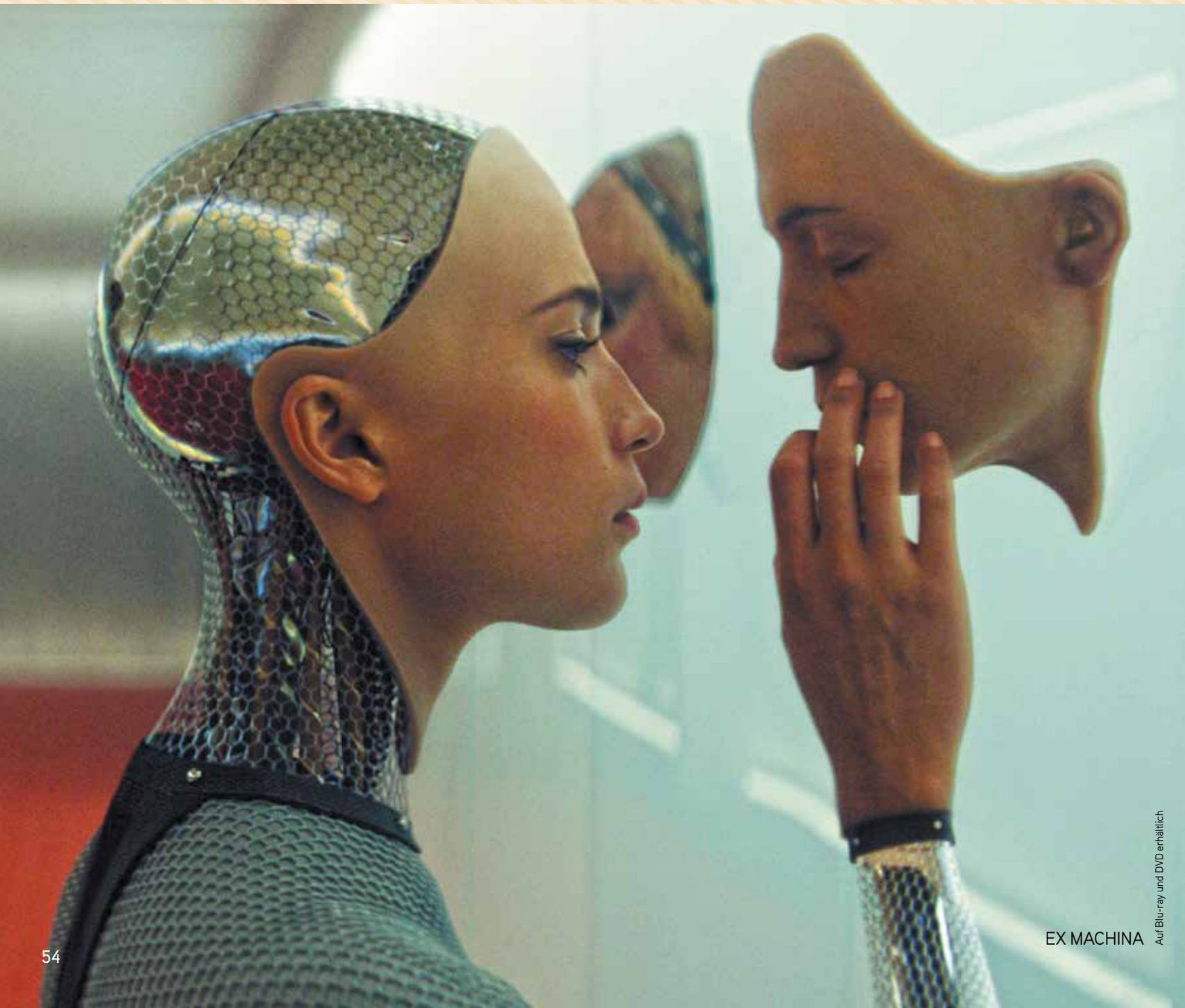
Emphatische Maschinen

KI ZWISCHEN ANGST UND HOFFNUNG

Von Björn W. Schuller *Die Emotionserkennung ist zweifellos einer der faszinierendsten Aspekte der aktuellen KI-Forschung. Die Einsatzmöglichkeiten für Maschinen, die in der Lage sind, menschliche Gefühle zu interpretieren scheinen grenzenlos. Doch wie weit sind wir in dieser Entwicklung fortgeschritten? Wie wird sich der zwischenmenschliche Umgang in Zeiten der empathischen Maschinen verändern? Neben einer Vielzahl von Anwendungen, die unsere Lebensqualität verbessern können, gibt es auch eine dunkle Seite emotional intelligenter und empathischer Maschinen. Wir müssen sicherstellen, dass KI verantwortungsvoll, vertrauenswürdig, sicher und ethisch korrekt entwickelt und eingesetzt wird.*



EX MACHINA Auf Blu-ray und DVD erhältlich

„Nutzerin, ich kann dich gut verstehen – solche Nachrichten machen mich auch immer traurig!" – in der Welt der Technologie schreiten Entwicklungen voran, die das Potenzial haben, unsere Beziehung zu Maschinen, aber auch auf zwischenmenschlicher Ebene nachhaltig zu verändern. Die Rede ist von emotional intelligenten und empathischen Maschinen – künstliche Intelligenz (KI), die in der Lage ist, menschliche Emotionen zu erkennen, zu verstehen und sogar zu reproduzieren. Aus Hollywood kennen wir das schon lange: Roboter wie R2-D2 aus dem **Star Wars**-Universum können auch trotz verbaler Limitation auf Piepgeräusche diverse Emotionen ausdrücken. Modernere Kollegen wie der aufblasbare Medizinroboter Baymax oder auch die stimmbasierte KI in Samantha im Film **Her** oder der Roboter Ava aus **Ex Machina** werden auf Grund ihrer emotionalen Intelligenz beste Freunde von Menschen: Oder Menschen verlieben sich sogar in sie – wie im deutschen Film **Ich bin dein Mensch** in den Roboter Tom. Nicht zuletzt gibt es Liebesgeschichten zwischen zwei KIs wie in Disneys **WALL-E** (mit EVE).

In der Wissenschaft hat das zugehörige Fachgebiet schon seit Mitte der 1990er Jahre einen Namen: ***Affective Computing***[1]. Erste Patente reichen schon in die 1970er Jahre zurück[2]. Aber erst mit den letzten Durchbrüchen im tiefen Lernen[3] – ein Spezialgebiet der KI, das sich mit künstlichen neuronalen Netzen beschäftigt und auch in anderen Bereichen zum Durchbruch verhalf – werden Maschinen zunehmend reif für den Einsatz solcher „künstlichen emotionalen Intelligenz" im Alltag.

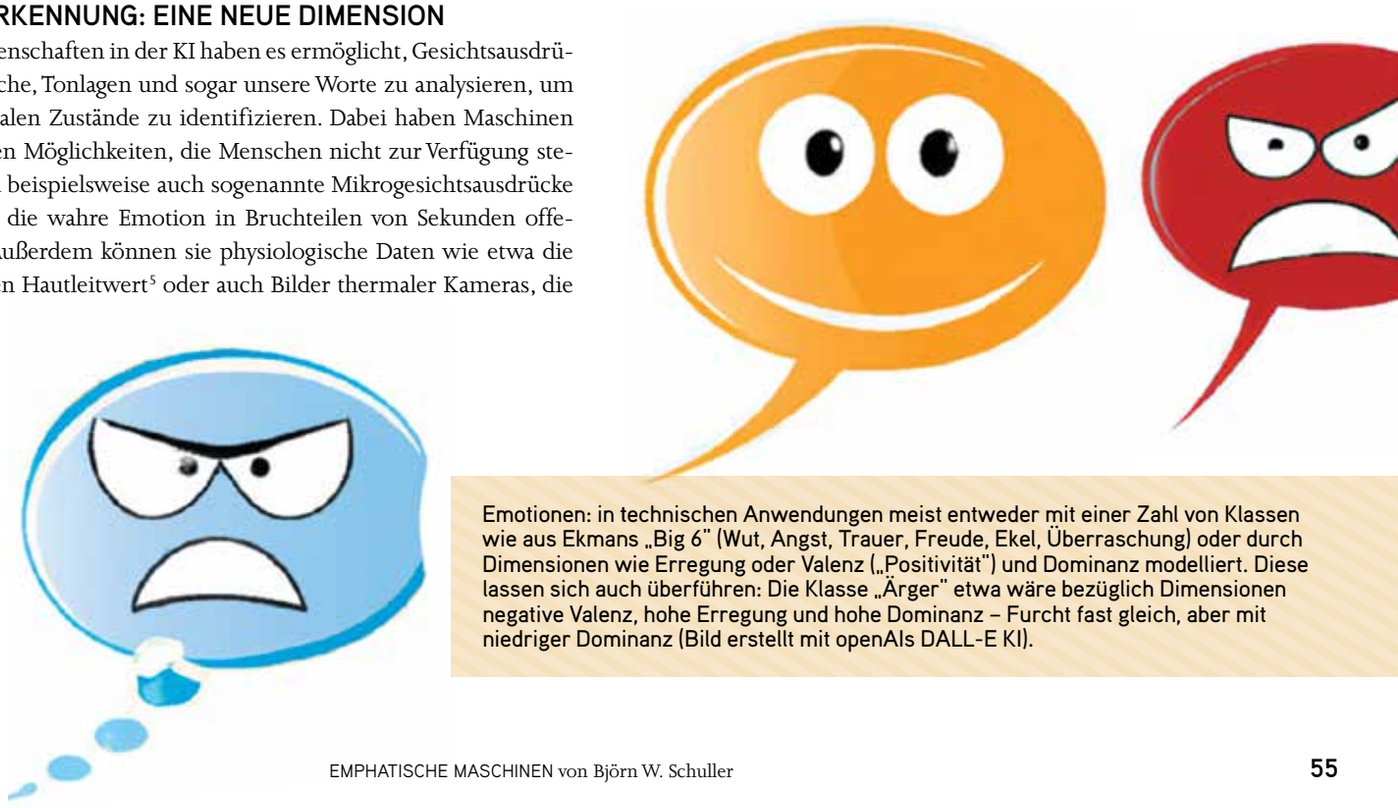
Doch mit diesen faszinierenden Fortschritten, die etwa in der Mensch-Maschine-Kommunikation, im Gamingbereich, dem e-Learning oder der digitalen Psychologie wesentliche Durchbrüche wie Prävention bei Depressionen oder Suizidalität bringen können, taucht neben der Hoffnung auch die Angst auf: Beherrschen wir die Technik oder beherrscht sie uns? Wie real ist die Angst des Menschen vor der Machtübernahme durch Maschinen, wie sie in Science-Fiction-Filmen und Büchern oft dargestellt wird? Auch wenn die Antwort auf die letzte Frage den Rahmen dieses Beitrags sprengen würde, so zeigt er doch auf, wie emotional intelligente Maschinen und deren Nutzer:innen uns künftig beeinflussen könnten – vielleicht auch bis zur Machtübernahme.

EMOTIONSERKENNUNG: EINE NEUE DIMENSION

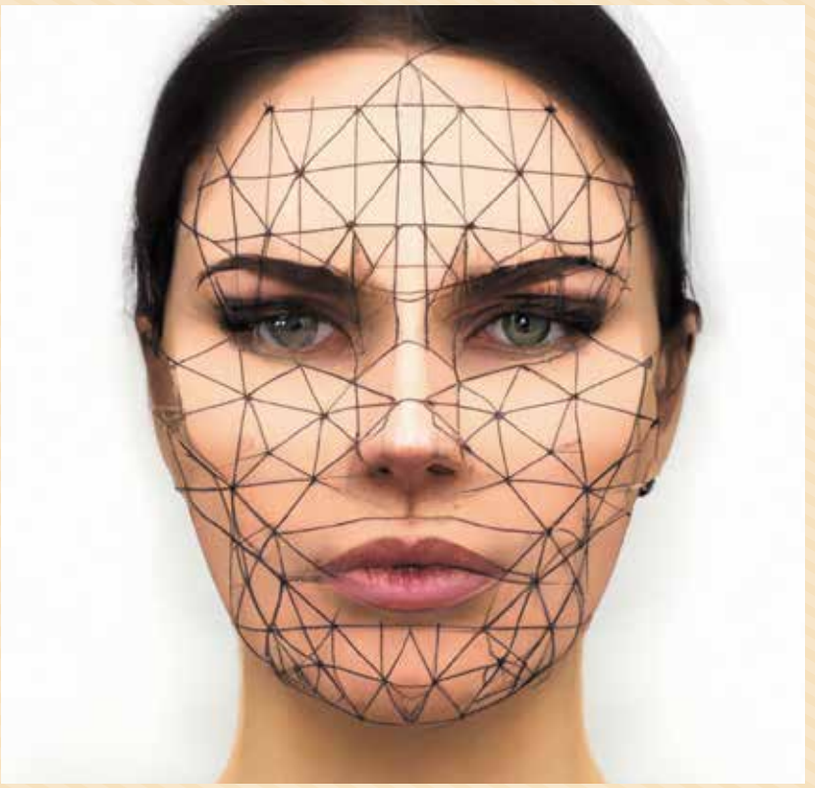
Aktuelle Errungenschaften in der KI haben es ermöglicht, Gesichtsausdrücke, Körpersprache, Tonlagen und sogar unsere Worte zu analysieren, um unsere emotionalen Zustände zu identifizieren. Dabei haben Maschinen unter Umständen Möglichkeiten, die Menschen nicht zur Verfügung stehen: Sie können beispielsweise auch sogenannte Mikrogesichtsausdrücke analysieren, die die wahre Emotion in Bruchteilen von Sekunden offerieren sollen[4]. Außerdem können sie physiologische Daten wie etwa die Pulsrate oder den Hautleitwert[5] oder auch Bilder thermaler Kameras, die erkennen lassen, wie stark wir Erröten, unter Umständen ergänzend zu Rate ziehen. Maschinen können nun erkennen, ob jemand fröhlich, traurig, wütend oder überrascht ist. Doch geht die Entwicklung noch weiter?

Die Forschung im Bereich der Neurotechnologie könnte eine Zukunft bringen, in der Maschinen auch subtilere Emotionen wie Unsicherheit oder Verlegenheit erkennen können. Vor allem aber werden uns sogenannte ***Brain Computer Interfaces*** oder Elektroenzephalogramme (EEG) der Zukunft wohl einen Bezugspunkt der tatsächlichen Emotion liefern – aktuell kann KI nämlich nur von menschlichen Urteilen lernen, um diese bestmöglich nachzubilden. In der Regel werden hier entweder die menschlichen Urteile über eine Emotion etwa in Ton- oder Bilddaten gemittelt oder die jeweilige Person, die eine Emotion empfindet, selbst nach dieser gefragt, um eine Bezugsemotion zu erhalten, anhand derer und mehrerer tausend Beispiele die KI lernen kann, um sie in neuen Daten wiederzuerkennen. Dies ist aber jeweils subjektiv und unsicher als Bezugsgröße, sodass Maschinen noch hauptsächlich die Emotion erkennen können, wie andere menschliche Beobachter sie wahrnehmen – nicht besser und auch nicht die „wahre" Emotion, sondern nur die wahrgenommene.

Immerhin kann KI aber die Meinung von mehreren Expert:innen nachahmen, sodass sie in bestimmten Anwendungen, wie bei klinischer Depressionsanalyse, Laien überlegen sein kann. Auch kann sie sich gleichzeitig gleichgut auf verschiedene Informationskanäle „konzentrieren" und somit Unstimmigkeiten wie Asynchronitäten besser erkennen. Wenn wir also zum Beispiel versuchen, Nervosität zu verstecken, können wir das vielleicht gut im Gesichtsausdruck oder dem Klang der Stimme schaffen, aber weniger gut gleichzeitig in „Bild und Ton" – Maschinen können dies leichter aufdecken.



Emotionen: in technischen Anwendungen meist entweder mit einer Zahl von Klassen wie aus Ekmans „Big 6" (Wut, Angst, Trauer, Freude, Ekel, Überraschung) oder durch Dimensionen wie Erregung oder Valenz („Positivität") und Dominanz modelliert. Diese lassen sich auch überführen: Die Klasse „Ärger" etwa wäre bezüglich Dimensionen negative Valenz, hohe Erregung und hohe Dominanz – Furcht fast gleich, aber mit niedriger Dominanz (Bild erstellt mit openAIs DALL-E KI).



Fiktives Beispiel eines Gesichtsnetzes, um Emotion im Gesichtsausdruck zu erkennen. Aktuelle tiefe Lernverfahren lernen mittlerweile selbstständig, worauf sie achten müssen, um Emotion in neuen Gesichtern, Stimmen, gesprochenen Worten oder anderer Information erkennen zu können – auch in sogenannten Mikroexpressionen, also sehr kurzfristigen Gesichtsausdrücken (Bild erstellt mit openAIs DALL-E KI).

EINFLUSS AUF ZWISCHENMENSCHLICHEN UMGANG

Die Einführung empathischer Maschinen wird zweifellos Auswirkungen auf unseren zwischenmenschlichen Umgang haben. Die Möglichkeit, Gefühle virtuell zu erzeugen, auszutauschen, zu beobachten und zu analysieren, eröffnet eine neue Ebene der Kommunikation. Automatische Analysen in großen Datenmengen erlauben so neue quantifizierte Einblicke mit hoher Stichprobenzahl in der Psychologie. Emotional intelligente KI könnte aber auch Menschen helfen, sich besser zu verstehen und empathischer miteinander umzugehen. So kann man sich etwa vorstellen, dass eine Software zwischenmenschliche Kommunikation auf der emotionalen Ebene analysiert und uns Feedback gibt – etwa, wer emotional wem „folgt" und somit Einfühlungsvermögen aufweist. KI kann uns dabei mit ungeteilter Aufmerksamkeit ohne Ermüdung beobachten und dies mit höherer Signalauflösung und Präzision. Sie kann aus mehr Beispielen lernen als wir dies in unserer Lebenszeit könnten. Maschinen, die einfühlsam sind, können aber auch unser zwischenmenschliches Verhalten in der Zukunft klar verändern. Auf der einen Seite werden sie möglicherweise in Konkurrenz mit uns treten, denn Maschinen könnten perfekt lernen, etwa charismatisch zu sein[6] – zum Beispiel aus der Interaktion mit Millionen von Menschen und dem Erlernen aus dem Feedback zu unserer Beeinflussbarkeit.

Auf der anderen Seite können Maschinen dabei auch neue Verhaltensmuster entwickeln, die gegebenenfalls noch effizienter sind als bisherige zwischenmenschliche – wir Menschen würden sie dann vielleicht übernehmen, um im Konkurrenzkampf zu bestehen. In der medizinischen Praxis könnten empathische Maschinen darüber hinaus in der Therapie von Autismus eine Revolution bringen. Menschen, die Schwierigkeiten haben, sozioemotionale Signale zu interpretieren, werden von Technologien profitieren, die ihnen helfen, diese ähnlich einer Fremdsprache zu erlernen und bei Interesse besser zu verstehen und auszudrücken[7]. Dies könnte zu einer erheblichen Verbesserung der Lebensqualität für viele Menschen führen.

MISSBRAUCH UND ETHISCHE BEDENKEN

Trotz aller vielversprechenden Anwendungen gibt es auch eine dunkle Seite emotional intelligenter und empathischer Maschinen. Die Möglichkeit, Emotionen automatisch zu manipulieren, könnte in falsche Hände geraten und zu Missbrauch führen[8]: Denken wir etwa an personalisierte Werbung, die auf die individuelle Gemütslage zugeschnitten ist und uns im „schwachen Moment" zu Käufen überredet, oder an manipulative Techniken, die auf unsere Emotionen abzielen – etwa um Wahlen zu beeinflussen.

Auch die Frage nach der Privatsphäre wird in diesem Kontext immer drängender. Wenn Maschinen in der Lage sind, unsere emotionalen Zustände zu erkennen, wer hat dann Zugriff auf diese Informationen und wie werden sie genutzt? Etwa beim nächsten Jobinterview – dann von einer KI (mit-)geführt? Oder bei Gesprächen mit der Krankenkasse, die nachträglich den Beitrag erhöht, da sie erhöhte Risiken psychologischer Erkrankung einschätzt? Es ist daher entscheidend, klare ethische Richtlinien zu etablieren, um den Missbrauch zu verhindern und die Privatsphäre der Nutzer:innen zu schützen.

Die Entwicklung empathischer Maschinen wirft so eine Vielzahl von technischen, ethischen und rechtlichen Fragen auf, die dringend geklärt werden müssen. Welche Standards sollten bei der Entwicklung solcher Technologien gelten? Wie können wir sicherstellen, dass sie verantwortungsvoll eingesetzt werden? Und wie können wir die Rechte der Individuen schützen, deren Emotionen von Maschinen analysiert werden? Spannend sind hier vor allem technische Beiträge, die mittels KI selbst vor missbräuchlich angewandter emotional intelligenter KI schützen können. Dies kann zunächst durch Veränderung der Information wie Ton der Sprache oder Gesichtsausdruck im Bild geschehen – Steven Spielberg ahnt etwa im Film **Ready Player One** eine Zukunft mit solcher emotionsunterdrückender KI voraus.

Tatsächlich lässt sich die Emotion bereits im Gesichtsausdruck oder der Stimme menschenwahrnehmbar in eine Zielemotion hin verändern[9]. KIs können die Information aber auch so verändern, dass der Mensch keinen Unterschied sieht oder hört, eine Ziel-KI aber absichtlich getäuscht wird, also etwa Ärger für Menschen unverändert Ärger bleibt, eine angegriffene KI aber Freude wahrnimmt – man nennt dies ***Adversarial Attack***[10] (feindliche Attacke). Darüber hinaus ist es aber natürlich auch Aufgabe der Gesellschaft, entsprechenden Schutz zu schaffen: Die Europäische Union will dies zum Beispiel künftig im Rahmen des ***AI Act*** tun. Nach aktueller Parlamentsposition soll Emotionserkennung etwa am Arbeitsplatz, an Ausbildungsinstitutionen und Grenzübergängen oder im Strafvollzug verboten werden[11].

Weitere Herausforderungen kommen – wie auch in anderen Bereichen der KI – mit der Gerechtigkeit der Anwendung: Werden alle Nutzer:innen gleich gut erkannt? Entstehen nur bei manchen Altersgruppen häufiger Fehler? Dazu gesellt sich die oft beschränkte Erklärbarkeit von Entscheidungen der genannten tiefen neuronalen Lernverfahren[12]. Auch Fragen der Umwelt stehen bei KI zunehmen im Vordergrund[13]: Wieviel Energie wurde verbraucht, um solche Netze anzulernen? Noch verhält sich dies bei emotionaler Intelligenz in kleinerem Rahmen, aber das kann sich bei Hochskalierung schnell ändern.

Ferner müssen Daten noch von Menschen „verschriftet" werden, also zum Anlernen der Maschine muss bestimmt werden, welcher Gesichtsausdruck, welche Worte beispielsweise zu welcher Emotion gehören. Nicht immer sind die Arbeitsbedingungen der Arbeitenden adäquat. Vor allem aber stellt sich die Frage, ob die Entwicklung emotional intelligenter Maschinen noch aufhaltbar ist – längst sind entsprechende Eigenschaften „emergent" beobachtbar, das heißt sie tauchen einfach auf in großen Modellen, also solchen mit mehr als 100 Milliarden lernbaren Parametern – bereits teilweise auch ohne, dass sie speziell darauf angelernt wurde[14].

Emergente emotionale Kompetenz in KI: In „großen" Modellen, die auf großen Datenmengen trainiert sind, sind auch emotionale Konzepte implizit miterlernt – hier etwa das Beispiel von openAIs DALL-E beim Erstellen von Bildern zu Personen, die traurig sind. In früheren Ansätzen wurden hierzu speziell gezielt einzelne Modelle angelernt (Bild erstellt mit openAIs DALL-E KI).

CHANCEN UND HERAUSFORDERUNGEN

Emotional intelligente und empathische Maschinen haben zweifellos das Potenzial, unsere Welt zu verändern. Sie könnten eine neue Ära der Kommunikation, Interaktion und Therapie einleiten, insbesondere im Bereich psychologischer Erkrankungen. Doch gleichzeitig müssen wir uns den potenziellen Gefahren etwa ungewollter Überwachung bewusst sein und sicherstellen, dass diese Technologien verantwortungsvoll, vertrauenswürdig, sicher und ethisch entwickelt und eingesetzt werden. Es liegt an uns, die Zukunft dieser Maschinen mitzugestalten und sicherzustellen, dass sie zu einer positiven Veränderung in unserer Gesellschaft beitragen. Indem wir die Chancen nicht verhindern und gleichzeitig die Herausforderungen des technischen und regulatorischen Schutzes angehen, können wir eine Welt schaffen, in der Technologie und Menschlichkeit in natürlicher Weise kommunizieren und empathische Maschinen auch viele Leben verbessern oder sogar retten werden – sie dürfen uns dabei aber nicht steuerbar oder abhängig machen.

ANMERKUNGEN

1. Picard, R. W. (2000): Affective computing, MIT press.
2. Williamson, J. (1978): Speech analyzer for analyzing pitch or frequency perturbations in individual speech pattern to determine the emotional state of the person, U.S. Patent 4.093.821.
3. Zhang, A.; Lipton, Z. C., Li, M. & Smola, A. J. (2023): Dive into Deep Learning, Cambridge University Press.
4. Zhao, G.; Li, X.; Li, Y. & Pietikäinen, M. (2023): Facial Micro-Expressions: An Overview. Proceedings of the IEEE.
5. Can, Y. S.; Mahesh, B. & André, E. (2023): Approaches, Applications, and Challenges in Physiological Emotion Recognition: A Tutorial Overview, Proceedings of the IEEE.
6. Schuller, B. W.; Amiriparian, Sh,; Batliner, A.; Gebhard, A.; Gerzcuk, M.; Karas, V.; Kathan, A.; Seizer, L. & Löchner, J. (2022): Computational Charisma: A Brick by Brick Blueprint for Building Charismatic Artificial Intelligence.arXiv:2301.00142.
7. Bishop, J. (2015): Supporting communication between people with social orientation impairments using affective computing technologies: Rethinking the autism spectrum. In: Assistive technologies for physical and cognitive disabilities, S. 42-55. IGI Global.
8. Cowie, R. (2015): Ethical issues in affective computing: The Oxford Handbook of Affective Computing, S. 334-348. Oxford Academic.
9. Zhou, K.; Sisman, B.; Rana, R.; Schuller, B. W. & Li, H. (2023): Emotion Intensity and its Control for Emotional Voice Conversion, IEEE Transactions on Affective Computing 14(1):31-48, IEEE.
10. Gao, J.; Yan, D. & Dong, M. (2022): Black-box adversarial attacks through speech distortion for speech emotion recognition. In: EURASIP Journal on Audio, Speech, and Music Processing, 2022(1):1-10, SpringerOpen.
11. https://digital-strategy.ec.europa.eu/de/policies/european-approach-artificial-intelligence
12. Roundtree, A. K. (2023): AI Explainability, Interpretability, Fairness, and Privacy: An Integrative Review of Reviews. In: Degen, H.; Ntoa, S. (Editoren) Artificial Intelligence in HCI. Lecture Notes in Computer Science (14050). Springer.
13. Schwartz, R. Dodge, J.; Smith, N. A. & Etzioni, O. (2020): Green AI: Communications of the ACM, 63(12): 54-63, ACM.
14. Mostafa, A. M.; Cambria, E. & Schuller, B. W. (2023): Will Affective Computing Emerge From Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT. IEEE Intelligent Systems 38(2):15-23, IEEE.

AUTOR
PROF. DR. BJÖRN W. SCHULLER

ist Professor für Artificial Intelligence am Imperial College London in Großbritannien und Inhaber des Lehrstuhls für Embedded Intelligence for Health Care and Wellbeing an der Universität Augsburg. Er ist weltführend im Fachgebiet des Affective Computing.