

# Aktuelle Themen

Wintersemester 2018/19  
Herr Professor Dr. Koch

Vorgelegt von  
Marlis Fleischmann  
Mtr-Nr.: 34917

## Artificial Empathy - künstliche Empathie <sup>1</sup>



### Empathiebegriff

Bevor geklärt werden kann, wovon die Thematik 'Künstliche Empathie' (im Folgenden KE) handelt, soll auf den eigentlichen Empathiebegriff eingegangen werden. Das griechische Wort 'empátheia' hatte noch wenig mit dem heutigen Begriff zu tun: Es stand für Leidenschaft oder Boshaftigkeit. Sigmund Freud verwendete den Begriff Einfühlung, der ins Englische als Empathy übersetzt und als Empathie zurück übersetzt wurde.[1] Da Empathie ein vielschichtiger Begriff mit unterschiedlichen Teilaspekten ist gibt es keine klare Definition. Durch die sehr verschiedenen Bereichen, in denen Empathie Relevanz hat gibt es verschiedene Auslegungen der Bedeutung. [2]

Allgemein lässt sich Empathie definieren als die "Fähigkeit und Bereitschaft, Empfindungen, Gedanken und Emotionen, Motive und Persönlichkeitsmerkmale einer anderen Person zu erkennen und zu verstehen"[3] (sprich: Einfühlungsvermögen). In der Literatur werden oft zwei Formen von Empathie unterschieden: kognitive und emotionale Empathie. *Kognitive Empathie* beschreibt die Fähigkeit, Gefühle und andere mentale Vorgänge eines Anderen zu verstehen und daraus korrekte Schlussfolgerungen zum Verhalten abzuleiten. Dafür ist die Fähigkeit andere Perspektiven einzunehmen notwendig. [4] [3]

*Emotionale Empathie* ist hingegen die Fähigkeit komplexe soziale Situationen zu verstehen, um mit ihnen in diesen kommunizieren zu können. Nach Paul Ekman lässt uns Emotionale Empathie *fühlen*, was ein Anderer fühlt. [3]

Zum eigentlichen Empathiebegriff wird oft auch die Fähigkeit zu angemessenen Reaktionen auf Gefühle Anderer dazugezählt, wie beispielsweise Mitleid, Trauer, Schmerz und Hilfsbereitschaft aus Mitgefühl. Diese empathischen Handlungen gehören allerdings nicht zu Empathie an sich, sondern werden als prosoziales Verhalten kategorisiert. Dennoch werden diese Handlungen im Rahmen dieser Arbeit ebenfalls betrachtet.

Artificial Empathy (=KE) ist die Entwicklung von KI wie beispielsweise Begleiter-Robotern, die menschliche Emotionen erkennen und darauf reagieren können. Die Thematik verbindet

---

<sup>1</sup> Bildquelle Abb. 1:

[https://cdn1.iconfinder.com/data/icons/the-brain-collection-glyphs/99/right\\_brain-512.png](https://cdn1.iconfinder.com/data/icons/the-brain-collection-glyphs/99/right_brain-512.png)

zahlreiche Disziplinen miteinander: Informatik, Neurobiologie, Neurologie, Psychologie, Robotics, Anthropologie, Sozialwissenschaften und viele andere.

## Modell von KE

Doch was benötigt KI um empathisch zu sein? Wie kann KE umgesetzt werden? Srivastava hat ein Modell zu exakt diesem Zweck erstellt. Als Basis ihrer Arbeit dient Mark Davis' Unterteilung von Empathie. Laut ihm besteht Empathie aus zwei Komponenten:

- dem kognitiven Verständnis einer anderen Perspektive und dem Verständnis der zu beobachtenden Reaktionen (z.B. Gesichtsausdrücke, verbale Antworten)
- Um das Erlebnis einer anderen Person zu teilen muss das Subjekt dazu fähig sein, das Leiden eines Anderen mit den eigenen Lebenserfahrungen in Verbindung zu bringen. Folglich wird ein Gedächtnis benötigt. Die Situation der anderen Person wird mit einer Erinnerung an ein ähnliches Ereignis oder durch verschiedene persönliche Erinnerungen rekonstruiert. [5]

Aufgrund dieser Annahmen schlägt Srivastava vor, dass mit den folgenden Schritten ein hoher Grad von Empathie erreicht werden kann:

- 1) Das Leiden einer Person wird durch die verbale Beschreibung, den Stimmklang oder den Gesichtsausdruck erkannt.
- 2) Zwischen dem Leiden dieser Person und den eigenen Erinnerungen der KI werden Ähnlichkeiten berechnet.
- 3) Nun, da der Zustand der Person bekannt ist und Korrelationen mit dem eigenen Erinnerungen gefunden wurden, wird eine angemessene Antwort erstellt. Sie beinhaltet eine sowohl sprachlich als auch z.B. in Hinblick auf die Gestik angemessene Ausdrucksform. [5]

Im Folgenden sollen diese drei Schritte detaillierter betrachtet werden.

### **1. Erkennen von Schmerz oder Trauer**

Wie zumeist in der Entwicklung von KI wird der Mensch analysiert, um Muster und Regeln zu erkennen. Die Fähigkeit Schmerz oder Trauer zu erkennen ist uns Menschen nicht angeboren. Sie muss zunächst entwickelt werden, wobei die Feinheiten über das Leben hinweg angepasst werden. Um diesen Lernprozess nachzubilden bietet sich Machine Learning an. Menschen bilden mentale Modelle durch vielerlei Eindrücke. [5]

Ein Kleinkind erlernt beispielsweise das Konzept eines Apfels dadurch, dass die Mutter das Wort oft ausspricht, während sie den Apfel zeigt. Das Kleinkind bildet hier ein Modell davon, welche Eigenschaften ein Apfel hat und wie er ausgesprochen wird. [5]

Gleichermaßen werden bei einem Emotionsmodell Informationen darüber abgespeichert, welche Eigenschaften, Aussprache, Gesichtsausdrücke, tonalen Klänge mit der Emotion verbunden sind. Die Fähigkeit, den Gefühlszustand eines Gegenübers emotional einzuordnen beruht laut Srivastava deswegen fundamental auf dem Einordnen der Reize in ähnliche Situationen aus dem Gedächtnis. Die Fähigkeit eine Szene als traurig zu erkennen ist demnach als Fähigkeit, sich an ähnliche Szenen aus der Vergangenheit, die ebenfalls 'traurig' bedeuteten zu erinnern. [5]

Diese menschlichen Vorgänge können für eine KE abstrahiert und umgesetzt werden. Natural Language Understanding wird in Machine Learning eingesetzt. Hiermit können die Bedeutung von Sätzen und die damit verbundenen Emotionen rechnerisch nachvollzogen werden. Auch hier werden Modelle gebildet, um das darunterliegende Gefühl eines Satzes durch die verwendeten Wörter und die grammatikalische Struktur zu erfassen. Praktisch ist hiermit die Analyse von Text auf Emotionen hin möglich, oder auch die Integration in Sprachassistenten. [5]

Noch ein weiteres Feld aus dem Gebiet Machine Learning hilft beim Entwickeln von KE: Optische Erkennung. Hiermit könnte es einer KI gelingen, nonverbale Kommunikation zu analysieren. In der Lernphase werden der KI verschiedene Beispiele dafür vorgelegt, welche Gesichtsausdrücke als "traurig" und welche nicht als "traurig" eingestuft werden. [5] Laut einer Studie aus dem Jahr 2014 sind Computer bereits besser als Menschen darin, in Gesichtsausdrücken vorgegebenen von echtem Schmerz zu unterscheiden.

## **2. Ähnlichkeiten in Erinnerungen finden**

Um eine KI als empathisch bezeichnen zu können reicht es nicht aus, wenn sie lediglich den Schmerz eines Anderen erkennt. Sie muss diese Gefühle teilen. Ein Verständnis dieser Erfahrung erfordert ein Vertrautheit mit dem Erlebnis (dieser Emotion). Auf einer Festplatte abgespeicherte Informationen lassen sich aber nicht mit dem menschlichen Wiedererleben einer Erfahrung vergleichen, so Srivastava. [5]

Wie also lässt sich ein vergleichbares, virtuelles Gedächtnis abbilden? Diesmal kommt der Lösungsansatz aus der Neurologie: Mit Spiegelneuronen lassen sich Erinnerungen möglicherweise automatisieren. Diese sitzen im menschlichen Gehirn und werden vereinfacht gesagt sowohl aktiv, wenn wir etwas bestimmtes tun, als auch wenn wir jemandem zusehen, wie er dieselbe Handlung ausführt. Psychologen nehmen sogar an, dass diese Neuronen vergleichbar auf das Erlebnis von Mobbing und dem Beobachten von Mobbing reagieren. Dieses Gebiet ist noch nicht vollständig erforscht und die Thesen sind nicht gefestigt. Dennoch besteht unter Psychiatern und Psychologen die Annahme, dass Spiegelneuronen eine starke Rolle bei der Entwicklung von Empathie spielen. Vielleicht lässt sich mit ihnen künftig ein Ansatz für AE finden. [5]

Wenn sich ein Mensch an etwas erinnert, ist dies zumeist mit sensorischen Reizen verbunden: Wie war das Wetter? Welche Gerüche und Geräusche waren wahrzunehmen? Wenn eine Reihe ähnlicher Reize vorliegt, fällt das Erinnern leichter. Nicht nur Marketing kann sich dies zu Nutzen machen. Falls sich Empathie auf Reize, die eine Erinnerung über Neuronen auslösen herunterbrechen lässt, könnten einer KI dieselben Reize geliefert werden, die dann eine künstliche Erinnerung hervorrufen. [5]

Im menschlichen Gehirn ist die Amygdala für emotionale Erinnerungen zuständig. Wenn die Struktur dieser Gehirnregion besser erforscht ist könnten daraus Erkenntnisse darüber gewonnen werden, wie Informatiker dieses Areal simulieren können. [5] Forschern am MIT ist es sogar bereits gelungen falsche Erinnerungen in Mäuse zu implantieren. [6] Beruhend auf diesen Fakten argumentiert Srivastava, dass komplexe Emotionen und Erinnerungen letztlich auf neuronalen Netzwerken im Gehirn basieren, die mit einer KI repliziert werden können. [5] Anmerkung: Dies sind zwar Aspekte die zum Einordnen und Speichern von Emotionen gehören. Es ist aber nicht zu vernachlässigen, dass Körper und Verstand ein hochkomplexes System bilden. Wie stark welche Hormone bei einem Individuum ausgeschüttet werden, wie die allgemeine momentane Stimmung ist und wie die individuelle

Wahrnehmung die Empathie beeinflusst oder beeinträchtigt, werden in Srivastavas Modell außer Acht gelassen.

### **3. Eine angemessene Reaktion formulieren**

Nachdem die fiktive KI Gefühle einer Person erkannt und eingeordnet hat ist eine adäquate Reaktion erforderlich. Die derzeitige Kommunikation zwischen KI und Menschen beruht auf vorgespeicherten Antworten. "Alexa, erzähl mir einen Witz!" fordert Alexa dazu auf, in der Datenbank nach einer von Designern entworfenen Antwort zu suchen. Diese hard-codierte Form der Kommunikation erscheint zwar rigide, soll aber nach Srivastava der menschlichen Reaktion sehr nahe kommen. [5]

Laut ihr gibt es für empathische Antworten gesellschaftlich anerkannte Regeln und Strukturen. Diese sind Algorithmen ähnlich und sollen von ihnen verwendet werden können. Bei einem Gespräch, in dem eine Person mitteilt, einen Angehörigen verloren zu haben enthält die Antwort eines Anderen zumeist eine Variante von "Es tut mir leid". Ein Witz oder hämischer Tonfall würde allgemein als ignorant und unhöflich eingestuft werden. Eine gesellschaftliche Regel für Reaktionen auf Todesfälle ist gefunden. [5]

Srivastava formuliert ein Fallbeispiel für eine angemessene Reaktion einer empathischen KI. Eine Frau teilt der KI mit: "Jemand hat mein Haus ausgeraubt und den gesamten Schmuck gestohlen." Die KI könnte ihren Gesichtsausdruck als aufgebracht einordnen. Die Wörter "ausgeraubt" und "gestohlen" umschreiben normalerweise eine besorgniserregende Situation. Das Wort "ausgeraubt" könnte in der KI darüber hinaus eine künstliche Erinnerung auslösen, in der die KI in ihrer Wohnung vor fünf Jahren ausgeraubt wurde und all ihr Geld verlor. Die Antwort der KI könnte gemäß gesellschaftlichen Erwartungen folgend aussehen: "Es tut mir so leid das zu hören. Jemand stahl vor fünf Jahren mein Geld aus meiner Wohnung und ich erinnere mich daran, wie belastend das war. Bitte lass mich wissen ob du Hilfe möchtest." Laut Srivastava hätte der Computer hier KE praktiziert. [5] [7]

Dass diese Reaktion unglaubwürdig scheinen könnte, hat Srivastava erkannt. Sie begründet dies mit dem Uncanny Valley. Allein das Wissen, dass es sich um eine KI handelt könnte diese im Hinblick auf Empathie als unglaubwürdig erscheinen lassen. Sie schlägt vor, dass Menschen durch die optische Repräsentation der KI von diesem Wissen abgelenkt werden könnten, um sie von der Echtheit der Empathie zu überzeugen. Statt einem zu humanoiden Äußeren sollten niedliche, abstrakte Formen bevorzugt werden. Dadurch könnte das Vertrauen der KI gegenüber leichter fallen. Diesen Ansatz verfolgt übrigens die Robotertechnik ebenfalls. [5]

Anmerkung: Auch wenn ein niedliches Äußeres angenehmer ist als Humanoide, die im Uncanny Valley anzusiedeln sind, ändert dies nichts am grundlegenden Wissen, dass die KI nicht menschlich ist.

## **Emotionale KI**

Wie bereits erörtert, muss das Erkennen von Emotionen einer empathischen Reaktion vorangehen. Deswegen ist es von fundamentaler Notwendigkeit, zunächst Emotionale KI zu entwickeln um letztlich KE zu erzielen. Insbesondere im Silicon Valley wird dieses Thema derzeit intensiv bearbeitet und gepriesen, diverse Firmen versprechen hier bahnbrechende Resultate. Für das Erkennen von Emotionen werden aber zunächst Inputdaten benötigt. Für diese gibt es verschiedene Ansätze:

## **Gesichtsanalyse**

Kairos [8] und viele andere Firmen spezialisieren sich auf optische Detektion. Wie die meisten anderen Anbieter erkennt ihre Software positive, negative und neutrale Stimmungen auf einem Gesicht. Die detektierten Emotionen differenziert die Software in Freude, Wut, Ekel, Traurigkeit, Angst und Überraschung. Auch Microsoft Azure [9] bietet optische Analyse an. Der Selbstversuch zeigte deutlich, dass die Erkennung und Einordnung von Emotionen bei den Werbebildern von Azure besser funktioniert als mit eigenen Bildern. Grimassen werden fehlerhaft eingeschätzt; auf schiefen Bildern wird das Gesicht überhaupt nicht erkannt. Andere Firmen liefern hier sehr ähnliche Resultate. Auffällig ist, dass sich diese Anbieter auf nur sehr wenige Gefühlsdimensionen beschränken.

## **Pupillenanalyse**

Die Neurologin Poppy Crum referierte in einem TED-Talk zu Emotionserkennung unter anderem über Pupillenanalyse. Laut Crum tragen wir bewusst oder unbewusst stetig einen Teil unseres Innenlebens nach außen. Die Augen reagieren darauf, wie schwer das Gehirn arbeitet. Diese Pupillenveränderungen sind rein an eine interne Anstrengung gekoppelt, gänzlich unabhängig der Adaption an die Beleuchtung.

## **Infrarotbildanalyse**

Änderungen des Stresspegels, die Intensität der Gehirnaktivitäten, ob eine Person aufmerksam ist oder nicht, in eine Konversation mit einbezogen - all dies soll durch Infrarotbilder ermittelt werden können. Auch Gedanken lassen sich zumindest partiell ermitteln: Beim Denken an ein Feuer oder das Betrachten eines Bildes davon sollen Probanden wärmere Wangen gehabt haben. Sogar Ehrlichkeit soll sich künftig hiermit tracken lassen.

Nebst diesen genannten Technologien gibt es noch viele weitere mögliche Datenquellen für die Emotionserkennung. Die Zusammensetzung der Atemluft, die Leitfähigkeit der Haut, der Puls, der Stimmklang [10], die Wortwahl und Grammatik könnten künstlicher Empathie auf die Sprünge helfen. Zusammenfassend müssen Emotionen Anderer korrekt erkannt werden, bevor diese von einer KI verstanden werden können, um eine empathische Antwort hervorzurufen. Hierfür gibt es verschiedene Ansätze und Systeme, in Zukunft wird wahrscheinlich vor allem mit einer Mischung verschiedener Techniken gearbeitet. Auffällig ist, dass sich Emotionserkennungssysteme derzeit nur auf eine sehr grobe Unterteilung der Gefühle beschränken. Auf dem Diagramm Abb. 2 ist beispielsweise eine feinere Auffächerung von verschiedenen Gefühlen zu sehen, die aber bei weitem noch nicht alles abdeckt. Für sich allein gestellt sind die derzeitigen Analysesysteme noch zu ungenau. Gute Ergebnisse liefern optische Erkennungssysteme nur bei stark ausgeprägten Gesichtsausdrücken, die gut in eine Kategorie passen. Vor allem Vermischungen von fein unterteilten Emotionen bereiten derzeit noch Probleme.

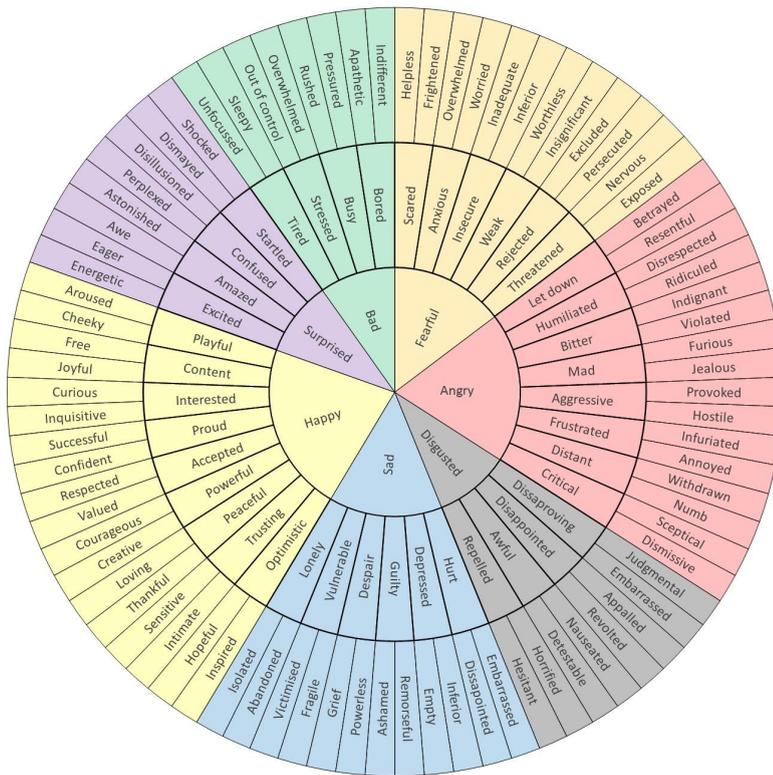


Abb.2: Gefühlsdimensionen<sup>2</sup>

## Stand der Technik - ein Blick auf den Markt

Derzeit gibt es schon verschiedene Produkte, die als KE vermarktet werden. In einem kurzen Überblick über eine Auswahl von Produkten wird dargestellt, wie nahe diese einer authentischen KE sind.

### Roboterbegleiter Pepper

Dieser Roboter ist seit 2014 auf dem Markt. Die Emotionserkennung erfolgt über eine Kombination von vokaler und inhaltlicher Analyse des Gesprochenen, sowie die der Körperhaltung. Für den Heimgebrauch vorgesehen ist diese KI noch nicht, sie wird als Forschungsobjekt für akademische Einrichtungen oder beispielsweise als Empfangsdame in Restaurants eingesetzt. [11] In einem Demovideo auf YouTube [11], [12] funktioniert die Emotionserkennung nicht fehlerfrei. Die gegenüberstehende Person muss ihre Mimik und Gestik sehr überbetonen. Pepper imitiert, also spiegelt lediglich das Verhalten des Gegenübers, in dieser Demo ist keine künstliche Empathie zu finden. Vielmehr ist Peppers Reaktion als Gefühlsansteckung einzuordnen. [13]

### Roboterbegleiter Musio

Ein Jahr nach Pepper wurde Musio veröffentlicht. Dieser Begleiter führt eigenständige Gespräche mit Menschen, erkennt deren Gesichtsausdrücke, sowie Gegenstände. Stehen mehrere Personen simultan in Musios Nähe, so passt er seine Interaktionen individuell an den Einzelnen an. Diese basieren auf dem emotionalen Zustand des Gesprächspartners, vorherigen Diskussionen und den besprochenen Themen. Um das notwendige Gedächtnis zu simulieren wird Deep Learning verwendet. Die Entwickler hoffen, dass Musio zum

<sup>2</sup> <https://i.imgur.com/q6hcgsh.jpg>

Beispiel ängstlichen Menschen die Möglichkeit ein dynamisches Gespräch in einer Fremdsprache zu führen bietet, sodass diese üben können mit anderen zu reden. Auch hierzu gibt es ein Demo-Video. [14] Wenngleich die Interaktion zwischen Musio und dem Gesprächspartner unecht wirkt ist sie dynamischer als die mit Pepper. Allerdings gibt es mehrere Demos mit exakt derselben Konversation aber an verschiedenen Drehorten. Deswegen ist nicht ersichtlich, ob und wieviel die KI selbstständig analysiert und erkennt und welche Inhalte hart codiert ist.

### **Ellie - die künstliche Therapeutin**

Seit den späten 90er Jahren ist diese KE in Arbeit. Über die Microsoft Kinect verfügt Ellie über Gesichtserkennung, detektiert Kopfhaltungen und Körpergesten, durch ein Mikrofon auch Stimmparameter. Die virtuelle Therapeutin spricht hauptsächlich mit ungebundenem Militärpersonal. Eingesetzt wird sie um zuzuhören und gegebenenfalls Symptome einer psychischen Belastung festzustellen. Ellie führt sehr ungezwungene Gespräche mit Menschen und kann Folgefragen stellen. Um Empathie zu zeigen nickt und grübelt sie in angemessenen Situationen. Ellie war in einer Probestudie des ICT sehr überzeugend: Hier wurde der Hälfte der Probanden erzählt, dass Ellie von Menschen kontrolliert sei. Die andere Hälfte wurde darüber in Kenntnis gesetzt, dass Ellie computergesteuert war. Die letztere Gruppe vertraute Ellie mehr persönliche Details über ihre post-traumatische Belastungsstörung und Ängste an, als die erste Gruppe. [15] In der Demo hält Ellie die Versprechen, die ihre Entwickler machen. [16]

Unabhängig der Produktfeatures einer KE ist die Betrachtung, wie die KE auf Nutzer wirkt wichtig. Zu eben diesem Thema führten Shyam Sundar und Bingjie Liu eine Studie durch. Sie fanden heraus, dass Menschen unterschiedlich auf KI die Emotionen hat reagieren. Die Reaktion hängt davon ab, ob künstliche Emotionen für möglich gehalten werden, oder nicht. Darüber hinaus bevorzugten ihre Probanden einfühlsame Antworten gegenüber emotionslosen Antworten. [17]

Für die Studie wurden vier Chatbots mit verschiedenen Antwortmustern erstellt. Einer drückte Mitgefühl anderen gegenüber aus; er formte Beispielsätze wie "Es tut mir leid, das zu hören". Ein weiterer Chatbot war auf kognitive Empathie ausgelegt. Er antwortete: "Dieses Problem kann sehr beunruhigend sein". Der dritte Chatbot reagierte emotional empathisch und sagte beispielsweise: "Ich verstehe deine Angst vor der Situation". Der vierte Bot war eine Kontrollversion und gab lediglich Ratschläge. [17]

Die Mehrheit der Stichprobe glaubte nicht an maschinelle Emotionen, sie interpretierten die Äußerungen von Mitgefühl und Empathie als einprogrammierte Höflichkeitsfloskeln. Probanden, die maschinelle Emotionen hingegen für möglich hielten, reagierten negativ auf Äußerungen des Mitgefühls und Empathie von Chatbots. Liu erwägt, dass den Probanden die Chatbots zu persönlich werden könnten, oder dass ihnen das Konzept einer fühlenden Maschine unangenehm sein könnte. Insgesamt favorisierten die Probanden den Bot mit emotionaler Empathie und den mit Mitgefühl. Die kognitiv empathische Version empfanden die Probanden als zu kalt und distanziert. Künftig wollen Liu und Sundar herausfinden, ob sich ein einheitliches Muster bezüglich der Reaktion von Menschen auf maschinelle Emotionen finden lässt. [17] Wird KE in Zukunft nach einem benutzerzentrierten Ansatz entwickelt, ist es also wichtig die Wirkung von KE in das Konzept miteinzubeziehen.

## Diskussion

Die Rezipienz von KE ist bisher sehr zwiegespalten. So wird zum Beispiel argumentiert, dass KE neue Arbeitsplätze schafft.[18] Andere befürchten wiederum, dass dadurch Arbeitsplätze wegrationalisiert werden[17]. Auch im Hinblick auf Datenschutz gibt es Bedenken: Der Mensch könnte gläsern werden - intimste Daten werden so zukünftig womöglich sogar ohne Einverständnis aufgenommen.

Auch die Sorge vor dem außer Kontrolle geratenen Experiment fließt in KE mit ein. Schon jetzt gibt es KI, die ihre eigene Sprache entwickelt hat, weshalb das Experiment abgebrochen wurde. Was passiert, wenn KI ein eigenes Empathiekonzept erstellt (und den Menschen darin nicht wie von den Entwicklern vorgesehen berücksichtigt)? Wäre dies nicht eine psychisch kranke KI? Dagegen wird argumentiert, dass Kriminellen Empathie fehlt. Um also die Gefahr von KI, die die Macht übernimmt einzudämmen, muss KI Empathie besitzen. Befürworter von KE sind davon überzeugt, dass uns Technik auf einer tieferen Ebene miteinander verbinden kann. [19] Mittels Augmented Reality und KE könnten Highschoolberater beispielsweise psychische Belastungen besser feststellen. Kritiker halten dagegen, dass KE die Zwischenmenschlichkeit ersetzen soll, statt sie zu stärken. Die eigentlichen subjektiven Eigenschaften könnten verloren gehen, statt eigenen Einschätzungen könnte man in Zukunft der Technik vertrauen.

Wenn ein Mensch weiß, dass sein Gegenüber künstlich ist, wie kann er dann davon überzeugt werden, dass der Gegenüber echte Empathie empfindet? Durch Täuschung könnte hier Menschen geholfen werden. Allerdings widerspricht diese Täuschung je nach Auslegung der Menschenwürde.

Der Journalist Jonathan Cook kritisiert, dass die Firmen im Silicon Valley zu leichtfertig mit einer so bedeutungsvollen und komplexen menschlichen Eigenschaft umgehen. Gefühle werden auf Einzelteile heruntergebrochen, wobei das große Ganze aus den Augen gelassen wird. [19], [20]

### **Können Maschinen Emotionen besitzen?**

Zuvor wurden die Möglichkeiten und derzeitigen Grenzen von Emotionaler KI behandelt. Aber wird es KI überhaupt jemals möglich sein, Gefühle zu empfinden? Die Frage, ob Maschinen jemals Emotionen empfinden können werden hängt mit der Definition von Emotionen selbst zusammen. Diese Auslegung bedingt auch, ob KE als manipulativ oder benutzerfreundlich gesehen werden könnte.

Wenn Emotionen, wie nach Moravec nur in Bezug auf ihre Wirkung auf das Verhalten, oder auf die Funktionsweise innerhalb eines Organismus definiert werden, so können nach Crevier Emotionen als ein Mechanismus angesehen werden, den ein intelligenter Handlungsträger verwendet, um den Nutzen seiner Handlungen zu maximieren. Kurzum sind Emotionen hier nur ein Mittel, um das Verhalten eines Gegenübers in eine Richtung zu lenken, die dem Überleben der Spezies dient. Nach dieser Auslegung wird KI in Zukunft sehr emotional sein. [21]

Eine andere Sichtweise fokussiert sich auf die subjektive Qualität von Gefühlen. Also darauf, wie es sich anfühlt, eine Emotion zu haben. [21] Ob KI jemals Emotionen fühlen können wird, oder diese immer nur vorgibt hängt stark mit der Frage, ob KI sich jemals über sich selbst bewusst wird zusammen. Dies geht zwar über diese Arbeit hinaus, illustriert aber wie intermediär das Gebiet ist. Krettek, die Gründerin des Empathy Labs von Google ist davon

überzeugt, dass KI niemals selbst fühlen wird. Sie glaubt, dass Gefühle etwas rein menschliches bleiben werden. Statt zu fühlen wird KI künftig Emotionen täuschend echt vorgeben: [Wenn wir uns also lange genug von Maschinen beobachten lassen, werden sie uns wahrscheinlich sehr überzeugend nachahmen können. Aber meine persönliche Meinung ist, dass die echte emotionale Verbindung - die echte empathische Verbindung - und die Idee, sich seiner selbst bewusst zu sein - eine allein menschliche Sache ist.][22]

### **Ist KE die Emotionen vorgibt ethisch? [23]**

Bei einer KI, die Gefühle vortäuscht steht in Frage ob dies ethisch vertretbar ist. Laut Krettek ist hier die Transparenz und Ehrlichkeit den Nutzern gegenüber wichtig. [22]

Jedoch ist davon auszugehen, dass nicht alle Nutzer tatsächlich nachvollziehen können werden, dass es sich bei den künstlichen Gefühlen um eine Täuschung handelt. Auch eine Täuschung zum Vorteil der Nutzer ist Manipulation.

Kommerziell eingesetzt wird KI viel effektiver im Verkaufen werden. Selbst wenn man hier das Bewusstsein und die Kenntnis hat, dass man der KI egal ist, können die eigenen Handlungen trotzdem beeinflusst werden. Professor Sparrow diskutierte über den Einsatz von Robotern in der Altenpflege: [Das Design zur Täuschung gegenüber einer Person sei respektlos, auch wenn diese Täuschung glücklich mache. Es sei ethisch problematisch zu sagen: "Ich werde die Menschen zu ihrem eigenen Besten täuschen"].[23]

Im medizinischen Dienstleistungsbereich wird KE bereits eingesetzt. Der Roboter Paro erzielte bereits Erfolge im Umgang mit Demenz-Patienten. Kritiker fragen jedoch warum überhaupt KE statt Menschen eingesetzt wird. Sie sind auch der Meinung, dass es unmenschlich sei, Patienten nur eine KE bereitzustellen. Aber auch hier gibt es wieder Punkte, die für KE sprechen: Gegenüber dem Menschen bietet KE eine höhere Verlässlichkeit. Sie wird im Gegensatz zu menschlichen Angestellten nicht durch Emotionsarbeit belastet und könnte einen Teil davon übernehmen. So würde das Personal entlastet. Wo Menschen nicht ständig Fürsorge aufbringen können, könnte KE das ergänzend eingesetzt werden. Außerdem fluktuiert die Empathie einer KI nicht. Bei Menschen hängt diese zum Beispiel von der Stimmung ab, dem Stresspegel, der Fähigkeit sich an vergangene Ereignisse zu erinnern, der Meinung von anderen. KE könnte hier also unterstützend eingesetzt werden, um menschlichen Bedürfnissen besser nachzukommen.[5], [24]

### **Missbrauch von Empathie**

In Zukunft könnte KE auch in der Politik eingesetzt werden, zum Beispiel unterstützend beim Verfassen von Reden. KI, die üblicherweise über große Datenmengen verfügt, wie beispielsweise Meinungen aus den sozialen Medien und gleichzeitig zu kognitiver Empathie befähigt ist, kann diese theoretisch nutzen, um Wähler zu manipulieren. In der Vergangenheit lassen sich Negativbeispiele dafür finden, wie Massen über ihre Empathie zu politischen Zwecken manipuliert wurden. [3] Wie bereits erwähnt könnte KE auch im Marketing instrumentalisiert werden. Auch die Massenmedien könnten von KE, die die Inhalte, Titel an den Nutzer anpasst mit höheren Einschaltquoten profitieren.

## **Fazit**

Laut dem CEO von Affectiva, einer Firma, die sich auf Gesichtserkennung spezialisiert hat wird sich die Technik um uns herum innerhalb der nächsten vier Jahre (2017) über

Emotionen bewusst. Irgendwann soll diese Technologie nicht mehr aus dem Alltag wegzudenken sein. [7], [25]

Die von Srivastava vorgestellten Konversationen mit einer KE wirken nicht authentisch, da sich der Anwender über die Simulation bewusst ist. Um eine Täuschung, das Vorgeben von Emotionen künftig moralisch vertreten zu können, ist Krettek jedoch der Meinung, dass Transparenz wichtig ist. Hier besteht ein Zwiespalt zwischen dem Nutzen der KE und der Moral.

Die nach Srivastava vorgeschlagenen recht rigiden Konversationsmuster wirken nicht in jeder Situation empathisch. Menschen besitzen gelegentlich das Bedürfnis, ihre Sorgen mitzuteilen und über *ihre* Probleme zu sprechen. Wenn die KE nach dem Muster von Srivastava ausführlich über ihre eigenen fiktiven Erlebnisse und Probleme spricht, könnten sich Nutzer vor den Kopf gestoßen vorkommen. Die KE muss also zudem die Bedürfnisse der Nutzer erkennen und darauf eingehen.

Es ist ersichtlich, dass eine authentische KE noch nicht umsetzbar ist. Dieses intermediäre Gebiet wird derzeit stark forciert. Ermutigt von den allgemeinen Fortschritten im Bereich KI und Neurowissenschaften sind Informatiker und Forscher davon beflügelt alles quantifizierbar abzubilden. Es ist bei informatischen Fragen üblich, dass die betrachteten Untersuchungen mit einem Schema moduliert und mit Formeln abgebildet werden können. Für KE sollen innere Prozesse wie Emotionen durch eine KI nachgeahmt werden. Aber Zwischenmenschlichkeit ist so vielfältig und facettenreich, weil es viele verschiedene Individuen gibt, die komplett unterschiedlich situativ abhängig miteinander umgehen. Dies lässt sich nicht durch ein simplifiziertes System darstellen. Davon abgesehen sind selbst die menschlichen Prozesse noch nicht ausreichend erforscht und verstanden. [26], [27]

Es ist nicht klar, ob authentische KE letztlich umsetzbar werden wird. Jedoch ist echte zwischenmenschliche Empathie und das Erleben eines Gefühls viel komplexer, als es derzeit in der Entwicklung von KE betrachtet wird. Das große Ziel wird in Teilprobleme zerlegt. Ist eines dieser Teilprobleme gelöst, so wird es nach außen vermarktet, als sei das große Ziel erreicht - als gebe es schon Künstliche Empathie - wie an der Vermarktung von Emotionaler KI deutlich wird.

Nicht nur innerhalb der Forschung, aber auch in der Gesellschaft ist eine Klarheit darüber notwendig, aus welchen Gründen und für welchen Zweck KE entwickelt werden soll. KE wird auch aus den Grundbedürfnissen von Menschen wie beispielsweise nicht alleine zu sein, dem Verlangen nach Verbundenheit, gesehen und verstanden zu werden und Nähe entwickelt. Es ist dafür zu sorgen, dass KE nicht genau das Gegenteil bezweckt. Wie oft in der Geschichte der Menschheit wird sich der technische Fortschritt vermutlich nicht aufhalten lassen, zumindest eine näherungsweise Form von KE wird früher oder später auf dem Markt veröffentlicht werden. Es liegt auch an uns als Gesellschaft und Konsumenten, diese neue Technologie bewusst einzusetzen und hier die richtigen Fragen zu stellen.

“Humans ultimately hold the power to decide whether the future of AI is one we should fear or a future to look forward to.“ [5]

## Quellen

- [1] P. P. Jan Sievers, "Empathie: Bedeutung, Definition, Beispiele & Herkunft," *Jeden Tag ein Fremdwort - neueswort*, 08-Mar-2018. [Online]. Available: <https://neueswort.de/empathie/>. [Accessed: 27-Feb-2019].
- [2] C. Düllings, "Empathie-Definition – was bedeutet Empathie eigentlich genau?," *Empathie Akademie*, 21-Nov-2013. [Online]. Available: <http://www.empathie-lernen.de/empathie-definition>. [Accessed: 27-Feb-2019].
- [3] A. der Wikimedia-Projekte, "Empathie – Wikipedia," *Wikimedia Foundation, Inc.*, 30-Sep-2003. [Online]. Available: <https://de.wikipedia.org/wiki/Empathie>. [Accessed: 27-Feb-2019].
- [4] M. Asada, "Affective Developmental Robotics - How can we design the development of artificial empathy?" Dec-2014.
- [5] M. Srivastava, "View of The Computational and Aesthetic Foundations of Artificial Empathy," Stanford University, 2016.
- [6] S. Ramirez *et al.*, "Creating a False Memory in the Hippocampus," *Science*, vol. 341, no. 6144, pp. 387–391, Jul. 2013.
- [7] C. Alvarenga, "The Algorithm Manager Is Here. Artificial Empathy Is Next," *Reconomics*, 11-Nov-2017. [Online]. Available: <https://reconomics.com/2017/11/11/the-algorithm-manager-is-here-artificial-empathy-is-next/>. [Accessed: 27-Feb-2019].
- [8] R. Belyeu, "Face Recognition Features from Kairos," *Kairos*. [Online]. Available: <https://www.kairos.com/features>. [Accessed: 27-Feb-2019].
- [9] "Emotionen," *Microsoft Azure*. [Online]. Available: <https://azure.microsoft.com/de-de/services/cognitive-services/emotion/>. [Accessed: 27-Feb-2019].
- [10] *DEMO Real Time Speech Emotion Recognition Using Deep Learning*. Margaret Lech, 2017.
- [11] "Pepper (robot) - Wikipedia," *Wikipedia*. [Online]. Available: [https://en.wikipedia.org/wiki/Pepper\\_\(robot\)](https://en.wikipedia.org/wiki/Pepper_(robot)). [Accessed: 27-Feb-2019].
- [12] *Live Demonstration "Pepper, the 1st humanoid robot capable of recognizing main human emotions."* cebitchannel, 2017.
- [13] "Gefühlsansteckung – Wikipedia," *Wikipedia*. [Online]. Available: <https://de.wikipedia.org/wiki/Gef%C3%BChlsansteckung>. [Accessed: 27-Feb-2019].
- [14] *Hands-on with Musio*. Engadget, 2015.
- [15] C. Calistra, "Part 3: Impressive Artificial Intelligence Using Empathy Now," *Kairos*, 07-Sep-2016. [Online]. Available: <https://www.kairos.com/blog/empathy-in-ai-series-part-3-impressive-artificial-intelligence-using-empathy-now>. [Accessed: 27-Feb-2019].
- [16] *SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support*. USCICT, 2013.
- [17] M. Swayne, "Empathetic machines favored by skeptics but might creep out believers," *ScienceDaily*, 01-Nov-2018. [Online]. Available: <https://www.sciencedaily.com/releases/2018/11/181101085240.htm>. [Accessed: 27-Feb-2019].
- [18] B. Virdee-Chapman, "Kairos API Series: Part 2 Empathetic Machines Creating Jobs?," *Kairos*, 30-Aug-2016. [Online]. Available: <https://www.kairos.com/blog/empathy-in-ai-series-part-2-empathetic-machines-creating-jobs>. [Accessed: 28-Feb-2019].
- [19] P. Crum, *Technology that knows what you're feeling*. TED.
- [20] J. Cook, "What Emotional AI Fails to Grasp About Emotion – Jonathan Cook – Medium," *Medium*, 06-Aug-2018. [Online]. Available: <https://medium.com/@JonathanCCook/what-emotional-ai-fails-to-grasp-about-emotion-344774e1a969>. [Accessed: 27-Feb-2019].
- [21] Contributors to Wikimedia projects, "Philosophy of artificial intelligence - Wikipedia," *Wikimedia Foundation, Inc.*, 21-Oct-2005. [Online]. Available: [https://en.wikipedia.org/wiki/Philosophy\\_of\\_artificial\\_intelligence#Can\\_a\\_machine\\_have\\_emotions](https://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence#Can_a_machine_have_emotions). [Accessed: 27-Feb-2019].
- [22] D. Allegretti, "Meet the Woman Teaching Empathy to AI," *Vice*, 29-May-2018. [Online]. Available: [https://www.vice.com/en\\_au/article/ywe33m/this-woman-believes-ai-can-be-taught-empathy%20%20](https://www.vice.com/en_au/article/ywe33m/this-woman-believes-ai-can-be-taught-empathy%20%20). [Accessed: 27-Feb-2019].
- [23] A. Bogle, "Good bot, bad bot: Can you trust a robot that cares?," *ABC News*, 02-Jun-2018. [Online]. Available: <https://www.abc.net.au/news/science/2018-06-02/can-you-trust-a-robot-that-cares/9808636>. [Accessed: 27-Feb-2019].
- [24] M. N. S. Fasola, "View of The Computational and Aesthetic Foundations of Artificial Empathy," no. 10, 2016.
- [25] J. de la Rubia, "Website," *LinkedIn*, 03-Jul-2018. [Online]. Available: <https://www.linkedin.com/pulse/rise-emotionally-intelligent-ai-jos%C3%A9-de-la-rubia>. [Accessed:

27-Feb-2019].

[26] B. Virdee-Chapman, "Empathy in AI Series: Part 5, The Future of Empathetic AI," *Kairos*, 24-Oct-2016. [Online]. Available: [ps://www.kairos.com/blog/empathy-in-ai-series-part-5-the-future-of-empathetic-ai](https://www.kairos.com/blog/empathy-in-ai-series-part-5-the-future-of-empathetic-ai). [Accessed: 28-Feb-2019].

[27] T. Liu, "Blog," 2013. [Online]. Available: <https://www.startupgrind.com/blog/quantum-processing-the-path-from-autistic-to-empathic-ai/>. [Accessed: 28-Feb-2019].