



Catrin Misselhorn

# **Künstliche Intelligenz und Empathie**

Vom Leben mit Emotionserkennung, Sexrobotern &  
Co.

**Reclam**

E-Book-Leseproben von einigen der beliebtesten Bände unserer Reihe [Was bedeutet das alles?] finden Sie [hier](#) zum kostenlosen Download.

2021 Philipp Reclam jun. Verlag GmbH,  
Siemensstraße 32, 71254 Ditzingen

Covergestaltung: Cornelia Feyll, Friedrich Forssman

Gesamtherstellung: Philipp Reclam jun. Verlag GmbH,  
Siemensstraße 32, 71254 Ditzingen

Made in Germany 2021

RECLAM ist eine eingetragene Marke der Philipp  
Reclam jun. GmbH & Co. KG, Stuttgart

ISBN 978-3-15-961840-1

ISBN der Buchausgabe 978-3-15-014047-5

[www.reclam.de](http://www.reclam.de)

# **Inhalt**

## **Vorwort**

### **1. Emotionale Künstliche Intelligenz**

**Starke und schwache KI**

**Von der künstlichen Intelligenz zur emotionalen künstlichen Intelligenz**

**Was sind Emotionen?**

**Basale Emotionen**

### **2. Wie künstliche Systeme wissen, was wir fühlen**

**Gesichtsbasierte Emotionserkennung**

**Stimmbasierte Emotionserkennung**

**Sentimentanalyse**

**Bio-Sensoren**

**Multimodale Ansätze**

**Technische Herausforderungen und ethische Probleme**

### **3. Menschliche und artifizielle Empathie**

**Kognitive oder affektive Empathie?**

**Notwendige und hinreichende Bedingungen für Empathie**

**Perzeptuelle Empathie als Resonanz**

**Empathie im menschlichen Zusammenleben**

**Empathie und moralisches Urteilen**

**Artifizielle Empathie**

**Theoretische Grundlagen artifizieller Empathie**

#### **4. Von empathischen Pflegerobotern und virtuellen Seelenklempnern**

**NICA - ein Pflegeroboter mit Empathie**

**Psychologische Diagnose mit Ellie**

**Wie empathisch sind künstliche Systeme?**

#### **5. Können Roboter Schmerzen empfinden? - Neue Tendenzen der Biorobotik**

**Schmerzen als Grundlage von Empathie**

**Steht das Zeitalter der »Blade Runner« vor der Tür?**

**Ist der Dualismus das Problem?**

**Soft Robotics und homöostatisches Wohlbefinden**

**Programmierbare Organismen - eine neue Lebensform?**

## **6. Arme Alexa! - Empathie mit Robotern und virtuellen Agenten**

**Ein kleiner Ausflug ins Reich der Science-Fiction**

**Empathie und Menschenähnlichkeit**

**Ist Empathie mit Robotern moralisch relevant?**

**Empathie mit Robotern und moralisches Urteilen**

**Empathie mit Robotern und moralische Motivation**

**Empathie mit Robotern und Moralentwicklung**

**Konsequenzen der Argumentation**

## **7. Freundschaft, Liebe und Sex mit Robotern**

**Soziale Roboter als relationale Artefakte**

**Sexroboter auf dem Vormarsch**

**Sexroboter als Pornographie**

**Sexroboter als Ersatz für Frauen?**

**Solipsistische Liebe**

**Solipsismus und Selbstverdinglichung**

**Ausblick**

**Literaturhinweise**

**Zur Autorin**

# Vorwort

Der Siegeszug der künstlichen Intelligenz schreitet unaufhaltsam fort. Mehr und mehr dringt die Technologie in Bereiche vor, die bislang dem Menschen vorbehalten zu sein schienen. Zwar übertraf die künstliche Intelligenz immer wieder punktuell bestimmte menschliche kognitive Leistungen wie etwa im Go-Spiel. Doch der Weg zu solchen kognitiven Höchstleistungen bestand stets in der überlegenen Rechenfähigkeit der künstlichen Intelligenz. Menschliche Kernkompetenzen wie Emotionen und Empathie blieben außen vor. Das ändert sich zunehmend. Seit den 1980er Jahren macht sich der KI-Visionär Marvin Minsky dafür stark, auch Emotionen bei der Entwicklung künstlicher Intelligenz zu berücksichtigen, um komplexe Aufgaben besser zu bewältigen. **1**

Dies gilt umso mehr, wenn künstliche Intelligenz Menschen verstehen und mit ihnen interagieren soll. Entsprechende Fähigkeiten gewinnen zunehmend an Bedeutung, weil in der Wirtschaft in den letzten Jahrzehnten ein Umdenken stattgefunden hat. Das Paradigma des Menschen als Homo Oeconomicus, der nur darauf aus ist, seinen Nutzen rational zu maximieren, ist in die Kritik geraten. Stattdessen hat sich herausgestellt, wie sehr Menschen sich auch in ihren ökonomischen Entscheidungen von Emotionen und Affekten leiten lassen.

Geschäftsanwendungen künstlicher Intelligenz versuchen diesen Wandel abzubilden, um den Menschen als Kunden, aber auch als Mitarbeiter besser zu erfassen.

Durch die zunehmende Automatisierung wandelt sich auch das Erscheinungsbild menschlicher Arbeitsplätze. Soziale Interaktion und Kommunikation stehen immer mehr im Vordergrund. Doch selbst in diesem Feld sind die Menschen nicht mehr unter sich. Vielmehr lautet das neue Stichwort der sogenannten Arbeitswelt 5.0 *Kooperation zwischen Mensch und Maschine*. Künstliche Intelligenz soll nicht nur Routineaufgaben im Büro erledigen, Versicherungsfälle abwickeln oder Teile der Kundenkommunikation führen. Sie soll auch anspruchsvollere soziale Tätigkeiten übernehmen, etwa die Motivation von Schülern einschätzen oder auf psychische Störungen aufmerksam machen. Um diese Aufgaben zu erfüllen, müssen die Systeme in der Lage sein, menschliche Emotionen zu erkennen und angemessen auf diese zu reagieren.

Künstliche Systeme und Roboter sollen darüber hinaus Helfer, soziale Gefährten und Liebhaber sein. Kameras, Mikrofone und Sensoren im Haus sollen die Stimmung der Nutzer erkennen und etwa Beleuchtung und Musik entsprechend anpassen. Pflegeroboter sollen sich um alte Menschen kümmern, soziale Roboter Einsamen Gesellschaft leisten und Sexroboter sollen nicht nur die üblichen sexuellen Bedürfnisse erfüllen, sondern auch

solche Vorlieben, die in der realen Welt moralisch anstößig oder gar gesetzlich verboten sind.

Dieses Buch gibt einen Überblick über die noch recht junge Disziplin der emotionalen künstlichen Intelligenz und will zu einer kritischen Auseinandersetzung anregen. Der Schwerpunkt liegt auf Technologien, die ökonomisch relevant sind und unseren Alltag grundlegend verändern könnten:

Im ersten Kapitel werden die Grundbegriffe der emotionalen künstlichen Intelligenz erläutert. Das zweite Kapitel stellt die wichtigsten Methoden und Entwicklungen zur Erkennung von Emotionen durch künstliche Systeme dar und diskutiert sie aus ethischer Sicht. Wie sich zeigt, zeichnen sich diese Technologien durch einen grundlegenden Mangel an Empathie aus. Deshalb wendet sich das dritte Kapitel dem Schlüsselbegriff der Empathie zu und thematisiert die Bedeutung zwischenmenschlicher Empathie sowie die Möglichkeit, sogenannte artifizielle Empathie in künstlichen Systemen zu realisieren.

Das vierte Kapitel beschäftigt sich mit zwei Beispielen künstlicher Systeme, die den Anspruch haben, eine empathische Interaktion mit den Nutzern zu ermöglichen. Das eine stammt aus der Altenpflege und das andere aus der klinischen Psychologie. Beide werden zwar von den Nutzern durchaus als empathisch interpretiert, sie empfinden jedoch nicht wirklich Empathie. Um dies zu erreichen, wäre es notwendig, künstliche Systeme zu

konstruieren, die selbst Emotionen haben können. Dieser Möglichkeit geht das fünfte Kapitel am Beispiel des Schmerzempfindens nach, das für die Entwicklung artifizieller Empathie besonders bedeutsam ist.

Das sechste Kapitel wendet sich der Beobachtung zu, dass Menschen nicht nur mit Menschen, sondern auch mit Robotern Empathie empfinden. Zwar gleicht dieses Gefühl einer optischen Täuschung, dennoch muss man es in moralischer Hinsicht ernst nehmen und sich fragen, in welchen Bereichen solche Technologien sinnvoll und wünschenswert sind.

Diese Frage stellt sich mit besonderer Dringlichkeit für soziale Roboter, die dafür geschaffen werden, uns als Freunde, Sexualpartner oder sogar Geliebte zu dienen. Im abschließenden siebten Kapitel wird geklärt, worin sich soziale Roboter von anderen unbelebten Objekten unterscheiden, zu denen wir eine emotionale Bindung aufbauen können. Heiß umstritten sind vor allem Sexroboter. Wir werden diskutieren, inwiefern sie moralisch problematisch sind und ob der Kampagne zum Stopp von Sexrobotern recht zu geben ist. Doch Roboter versprechen nicht nur Sex, sondern auch die perfekte Liebe. Ob sie dieses Versprechen einzulösen vermögen oder ob die Liebe die letzte Bastion des Menschlichen ist, in die die künstliche Intelligenz nicht vordringen kann, wird sich am Ende des Buchs zeigen.

# **1. Emotionale Künstliche Intelligenz**

Das Forschungsgebiet der künstlichen Intelligenz (KI) stellt sich die Aufgabe, intelligentes Problemlösungsverhalten durch Maschinen nachzubilden oder zu simulieren, beispielsweise visuelle Wahrnehmung, Spracherkennung und -produktion, automatisiertes Schließen und Entscheiden. Je nachdem, wie stark der damit verbundene Anspruch ist, lassen sich zwei Formen der KI voneinander unterscheiden.

# Starke und schwache KI

*Schwache KI* ist auf die Lösung konkreter Anwendungsprobleme bezogen, menschliche Intelligenz wird im besten Fall nachgeahmt. *Starke KI* wird hingegen mit dem Anspruch verbunden, eine dem Menschen vergleichbare allgemeine und flexible Form der Intelligenz hervorzubringen. Die starke KI wirft die Frage auf, wann man einem System Intelligenz in einem menschlichen Sinn zuschreiben sollte.

In diesem Zusammenhang hat eine Versuchsanordnung namens *Turingtest* Berühmtheit erlangt. Alan Turing schuf mit der Turingmaschine die theoretischen Grundlagen des modernen Digitalcomputers. In den 1950er Jahren stellte er die These auf, dass ein künstliches System in einem menschlichen Sinn als intelligent zu gelten habe, wenn es ihm gelänge, einen menschlichen Gesprächspartner in einem Imitationsspiel davon zu überzeugen, dass es sich bei ihm um einen Menschen handelt. **2** Da Mensch und Computer nur über ein Display miteinander kommunizieren, wird von den äußerlichen Merkmalen der Beteiligten in dieser Situation abstrahiert. Dies soll Vorurteilen vorbeugen, die sich aus der mechanischen Erscheinungsweise einer Maschine ergeben. Bis heute gibt

es allerdings kein künstliches System, das den Turingtest zweifelsfrei bestanden hat.

Trotzdem gilt der Test manchen sogar als zu schwach, da er allein auf der Verhaltensbeobachtung in einem bestimmten Zeitfenster basiert. Das öffnet die Tür für allerlei Täuschungsmöglichkeiten und Tricks. So trat im Jahr 2014 das russische Computerprogramm *Eugene* mit dem Anspruch auf, den Test bestanden zu haben. Allerdings dauerte der Test nur fünf Minuten, und das System gab sich als einen 14-jährigen Jungen aus der Ukraine aus, um sprachliche Unzulänglichkeiten und mangelndes Weltwissen zu kaschieren. Wie wir sehen werden, sind wir allzu leicht bereit, künstliche Systeme als emotionale Wesen wahrzunehmen, auch dann, wenn sie über keinerlei Gefühle verfügen.

Ursprünglich bezog sich der Begriff der künstlichen Intelligenz auf Computer allgemein. Heutzutage wird der Begriff jedoch häufig in einem engeren Sinn zur Bezeichnung neuronaler Netze verwendet, die den Neuronenverbänden im menschlichen Gehirn nachempfunden sind. Hierbei handelt es sich eigentlich um eine Methode der KI, die auch unter dem Begriff des maschinellen Lernens firmiert, jedoch oft als Synonym für künstliche Intelligenz allgemein steht. Im Wesentlichen fußt das Verfahren darauf, mit Hilfe neuronaler Netze Muster in großen Datenmengen zu erkennen. Obwohl künstliche neuronale Netze mutmaßlich anders lernen als

Menschen, ist es dennoch aufgrund ihrer hohen kognitiven Leistungsfähigkeit gerechtfertigt, von künstlicher Intelligenz im schwachen Sinn zu sprechen. Allerdings handelt es sich nicht um eine allgemeine Intelligenz im menschlichen Sinn, die auf die unterschiedlichsten Problemfelder anwendbar ist, wie sie die starke künstliche Intelligenz anstrebt.

# Von der künstlichen Intelligenz zur emotionalen künstlichen Intelligenz

Der Begriff der emotionalen künstlichen Intelligenz (in Englisch auch *affective computing*) wurde in den 1990er Jahren im Zuge der Debatte um die emotionale Intelligenz geprägt. **3** Sie lässt sich analog zur künstlichen Intelligenz als diejenige Disziplin definieren, die es mit der Nachbildung oder Simulation von Emotionen zu tun hat, also damit, ob und wie künstliche Systeme Emotionen erkennen, ausdrücken und vielleicht sogar selbst aufweisen können.

Die Notwendigkeit emotionaler KI erwuchs aus der Einsicht, dass Emotionen eine wichtige Rolle für intelligentes Verhalten spielen. So gibt es neurowissenschaftliche Evidenzen dafür, dass emotionale Störungen aufgrund von Hirnverletzungen zu irrationalem Verhalten führen. **4** Das kann selbst dann vorkommen, wenn die kognitiven Fähigkeiten im engeren Sinn, die sich durch einen klassischen IQ-Test messen lassen, nicht beeinträchtigt wurden.

Auch theoretisch ist die emotionale KI von großem Interesse, da sich die Kognitionswissenschaften erhoffen, mit Hilfe künstlicher Systeme besser zu verstehen, wie der Zusammenhang zwischen Emotionen und Kognition bei

Menschen beschaffen ist. Zwar spielt die Implementation von Emotionen in künstlichen Systemen in Anwendungskontexten bislang noch keine große Rolle, doch könnte sich dies ändern, sollte sich die Hypothese bestätigen, dass Emotionen entscheidend für intelligentes Verhalten sind. **5**

Von besonderer Bedeutung sind Emotionen für soziale Kommunikation und Kooperation. Dies führt zu einem weiteren Motiv, emotionale KI zu betreiben. Da das menschliche Sozialverhalten stark auf Emotionen beruht, ist es naheliegend, dass auch künstliche Systeme Emotionen berücksichtigen müssen, sofern sie mit Menschen zu tun haben. Wenn Mensch und Maschine sozial miteinander interagieren sollen, kommt es darauf an, dass künstliche Systeme menschliche Emotionen zuverlässig erkennen, einordnen und sozial angemessen darauf reagieren können. Sie brauchen jedoch nicht zwangsläufig selbst über Emotionen zu verfügen, um diese Ziele zu erreichen.

In der sozialen Interaktion spielen auch Faktoren wie die visuelle Erscheinung, Mimik und Gestik eine große Rolle. Deshalb liegt es nahe, auf Roboter zu setzen, die über einen Körper, Sensoren und Aktoren verfügen. Zwar muss insbesondere emotionale KI im schwachen Sinn nicht zwangsläufig auf die gleichen Prozesse zurückgreifen, die menschlichen Emotionen zugrunde liegen. Dennoch ist eine gewisse Vertrautheit mit den Emotionstheorien für

biologische Lebewesen für das Verständnis der emotionalen  
KI hilfreich.

# Was sind Emotionen?

Gehen wir von unserem Alltagsverständnis von Emotionen aus, so scheint es sich um *episodische Reaktionen* zu handeln, denen wir weitgehend passiv unterliegen, etwa wenn wir von Panik ergriffen, vor Freude überwältigt oder von Schuldgefühlen erdrückt werden. Mit *episodisch* ist gemeint, dass sich Anfang und Ende einer Emotion relativ eindeutig bestimmen lassen.

Außerdem umfassen Emotionen zumeist Formen *körperlicher Erregung*. Dazu gehören etwa erhöhter Puls, Erröten, Schweißbildung oder Tränenfluss. Im Unterschied zu anderen körperlichen Vorgängen wie etwa der Verdauung sind diese körperlichen Erregungszustände typischerweise mit einer bestimmten *subjektiven Erlebnisqualität* verbunden. Es fühlt sich auf eine bestimmte Art und Weise an, Freude zu empfinden, traurig oder wütend zu sein. Diese subjektive Erlebnisqualität wird in der philosophischen Debatte auch als *phänomenales Bewusstsein* bezeichnet und ist nicht zwangsläufig an das Auftreten körperlicher Erregung gebunden.

Außerdem beziehen sich Emotionen auf ein Objekt. Man kann fragen: Worüber ist Anne wütend? Wovor fürchtet sich Julian? Worüber freut sich Paul? Die Antworten könnten beispielsweise lauten: Anne ist wütend auf Franz,

weil er ihren Stift absichtlich zerbrochen hat. Julian fürchtet sich vor dem Hund. Paul freut sich über das Buch, das Marie ihm zum Geburtstag geschenkt hat. Eine solche Bezugnahme auf ein Objekt wird in der philosophischen Emotionstheorie als *Intentionalität* bezeichnet.

Der Begriff der Intentionalität wird manchmal auch im Sinn von Absichtlichkeit verwendet. Doch damit hat der Objektbezug von Emotionen nichts zu tun. Der weitgehend passive Charakter von Emotionen schließt aus, dass wir sie absichtlich herbeiführen. Wir können nicht direkt beabsichtigen, uns zu freuen, wenngleich wir natürlich absichtlich nach Objekten suchen können, die Freude in uns auslösen.

Die Bezugsobjekte von Emotionen bezeichnet man als *intentionale Objekte*. Diese Objekte müssen nicht immer konkrete materielle Gegenstände sein: Man kann auch über die Klimakrise besorgt sein oder eine Idee bewundern. Wichtig ist jedoch, dass die Bezugsobjekte von Emotionen auf eine ganz bestimmte Art und Weise gegeben sind.

Nehmen wir an, Julia ist traurig darüber, dass Deutschland bei der Fußballweltmeisterschaft in der Vorrunde ausgeschieden ist. Das Objekt ihrer Emotion ist nicht nur das Ausscheiden der Nationalelf als solcher. So könnte ein Anhänger einer anderen Mannschaft über dasselbe Ereignis hoch erfreut sein. Sie erachtet es vielmehr auf eine ganz besondere Art und Weise als schlecht, nämlich im Sinn eines Verlusts. Ihre Traurigkeit

beinhaltet also eine negative *Bewertung* des Objekts als Verlust. Andere negative Bewertungen wären etwa Peinlichkeit (dann würde sie Scham empfinden), Kläglichkeit (dann würde sie Verachtung empfinden) oder Anstößigkeit (dann würde sie Ärger empfinden).

Die Art des Bezugsobjekts und seines Gegebenseins ist wichtig, um verschiedene Emotionen voneinander unterscheiden zu können. Die subjektive Empfindungsqualität allein reicht dafür nicht aus. So lässt sich der Unterschied zwischen der Scham und der Trauer über das Ausscheiden der Fußballmannschaft nicht allein anhand der subjektiven Erlebnisqualität ausmachen. Der Bezug auf ein intentionales Objekt ist auch verantwortlich dafür, dass Emotionen begründet oder unbegründet sein können. So wäre Julians Furcht vor dem Hund unbegründet, wenn dieser nicht gefährlich, sondern ganz sanftmütig ist.

Weiterhin sind Emotionen mit bestimmten Ausdrucksformen verbunden. Sie manifestieren sich in Sprache, Haltung, Gestik und Mimik und bewirken häufig bestimmte Verhaltensweisen. So führt Furcht zu Angriff, Flucht oder Erstarrung. Emotionen lassen sich anhand dieser Merkmale gut von anderen affektiven Phänomenen abgrenzen.

Stimmungen haben ebenso wie Emotionen eine subjektive Erlebnisqualität. Es handelt sich jedoch nicht um episodische Reaktionen, und sie verfügen nicht über

Bezugsobjekte. Sie gehen auch nicht mit bestimmten körperlichen Veränderungen, Ausdrucksformen oder Verhaltensweisen einher. Emotionen sind auch von körperlichen Empfindungen (wie einem Prickeln oder Ziehen) zu unterscheiden. Körperliche Empfindungen sind zwar episodisch, sie haben aber typischerweise kein Bezugsobjekt und sind nicht mit spezifischen körperlichen Veränderungen, Ausdrucksformen oder Verhaltensweisen verbunden, sondern zeichnen sich vor allem durch ihre Erlebnisqualität aus.

Charakterzüge wie Extrovertiertheit, Offenheit oder emotionale Verletzlichkeit stellen ebenfalls keine Emotionen dar. Sie weisen keine der für Emotionen charakteristischen Merkmale auf, auch wenn sie etwas damit zu tun haben mögen, zu welchen emotionalen Reaktionen jemand in bestimmten Situationen tendiert.

Emotionstheorien unterscheiden sich darin, welche Merkmale von Emotionen sie als vorrangig betrachten und wie sie deren Wechselspiel verstehen. So kommt es für den *Behaviorismus* besonders auf die Verhaltensreaktionen an. Demnach wäre beispielsweise Furcht gleichzusetzen mit der Tendenz zu Flucht, Erstarrung oder Kampf. *Somatische Theorien* hingegen halten Emotionen für Wahrnehmungen körperlicher Veränderungen. Der Philosoph und Psychologe William James, auf den dieser Ansatz zurückgeht, prägte hierfür Ende des 19. Jahrhunderts den Slogan: Wir weinen