejahend ausfallen, denn verschiedene Formen von KI kommunizieren, lernen und können sich an Diskussionen über Werte beteiligen.

Solidarität schließlich erweist sich als einer der herausforderndsten Aspekte der Anerkennung in den Beziehungen zwischen Mensch und KI. Denn bei Solidarität spielt in einem gewissen Maß das mit, ein Verständnis für oder eine Identifikation mit dem zu haben, in welchen Kämpfen die unterschiedlichen Gruppen oder Individuen stehen. Gemäß Honneths Beschreibung ist Solidarität erreichbar, wenn Menschen ihre Verbundenheit erkennen und die vielfältigen Rollen und Fähigkeiten wertschätzen, die für die Gesellschaft wesentlich sind. Die Frage ist also: Kann man Solidarität gegenüber KI empfinden? Unter Bezugnahme auf die Forschung von Singer (Singer 2010) schlägt Nolen Gertz diesbezüglich vor, dass in bestimmten Situationen starke Impulse zur Projektion und zur Identifikation in der Tat zu einem Gefühl der Solidarität auch mit einer Technologie führen können: die Projektion menschlicher Funktionen, die für die Gesellschaft wichtig sind, und die persönliche Identifikation mit diesen Funktionen. Ein Beispiel für dieses Verhalten könne in der Art und Weise beobachtet werden, wie Soldaten Roboter auf dem Schlachtfeld behandeln, wo scheinbar irrationale Handlungen, wie das Risiko eigenen menschlichen Lebens, um einen Roboter zu retten, auftreten. Die Erklärung für dieses Verhalten läge genau im Gefühl der Solidarität – Solidarität in einer mit dem technischen Artefakt gemeinsamen Schicksalssituation. Die Soldaten nehmen den Roboter als »Kameraden« wahr und sehen darin ein Spiegelbild von sich selbst. Folglich reagieren sie in einer Weise, die mit dem Verhalten übereinstimmt, das sie ihrerseits von allen anderen Individuen (und den eingesetzten Maschinen, auf die sie sich verlassen) im Krieg erwarten. (Gertz 2018)

Die Beispiele zu Liebe, Recht und Solidarität, obwohl kurz dargestellt, liefern starke Gründe, um das Thema der Anerkennung auch in der zeitgenössischen Philosophie der Technologie, insbesondere im Zusammenhang mit KI anzugehen. Wir stehen vor dem Erfordernis, unser genaues Verständnis von Selbstvertrauen, Selbstachtung und Selbstwertgefühl in unserer gegenwärtigen modernen Gesell-

schaft zu erforschen, seit diese Vorstellungen nicht mehr nur von Interaktionen zwischen Menschen geprägt werden, sondern auch dem Einfluss von Interaktionen zwischen Mensch und KI unterstehen. Um dazu einen Ansatz vorzuschlagen, der eine Diskussion über KI im Kontext der Anerkennung ermöglicht, möchte ich die Theorie von Arto Laitinen einbringen.

## 2.2 Die Theorie von Laitinen und verschiedene Formen der Anerkennung

Laitinen diagnostiziert, dass es verschiedene Dimensionen des Konzepts der Anerkennung gibt, die berücksichtigt werden sollten. Das beginnt damit, dass Anerkennung eindimensional oder dagegen multidimensional verstanden werden kann. Eindimensionale Anerkennung ist zum Beispiel in der Care-Ethik gegeben, bei der in der Struktur ›A anerkennt B als Z‹ dieses ›Z‹ konstant ist (beispielsweise in der Beziehung zwischen Arzt und Patient), während sich das ›Z‹ in der multidimensionalen Anerkennung je nach Veränderung des Status der Person ändern kann (Laitinen 2002: 464f.). Zweitens kann Anerkennung eine praktische oder symbolische Bedeutung haben. Wenn jemand anerkannt wird, weil er jemand ist oder bestimmte Eigenschaften hat, bedeutet dies, dass er\* sie ›entsprechend behandelt werden sollte‹, was dabei im Konkreten von verschiedenen praktischen und symbolischen Kontexten abhängt. Drittens aber kann Anerkennung auch als ein entweder weites Konzept verstanden werden, bei dem verschiedene Situationen Verschiedenes, ihr Genüge zu tun, auslösen können, oder dagegen als strikte Auffassung, bei der die Regeln der Anerkennung in der Gesellschaft bereits existieren und wir die Bedeutung, was in einem Fall die Anerkennung ist, den verschiedenen eintretenden Situationalitäten als solchen zuordnen.

Diese verschiedenen Ebenen der Anerkennung führen nach Laitinen zu zwei allgemeinen Modellen der Anerkennung: einem generativen Modell und einem responsiven Modell. Im generativen Modell wird Anerkennung als *Handlung* betrachtet – eine Handlung, die normative Gründe etabliert und die soziale Realität effektiv verändert oder, mit anderen Worten, als *performativer Akt*, der Veränderungen in der Ontologie der Realität bewirkt. Das generative Modell sieht Anerkennung als einen Prozess, der, letztlich mit jeder Ereignung seines Akts, neue Attribute von anerkannten Objekten oder Subjekten enthüllt. Auf der anderen Seite das responsive Modell geht davon aus, dass innerhalb der Gesellschaft bereits normative Gründe für die Gewährung von Anerkennung an verschiedene Objekte oder Subjekte existieren. Es ist ein Modell der Zuteilung. Dieses Modell legt nahe, dass der Akt der Anerkennung eine Form des ›Benennens‹ oder ›Zuschreibens‹ ist, bei dem Rechte, Eigenschaften, Bedürfnisse usw. der rechtmäßig anerkannten Entität zugewiesen werden. Diese Prozesse verändern entsprechend die Realität nicht von sich aus, indem sie neue Bedeutungen von Anerkennung einführen würden, vielmehr werden sie allenfalls den Bereich der Objekte und Subjekte erweitern, die anerkannt werden sollten, also

was unter die gesellschaftlich gegebenen Bedeutungen fällt bzw. der betreffenden Zuschreibung zuteil werden kann.

Die Frage, die sich im Zusammenhang mit KI ergibt, lautet dann: Können *Objekte* ›anerkannt‹ werden, und kann ein Objekt eine Person in einem Sinne ›anerkennen‹, der über die (dem Objekt mitgegebene) bloße technische Funktionalität hinausgeht? – Laitinen weist darauf, dass es vier Arten von Objekten gibt, die einen Wert tragen und die es rechtfertigen, die Reaktionen, die sie basierend auf Gründen und Normen hervorrufen, jeweils zu analysieren. Diese Objekte sind 1. instrumentell wertvolle Objekte; 2. intrinsisch wertvolle Objekte; 3. lebende Wesen, die sich (in ihren Vitalprozessen) entfalten können, aber nicht unterscheiden können zwischen Verletzung und Beleidigung; 4. selbstbewusste Akteure, die in der Lage sind, Überzeugungen über die Überzeugungen anderer und über sich selbst zu bilden. (Laitinen 2002: 466) Wie zu sehen, erfüllt nur die letzte Art von Objekt die Kriterien, die normalerweise an Personen in einem vollen Sinne dieses Begriffs gestellt werden, und Laitinen betont, »[that] only persons can be recognizers« (Laitinen 2002: 465). Er macht jedoch keine solche Einschränkung für die oder das Anerkannte, die, wie aus der obigen Typologie ersichtlich, auch einen im weiteren Sinne objektmäßigen Charakter haben könnten. Allerdings gilt es zu bedenken, dass Technologien, wenn wir die oben erwähnte Analyse von Waelen berücksichtigen, nicht nur das Potenzial haben, Menschen (an)zuerkennen, sondern sie auch falsch anerkennen können, was zu normativen Konsequenzen führt. (S. auch Brinck/Balkenius 2020)

Man könnte jedoch anführen, dass eine einseitige Form der Anerkennung, die auf Objekte gerichtet ist oder die durch sie umgekehrt Menschen gezollt ist, nicht mit einem angemessenen Verständnis von Anerkennung übereinstimmt. Dafür gäbe es zwei Gründe. Zum einen, dass KI keine Intentionalität besitzt, um die Werte oder Identitäten, die von verschiedenen Individuen repräsentiert werden, zu schätzen – alles ist für sie gewissermaßen gleich, ein Faktum ihrer Umwelt. Und zum andern behandeln Menschen, außer in begrenzten Fällen, KI eben nicht gleichwertig mit anderen Menschen. – In den nächsten beiden Unterabschnitten werde ich argumentieren, dass die Annahme von Gegenseitigkeit der Anerkennung und Bewusstsein beim Anerkennenden/Anerkannten kein notwendiges Element der Theorie der Anerkennung ist. Was damit aber nicht berührt ist, ist, dass trotzdem die KI-Technologie, die auf begrenzter Anerkennung basiert, möglicherweise nicht unseren Erwartungen an ›wahre‹ Anerkennung entspricht. Deren normative Implikationen sind und bleiben weiterhin bedeutend.

### 2.3 Die Theorie von Ikäheimo und das Problem der Gegenseitigkeit der Anerkennung

H. Ikäheimo geht für die von ihm erörterten Fragen zurück auf die Hegelsche Tradition. Er hebt die Bedeutung von drei Dimensionen hervor, um gemäß dieser Tradi-

tion zu verstehen, was eine Person und ihre sozialen Prozesse ausmacht: Singularität (Einzigartigkeit), Autonomie (Universalität) und Partikularität. (Ikäheimo 2002) In Bezug auf die Singularität werden wir als Individuen mit einzigartigen Eigenschaften und unserer »suchness« anerkannt, wobei der Anerkennende sein Verhältnis zu uns an unserem Glück und Wohlbefinden ausrichtet. Als autonomen Wesen können uns Rechte gewährt werden, die dabei als Bedingung auf bestimmten universellen Werten basieren, die von allen Individuen geteilt werden. Als besondere Wesen schließlich wünschen wir uns, uns von anderen zu unterscheiden, und suchen Anerkennung für diese Partikularität. In diesem Sinne verkörpern wir, in unserem Uns-Unterscheiden, einen sozial geschätzten Wert. Ikäheimo argumentiert überzeugend, dass man diese drei Dimensionen bei der Anwendung auf Honneths Kategorisierung der Formen der Anerkennung, beinhaltend Liebe, Recht und Solidarität, berücksichtigt finden kann oder aber auch nicht.

Im Kontext der Liebe werden wir als einzigartige Wesen anerkannt, aber nicht ausschließlich als völlig autonom dabei (da wir immer noch auf andere angewiesen sind, um unsere Bedürfnisse zu erfüllen, was im Fall von Kindern deutlich wird) und auch nicht ausschließlich in unserer besonderen Partikularität (da von Eltern erwartet wird, dass sie ihre Kinder lieben, selbst wenn sie nicht einzigartig sind). Im Raum der Rechte werden alle Dimensionen unseres Menschseins berücksichtigt. Wir werden als Menschen mit unseren inhärenten Bedürfnissen anerkannt, ebenso als autonome Wesen, die zu verantwortungsbewusstem Verhalten fähig sind, und nicht zuletzt als einzigartige Individuen, insbesondere wenn wir eine Minderheit oder Ethnie repräsentieren, die um Anerkennung kämpft. Schließlich Solidarität, die wiederum einen speziellen Fokus hat, nicht alle Dimensionen gleichermaßen in sich einbefasst. Sie zielt in erster Linie auf unsere Besonderheit und betont unsere einzigartigen Fähigkeiten, die Werte, die wir verkörpern, und unseren Beitrag zur Gesellschaft.

In Ikäheimos Interpretation der Theorie der Anerkennung liegt der Schwerpunkt darauf, verschiedene Identitätsdimensionen auf verschiedene Momente der Anerkennung anzuwenden. Anstatt Liebe, Rechte und Solidarität als separate Analysebereiche zu betrachten, könne die Theorie der Anerkennung aus der dreifachen Perspektive von Einzigartigkeit, Autonomie und Besonderheit dargelegt werden. Wenn jedoch die Anerkennung durch diese Konzepte untersucht wird, ist es nicht mehr immer notwendig, dass klassischerweise eine *Gegenseitigkeit* der Anerkennung stattfindet.

Die Singularität, die beinhaltet, ein menschliches Wesen in seiner\* ihrer »suchness« wahrzunehmen, wird in den Bereichen Liebe und Rechte erfüllt. Diese Formen der Anerkennung betonen den Wert einer Person als Person, ohne dabei das, dass sie bestimmte spezifische Eigenschaften besitzen muss. Schon bei der Singularität wird deutlich, dass nicht immer eine gegenseitige Anerkennung erforderlich ist. Ich kann als Mensch, als ein menschliches Wesen, von meiner Familie oder durch das

Rechtssystem anerkannt werden, auch ohne dass ich mir dessen bewusst bin. Ich kann Menschenrechte haben, ohne vollständig zu verstehen, wie sie im internationalen Recht formuliert sind. Darüber hinaus sollte ich selbst dann, wenn ich gegen das gehandelt habe, was mir umgekehrt entgegengebracht wird, also wenn ich ein Verbrechen begangen habe und bestraft werde, immer noch als Mensch in meiner wesentlichen Natur behandelt werden. Dies verdeutlicht, dass die willentliche Anerkennung, andere »as such« anzuerkennen, nicht unbedingt eine Voraussetzung ist. Singularität der Person als Form der Anerkennung kann gegenseitig sein, kann aber auch ohne Gegenseitigkeit existieren.

Autonomie, wie sie auf dem universellen Verständnis basiert, dass jeder Mensch Vernunft und Würde besitzt, ist der umfassendste Aspekt und ist möglicherweise das einzige, in dem Gegenseitigkeit wirklich Berücksichtigung findet. Als autonomes Wesen kann ich Unabhängigkeit von anderen erreichen, weil sie es mir garantieren, und im Gegenzug behandle ich sie als ebenso autonom. Autonomie ist eng mit Verantwortung und Selbstachtung verbunden; verlässlich real verwirklicht kann sie nur durch das Rahmenwerk der Rechte werden.

Menschen in ihrer Partikularität zu verstehen, die sich auf die Einzigartigkeit von Fähigkeiten, Verhalten, dabei Individuen und Gruppen bezieht, manifestiert sich durch Schätzung, die sowohl in Rechten als auch in Solidarität zu finden ist. Diese Perspektive entspricht den theoretischen Grundlagen der ›Politik der Differenz‹ (Taylor 1992) und dient als Grundlage für die Wertschätzung von Individuen aufgrund ihres Beitrags zur Gesellschaft. Auch hier gilt es zu beachten, dass bei dem, jemanden aufgrund bestimmter Fähigkeiten zu schätzen, Gegenseitigkeit nicht unbedingt erforderlich ist. Entsprechend unrealistisch ist, zu erwarten, dass es eine gegenseitige Solidarität mit jedem Individuum, jedem Beruf und jeder Facette des Lebens der Menschen gäbe.

Anerkennung also ist nicht immer gegenseitig. Damit stellt sich die Folgefrage: Ist es hingegen notwendig anzunehmen, dass jedes Subjekt oder Objekt der Anerkennung dabei *bewusst* ist? – Um diese Frage zu beantworten, legt es sich nahe, sich noch einmal dem Urtext, Hegels *Phänomenologie des Geistes* zuwenden. Hegels Grundmodell beschreibt die gegenseitige Anerkennung von zwei Selbstbewusstseinen – Selbstbewusstwerdung, die von einem anderen Selbstbewusstsein anerkannt wird – und impliziert dabei, dass Gegenseitigkeit diesem Prozess konstitutiv innewohnt (das also, was wie gesehen nicht immer wesentlich erforderlich ist). Aber, geht Hegel davon aus, dass jedes solche Selbst tatsächlich Bewusstsein haben muss? Und haben wir irgendwelche Evidenzen, dies anzunehmen? N. Gertz hat dafür einen tiefgehenden Ansatz vorgeschlagen.



## 2.4 Das Problem des Bewusstseins: ein Ansatz von Gertz

Gertz schlägt vor, das Problem der Anerkennung so zu untersuchen, dass man die Möglichkeit in Betracht zieht, dass auch Technologie in die Position des Anderen in der Anerkennungsbeziehung gestellt wird. (Gertz 2018) Diese alternative Perspektive auf Anerkennung führt zu neuen Implikationen für unser Verständnis dieses Konzepts. »The actual moments of the encounter as described by Hegel [in the *Phenomenology of Spirit* – NJ]«, so Gertz, »leave open the possibility that the other need to be a consciousness, but only appear to be a consciousness« (Gertz 2018: 142). Wie Gertz argumentiert, klassifiziert das Selbstbewusstsein, im Moment der Begegnung mit »dem Anderen«, diesen als »Selbstbewusstsein«, durch *Projektion* – die Projektion seiner eigenen Wünsche und Bedürfnisse (Gertz 2018: 143); und begrenzt damit eigentlich die Bedeutung der Anerkennung auf die Bedingungen seiner eigenen Selbstkenntnis. Die »Lösung« für diese Beschränkung, Selbstbewusstsein als die einseitig ausgehende Projektion eines Bewusstseins zu behandeln, besteht nichtsdestoweniger im bekannten »Kampf auf Leben und Tod«, der den Unterschied in den Wünschen und Bedürfnissen enthüllt.

Es ist jedoch ein attraktiver Gedanke, im Einklang mit dem, was Gertz den »illicit move« jener Projektion nennt (Gertz 2018: 142), gegenläufigerweise zu erwägen, dass bei Übertragung der Anerkennungsbeziehung vom Bewusstsein auf Technologie der anfängliche Moment der Begegnung zweier Selbst im »Kampf um Anerkennung« eben kein Wissen über den Anderen erfordert, sondern nur über uns selbst. Dies lässt die Möglichkeit zu, einem Objekt Bewusstsein *zuzuschreiben*, das sich so verhält, als ob es bewusst wäre, auch wenn es dies nicht ist. Zwar könnte man argumentieren, dass ohne wahres Bewusstsein beim Anderen die »echte« Anerkennung in dieser Beziehung fehle. Aber ist dies immer der Fall? Wenn man zum Beispiel mit einem Chatbot in sozialen Medien spricht, ohne zu wissen, dass es keine echte Person ist, »gibt« es dann keine Anerkennung? Ebenso, wenn man einem Instagram-Profil einer KI-generierten Person folgt, ohne den Unterschied zu echten Individuen zu erkennen, wirkt sich dies etwa auf das Konzept der Anerkennung aus?

Wie darauf zu antworten wäre, ließe sich im Anschluss an Waelen lernen. Waelen folgt Laitinen und führt einen Ansatz der angemessenen Berücksichtigung (»adequate regard«) ein, der sich ausschließlich auf die Haltung von A gegenüber B stützt, nicht jedoch das Umgekehrte. (Waelen 2022; s. Laitinen 2010) Wenn ich etwas als jemanden anerkenne und es versäume, zwischen einer Technologie und einem lebendigen Subjekt zu unterscheiden, ist dies zweifellos ein Fehler, indem ich der in der Technologie wirkenden KI menschliche Qualitäten zuschreibe. Doch kann dasselbe über die praktischen Konsequenzen dieser Anerkennung gesagt werden? »If we follow the adequate regard account [...], it does not matter if the person considers the technology to be capable of recognizing them. All the matters under this understanding of recognition is the effect the system has on the person's

self-development.« (Waelen 2022: 221) KI in falscher Weise als Person anzuerkennen, wirkt sich zwar immer noch darauf aus, wie der Anerkennende sich selbst in Bezug auf die KI wahrnimmt und sie effektiv anstelle von Beziehungen zu echten Personen nimmt. Ansonsten aber gilt: »Mutual [sic!] recognition has been achieved when one individual shows that its behavior can be influenced by the other.« (Brinck/Balkenius 2020: 65) Wie deutlich, basiert das Argument hier auf dem Einfluss auf das Verhalten einer (menschlichen) Person, was ausreichend ist, um von Anerkennung zu sprechen, und was eben auch durch solche Technologien wie KI ermöglicht werden könnte. – Das hätte nun weitreichende theoriekonzeptionelle Folgen. Nach meiner Überzeugung ist es notwendig, dafür nicht nur bestehende Theorien der Anerkennung zu reformulieren, um dieses in der Sozialphilosophie so fruchtbare Konzept angemessen auch im Hinblick auf die Realität der intersubjektiven Beziehungen, die Menschen mit KI aufbauen, zu fassen. Sondern es muss auch konzeptionell mit Erkenntnissen vonseiten der Philosophie der Technik erweitert werden. Dazu ein letzter Abschnitt.

### **3. Neue Herausforderungen für eine KI mit einbeziehende Sozialphilosophie**

Im gegenwärtigen Diskurs der den Entwicklungen und Veränderungen durch KI gewidmeten Philosophie hat sich tendenziell die Neigung zu zwei gegensätzlichen Lagern gebildet: entweder eine technologiezentrierte Perspektive zu betonen oder eine menschenzentrierte. (Peeters et al. 2021) Erstere legt den Schwerpunkt auf die fortschreitenden Fähigkeiten von KI – Fähigkeiten, die nicht nur das Potenzial haben, Menschen in rechnerischen Aufgaben zu ersetzen, sondern auch als eine Alternative in Situationen dienen könnten, in denen Skepsis bezüglich der menschlichen Entscheidungsfindung besteht, wie beispielsweise in Fragen der Gerechtigkeit (Peeters et al. 2021: 220). Die letztere Perspektive lenkt demgegenüber die Aufmerksamkeit auf die Probleme, die sich aus der KI selbst ergeben, wie etwa Bedenken, wie es Menschen dabei ergeht, oder die sich abzeichnende Zukunft in ethischer Hinsicht zu beurteilen. Dies sind Probleme, die KI aus sich heraus, KI allein nicht lösen kann. Sie erfordern vielmehr, dass es verantwortungsbewusste menschliche Entscheidungsfindung auf verschiedenen Ebenen moralischer, rechtlicher und sozialer Komplexität gibt.

Ansichts der fundamentalen Divergenz zwischen den Perspektiven führen Peeters et al. einen dritten Ansatz ein, den sie als den der kollektiven Intelligenz umreißen (»Collective Intelligence«). Das zentrale Argument dieses Ansatzes besteht darin, dem Rechnung zu tragen, dass »Intelligenz«, ob menschlich oder künstlich, nicht isoliert existiert. Um Prozesse und Leistungen von Intelligenz im Einklang mit dem, was »Intelligenz« wesensmäßig ausmacht, tatsächlich umfassend zu ana-

lysieren, ist es notwendig, sie auf Ebene von Gruppierungen und Vernetzungen zu untersuchen – und sich dann heute auch auf das Zusammenwirken zwischen Menschen und Maschinen zu konzentrieren (Peeters et al. 2021: 222). Nach meinem Dafürhalten ist es dabei in Folge auch entscheidend, nicht nur die kognitive Dimension des Menschen im Kontext dieser Kollektivität zu betrachten, sondern ebenso die sozialen Aspekte. Allerdings ergibt sich bei solcher Betrachtung eminent die Frage nach der Verantwortung.

Taddeo und Floridi argumentieren, dass Entscheidungen, die auf KI basieren, das Ergebnis von verteiltem Handeln sind: Interaktionen zwischen verschiedenen menschlichen und nicht-menschlichen Akteuren wie Designern, Entwicklern, aber auch Software und Hardware verschiedener Technologien. Dies würde zu *verteilter Verantwortung* (»distributed responsibility«) führen. (Taddeo/Floridi 2018) Das wäre für die Ethik schon grundlegend ein neues Phänomen; es gäbe nicht mehr das eine Subjekt einer betreffenden Verantwortung. Ethik war traditionell auf individuelle Handlungen und Absichten ausgerichtet, jedoch nicht auf kollektive und aufgeteilte. Diese neue Herausforderung für die Ethik führt zu einer Idee der Verantwortung, die diese nicht an die Intentionen der Menschen koppelt – ob jemand etwas so »gewollt« hat bei seinem\*ihrem Tun oder Unterlassen –, sondern vielmehr berücksichtigt, was etwas als die Ergebnisse (und dabei nicht immer beabsichtigte) eines betreffenden Konglomerats von Menschlichem und Technologischem ist. Und seit es dazu gekommen ist, dass KI eine »lernende« Technologie ist, können nicht alle Ergebnisse dieses Prozesses überhaupt vorhergesagt werden. Nichtsdestoweniger aber erfordert es Nachänderungen, wenn die KI auf mit einem Bias behafteten Daten trainiert wird oder wurde. (S. Burrell 2016; Jacobsen 2023) – Dies beschreibt eigentlich nur, was ohnehin alltäglich geschieht, nur die Theorie hat dem bisher noch mit einseitigen Modellen nicht entsprochen. Fast jeden Tag entscheiden wir alle, welche Technologie wir haben möchten, welche Praktiken, die durch Technologie angeregt werden, wir akzeptieren und welche davon geändert werden müssen, sind wir doch Nutzer verschiedener Technologien, und ein Teil dieser verteilten Verantwortung liegt allenthalben auch in unseren Händen.

Das Konzept der verteilten Ethik hat Auswirkungen auf das Verständnis von Anerkennung, das, sosehr in ihm Anerkennung als ein Prozess und als sich im Laufe der Zeit entwickelnd gefasst, traditionell die Prozesse beschrieben hat, die mit der Entwicklung von Selbstwissen, persönlicher und gruppenbezogener Identität, dem normativen Rahmen von sozialen Konflikten usw. verbunden sind. Die Anerkennungstheorie hat konkrete Konzepte bereitgestellt, um die verschiedenen Momente innerhalb dieser Prozesse aufzuspüren und zu verstehen. Die zwei wesentlichsten Punkte, die heute an die bisherigen Horizonte rühren, sind zum einen das, dass positive Dimension der Anerkennung und negative Dimension der Missachtung nicht mehr so klar in ihrer Bedeutung sind, wenn, wie zunehmend eingetreten, der Prozess der Anerkennung *verteilt* ist und dies über das (Zwischen-)Menschliche

hinausgeht. Und zum anderen eben das, dass der Anerkennende wie der Anerkannte nicht unbedingt menschlich und darin ›volles‹ Personensubjekt sein müssen, sondern Hirvonen gemäß eine »partial person« sein können. Daher stellt sich nicht nur die Frage, ob KI als (teilweise) Person anerkannt werden kann, KI als »der Andere« im Kampf um Anerkennung, und wie sich dies tatsächlich auf Liebe, Recht und Solidarität auswirken würde, sondern auch, was es für ein menschliches Wesen bedeutet, von KI ›anerkannt‹ zu werden.

Die Fragen reichen bis in die Struktur des Konzepts der Anerkennung. Auf KI als solche wird in der Regel wohl das responsive Modell der Anerkennung (s.o.) angewendet, das auf der Zuweisung von anerkannten Merkmalen beruht. Die KI muss dabei dann in einer Weise gestaltet sein, dass Benutzer\*innen (Menschen) sie als Technologie (an)erkennen können, die konkrete Bedürfnisse erfüllt, oder als fähig dazu, Muster im menschlichen Verhalten zu erkennen. Im Falle von Menschen jedoch, insbesondere in ihren Beziehungen zu KI, könnte das generative Modell der Anerkennung (s.o.) eindeutig sachhaltiger sein, auch wenn noch nicht ganz klar ist, wie sich soziale Beziehungen, menschliches Verhalten und die emotionalen und intimen Dimensionen des menschlichen Lebens unter dem Einfluss von KI verändern können. Die in diesem Artikel verwendeten Beispiele, um den Einfluss von KI auf unsere täglichen Interaktionen zu veranschaulichen, enthüllen auch die Potenziale, wie KI uns Menschen als Träger der Anerkennung wahrscheinlich verändern könnte. KI wird schon jetzt zunehmend nicht mehr nur als Werkzeug betrachtet, das menschlichen Zielen dient, sondern auch als Partner oder Freund (s. Brinck/Balkenius 2020: 54), der die Art und Weise beeinflusst, wie Individuen *sich* im Akt der Anerkennung wahrnehmen. Dies ließe sich als ein Bereich aufkommender neuer Formen der Anerkennung verstehen, die der empirischen Forschung bedürfen.

## Was sich festhalten lässt

Die große und einflussreiche Leistung der Theorie der Anerkennung war, verschiedene Formen des Aufbaus menschlicher Identität, des Selbstbewusstseins, der Gegenseitigkeit und der Intersubjektivität darzulegen. Solange sie traditionell verschiedene Phänomene im sozialen Leben explizierte, wie den Kampf um Anerkennung, verschiedene Formen sozialer Fehlzureisungen und die Notwendigkeit, dass Menschen sich emotional, in ihren Rechten und im ökonomischen Wohlergehen entwickeln (können), wurde sie hauptsächlich auf Beziehungen unter Menschen in Ansatz gebracht. Mit den sich ändernden technologischen Bedingungen des sozialen Lebens und der zunehmenden Rolle nicht-menschlicher Akteure (Latour 2005) und nicht-menschlicher Handlungen (Bowden 2015) fordert das Feld der Beziehungen von Menschen und KI dabei eine weitere Forschung unter dem Gesichtspunkt der Theorie.

Ich habe dafür argumentiert, dass Anerkennung dabei nicht allein durch die instrumentelle Fähigkeit der KI verstanden werden kann, verschiedene menschliche Merkmale anzuerkennen bzw. fehlzuinterpretieren, und auch nicht schon dann, wenn sie soziale Hintergründe erkennen kann. Die Beziehungen von Menschen und KI sollten vielmehr konsequent im Kontext der Sozialphilosophie analysiert werden (s. Jacobs 2024). Das war die Motivation dabei, Honneths Theorie der Anerkennung einzuführen und sie in der Folge durch Interpretationen wie von Laitinen und Ikäheimo zu erweitern. Auf dieser Basis lassen sich gezeigterweise verschiedene Formen der Anerkennung (Liebe, Recht und Solidarität) und mögliche Fehleinschätzungen, verschiedene Analyseebenen (Singularität, Autonomie und Partikularität) und Modelle (ein responsives Modell und ein generatives Modell) identifizieren. Die dabei zu gewinnenden Differenzierungen erhöhen stark das Potenzial der Theorie. Die Untersuchung all dieser Aspekte könnte in diesem Sinne dazu beitragen, einen neuen Ansatz zur Beschreibung der Beziehungen von Menschen und KI zu entwickeln.

Neben der theoretischen Analyse auch empirische Beispiele für die Argumente einzubringen, stärkt dies noch weiter. Das zeigte sich hier an den eingeführten Konzepten der »partial persons«, der verteilten Verantwortung und der kollektiven Intelligenz. Mit ihnen ist man nicht mehr an Denkschemata gefesselt, die eindeutige – auch qualitativ eindeutige – Grenzen zwischen rein menschlichen und rein künstlichen Aktivitäten ansetzen. Von dem aus spricht vieles heute dafür, dass die Theorie der Anerkennung auch einer Neudeutung bedarf und zu einer *verteilten Anerkennung* werden muss. Erst so wird sie den sich ändernden Bedingungen des sozialen Lebens gerecht werden. Statt des responsiven Modells könnte in der Gesellschaft im Zeitalter der künstlichen Intelligenz das generative Modell der Anerkennung die größere Erkenntnis bieten: Anerkennung als performativer Akt, als eine Handlung, die normative Gründe etabliert und die soziale Realität effektiv verändert.

## Literatur

- Avnoon, N.; Kotliar, D.M.; Rivnai-Bahir, S. (2023): Contextualizing the ethics of algorithms. A socio-professional approach, in: *New Media & Society* [<https://doi.org/10.1177/14614448221145728>].
- Bowden, S. (2015): Human and Nonhuman Agency in Deleuze, in: Roffe, J.; Stark, H. (Hg.), *Deleuze and the Non/Human*, London: Palgrave Macmillan, 60–80.
- Brandom, R.B. (2007): The structure of desire and recognition. Self-consciousness and self-constitution, in: *Philosophy & Social Criticism*, 33(1), 125–148.
- Breazeal, C. (2002): *Designing sociable robots*, Cambridge (MA): The MIT Press.
- Brinck, I.; Balkenius, C. (2020): Mutual Recognition in Human-Robot Interaction. A Deflationary Account, in: *Philosophy & Technology*, 33, 53–70.

- Butler, J. (1987): *Subjects of Desire. Hegelian Reflections in Twentieth-century France*, New York: Columbia University Press.
- Cappuccio, M.; Peeters, A.; McDonald, W. (2020): Sympathy for Dolores. Moral Consideration for Robots Based on Virtue and Recognition, in: *Philosophy & Technology*, 33, 9–31.
- Coeckelbergh, M. (2020): *AI Ethics*, Cambridge (MA): The MIT Press.
- Dautenhahn, K. (2007): Socially intelligent robots. Dimensions of human–robot interaction, in: *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 679–704.
- Floridi, L.; Cowl, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; Vayena, E. (2018): AI4People – An Ethical Framework for a Good AI Society. Opportunities, Risks, Principles, and Recommendations, in: *Minds and Machines*, 28(4), 689–707.
- Gertz, N. (2018): Hegel, the Struggle for Recognition, and Robots, in: *Techné. Research in Philosophy and Technology*, 22(2), 138–157.
- Gunkel, D.J. (2014): A Vindication of the Rights of Machines, in: *Philosophy and Technology*, 27(1), 113–132.
- Habermas, J. (1968): Arbeit und Interaktion. Bemerkungen zu Hegels Jenenser ›Philosophie des Geistes‹, in: Ders., *Technik und Wissenschaft als ›Ideologie‹*, Frankfurt a.M.: Suhrkamp, 9–47.
- Hirvonen, O. (2017): Groups as Persons? A Suggestion for a Hegelian Turn, in: *Journal of Social Ontology*, 3(2), 143–165.
- Honneth, A. (1994): *Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte*, Frankfurt a.M.: Suhrkamp.
- Ikäheimo, H. (2002): On the Genus and Species of Recognition, in: *Inquiry*, 45(4), 447–462.
- Jacobs, K. (2024): Roboter gegen Einsamkeit? Zur Reproduktionsdynamik falscher und mangelnder Anerkennung durch »soziale« KI, in: Adophi, R.; Alpsancar, S.; Hahn, S.; Kettner, M. (Hg.), *Philosophische Digitalisierungsforschung. Verantwortung, Verständigung, Vernunft, Macht*, Bielefeld: transcript, 219–256.
- Kloc-Konkolowicz, J. (2015): Anerkennung als Verpflichtung. Klassische Konzepte der Anerkennung und ihre Bedeutung für die aktuelle Debatte, Würzburg: Königshausen & Neumann.
- Kurzweil, R. (2005): *The Singularity is Near. When Humans Transcend Biology*, New York: Viking Press.
- Laitinen, A. (2002): Interpersonal Recognition. A Response to Value or a Precondition of Personhood?, in: *Inquiry*, 45(4), 463–478.
- Laitinen, A. (2010): On the scope of ›recognition‹. The role of adequate regard and mutuality, in: Schmidt am Busch, H.-C.; Zurn, C.F. (Hg.), *The Philosophy of Recognition. Historical and Contemporary Perspectives*, Lanham: Rowman & Littlefield, 319–342.

- Latour, B. (2005): *Reassembling the social. An introduction to actor network theory*, Oxford: Oxford University Press.
- Peeters, M.M.M.; van Diggelen, J.; van den Bosch, K.; Bronkhorst, A.; Neerincx, M.A.; Schraagen, J.M.; Raaijmakers, S. (2021): Hybrid collective intelligence in a human–AI society, in: *AI & Society*, 36(1), 217–238.
- Rafanelli, L.M. (2022): Justice, injustice, and artificial intelligence. Lessons from political theory and philosophy, in: *Big Data & Society*, 9(1) [<https://doi.org/10.1177/20539517221080676>].
- Rodrigues, R. (2020): Legal and human rights issues of AI. Gaps, challenges and vulnerabilities, in: *Journal of Responsible Technology*, 4(3), 100005.
- Singer, P.W. (2010): *Wired for War*, New York: The Penguin Press.
- Taddeo, M.; Floridi, L. (2018): How AI can be a force for good. An ethical framework will help to harness the potential of AI while keeping humans in control, in: *Science*, 361(6404), 751–752.
- Taylor, C. (1992): The politics of recognition, in: Gutmann, A. (Hg.), *Multiculturalism. Examining the Politics of Recognition*, New Jersey: Princeton University Press, 25–73.
- Vallor, S. (2011): Carebots and Caregivers. Sustaining the Ethical Ideal of Care in the Twenty-First Century, in: *Philosophy & Technology*, 24, 251–268.
- Van Dijck, J. (2013): ›You have one identity‹. Performing the self on Facebook and LinkedIn, in: *Media, Culture & Society*, 35(2), 199–215.
- Waelen, R.A. (2022): The struggle for recognition in the age of facial recognition technology, in: *AI and Ethics*, 3(1), 215–222.
- Waelen, R.; Wieczorek, M. (2022): The Struggle for AI's Recognition. Understanding the Normative Implications of Gender Bias in AI with Honneth's Theory of Recognition, in: *Philosophy & Technology*, 35(2), article 53.