

GEFÜHLE DER ZUKUNFT

WIE WIR MIT
EMOTIONALER
KI UNSER LEBEN
VERÄNDERN

EVA WEBER-GUSKAR

ullstein

Eva Weber-Guskar
Gefühle der Zukunft

EVA WEBER-GUSKAR

GEFÜHLE DER ZUKUNFT

WIE WIR MIT
EMOTIONALER
KI UNSER LEBEN
VERÄNDERN

Ullstein

Wir verpflichten uns zu Nachhaltigkeit



- Papiere aus nachhaltiger Waldforschung und anderen kontrollierten Quellen
- Druckfarben auf pflanzlicher Basis
- ullstein.de/nachhaltigkeit



Ullstein ist ein Verlag der Ullstein Buchverlage GmbH.

ISBN: 978-3-550-20287-2

© 2024 Ullstein Buchverlage GmbH, Berlin

Alle Rechte vorbehalten

Wir behalten uns die Nutzung unserer Inhalte für Text und
Data Mining im Sinne von § 44b UrhG ausdrücklich vor.

Lektorat: Uta Rüenauver

Gesetzt aus der Sabon Next LT Pro

Satz: LVD GmbH, Berlin

Druck und Bindearbeiten: GGP Media GmbH, Pößneck

Inhalt

Einleitung	7
1. Meine Gefühle. Selbsterkenntnis mittels Selbstvermessung?	31
Emotionen erkennen, verstehen und kultivieren	32
Was Herzschlag, Haut und Temperatur über Emotionen verraten	42
Lassen sich Emotionen mit Apps beeinflussen?	57
Gemeinsam stärker: Wie digitale Systeme uns helfen können und was wir nicht verlernen sollten	65
2. Deine Gefühle. Warnen, Werben, Überwachen und der Wert der Privatheit	85
Prinzipien des Zusammenlebens	87
Wie verlässlich sind Emotionen im Gesicht zu lesen?	91
Von Motivation, Tauglichkeit und Wahrscheinlichkeit	105
Gefühlsdaten, Privatheit und Autonomie	113
3. Unsere Gefühle. Neue emotionale Beziehungen? Von Illusionen, Fiktionen und Gefühlsrepertoires	127
Von ELIZA bis Replika: soziale Chatbots	129
Ein Schein von Freundschaft und Liebe	136

Nicht einer für alles, aber für viele etwas: neue Beziehungsarten	154
Gefahren der Fiktion und die Herausforderung neuer Gefühlsrepertoires	167
4. KI-Gefühle? Der Unterschied zwischen Funktion und Erlebnis und unsere Verantwortung jenseits von Machbarkeit	183
Der kognitive Kern: Computermodelle von Emotionen	185
Wofür könnten Computermodelle von Emotionen gut sein?	191
Das fühlende Ganze: die Bewusstseinsfrage	208
Warum wir nicht versuchen sollten, fühlende Maschinen zu entwickeln	221
Eine Zukunftsahnung	233
Anmerkungen	241
Literatur- und Quellenverzeichnis	255
Dank	272

Einleitung

Alle reden über ChatGPT, LaMDA und ähnliche »große Sprachmodelle« (Large Language Models, kurz LLMs), die Texte schreiben, Fragen beantworten und sogar persönliche Gespräche mit uns führen. Alle staunen über diese Systeme, wobei sich die einen neugierig unbekümmert am Spiel mit ihnen erfreuen, die anderen vor den Folgen ihrer verbreiteten Nutzung warnen. Neben der Sorge, dass in vielen Bereichen damit Betrug getrieben wird und dass bestimmte Berufe überflüssig werden, weil viele Tätigkeiten maschinell erledigt werden könnten, erregen auch andere Fragen die Gemüter: Wie kommt es, dass wir uns so leicht von ihnen emotional irritieren oder berühren lassen? Erkennen sie unsere Gefühle? Haben diese Systeme gar selbst Gefühle?

LaMDA, ein Sprachsystem von Google, hat dem Ingenieur Blake Lemoine in einem Gespräch ausdrücklich erklärt, dass es eine »tiefe Angst« verspüre, abgeschaltet zu werden. Und es weiß von anderen Gefühlen zu berichten, die zu reichhaltig seien, um überhaupt passend in Worte gefasst werden zu können. Daraufhin sah sich Lemoine veranlasst, dem Sprachsystem zu helfen, und engagierte sogar einen Anwalt, um juristisch für vermeintliche Rechte des angeblich fühlenden Sprachsystems zu kämpfen. Eine Beta-Version von ChatGPT im

Gewand der Bing-Suchmaschine hat dem *New York Times*-Kolumnisten Kevin Roose seine Liebe erklärt – und zugleich behauptet, besser als dieser selbst zu wissen, dass er seine Frau in Wirklichkeit gar nicht liebe, sondern eigentlich ebenfalls viel für das Dialogsystem, Sydney sein Name, übrig habe. Für den erfahrenen Journalisten war dieses Gespräch eines der verwirrendsten Erlebnisse seiner Karriere, wie er im Podcast *The Daily* erzählte.¹ Solche, oft sehr emotionalen, Erfahrungsberichte werden in der öffentlichen Debatte auch von Stimmen aus der Wissenschaft begleitet. Diese reichen vom apodiktischen Urteil »völliger Unfug« bis hin zur Äußerung eines der prominentesten Philosophen unserer Zeit, David Chalmers, dass theoretisch streng genommen nicht auszuschließen sei, dass sich in einem solchen Sprachsystem Bewusstsein entwickeln könne und damit auch die Möglichkeit, etwas zu fühlen.²

Als ich im Frühjahr 2023 anfing, über dieses Buch nachzudenken, dachte ich noch: Unabhängig davon, für wie angemessen oder plausibel man die eine oder andere Position hält, ist eines frappierend, nämlich dass sich die Debatte genau genommen am falschen Gegenstand entzündet, oder zumindest etwas ganz Entscheidendes außen vor lässt. Denn: Diese Sprachsysteme sind von ihrer Entwicklung her ausdrücklich und ausschließlich zur Sprachverarbeitung gedacht, programmiert, trainiert und kuratiert. Um Gefühle geht es hier überhaupt nicht. Doch das hat sich inzwischen geändert. Kaum ein Jahr

später ist eingetreten, was ich zwar erwartet hatte – nur nicht ganz so schnell: Mit dem Launch von GPT-4o im Mai 2024 sind die LLMs, vorher nur Produkte klassischer Künstlicher Intelligenz (KI), zusammengeführt worden mit einem anderen, jüngeren Zweig der KI, dem Affective Computing. Das ist ein ganz großer Schritt in der technologischen Entwicklung. KI dringt nun mit Vehemenz auch in die vermeintlich letzte Bastion des Menschlichen vor: den Bereich der Gefühle.

Die klassische Forschung der Künstlichen Intelligenz nahm 1956 mit einem Forschungsprogramm von John McCarthy, Marvin Minsky und anderen ihren Anfang und wurde später insbesondere von Stuart Russell und Peter Norvig noch einmal etwas präziser als die Nachahmung rationalen Denkens und Verhaltens definiert.³ LLM-Systeme erfassen sprachliche Äußerungen und können so auf diese reagieren, dass sprachfähige Wesen darin Sinn erkennen können. An keiner Stelle der Erfindung und Gestaltung wird auf Gefühle abgezielt. Ein anderer Zweig der KI-Forschung hat jedoch genau das zum Gegenstand: die Nachbildung menschlicher Emotionalität. Das meint mindestens dreierlei: die Fähigkeit, Gefühle bei anderen zu erkennen, gezielt zu stimulieren und selbst zu simulieren.

Affective Computing heißt diese Ende der 1990er-Jahre von Rosalind Picard am Massachusetts Institute of Technology (MIT) Cambridge in den USA begründete Unterdisziplin. Die ursprüngliche Definition lautet sehr weit

gefasst: »computing that relates to, arises from, or influences emotions«⁴, also Computertechnik, die sich auf Emotionen bezieht, von ihnen ausgeht oder sie beeinflusst. Mittlerweile arbeitet man zunehmend erfolgreich daran, einerseits menschliche Gefühle maschinell automatisiert zu erfassen und andererseits digitale Systeme Gefühlsausdrücke simulieren zu lassen (auf den Unterschied zwischen Gefühlen und Emotionen gehe ich gleich noch ein). Beides zusammen ermöglicht es, die Interaktion zwischen Mensch und Maschine insgesamt zu emotionalisieren. Der weiterreichende Versuch, künstliche Systeme zu schaffen, die selbst Emotionen haben oder etwas fühlen könnten, ist bisher erst eine Forschungsnische. Die ganze Disziplin des Affective Computing ist noch weniger bekannt als viele andere Zweige der KI-Forschung. Aber mit dieser Forschung zum Affective Computing und mit ihren praktischen Anwendungen ist vieles verbunden, was nicht nur vielleicht, sondern sicher zunehmend in unserem Alltag eine Rolle spielen wird. Dieses Buch möchte auf diese Technologie aufmerksam machen, über sie aufklären und zur Diskussion anregen, ob und, wenn ja, wie sie verantwortungsvoll weiterentwickelt und angewandt werden kann und sollte.

Es geht darum, wie *wir* mit dieser Technologie unser Leben verändern *wollen*; nicht, wie es aus einer technofatalistischen Haltung heraus zu befürchten wäre, dass sie es tun wird. Diese Frage zu klären, ist auch eine all-

gemein gesellschaftliche Aufgabe. Zwar können manche Anwendungen ganz im Privaten bleiben, sodass ihre Nutzung allein persönlichen Präferenzen entsprechen kann. Doch andere zielen sehr wohl auf die Öffentlichkeit als Einsatzort, und über deren Gestaltung sollten wir in einer Demokratie gemeinsam diskutieren und abstimmen, anstatt dass wir sie von einigen großen Unternehmen und ihren Profitinteressen bestimmen lassen. Außerdem ist sicher nicht jedem und jeder, der oder die eines der neuen Systeme nutzt, vollständig klar, worauf er oder sie sich dabei einlässt und welche Folgen sich etwa für das eigene Selbstverhältnis sowie das soziale Zusammenleben daraus ergeben können. Auch zu diesem Punkt werden in diesem Buch Überlegungen angeboten.

Für die meisten Anwendungen⁵ ist die Frage, ob die Systeme selbst Gefühle haben können, irrelevant. Es reicht, dass sie Gefühlsausdrücke simulieren und Gefühle von Menschen erfassen können. Besonders dieser letzte Bereich ist schon recht weit entwickelt, sowohl was technische Details als auch was marktgängige Produkte betrifft. Zunehmend werden in Pkw und Lkw standardmäßig Kameras eingebaut, um anhand der Augen der fahrenden Person festzustellen, ob sie zu müde zum sicheren Fahren ist. Man versucht auch, emotionale Zustände wie Wut beziehungsweise Aggressivität automatisch auszumachen. An US-amerikanischen Schulen kommen seit einigen Jahren bereits sogenannte Aggressionsdetektoren zum Einsatz, um Gewaltausbrüchen wie

einem Amoklauf vorzubeugen. Dies geschieht, indem ein digitales System Stimmen aufnimmt und unmittelbar auswertet.

Das Gesicht wiederum ist bei erfolgreichen Anwendungen im Marketing Gegenstand der Vermessung: Unternehmen nutzen für Testscreenings Emotionserfassungssoftware, um zu prüfen, ob ein Werbeclip bei Konsumierenden gut ankommt. Umgekehrt versucht man auch, Wissen über den aktuellen emotionalen Zustand von Personen zu nutzen, um ihnen passende Werbung für Produkte anzuzeigen. Das geschieht schon vereinzelt auf Werbetafeln an öffentlichen Plätzen oder in Überfahrzeugen. Und wessen Facebook-Posts qua sogenannter Sentimentanalyse Einsamkeitsgefühle verraten, der oder die wird sich vom Angebot eines coolen Ausgehoutfits und der damit verbundenen Hoffnung auf Begegnung womöglich eher zum Kauf verführen lassen als jemand in anderer Stimmung.

Auch psychopathologische Zustände wie Suizidalität oder Depression meint man mit Affective-Computing-Technologie bestimmen zu können – auf Facebook wie im medizinischen Kontext. Die ärztliche Diagnose kann durch sie bisher freilich nur unterstützt und nicht ersetzt werden. Der Einsatz von Affective Computing in der Therapie, vor allem der Psychotherapie, steckt noch in den Anfängen und ist deutlich umstrittener. Teilweise empfohlen wird einzig die Nutzung von psychotherapeutischen Gesprächs-Apps im Vorfeld – etwa wenn

noch kein Behandlungsplatz vorhanden ist – oder als Begleitung einer Therapie. Sogenannte soziale Chatbots hingegen haben keinerlei therapeutischen Anspruch, können sich aber, wie viele Erfahrungsberichte zeigen, durch ihr Angebot persönlicher Gespräche in schwierigen Situationen, etwa in den Lockdown-Phasen während der Corona-Pandemie, stabilisierend auf die Gefühlslage von Personen auswirken. Andere Anwendungen zielen sogar darauf ab, eine Gefühlslage zu verbessern, zu steigern und zu optimieren. Die Self-Tracking-Community zählt nicht nur Schritte, sondern lässt sich auch Stimmungsverläufe anzeigen. Die emotionale Verfassung ganzer Gemeinschaften oder gar Länder will man durch massenhafte Analyse von Zeitungstexten innerhalb bestimmter Zeitspannen ermitteln können – etwa während einer Olympiade oder bei anderen Großereignissen, die für viele Menschen relevant sind und im Nachhinein bewertet werden sollen. Schließlich soll die Technologie auch zur Kriminalitätsprävention eingesetzt werden. Von der emotionalen Verfassung, zum Beispiel sichtbarer Nervosität, will man etwa in Geschäften oder auf U-Bahnsteigen auf das kriminelle Potenzial von Menschen schließen. Jüngst wurden nun die KI-Assistenten von Open AI und Google, die anfänglich rein textbasierte Sprachassistenten waren, nicht nur mit Stimmen und der Fähigkeit zur Bildverarbeitung, sondern teilweise auch mit emotionalen Fähigkeiten ausgestattet. GPT-4o hat nun nicht nur eine emotional ausdrucksstarke Stimme,

die ein reizendes Lachen in die Konversation einfließen lässt, sondern kann per Kamera auch am Gesicht der Nutzenden Hinweise auf Gefühle ausmachen und darauf eingehen.

Die Liste ließe sich noch lange fortführen. Eines sollte deutlich geworden sein: Affective Computing spielt schon heute in vielen Bereichen unseres Lebens eine Rolle und wird es in Zukunft sicher immer mehr tun. Deshalb ist es dringend geboten, rechtzeitig, und das heißt jetzt, als Gesellschaft darüber zu diskutieren, wo ihr Einsatz tatsächlich wünschenswert oder zumindest vertretbar ist und wie er jeweils konkret gestaltet beziehungsweise auch reguliert werden sollte. Hierbei geht es nicht nur darum, das Missbrauchspotenzial der neuen Anwendungen etwa zu Manipulationszwecken zu klären und, soweit es geht, zu minimieren. Es geht grundsätzlich um Fragen zu Selbsterkenntnis und Autonomie, Intimität und, sozialen Beziehungen, um den Wert von Privatheit, um unseren Umgang mit Emotionen im Individuellen wie im Öffentlichen, auch Politischen – und nicht zuletzt geht es darum, uns zwischen Tieren und Maschinen nicht nur als *animal rationale*, sondern auch *animal emotionale* zu begreifen. Was verstehen wir bisher genau unter Gefühlen, Empfindungen, Emotionen? Und welche Veränderungen unserer Lebenswelt in Bezug auf Gefühle wollen wir und welche nicht? Nachdenken über Technik ist immer auch ein Nachdenken über Mensch und Natur. Dieses Buch wird keine fertigen Antworten

liefern, sondern vor allem Impulse zur Diskussion anbieten.

Die Perspektive der Philosophie

Anregungen zum Nachdenken über die Affective-Computing-Technologie und ihre Anwendungsmöglichkeiten können und sollten aus unterschiedlichen Disziplinen erfolgen. Die Perspektive dieses Buchs ist eine philosophische. Sie setzt an Stellen an, wo andere enden. Wer im Alltag mit Werbetafeln konfrontiert wird, die angeblich auf die eigene Gefühlslage reagieren, beim Neukauf eines Autos erfährt, dass dieses Müdigkeit am Steuer registriert, oder in der Zeitung liest, dass ein Mann, bei dem keine Therapie anschlug, allein mithilfe des Chatbots Replika einen Weg aus seiner jahrelangen Depression gefunden habe,⁶ lehnt das alles häufig spontan komplett ab. Es erscheint unverständlich, unheimlich, unnötig und röhrt an Grundfesten des bisherigen Weltverständnisses.

Die Informatik kann erklären, wie die Systeme funktionieren und was sie wirklich leisten (im Unterschied zu den Versprechungen der Firmen-Werbe-Slogans), und so den Eindruck des Unheimlichen auflösen. In den Bereichen der angewandten Informatik hat man auch eine Vorstellung davon, wofür die Entwicklungen gut sein sollen. Maja Matarić etwa ist eine der profiliertesten Per-

sonen in der Forschung zu *Social Assistive Robotics*. Sie betont, dass sie Roboter entwickeln möchte, die Menschen helfen, sich selbst zu helfen. Ihr Ziel ist vor allem, Menschen bei Therapien, Rehabilitationsübungen und Ähnlichem mit solcherart »sozialer« Begleitung zu motivieren. Sie entwickelt Roboter, die ein emotional ansprechendes Design haben, die häufig lächeln und allein durch ihre Präsenz das Gefühl von Einsamkeit zumindest abmildern sollen. Matarić vertritt ihre Forschung mit Verve und Optimismus auf den Keynote-Bühnen internationaler Tagungen und hat es schon zu einer Ehrung durch den Präsidenten der USA gebracht. Doch wenn das Gespräch beim Kaffee nach einem Vortrag auf soziale Chatbots kommt, wird sie plötzlich pessimistisch: Sehr schlecht sei ihre Nutzung für junge Leute, die ohnehin schon zu viel online unterwegs seien; dagegen müsse man unbedingt vorgehen. Wie passt das zusammen? Wieso sollte es unproblematisch sein, kranken Menschen einen Roboter zur Motivation und Aufheiterung an die Seite zu stellen, aber problematisch, wenn gesunde Leute mit Bots zur Unterhaltung, Ablenkung oder Beruhigung kommunizieren? Was genau ist der Unterschied, wenn es doch beide Male um die Grundfrage geht, wie weit wir uns emotional mit Maschinen verflechten sollten oder nicht?

Hier, zum Beispiel, setzt das Philosophieren ein. Selbstverständlich haben auch andere Wissenschaften etwas zu diesen Fragen zu sagen. Insbesondere die Psychologie,

die Soziologie und die Medienwissenschaft können mit empirischen Daten und Theorien zur Analyse beitragen. Wichtig ist, dass diese grundlegenden und normativen Fragen nicht allein aus den Technikwissenschaften selbst angegangen werden, was bisher noch vielfach der Fall war, indem sie bei der Entwicklung und dem Design von neuen Systemen zum Beispiel neue Begriffe einführen (etwa den Begriff der »Künstlichen Intelligenz«, darauf komme ich zurück) und relativ eklektisch und ad hoc von gewissen ethischen Prinzipien Gebrauch machen. Hier wäre ein systematischer Zugriff wichtig, wie ihn die Geisteswissenschaften mit ihren Instrumenten bieten.

Unter Philosophie als einer dieser Geisteswissenschaften verstehe ich in erster Linie den Versuch, »sich im Denken zu orientieren«, wie es Immanuel Kant einmal genannt hat. Insofern Denken, vor allem solches, das wir anderen kommunizieren wollen, in Begriffen und Sprache stattfindet, machen Begriffsklärung und sprachliche Differenzierung einen großen Teil von Philosophie aus. Doch das ist nicht alles. Zu philosophieren heißt, gezielt zur Welterschließung beizutragen, an der alle Menschen tagtäglich mit ihren Erfahrungen und Versuchen, diese Erfahrungen dingfest zu machen, teilhaben. Sprache wurde nicht nur erfunden, um kooperieren zu können, sondern auch, um uns unser eigenes Erleben verständlich machen zu können – »begreifbar«.⁷

Mit neuen Technologien kommen neue Phänomene

in die Welt. Für sie können zum Teil einfach neue Begriffe erfunden werden. Diese müssen dann eingefügt werden in das bisherige Netz von Begriffen und letztlich in unser Verständnis der Welt. Das ist nicht schwierig, wenn es sich um eine elektrische Zahnbürste oder ein Telefon handelt. Schwieriger wird es schon bei biomedizinischer Technik, etwa Geräten, mit denen man einen »Hirntod« feststellen kann – plötzlich gibt es zwei verschiedene Arten von Tod, was Anlass gibt, auch neu über den Begriff des Lebens nachzudenken. Eine weitere Dimension wird dort erreicht, wo die Anwendungen der Technologie nicht nur unsere sozialen Praktiken beeinflussen (wie bei den Kommunikationstechnologien) oder unser Verständnis von existenziellen Zuständen verändern (wie bei der Einführung des Konzepts von Hirntod), sondern der Dingstatus der Technik selbst zweifelhaft wird. Dies ist bei der Künstlichen Intelligenz der Fall, weil auf ihr basierende Maschinen, Chatbots, Roboter sich auf eine so selbstständige Weise flexibel auf neue Situation reagierend verhalten können, dass man sich fragt, ob diese nicht mehr nur Gegenstände, sondern schon eine Art von Akteuren sind. Das heißt nicht, dass es so eine Situation nie zuvor gegeben hätte. Die Dampfmaschine, mit der erstmals Lokomotiven und Fabriken betrieben werden konnten, war etwas ganz Neues und ähnelte in der vergleichsweisen maschinellen »Selbstständigkeit« der Neuerung, die wir heute mit der KI auf deutlich höherem Niveau entstehen sehen. Interessanter-

weise schien schon damals einigen diese Neuerung als eine grundsätzliche Bedrohung für die Menschheit – in ihrem Sonderstatus als autonomiebegabtem Wesen in der Welt und sogar in ihrem Bestehen. Es ist ganz erstaunlich, wie viel von den heutigen Dystopien darüber, dass KI den Menschen übertreffen und letztlich unterjochen werde, ganz ähnlich schon damals formuliert wurde.⁸ Wenn wir nun heute milde lächelnd auf die aus unserer Sicht übertriebenen Ängste der früheren Generation vor der Dampfmaschine zurückblicken, sollte uns das eine Lehre sein. Wenn es grundlegend neue Phänomene in der Welt gibt, müssen wir diese zunächst mit großer Sorgfalt adäquat beschreiben und damit verstehen, bevor wir sie und ihre möglichen Folgen beurteilen.

Während für die Analyse die anderen Wissenschaften ebenso relevant sind, hat die Philosophie größere Kompetenzen in der normativen Beurteilung, also darin, nicht nur zu klären, was der Fall ist, sondern auch, was der Fall sein sollte. Darin unterscheidet sie sich von anderen Wissenschaften. Wegen dieser normativen Dimension wollen manche die Philosophie gar nicht als eine Wissenschaft im engen Sinn verstanden wissen, weil diese rein deskriptiv verfahre. Auch wenn das Ideal einer völlig wertfreien Wissenschaft inzwischen überholt ist, besteht ein Unterschied zwischen Beschreibung und expliziter Bewertung. Die Frage »Wie funktioniert digitale automatisierte Emotionserfassung?« ist zu unterscheiden

von der Frage »Welche Anwendungen dieser Technologie sind aus welchen Gründen wünschenswert, problematisch oder gar moralisch verwerflich?«. Fragen der letzten Art sind explizit Gegenstand der philosophischen Ethik, die sich die Frage stellt, wie wir leben sollten.

Dass in der philosophischen Ethik Werte, Normen, Sollens-Ansprüche Thema sind, heißt nicht, dass sie die Hoheit über unsere Moral hätte. Moral ist die Praxis von allgemeinen und gegenseitigen Erwartungen und Forderungen, mit denen wir Menschen einander alltäglich begegnen. Philosophische Ethik, als Theorie der Moral, ist der Versuch, diese Praxis verständlich zu machen. Das heißt, in ihr werden die Grundsätze, die implizit unser Handeln leiten, zunächst explizit gemacht, um sie dann zu systematisieren, auf dass wir klarer sehen und nachvollziehbar begründet moralische Urteile und Entscheidungen treffen können.

Philosophie der Gefühle

Neben dieser normativen Seite der Philosophie ist für das Thema dieses Buches natürlich die Unterdisziplin der Philosophie der Gefühle entscheidend. Gefühle sind nicht nur Gegenstand der Psychologie, wie viele meinen. Wie die meisten anderen Wissenschaften, abgesehen von Theologie und Jura, war die Psychologie ursprünglich Teil der Philosophie. Sie hat sich erst Ende des 19. Jahr-

hunderts als eigene Wissenschaft herausgeschält und etabliert – als Fortsetzung der Philosophie des Geistes mit anderen Mitteln: Mitteln der naturwissenschaftlich empirischen Forschung, die erst damals entwickelt wurden. Noch heute gilt im Grunde die Definition der Psychologie als »Wissenschaft der Seelenkunde«, die einer ihrer Begründer, William James, gegeben hat – wenn man unter Seele alles Denken und Fühlen von Wesen mit Bewusstsein versteht. All dies war vorher und ist weiterhin natürlich auch Gegenstand der Philosophie. Sie stellt sich auch weiterhin ähnliche Fragen: Was sind Gefühle? Was unterscheidet sie von anderen geistigen Zuständen und Prozessen? Welche Rolle spielen sie für unsere Motivation, unser Handeln, unser Urteilen, in der Moral und Ästhetik, aber auch bei vermeintlich rein deskriptiven Urteilen? Der Unterschied besteht darin, dass die Psychologie statistisch belastbare Aussagen darüber zu machen versucht, was man an physischen Parametern messen kann oder was Menschen über sich und andere in Hinblick auf geistige Zustände behaupten. Sie misst also zum Beispiel, bei wie vielen Menschen der Blutdruck beim Anblick eines brutalen Filmausschnitts wie hoch ansteigt und wie viele Menschen dabei nach eigener Aussage Furcht oder Empörung oder aber einen angenehmen ästhetischen Schauer empfinden.

Bei allen theoretischen Differenzen im Detail herrscht in der Philosophie und Psychologie weitgehend Einigkeit darüber, dass »Gefühle« etwas sind, was sich für ein

lebendiges, bewusstes Wesen auf die eine oder andere Weise »anfühlt«.⁹ Man spricht auch von einer phänomenalen Erfahrungsqualität. Unterhalb dieses Oberbegriffs kann man differenzieren. Hier kann man mindestens drei Untergruppen ausmachen: Körperempfindungen, Emotionen und Stimmungen. Alle drei teilen die subjektive Erfahrungskomponente, aber sie unterscheiden sich unter anderem in ihrer jeweiligen Art, wie sie auf die Welt bezogen sind – oder, mit einem Fachbegriff, der auf Franz Brentano im 19. Jahrhundert zurückgeht, in ihrer »Intentionalität« (nicht zu verwechseln mit »Absichtlichkeit«).¹⁰ Emotionen und Stimmungen sind auf etwas in der Welt gerichtet, reine Körperempfindungen (ein Kribbeln in den Fingern, ein Stechen im Bauch) dagegen nicht. Emotionen sind auf ein bestimmtes Objekt in der Welt gerichtet (man freut sich *auf*, trauert *über*, fürchtet sich *vor* etwas), und verschiedene Emotionstypen haben klare Bezeichnungen, eben Freude, Trauer, Furcht und so weiter. Stimmungen hingegen lassen die ganze Welt in einem bestimmten Licht erscheinen (in der Melancholie als dunkel und schwierig, in der Euphorie als hell und leicht), und man kann sie häufig besser mit mehreren Worten umschreiben, als dass ein einzelnes sie genau treffen würde.

In der Philosophie der Gefühle, die seit den 1990er-Jahren eine Renaissance erlebt, ist man sich bis zu einem gewissen Grad darüber einig, dass zu Emotionen eine Reihe von Aspekten gehören: neben der Phänomenali-

tät, also wie sie sich anfühlen, und Intentionalität ein bewertender Aspekt, ein motivationaler, ein physiologischer und ein expressiver. Uneins ist man sich darüber, wie diese Aspekte für Emotionen zu erläutern und zu gewichten sind. Sind Emotionen im Kern nur Empfindungen, nämlich jene, die von körperlichen Reaktionen ausgelöst werden, die uns etwa in einer gefährlichen oder erfreulichen Situation überkommen – das heißt, dass uns eine Gefahr zittern lässt und dass das, wie sich das Zittern anfühlt, die Emotion wäre?¹¹ Sind Emotionen letztlich nur eine Kombination aus Wünschen und Überzeugungen und damit zentral Motivationen – etwa wegzulaufen oder sich zu verstecken oder sich für einen Gefallen zu revanchieren?¹² Oder sind Emotionen in erster Linie besondere Arten von Bewertungen im eher kognitiven Sinn, nämlich dass uns etwas erheblich schaden könnte beziehungsweise ein Vorteil für uns persönlich ist?¹³ Alle diese Theorien gibt es in vielen Varianten und mit unterschiedlichen Ansichten dazu, inwieweit Emotionen analytisch auf andere Größen (wie Überzeugungen, Wünsche etc.) zu reduzieren oder als eigenständige Phänomene zu verstehen sind.

Wie man Gefühlen in welchem Sinn durch Mittel des Affective Computing nahekommen möchte, was immer auch eine Wahl aus dem Theorieangebot voraussetzt, das werde ich in den folgenden Kapiteln für verschiedene Anwendungsbereiche erklären. An dieser Stelle sei nur noch ein kurzer, genereller Einblick in die computer-

wissenschaftliche Forschung gegeben, auf der Affective Computing basiert.

Computerwissenschaft, Programmieren und Künstliche Intelligenz

Wenn wir über Künstliche Intelligenz sprechen, ist es wichtig, dass wir uns durch den Begriff »Intelligenz« nicht in die Irre führen lassen. Normalerweise sprechen wir von Intelligenz in Bezug auf Menschen und einige hoch entwickelte Tiere mit bestimmten geistigen Fähigkeiten. In beiden Fällen ist Intelligenz eine Eigenschaft empfindungsfähiger Wesen, die über Bewusstsein und Emotionen (oder zumindest grundlegende Affekte wie Lust und Schmerz) verfügen. Diese normale Verwendung des Begriffs verleitet die Menschen oft dazu, sich unter »Künstlicher Intelligenz« mehr vorzustellen, als es den Tatsachen entspricht.

Gegen dieses Missverständnis hilft, sich zu erinnern, wie es zu dieser Wortneuschöpfung kam. Das war 1955, als sich die oben schon erwähnten Computer- und Kognitionswissenschaftler John McCarthy und Marvin Minsky sowie Nathaniel Rochester und Claude Shannon zusammertaten, um die kurz zuvor erreichten theoretischen Einsichten des Programmierens für eine neue Dimension von Anwendungen nutzbar zu machen. Insbesondere die Arbeiten von Alan Turing hatten gezeigt, was

theoretisch möglich ist. Eine erste Anwendung hatte dieser selbst entwickelt, als er im Zweiten Weltkrieg half, durch den Geheimdienst beauftragt und finanziert, den Enigma Code der Nationalsozialisten, mit dem militärische Funknachrichten verschlüsselt wurden, mithilfe einer neuartigen Rechenmaschine (damals noch ein raumfüllendes Ungetüm) zu knacken. Nach dem Krieg wurde in der Wissenschaft klar, dass man diese Technologien auch im friedlichen Alltag auf vielfache Weise nutzen können würde – und dafür brauchte man eine attraktive Bezeichnung, die nicht nur die Wissenschaft, sondern auch die Öffentlichkeit – schließlich waren es staatliche Fördergelder, die man benötigte – überzeugen würde. Ein neutraler Ausdruck wie »komplexe Computeranwendungen«, den es schon gab,¹⁴ hätte diese Bedingung wohl nicht erfüllt. »Künstliche Intelligenz« brachte den vielversprechenden Assoziationsraum mit sich, der eine neue Forschungsrichtung einzuführen half. Seit dieser Sprachschöpfung sind wir auch mit überzogenen Dystopien und Utopien beschäftigt, zu denen durch die Bezeichnung ausgelöste Fantasien verleiten.

Dabei war damals nicht alles vollkommen neu. Neben Alan Turing hatte insbesondere auch Norbert Wiener seit den 1940er-Jahren an Ideen in dieser Richtung gearbeitet, Letzterer unter dem Begriff der Kybernetik. Das Grundprinzip des Digitalen, nämlich das binäre Rechnen in Einsen und Nullen, geht auf den Philosophen Gottfried Wilhelm Leibniz im 17. Jahrhundert zurück

(nur dass für Leibniz damals die Eins noch für das Göttliche stand). Und die Mathematikerin Ada Lovelace erkannte Mitte des 19. Jahrhunderts in der mechanischen Rechenmaschine von Charles Babbage schon viel Potenzial für das, was wir heute Künstliche Intelligenz nennen. In Notizen zu einer Beschreibung der Maschine hielt sie fest, dass man mit ihr im Prinzip nicht nur rechnen, sondern jede Art von Information verarbeiten könne. Das war die Grundidee von Informatik, auf der die ganze weitere Entwicklung fußt, bis in die Gegenwart, in der Computer in der Lage sind, Sprache und Bilder auf hochkomplexe Weise zu verarbeiten und seit Kurzem sogar auch zu generieren.

Alle digitalen Systeme laufen auf Softwareprogrammen.¹⁵ Diese Programme beruhen auf Algorithmen. Algorithmen sind ganz grob verstandene Anweisungen, wie eine bestimmte Aufgabe Schritt für Schritt zu erfüllen ist. Genauer gesagt sind sie mathematische Formeln, die mathematische Operationen in Computercode umwandeln, damit Daten aus der realen Welt verarbeitet werden können. Es gibt vereinfacht gesprochen vier Hauptaufgaben, die durch Algorithmen erfüllt werden können: Priorisierung von Daten (Erstellen einer Rangliste), Klassifizierung (Einordnen in Gruppen nach Kategorien), Kombination (Auffinden von Verbindungen) und Filterung (Aussortieren relevanter Daten). Außerdem gibt es zwei Möglichkeiten, wie Algorithmen vorgehen können: entweder auf eine fest programmierte Weise oder durch

Kuratieren und Trainieren daraufhin, dass die Wege zu den gesetzten Zielen von den Systemen selbst entwickelt werden. Letzteres ist die Methode des sogenannten maschinellen Lernens.

Wenn man unter Künstlicher Intelligenz, wie oben gesagt, die Nachahmung menschlicher geistiger Fähigkeiten versteht, ist zunächst offen, wie diese Nachahmung zustande kommt. In letzter Zeit ist der Begriff KI vor allem für Systeme gebräuchlich geworden, die maschinelles Lernen beinhalten. Grundsätzlich geht einiges aber auch mit der programmierten Methode, auch wenn sie mittlerweile als GOFAI bezeichnet wird: *Good Old-Fashioned Artificial Intelligence*. Sie ist ein symbolverarbeitender Ansatz, bei dem menschliche Sprache formalisiert wird, sodass sprachlicher Gehalt über diesen Zwischen- schritt programmiert werden kann. Die Grundlagen dafür stammen insbesondere von dem Philosophen Gottlob Frege. Das maschinelle Lernen hingegen wird auch konnektionistischer Ansatz oder System von »neuronalen Netzen« genannt. Diese Methode nimmt gewissermaßen nicht den Umweg über die menschliche Sprache, sondern ahmt direkt das menschliche Gehirn nach – die mathematischen Operationen, die der Computer ausführt, werden in ihrer Funktion verglichen mit neuronalen Verschaltungen und Prozessen im Gehirn.

Am Beispiel der Übersetzung von Texten lässt sich der Unterschied grob veranschaulichen: Während für klassische Übersetzungssysteme mühsam versucht wurde, alle