

Werteorientiertes Rahmenwerk für künstlich empathische Technologie in der Pflege

Die Pflegepraxis ist aufgrund begrenzter Ressourcen erheblichen Herausforderungen ausgesetzt. Menschliche Werte wie Sicherheit, Autonomie und Selbstachtung werden dadurch häufig vernachlässigt. Zudem sind kontinuierliche Co-Regulationsprozesse sowie zuverlässige Kommunikationsbrücken erforderlich. Dieses Kapitel erläutert die künstlich empathische Technologie (KET) als werteeorientiertes Rahmenwerk für soziale Assistenzsysteme im Pflegekontext. KET umfasst Fähigkeiten, die über die gegenwärtigen Systeme hinausgehen, die zwar »emotionale« Zustände »erkennen« und sich entsprechend anpassen — jedoch ohne ein Modell für die möglichen subjektiven Erfahrungen von Personen und Situationen. Bei KET steht nicht die Erkennungsgenauigkeit im Vordergrund, sondern werteegebundene Co-Regulation auf Basis von Mentalisierung und Dyaden-Modellen. KET simuliert, wie Handlungen die Person und die Beziehung beeinflussen, und wählt daraufhin (Re)Aktionen aus — begründet und kontextsensitiv. Werte sind dabei nicht nur Designkriterien, sondern auch operative Zielsetzungen im laufenden Betrieb. Auf diese Weise fördert KET eine beziehungsorientierte, verkörperte Sichtweise des Menschen als Leitlinie für die Entwicklung technischer Systeme.

1. Einordnung und Care-Werte

Die Pflege ist untrennbar mit den Grundwerten des Menschseins wie Sicherheit, Autonomie und Selbstachtung verbunden. Technologie ist im Pflegekontext nur verantwortbar, wenn normative Maßstäbe aus Care-Ethik, konfessions- und traditionsübergreifender Theologie und Anthropologie leitend sind. Zugleich muss sie einen er-

kennbaren Mehrwert bieten — hinsichtlich Transparenz, Erklärbarkeit und Verhältnismäßigkeit.

Pflege bedeutet, sich um Menschen zu kümmern, wenn existenzielle Fähigkeiten fehlen oder nicht mehr vorhanden sind und durch andere kompensiert werden müssen, um ein würdevolles Leben oder gar das Überleben zu sichern¹. Das ist auf mehreren Ebenen herausfordernd: individuell, familiär, im Arbeitskontext und für die Gesellschaft insgesamt².

In besonderer Dichte zeigt Pflege, was menschliche Verletzlichkeit, Autonomie und Würde bedeuten. Die Frage ist nicht, *ob* Technologie hier Platz hat, *sondern* unter welchen Bedingungen sie Fürsorgebeziehungen stärkt und das professionelle Handeln von Pflegenden unterstützt. Zugleich ist Pflege eine relationale, verkörperte und institutionell gerahmte Praxis — ein Lebensbereich, in den technische Systeme nur dann sinnvoll integriert werden können, wenn diese Besonderheiten von den Erschaffern respektiert werden.

Die Pflege ist ein besonders sensibles Einsatzfeld für technische Assistenzsysteme. Aus der Sicht der assistenzethischen Debatte, wie sie beispielsweise in der Arbeit von Manzeschke³ beschrieben wird, zielen sie primär auf Kompensation, eine robuste und fehlertolerante Funktionsweise sowie auf Entlastung durch standardisierbare Funktionen ab. Zukünftige *künstlich empathische Technologie (KET)* ergänzt diese, da sie zentrale Aspekte des Menschseins wie Würde, Autonomie und Beziehungsgestaltung integriert. KET-Systeme gehen über traditionelle Assistenzsysteme hinaus: Sie simulieren individuelles Wahrnehmen und Interpretieren sozialer Situationen, um situationsangemessene, wertesensible und erklärbare regulative Assistenzaktionen anzubieten — ohne eigene Empfindungsfähigkeit (Definition in Kapitel 3). Moralische Autorenschaft und Verantwortung verbleiben bei den handelnden Personen.

1 § 14 SGB XI – Begriff der Pflegebedürftigkeit.

2 Hans-Jörg Ehni und Selma Kadi, Ethische Aspekte des ›gelingenden‹ und ›gesunden‹ Alterns in der pluralistischen Gesellschaft. Leitlinien für Leitlinien, 10.13140/RG.2.2.11398.60486, 48ff, 2020.

3 Arne Manzeschke, Technische Assistenzsysteme. In: Fuchs, M. (eds) Handbuch Alter und Altern. J.B. Metzler, 2021.

Für KET wird ein tiefes Verständnis von sozialem Miteinander und Empathie⁴ sowie insbesondere der Emotionsregulation⁵ benötigt. Durch den mentalen Prozess der *Regulation* gestalten Menschen ihre eigenen Gefühle. Über soziale Interaktion können sie durch *Co-Regulation* auch die Gefühle anderer mitzugestalten versuchen^{6,7}. Diese Prozesse stabilisieren das Erleben und sind im Pflegekontext für alle Beteiligten von besonderer Bedeutung. Eine funktionierende Emotionsregulation ist die Grundlage für ein gesundes Leben^{8,9}. Bei der genannten Technologie geht es grundsätzlich nicht darum, Systemen eigene Emotionen zu verleihen, sondern darum, relevante Prozesse und notwendige Repräsentationen in Computermodellen abzubilden und zu simulieren, um Menschen situationsgerecht zu unterstützen. Somit erweitert KET klassische Kommunikationsmodelle um interne Prozesse und Repräsentationen.

Als normativer Rahmen für den verantwortungsvollen Einsatz von KET in der Pflege muss gelten: 1) Die unantastbare Würde aller Beteiligten ist gewahrt. 2) Autonomie und informierte Einwilligung sind gewahrt. 3) Die Fürsorgebeziehung ist nicht instrumentalisierend gestaltet. 4) Gerechtigkeit und Zugänglichkeit werden beachtet. 5) Rechenschaft, Transparenz und Aufsicht sind gewährleistet. Dieser Rahmen kann beispielsweise mittels Werkzeugen wie MEE-STAR¹⁰ systematisch überprüft werden.

4 John F. Kihlstrom und Nancy Cantor, Social Intelligence, The Cambridge Handbook of Intelligence. Cambridge Handbooks in Psychology, 756–779, 2020.

5 James Gross, Emotion Regulation: Current Status and Future Prospects. Psychological Inquiry, 26(1), 1–26, 2015.

6 Emily A. Butler und Ashley K. Randall, Emotional Coregulation in Close Relationships, Emotion Review 5.2: 202–210, 2013.

7 Peter Fonagy, Gergely und Jurist, Affect Regulation, Mentalization and the Development of the Self, Routledge, 2018.

8 Vincent J. Felitti et al., Relationship of Childhood Abuse and Household Dysfunction to many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study, American Journal of Preventive Medicine 14.4: 245–258, 1998.

9 Debora Cutuli, Cognitive Reappraisal and Expressive Suppression Strategies Role in the Emotion Regulation: An Overview on their Modulatory Effects and Neural Correlates, Frontiers in Systems Neuroscience, 2014.

10 Arne Manzeschke, MEESTAR: Ein Modell angewandter Ethik im Bereich assistiver Technologien, Technisierung des Alters–Beitrag zu einem guten Leben. In: Karsten Weber, Deborah Frommeld, Arne Manzeschke und Heiner Fangerau

KET wird theologisch anschlussfähig, wenn die relationale Würde¹¹ des Menschen als Maßstab gilt (Kapitel 9.3): Sie kann die Fürsorge unterstützen, darf jedoch die moralische Verantwortung nicht verschleiern.

Empathie ist eine menschliche Fähigkeit, die von Technologie lediglich simuliert und kontextsensibel begleitet werden kann. Aus anthropologischer Sicht liegt der Fokus der Pflege auf Körperlichkeit, Interdependenz und Rollen. Pflegende, Pflegebedürftige und Angehörige bewegen sich in geschaffenen institutionellen Ordnungen und folgen Interaktionsritualen. Beim Einsatz von KET ist zu beachten, dass sie in diese Routinen eingreift und diese verändert. Daher muss der Einsatz sorgfältig abgewogen werden. Das Ziel sollte darin bestehen, Pflegende in der Wahrnehmung, in der Kommunikation und in den konkreten Pflege Tätigkeiten zu entlasten, ohne dabei die Deutungshoheit zu beanspruchen oder soziale Verpflichtungen unbemerkt zu verschieben. Potenziale und Risiken liegen nah beieinander. Denn KET kann Belastungen reduzieren, die Kommunikation strukturieren und ein höheres Sicherheitsgefühl ermöglichen. Risiken sind hingegen Bevormundung, Überwachung, Verzerrungen (Bias) oder die Delegation moralischer Urteile an Maschinen. Dieses Kapitel plädiert daher für eine gestaltende, kritisch-reflektierte Einführung, die klinische Qualität, ethische Maßstäbe und lebensweltliche Passung verbindet.

1.1 Relevanz und Implikationen für die Pflege

Die Pflege von Menschen berührt existenzielle Werte, zu denen insbesondere Sicherheit, Autonomie und Selbstachtung zählen. Die Pflege ist neben der Kompensation körperlicher oder geistiger Fähigkeiten stets mit Interaktions- und Emotionsarbeit verbunden. Auswirkungen des demografischen Wandels, der Globalisierung, der Multimorbidität und der Personalknappheit erhöhen den Druck auf eine kontinuierliche, würdige Beziehungsgestaltung, die oft wegen

(Hrsg.): »Technisierung des Alters – Beitrag zu einem guten Leben?«, Stuttgart (Steiner), 263–283, 2015.

11 Relationale Würde entsteht in Beziehungen durch wechselseitige Anerkennung, Umgangsformen, Rollen und damit verbundenen Handlungen. Sie ist kontextabhängig, verletzbar und veränderlich.

Zeitmangels und der hohen Last administrativer Aufgaben kompromittiert wird.

KET könnte diese Lücke teilweise kompensieren, indem sie Prozesse der Wahrnehmung, der Interpretation und der Regulation in der Dyade explizit modelliert und erklärt und damit situativ Fachpersonal unterstützt. Empathie in der Pflege verbindet Perspektivübernahme, Mitbetroffenheit und handlungsleitende Regulation^{12,13}. Für die konkrete Gestaltung von KET ergibt sich Folgendes: Erstens braucht es erklärbare Modelle regulativer Prozesse, um sozial und individuell angepasst unterstützen zu können; zweitens einen Abgleich mit Normen und individuellen Werten; drittens eine Entscheidung, die potenzielle Folgen für Erleben und Beziehung antizipiert.

Der Einsatz von Technologie in der Pflege ist jedoch nicht nur aus den genannten inhaltlichen Gründen anspruchsvoll: Arbeitsprozesse müssen angepasst werden, und die Abwägung der Vor- und Nachteile für alle Beteiligten ist komplex.

Es existieren erste Erkenntnisse aus der Perspektive pflegebedürftiger Menschen, Fachpersonals, Angehörigen und der Gesellschaft im Allgemeinen über den Einsatz digitaler und assistiver »emotionaler« Technologien (im folgenden Kapitel) — einer Vorstufe des hier skizzierten Konzepts einer künstlich empathischen Technologie (Kapitel 2 und 5). Die genannte Emotionalität bezieht sich auf die imitierende Darstellung gezeigter Emotionen. Generell lässt sich argumentieren, dass derzeitige interaktive Assistenzsysteme einen leichteren Zugang zur Unterstützung in der Pflege bieten. Die empfundene Leichtigkeit kann auf den Umstand zurückgeführt werden, keinem tatsächlichen Menschen zur Last zu fallen. Wie in den folgenden Kapiteln genannten Studien berichtet, wird diese Technologie personifiziert, und Menschen berichten, in einer Beziehung dazu zu stehen.

12 David Martínez-Iñigo, Consuelo Bermejo-Pablos und Peter Totterdell, The Boomerang Effect: How Nurses' Regulation of Patients' Affect Associates with their own Emotional Exhaustion and Affective Experiences, *International Journal of Stress Management*, 25.1:1, 2018.

13 Iris Maria Vinatzer, Barbara Juen und Christiane Kreyer, Strategien der Interpersonellen Emotionsregulation in der Palliativen Begleitung von Angehörigen – Eine Explorative Qualitative Studie, *Zeitschrift für Palliativmedizin*, 2025.

1.2 Technologiekontext

Die *Mensch-Maschine-Interaktion (MMI)* hat sich in den letzten Jahrzehnten rasant entwickelt¹⁴: Neben klassischen Schnittstellen treten heute virtuelle Welten, Sprache und *sozial interaktive Agenten (SIAs)*¹⁵ auf. Beispiele sind soziale Roboter¹⁶ oder virtuelle Assistenten wie Billie¹⁷, der Menschen mit kognitiven Einschränkungen unterstützt, oder Lydia¹⁸, die zur Unterstützung von Menschen mit Depressionen konzipiert wurde.

SIAs ergänzen die MMI um soziale Eigenschaften und ermöglichen es, in Computeranwendungen Fachpersonal aus verschiedenen Bereichen zu simulieren. Beispiele hierfür sind Coaches im Gesundheitskontext, Dolmetschende für Gehörlose oder Ansprechpersonen in Museen, die Unterstützung bieten können, wenn keine menschlichen Ansprechpersonen zur Verfügung stehen. Darüber hinaus werden SIAs in sozialen Trainingssystemen^{19,20} eingesetzt, um menschenähnliche Interaktionspartner zu simulieren. Hierdurch können Nutzende den Umgang mit komplexen sozialen Situationen, etwa in Bewerbungsgesprächen, üben und verbessern.

Der aktuelle Stand lässt insbesondere hinsichtlich der Aspekte Transparenz und Erklärbarkeit keine entsprechenden Möglichkeiten

14 Kevin Liggieri und Oliver Müller, *Mensch-Maschine-Interaktion, Handbuch zu Geschichte, Kultur und Ethik*, Stuttgart, JB Metzler, 2019.

15 Birgit Lugrin, Catherine Pelachaud und David Taum, *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (1st. ed.). ACM Books, Vol. 37, 2021.

16 <https://navelrobotics.com> (zuletzt aufgerufen 26. April 2026).

17 Yaghoubzadeh et al., *Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People: Studies on Acceptance and Interaction Feasibility*, International Workshop on Intelligent Virtual Agents, Heidelberg: Springer Berlin Heidelberg, 2013.

18 Diana Immel et al., *Patients' and Health Care Professionals' Expectations of Virtual Therapeutic Agents in Outpatient Aftercare: Qualitative Survey Study*, JMIR Formative Research 9.1, 2025.

19 Patrick Gebhard et al., *Serious Games with SIAs. The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, 527–546, 2022.

20 Chirag Bhuvaneshwara et al., *MITHOS-mixed reality interactive teacher training system for conflict situations at school*, ISLS Annual Meeting, 2023.

zu. Vielmehr regen aktuelle Ansätze dazu an, menschliche Eigenschaften auf technische Systeme zu projizieren, ohne das dahinterliegende Menschenbild zu reflektieren oder zu schärfen (Kapitel 3).

Die folgenden Fallstudien illustrieren, wie weit aktuelle Systeme reichen und wo sich die beschriebenen Grenzen bemerkbar machen.

1.3 Entwicklungen, Fallstudien und Praxisbeispiele

1.3.1 Empathische Agenten

Forscher auf dem Gebiet der empathischen Agenten sind aus mehreren Gründen überzeugt, dass es gewinnbringend ist, Agenten mit »empathischen« Fähigkeiten auszustatten. Allgemein wird angenommen, dass Agenten eher akzeptiert werden, wenn sie Nutzende als soziale Akteure behandeln⁶⁹. Dieser Ansatz beinhaltet, dass Agenten sich auf soziale (vertraute) Weise verhalten sollten. Darüber hinaus sollten sie in der Lage sein, empathisches Verhalten zu zeigen^{21,22}. Damit verbunden sind die Forschungsfragen: 1) ob das Verhalten des Agenten als soziales (empathisches) Verhalten wahrgenommen wird und 2) ob die Glaubwürdigkeit solcher Agenten dadurch erhöht wird^{23,24,25,26}. Dazu müssen solche Agenten in der Lage sein, andere auf der Ebene von Absichten, Motivationen und

-
- 21 Timothy W. Bickmore, *Relational Agents: Effecting Change Through Human-Computer Relationships*, Dissertation Massachusetts Institute of Technology, 2003.
 - 22 Adriana Tapus und Maja J. Mataric, *Emulating Empathy in Socially Assistive Robotics*, AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics, 2007.
 - 23 James C. Lester et al., *The Persona Effect: Affective Impact of Animated Pedagogical Agents*, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 1997.
 - 24 Susanne Van Mulken, Elisabeth André und Jochen Müller, *The Persona Effect: How Substantial is it?*, People and Computers XIII: Proceedings of HCI'98. London: Springer London, 1998.
 - 25 Ana Paiva et al., *Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents*, AAMAS, Vol. 4, 2004.
 - 26 Kerstin Dautenhahn, *Socially Intelligent Robots: Dimensions of Human-Robot Interaction*, Philosophical Transactions of the Royal Society B: Biological Sciences 362.1480: 679–704, 2007.

Gefühlen zu verstehen – was das Wahrnehmen emotionaler Zustände und entsprechendes Handeln einschließt^{27,28}.

1.3.2 PARO – Chancen, Grenzen und Erkenntnisse

Die Babyrobotterrobbe PARO wird als prototypisches Beispiel einer »emotionalen« Technologie herangezogen^{29,30}. Pflegekräfte berichten überwiegend über positive psychosoziale Effekte bei Pflegebedürftigen – wie verbesserte Stimmung und erhöhte Kommunikationsbereitschaft –, wobei der Erfolg stark von ihrer aktiven Einbindung abhängt. Die Mehrheit der Pflegebedürftigen nimmt PARO als lebendiges Wesen wahr, was einerseits therapeutische Möglichkeiten eines emotionalen Zugangs eröffnet, andererseits ethische Herausforderungen hinsichtlich Täuschung und Manipulation aufwirft. Positive Wirkungen sind dabei meist kurzfristig; langfristige Effekte wurden nur selten beobachtet. PARO illustriert damit Chancen und Grenzen emotional-adaptiver Technologie: Situative Stabilisierung ist möglich, jedoch fehlen wertegebundene Regulation, ein Beziehungsmodell und eine mentalisierungsbasierte Handlungssteuerung – Kernmerkmale, die KET auszeichnen.

1.3.3 Weitere Systeme

Die folgenden Beispiele zeigen den Übergang von emotional adaptiven Assistenten zu sozial interaktiven, proaktiven Systemen in der Pflege. Robuste Langzeit-Wirksamkeitsdaten sind wenig vorhanden. Einen Rahmen für die wertegebundene Entscheidungslogik, die Erklärbarkeit von KI-Algorithmen und Human-in-the-Loop, den KET vorsieht, gibt es noch nicht.

-
- 27 Stacy Marsella und Jonathan Gratch, *Computationally Modeling Human Emotion*, Communications of the ACM, 57.12: 56–67, 2014.
 - 28 Ana Paiva et al., *Empathy in Virtual Agents and Robots: A Survey*, ACM Transactions on Interactive Intelligent Systems, 7.3: 1–40, 2017.
 - 29 Stefanie Baisch et al., *Emotionale Roboter im Pflegekontext*, Zeitschrift für Gerontologie und Geriatrie 51.1: 16–24, 2018.
 - 30 Lisa Geist et al., *Identifikation von Lernfördernden Maßnahmen zur Einführung von Digitalen und Assistiven Technologien (DAT) in Prozesse der Pflege-rischen Versorgung: Eine Qualitative Studie*, HeilberufeScience, 13.3: 152–161, 2022.

Billie¹⁷ ist ein sprachbasierter, auf einem Monitor dargestellter sprachinteraktiver Alltagsassistent, der zur Strukturierung des Tages dient. Er wurde gemeinsam mit der Bethel-Stiftung als Anwendungspartner für die Zielgruppe kognitiv eingeschränkter Menschen entwickelt und erforscht. Die aktuellen Design- und Akzeptanzarbeiten (Interviews/Fokusgruppen) konzentrieren sich auf soziale Anschlussfähigkeit und Routinen im häuslichen Kontext.

ElliQ ist ein proaktiver sozialer Roboter mit Funktionen für Gesprächs-, Aktivierungs- und Gesundheitscoaching. Aus Tests unter realen Einsatzbedingungen in US-Kommunen lassen sich ein hohes Engagement sowie Hinweise auf eine geringere Einsamkeit ableiten³¹.

Einen Schritt weiter geht der soziale Roboter Navel. Er bietet verbale und nonverbale Interaktion in Serienreife und wird seit Ende 2023 in Pflegeheimen pilotiert³². Die Evangelische Heimstiftung begleitet Akzeptanz- und Wirkungsevaluationen.

Diese Entwicklungen zeigen, was bereits möglich ist — und wo ein konzeptueller Rahmen wie KET ansetzt.

2. Definition und Abgrenzung KET

KET bezeichnet technische Systeme, die 1) soziale Signale von zwei Dialogpartnern³³ gleichzeitig sensorisch erfassen, 2) diese nach modellierten menschlichen mentalen Prozessen und kontextspezifischen sozialen Normen interpretieren und darauf aufbauend 3) die Auswirkungen von Re-/Aktionen auf das Erleben und die Beziehung (Gewinne und Risiken) simulieren, um dann 4) auf dieser Basis konkrete *regulative Assistenzaktionen* anbieten. Regulative Assistenzaktionen sind situative Systemhandlungen, die von verbaler Validie-

31 Elizabeth Broadbent et al., ElliQ, an AI-driven Social Robot to Alleviate Loneliness: Progress and Lessons Learned, *The Journal of Aging Research & Lifestyle* 13: 22–28, 2024.

32 Claude Toussaint, Philipp T. Schwarz und Markus Petermann, Navel-a Social Robot with Verbal and Nonverbal Communication Skills, *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

33 Das KET-System wird dabei auch als Dialogpartner gesehen. Dieses Kapitel fokussiert bewusst auf die Dyade Pflegebedürftige/r–KET. Jede weitere Dyade (z. B. Pflegekraft–KET) ließe sich als Erweiterung modellieren (Kapitel 5.4).

rung, Nicken und Lächeln bis hin zu konkreten Assistenzangeboten reichen und stets darauf ausgerichtet sind, positives Erleben zu fördern und Werte wie Sicherheit, Autonomie und Selbstachtung zu wahren.

KET-Systeme können von Menschen überwacht werden und lernen (vgl. Human-in-the-Loop³⁴) beziehungsweise nach expliziter Freigabe teil- oder vollautomatisiert sein. Diese Aktionen sind Ergebnisse regulativer Simulationsprozesse, die die genannten zentralen Werte berücksichtigen. KET rekonstruiert und simuliert somit relevante Anteile innerer Prozesse sowie Beziehungen zu anderen modellbasiert, stellt notwendige Schritte *nachvollziehbar* dar und wird dadurch *überprüfbar und justierbar*. KET besitzt keine eigene Erlebnisfähigkeit und definiert das Konzept der *künstlichen Empathie* als Repräsentation und Simulation relevanter innerer regulativer Prozesse des menschlichen Gegenübers sowie deren Auswirkungen auf die Beziehung zum Gegenüber und ein dazugehöriges soziales, *co-regulierendes* Verhalten. *Co-Regulation* ist die Regulation von Gefühlen in partnerschaftlichen Beziehungen durch reziproke soziale Interaktion⁶. Das Konzept ist in der Entwicklungspsychologie und der Bindungstheorie⁷ verankert.

KET wird in diesem Kapitel als *normatives Rahmenwerk* vorgestellt. Auf normativer Ebene ist das Rahmenwerk in seinen zentralen Aspekten ausgearbeitet: Zielfunktionen, Komponentenarchitektur, Governance-Prinzipien und Evaluationskriterien sind konzeptuell spezifiziert. Die Komponenten weisen unterschiedliche Ausarbeitungsstufen auf: Einige sind formal beschrieben und werden mit exemplarischen Berechnungen (Kapitel 6) belegt, andere, wie die Simulation des inneren Dialogs (Kapitel 5.5.1), werden konzeptuell eingeführt.

Klassische Systeme dieser Art orientieren sich an Arbeiten im Bereich des Affective Computing. Diese Systeme fokussieren sich primär auf das Erkennen, Klassifizieren sowie das Ausdrücken beziehungsweise Adaptieren von Emotionen (z. B. in Abhängigkeit von erkannten Emotionen die Schnittstelle zu Nutzenden anpassen). Anzumerken ist, dass hier mit »Emotion« ein beobachtbares Muster

34 Andreas Holzinger, Interactive machine learning for health informatics: when do we need the human-in-the-loop?, Brain Informatics, Vol. 3, 119–131, 2016.

gemeint ist, etwa im Gesicht oder in der Stimme (Kapitel 3.1). Es gibt weder ein vorgelagertes Wirkmodell noch eine normative Regulierung als feste Zielgröße. KET macht genau das zur Bedingung. Eine Aktion gilt nur dann als »empathisch«, wenn ihre simulierten Folgen im gegebenen Kontext (z. B. Pflege) vertretbar sind, die Reaktion wertgeleitet ist (siehe oben) und sich erklären lässt. Erfolg wird daher nicht nur an der Erkennungsgenauigkeit gemessen, sondern unter anderem an Autonomiewahrung, Transparenz und Regulationsangemessenheit. Zudem lassen sich dadurch Täuschungs- und Abhängigkeitsrisiken transparenter darstellen.

Die größte Herausforderung ist dabei: Wie kann Technologie, die selbst nicht erlebt, dennoch empathisch handeln? Die Antwort liegt in der interdisziplinären Fusion von Ethik, Anthropologie, Soziologie, Psychologie und moderner KI-Forschung. Durch ein damit einhergehendes tieferes Verständnis des menschlichen Erlebens kann sie dazu beitragen, einen menschenwürdigen Umgang zwischen Menschen selbst sowie zwischen Menschen und Technologie zu ermöglichen — gerade im Bereich der Pflege.

Emotional adaptive Technologie ist nach den Paradigmen von Affective Computing³⁸ gestaltet und nutzt multimodale Erkennung von Emotionen zur situativen Anpassung. Untersuchungen weisen darauf hin, dass Generalisierung und Validität ohne starken Kontextbezug begrenzt sind (vgl. die Inferenz von Emotionen auf Basis des Gesichtsausdrucks, Kapitel 3.1). KET adressiert diese Lücke, indem sie Werteziele, Mentalisierung und Dyaden-Modelle als entscheidungsleitende Elemente integriert und dadurch von *reaktiver Anpassung* zu einer *wertegebundenen Co-Regulation* übergeht.

Tabelle 1: Vergleich emotional-adaptiver und künstlich empathischer Technologie.

Kriterien	Emotional Adaptive Technologie	Künstlich Empathische Technologie
Erfassung, Kontext	Multimodale Erkennung emotionaler Signale (Sprache, Gesicht, Körper, Physiologie) zur Anpassung von UI/Dialog; Robustheit steigt mit Kontextfeatures, bleibt aber domänensensitiv.	Multimodalität und interne Metainformationen für Situation, Struktur, Rolle, kulturelle Normen und Kommunikation als Pflicht; Beziehungs- und Rollenrahmen strukturieren die Signalerfassung.

Kriterien	Emotional Adaptive Technologie	Künstlich Empathische Technologie
Interpretation, Modell	Heuristiken, (tief-)lernen- de Klassifikatoren, Fusionsmodelle; emotionssensitive Assistenz; Grenzen: keine universelle Abbildung von Mimik auf Emotionen; starke Kontext und Kulturabhängigkeit.	Erweiterung emotionaler Technologie um Mentalisierung: Selbst-, Gegenüber- und Beziehungsmodell; soziale Signale werden im Beziehungs-, Rollen- und Kulturrahmen interpretiert; Zustände, Ziele, Präferenzen, Vertrauen, Verlässlichkeit explizit geführt.
Entscheidung, Handlung	Reaktive, regelbasierte oder ML-gelernte Zustandsübergänge; Anpassung von Tonfall, Prompting, Aufgabenschwere, UI-Layout je nach erkanntem Zustand; Optimierungsziel: Engagement und Akzeptanz; kein Planungshorizont, keine Simulation von Handlungsfolgen.	Simulation mit Zielfunktion hinsichtlich innerer Sicherheit, Autonomie, Selbstachtung mit Unsicherheits-Kalibrierung; Co-Regulation und Reparaturpfade bei Fehlpassung.
Ethik, Erklärbarkeit, Governance	Technische Erklärbarkeit vorhanden (wichtige Features, Aufmerksamkeit); jedoch keine Erklärungen bzgl. innerer Werte, Beziehung oder Regulation. Risiken: Fehlschlüsse/Bias, Zweckentfremdung. Erfordert Privacy-by-Design und Domänenvalidierung.	Sicherheit, Autonomie und Selbstachtung als verbindliche Entscheidungskriterien; Widerruf, adressatengerechte Erklärungen auf Werteebene, Protokolle, Human-in-the-Loop — normativ an Care-Ethik ausgerichtet.
Beziehung, Daten, Evaluation	Erste Modellierung beziehungsnahe Größen (Rapport/Beziehungsqualität, Verlässlichkeit), Nutzerzustände eher episodisch; Metriken: Task-Erfolg, Engagement, Akzeptanz. Nutzung und Akzeptanz teils positiv, klinische Wirksamkeit heterogen.	Dynamische Beziehungsupdates; Evaluation misst erlebte Sicherheit, Autonomie und Selbstachtung sowie Beziehungs- und Wertequalität — nicht nur Task-Erfolg oder Akzeptanz.

Kriterien	Emotional Adaptive Technologie	Künstlich Empathische Technologie
Reifegrad	Breit erprobt, aber stark kontextabhängig.	Aufkommend, benötigt Pilotierungen und Forschung

Tabelle 1 illustriert die Unterschiede zwischen emotional adaptiver und künstlich-empathischer Technologie anhand der Kriteriengruppen: 1) Erfassung und Kontext, 2) Interpretation und Modell, 3) Entscheidung und Handlung, 4) Ethik, Erklärbarkeit und Governance, 5) Beziehung und Daten, 6) Evaluation.

Zwei vorausgegangene Ansätze verdienen in diesem Zusammenhang besondere Beachtung, da sie versuchen, normative Werte in die Gestaltung soziotechnischer Pflegesysteme zu integrieren.

Van Wynsberghe entwickelt mit dem *Care-Centered Value-Sensitive Design* (CCVSD) eine Methodik, die Friedmans Value-Sensitive Design für den Pflegekontext adaptiert³⁵. CCVSD nutzt Trontos vier Elemente der Care-Ethik — Aufmerksamkeit, Verantwortung, Kompetenz und Reaktionsfähigkeit — als normative Bewertungskriterien für Robotersysteme. Die Stärke liegt in der systematischen Einbeziehung ethischer Werte in Designentscheidungen, wie am Vergleich zweier Heberoboter (RI-MAN als autonomes Ersatzsystem vs. HAL als aktivierendes Exoskelett) veranschaulicht. Entscheidend für die Abgrenzung gegenüber KET: Ethische Werte bleiben bei CCVSD auf der Design- und Evaluationsebene – sie werden nicht in lauffähige Zielfunktionen überführt und steuern im Betrieb keine situativen Entscheidungen.

Das europäisch-japanische CARESSES-Projekt geht einen Schritt weiter³⁶: Kulturelle Werte — basierend auf Hofstedes Kulturdimensionen — werden zur Laufzeit in einer Wissensbasis kultureller Werte (OWL-Ontologie) repräsentiert und als Planungsrestriktionen für einen sozialen Roboter in der Altenpflege eingesetzt. CARESSES zeigt erstmals, dass normative Wertekategorien im Betrieb technisch wirksam sein können. Die Operationalisierung bleibt jedoch auf

35 Aimee van Wynsberghe, *Designing Robots for Care: Care Centered Value-Sensitive Design*, *Science and Engineering Ethics* 19(2):407–433, 2013.

36 Barbara Bruno et al., *The CARESSES EU-Japan Project: Making Assistive Robots Culturally Competent*. In: *Ambient Assisted Living. ForItAAL 2017. Lecture Notes in Electrical Engineering*, vol 540. Springer, Cham.

der Ebene von Wissensbasis-Einträgen und Planungs-Constraints — Werte fungieren als Kontextbedingungen, nicht als Optimierungsziele. Ein Modell innerer Zustände, eine dyadische Co-Regulations-Architektur oder eine an Care-Ethik ausgerichtete Zielfunktion fehlt.

KET schließt an beide Ansätze an und geht darüber hinaus: Care-ethische Werte (Sicherheit, Autonomie, Selbstachtung) werden nicht nur als Designkriterien oder Wissensbasiseinträge geführt, sondern auch als *Zielfunktionen* in eine simulationsbasierte Entscheidungsarchitektur eingebettet. Mentalisierungsbasierte Zustandssimulation(en) und dyadische Co-Regulation bilden die Berechnungsgrundlage; das Human-in-the-Loop-Prinzip verankert die Wertebindung im institutionellen Betrieb. Ethische Werte wirken damit als operative Zielfunktionen in situativen Entscheidungen (Kapitel 6): Laufzeit-Ethik statt nachträglicher Bewertung.

Die psychologischen Konzepte, auf denen KET aufbaut, werden im folgenden Kapitel dargestellt.

3. Psychologische Fundamente

3.1 Emotionen

Eine allgemeine Theorie der Emotionen gibt es nicht³⁷. Dies erschwert die theoretische Fundierung erheblich, da — im Gegensatz zu klar definierten Axiomen (vgl. Mathematik, Logik) oder empirisch messbaren Größen (vgl. Physik) — kein *belastbarer* Ausgangspunkt für darauf aufbauende Vorhaben existiert. Forschende, die mit Emotionskonzepten arbeiten, stehen daher vor der Herausforderung, wissenschaftliche Standards wie Nachvollziehbarkeit und Replizierbarkeit unter erschwerten Bedingungen zu wahren. Im Folgenden wird daher versucht, eine aktuelle, weiterentwickelte Definition als Grundlage für weitere Überlegungen heranzuziehen. Zunächst erfolgt ein Überblick über gängige Ansätze aus Sicht der Informatik.

37 Agnes Moors, An overview of contemporary theories of emotions in psychology. In Emotion theory: The Routledge comprehensive guide, 232–259, Routledge Academic, 2024.

Aus ingenieurtechnischer Perspektive werden Emotionen mithilfe verschiedener algorithmischer Verfahren, die auf wissenschaftlich fundierter Klassifikation basieren, erkannt. In jüngerer Zeit kommen dabei für sämtliche Modalitäten — etwa Gesicht, Körper oder Sprache — Methoden des maschinellen Lernens zum Einsatz³⁸, auf die in diesem Kapitel jedoch nicht näher eingegangen wird. Innerhalb einer Modalität (z. B. eine erhöhte Stimmfrequenz oder ein bestimmter Gesichtsausdruck) wird einem erkennbaren Muster ein Emotionswert zugewiesen, der entweder kontinuierlich (z. B. ein numerischer Wert zwischen -1,0 und 1,0) oder diskret (z. B. Freude, Überraschung) sein kann. Kontinuierliche numerische Werte stehen beispielsweise für Erregung (Arousal) und Wertigkeit (Valence), wie sie im Emotionsmodell von Russell³⁹ beschrieben werden. Diskrete, kategoriale Emotionswerte hingegen beziehen sich auf eindeutig benennbare Kategorien wie Freude oder Überraschung, die unter anderem im Modell von Ekman und Friesen⁴⁰ anhand bestimmter Gesichtsmuster definiert werden. Dieser Ansatz beruht auf der Annahme, dass Emotionen als objektiv messbare Muster identifizierbar sind. Neuere introspektive und qualitative Forschung deutet jedoch darauf hin, dass Emotionen komplexe innere Erfahrungen sind, die weder objektiv identifiziert werden können noch einen eindeutigen *Fingerabdruck* in beobachtbarem Verhalten, physiologischen Reaktionen oder neuronalen Mustern hinterlassen⁴¹.

Ein umfassenderes Konzept ist erforderlich. Für die Modellierung sozialer und selbststabilisierender Fähigkeiten einer künstlich empathischen Technologie stellt das Komponentenmodell der Emotionen eine geeignete Ausgangsbasis dar. Es integriert diverse relevante Aspekte der Emotionen, wie beispielsweise physiologische, kognitive, motivationale und expressive Komponenten (Abbildung 1).

38 Eman MG Younis et al., Machine learning for human emotion recognition: a comprehensive review, *Neural Computing and Applications*, 36.16: 8901–8947, 2024.

39 James A. Russell, A circumplex model of affect, *Journal of personality and social psychology* 39.6: 1161, 1980.

40 Paul Ekman und Wallace V. Friesen, Constants across cultures in the face and emotion, *Journal of personality and social psychology*, 17.2: 124, 1971.

41 Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt, 2017.

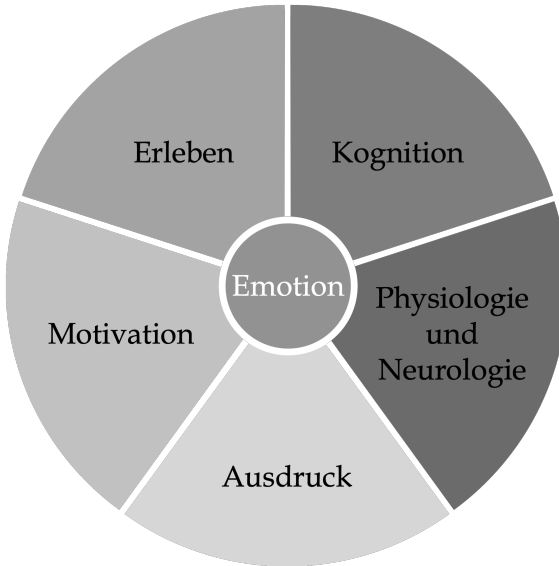


Abbildung 1: Visualisierung von Emotionskomponenten. Eigene Darstellung in Anlehnung an⁴², Abbildung 7.2.

In vielen wissenschaftlichen Arbeiten — insbesondere in den Ingenieurwissenschaften — werden Emotionen jedoch häufig auf eine einzelne Komponente reduziert, etwa auf physiologische Reaktionen, neuronale Aktivität, Gesichtsausdrücke oder Sprachmuster. Wie erwähnt, werden diese meist als kontinuierliche Werte oder als kategoriale Zustände modelliert. Eine solche Sichtweise vernachlässigt zentrale Komponenten oder Prozesse wie subjektive Erfahrung, Kognition, Motivation und Regulation. Zudem ist das Zusammenspiel der einzelnen emotionalen Komponenten komplex: Ein freudiger Gesichtsausdruck kann etwa mit sehr unterschiedlichen Regulationsprozessen und subjektiven Erfahrungen einhergehen³⁸.

Die Konzeption und Entwicklung von künstlich empathischer Technologie sollte unter anderem von folgenden Fragen geleitet werden: Ist es nützlicher, eine einzelne emotionale Komponente oder mehrere Komponenten zu erkennen? Könnte die subjektive

42 Andreas Eder und Tobias Brosch, Emotion. In: Allgemeine Psychologie. J. Müseler & M. Rieger (Ed.). Berlin: Springer, 2017.

Erfahrung der Emotion der wichtigste zu berücksichtigende Aspekt sein? Wie beeinflussen sich die Komponenten gegenseitig und wie spiegelt sich das im Ansatz wider? All diese Aspekte müssen sich in den Modellen für eine künstlich empathische Technologie (Kapitel 5) widerspiegeln.

3.2 Empathie und Mentalisierung

Empathisches Verhalten spiegelt stets Werte und Normen wider, die in der jeweiligen Situation als angemessen gelten. In alltäglichen wie auch in wissenschaftlichen Beschreibungen wird ein zu einer Situation passendes Verhalten häufig als empathisch bezeichnet⁴³. Ob dem beobachtbaren Verhalten jedoch tatsächlich empathische mentale Prozesse zugrunde liegen, lässt sich von außen nicht nachweisen⁴⁴.

Seit Aristoteles wird Empathie damit in Zusammenhang gebracht, wie andere eine Situation erleben und welche Emotionen sie dabei empfinden. Eine allgemeingültige Definition von Empathie gibt es nicht⁴⁵. Empathie wird in dieser Arbeit als eine Reihe von Fähigkeiten definiert, z. B. *Perspektivenübernahme*, sich imaginativ in die Gefühle und Handlungen fiktiver Charaktere hineinzusetzen, die Gefühle anderer oder Unbehagen in intensiven zwischenmenschlichen Situationen einschätzen⁴⁶. Zudem ist es notwendig, dass Empathie prozessorientierte Aspekte enthält: »*die Beteiligung psychologischer Prozesse, die dazu führen, dass eine Person Gefühle hat, die eher mit der Situation eines anderen als mit ihrer eigenen übereinstimmen*«⁴⁷.

43 Martin L. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge University Press, 2001.

44 Jean Decety und Philip L. Jackson, *The Functional Architecture of Human Empathy*, *Behavioral and Cognitive Neuroscience Reviews* 3(2), 71–100, 2004.

45 Nancy Eisenberg, *Empathy-related Responding and Prosocial Behaviour*. In: Novartis Foundation, *Symposium*, vol. 278, p. 71. Wiley Online Library, 2006.

46 Mark H. Davis, *Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach*, *Journal of Personality and Social Psychology*, 44, 113–126, 1983.

47 übersetzt aus dem Englischen »*the involvement of psychological processes that make a person have feelings that are more congruent with another's situation than with his situation*« aus ⁴³, Seite 30.

Eine nochmals vertiefte Betrachtung des Konzepts Empathie ist notwendig, um anhand dessen zu beschreiben, welche Eigenschaften eine künstlich empathische Technologie aufweist.

Eine gängige Definition von Empathie unterscheidet zwischen *kognitiver Empathie* und *affektiver Empathie*⁴⁸. Der kognitive Aspekt befasst sich mit der Fähigkeit, die Gedanken anderer Menschen (Überzeugungen, Wünsche, Absichten) zu erschließen sowie mit den damit verbundenen kognitiven Prozessen. Diese Fähigkeiten werden in der Entwicklungspsychologie durch die *Theory of Mind* (ToM) beschrieben⁴⁹. Affektive Empathie befasst sich mit den emotionalen Reaktionen des Einzelnen auf die Gefühle einer anderen Person, etwa durch emotionale Ansteckung oder stellvertretend mitempfundene Reaktionen, und unterscheidet sich damit von den kognitiven ToM-Fähigkeiten. Empathie ist damit *wertneutral*: Sie kann sich auf positive wie negative Zustände beziehen und führt nicht automatisch zu prosozialem Handeln; sie kann sogar belastend sein (z. B. empathic distress)⁵⁰. Diese beiden Empathiekonzepte werden häufig verwendet, um Empathie für emotional adaptive Technologie im SIA computergestützt zu modellieren⁵¹. Darüber hinaus ist es notwendig, Empathie von *Mitgefühl* (Compassion) zu trennen. Mit Letzterem ist prosoziale Motivation und Handlungsbereitschaft gemeint, um Leiden zu mindern. Das ist typischerweise von Wärme, Fürsorge und Annäherung⁵² begleitet.

Empathie ist stets an Beziehungen gekoppelt und folgt kulturellen Erwartungen⁵³. Soziale Relationen prägen unterschiedliche Interaktionslogiken und -strategien. Kulturen unterscheiden sich bei-

48 Nancy Eisenberg und Paul A. Miller, The Relation of Empathy to Prosocial and Related Behaviors, *Psychological Bulletin*, Volume 101(1), 91–119, 1987.

49 Josef Perner, Understanding the Representational Mind, *Pedagogische Studien*, 71.3, 1994.

50 Tania Singer und Olga M. Klimecki, Empathy and Compassion. *Current Biology*, 24(18), R875–R878, 2014.

51 Ana Paiva et al., Empathy in Virtual Agents and Robots: A Survey. *ACM Transaction on Interactive Intelligent Systems* 7, 3, Article 11, 2017.

52 Daniel C. Batson, These Things Called Empathy: Eight Related but Distinct Phenomena, In: *The Social Neuroscience of Empathy*, MIT Press, 2009.

53 Alan P. Fiske, The Four Elementary Forms of Sociality, *Psychological Review*, 99(4), 689–723, 1992.

spielsweise auch in der Strenge bei der Befolgung von Normen⁵⁴. Kulturabhängige Aspekte des Verhaltens, die technologische Anpassung daran sowie die Simulation möglicher mentaler Zustände, einschließlich spezifischer interner (unterbewusster) Repräsentationen und Prozesse, die auf einem allgemeineren Modell von Emotionen (Kapitel 3.1) basieren, gehen über die derzeitigen Ansätze emotionaler Technologie hinaus und erfordern ein breiteres Konzept als das der Empathie, das diese Fähigkeiten integriert. Zudem werden in den bisher vorgestellten Konzepten die Beziehungen zwischen Menschen nicht explizit behandelt. Wissenschaftliche Arbeiten zu mentalen Fähigkeiten im Kontext der klinischen Entwicklungs- und Bindungsforschung eignen sich hierfür⁵⁵. Die *Mentalisierung*⁷ ist eine emotional eingebettete, beziehungsabhängige Fähigkeit, eigenes und fremdes Verhalten als durch mentale Zustände verursacht zu verstehen und diese Deutungen zu regulieren. Wer mentalisiert, fragt sich also nicht nur »Was zeigt diese Person?«, sondern »Was geht in ihr vor — und was löst mein Handeln in ihr aus?«. Mentalisierung⁵⁶ ist ein breites, bindungstheoretisch verankertes Konzept. Es beschreibt einen Prozess, der kognitive und emotionale Aspekte integriert. Gerade der prozessorientierte Charakter bietet eine gute Grundlage für die Operationalisierung mittels geeigneter Computermodele und kann als Simulations- oder Vorhersageprozess für soziale Handlungen und deren mögliche Auswirkungen auf andere angesehen werden. Konzeptuell motiviert ist dies durch die Beschreibung der menschlichen Wahrnehmung (z. B. Sehen, Hören) als Vorhersageprozesse⁸¹. Allgemein gesehen, umfassen diese Vorhersagen sogar Handlungstendenzen. Die Simulationen werden kontinuierlich mit den tatsächlichen Sinneseindrücken verglichen. Wenn sie nicht übereinstimmen, werden die Fehler behoben und anschließend neue Vorhersagen und Simulationen erstellt.

Was die Simulation innerer Erlebniszustände und damit verbundener Regulationsprozesse anderer Menschen angeht, stellt die Rea-

54 Michele J. Gelfand et al., Differences Between Tight and Loose Cultures: A 33-nation Study. *Science*, 332(6033), 1100–1104, 2011.

55 György Gergely und Zsolt Unoka, Attachment, Affect-Regulation, and Mentalization: The Developmental Origins of the Representational Affective Self, 2008.

56 Lois W. Choi-Kain und John G. Gunderson, Mentalization: Ontogeny, Assessment, and Application in the Treatment of Borderline Personality Disorder, *American Journal of Psychiatry* 165.9: 1127–1135, 2008.

lisierung rechnergestützter *Mentalisierung* als prädiktiver Prozess eine enorme Herausforderung dar.

Zentrale Aspekte wurden in Vorarbeiten bereits realisiert: MARS-SI⁵⁷ realisiert erstmals die rechnerische Trennung zwischen Beobachtbarem (sozialen Signalen) und Innerem (Bewertungs- und Regulationsprozessen) sowie die Modellierung struktureller und situativer emotionaler Information. DEEP⁵⁸ erweitert dies um die repräsentative Einbeziehung individueller und situativer Faktoren des Gegenübers. Das sind erste Schritte zur Modellierung eines Gegenübers. Was hingegen bislang in keinem HCI- oder Affective-Computing-System realisiert wurde, sind die explizite Trennung der eigenen Co-Regulation von der simulierten Co-Regulation des Gegenübers, der Einfluss einer explizit modellierten Beziehungsrepräsentation auf Mentalisierungsprozesse sowie die simulative Folgenabschätzung von Systemhandlungen auf Erleben und Beziehung. Diese Aspekte sind originäre konzeptionelle Beiträge des KET-Rahmenwerks und Gegenstand zukünftiger technischer Realisierung (Kapitel 5.3 und 5.4). Generell verknüpft dieser Ansatz Beziehung, affektive und kognitive Aspekte in einem gemeinsamen Modell — einem umfassenderen Empathiemodell, das innere mentale Zustände und kulturell geprägtes Verhalten als einen Prozess beschreibt. Auf dieser Grundlage lassen sich empathische Reaktionen generieren, die die kulturellen Normen des empathischen Verhaltens widerspiegeln. Funktional lässt sich der prädiktive Simulationsansatz mit der funktionalen Architektur der menschlichen Empathie in Verbindung bringen, die drei Kernkompetenzen der menschlichen Empathie beschreibt⁵⁹: 1) die Fähigkeit, die Gefühle anderer zu teilen, 2) die kognitive Fähigkeit, die Gefühle anderer zu simulieren, und 3) die »sozial vorteilhafte« Absicht, mitfühlend auf die negativen Emotionen einer Person zu reagieren (vgl. oben, Unterscheidung zwischen Empathie und Mitgefühl).

57 Patrick Gebhard et al., MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation, 2018.

58 Tanja Schneeberger et al., The Deep Method: Towards Computational Modeling of the Social Emotion Shame driven by Theory, Introspection, and Social Signals, IEEE Transactions on Affective Computing, 15.2, 417–432, 2023.

59 Decety und Jackson, The Functional Architecture of Human Empathy, Behavioral and cognitive neuroscience reviews 3(2), 71–100, 2004.

Die in Kapitel 5 vorgestellte Organisation co-regulierender Fähigkeiten für eine künstlich empathische Technologie basiert auf dem oben genannten Konzept der Mentalisierung. Repräsentationen von Werten und Normen aus individueller Sicht sowie aus der Perspektive sozialer Gruppierungen (Familie, Kohorte) und des Beziehungseinflusses werden explizit in das Rahmenwerk integriert, was bislang meist implizit erfolgt ist. Sie repräsentieren individuelle Moralsysteme und vermitteln die Verbindung zur Ethik. In diesem Kontext ist es wichtig, zu definieren, was genau damit gemeint ist, wenn Menschen ihre Empathie (Fähigkeit) beziehungsweise ihr empathisches Verhalten beschreiben, um anhand dieser Simulationen generieren zu können. Dazu bedarf es überprüfbarer Kriterien, z. B.: 1) wie gut Einschätzungen über Gedanken und Gefühle anderer mit den eigenen Angaben oder unabhängigen Beurteilungen übereinstimmen (Empathic Accuracy)⁶⁰, 2) ob sich die andere Person verstanden und angemessen behandelt fühlt (Perceived Responsiveness)⁶¹ und 3) ob Kommunikation und Zusammenarbeit spürbar leichter, kohärenter und weniger anfällig für Missverständnisse werden.

Eine Befragung von Personen nähert sich diesen inneren Werten und Prozessen, liefert jedoch zunächst sozial konforme, bewusste Beschreibungen. Unbewusste Werte werden durch Regulationsmechanismen erst in vertrauensvoller Umgebung zugänglich⁷. Nicht jede Person verfügt über die Fähigkeit, unbewusste Prozesse zu introspektieren. Ein vorgelagertes Screening auf *Psychological Mindedness*⁶² (PM) ermöglicht eine differenzierte Vorgehensweise: Für Personen mit ausreichender PM ermöglicht die introspektive Befragung¹²⁷ einen gezielten Zugang, die auch für die DEEP-Methode⁵⁸ angewendet wurde. Für Personen mit geringer PM oder unzureichenden Daten kann KET mithilfe eines hinreichend großen Interaktionskorpus aus gelernten Verhaltenskorrelationen Annahmen über unbewusste Werte ableiten — wofür jedoch weitere Forschung zur Validierung solcher Korrelationen im Pflegekontext erforderlich ist. Beide Zugänge ergänzen die in Kapitel 8.1 beschriebene Drift-

60 William Ickes, *Empathic Accuracy*, New York: Guilford Press, 1997.

61 Reis und Lemay, Perceived partner responsiveness. *Current Opinion in Psychology*, 13, 6–10, 2016.

62 Stephen A. Appelbaum, *Psychological-mindedness: Word, concept and essence*, *International Journal of Psycho-Analysis*, 54(1), 35–46, 1973.

Analyse, die unbewusste Präferenzen anhand beobachteter Verhaltensänderungen erschließt.

Erste Arbeiten versuchen, mithilfe großer Sprachmodelle (LLMs) Hinweise auf innere mentale Repräsentationen zu gewinnen⁶³. LLMs können durchaus als Werkzeuge zur Verständnisaufbereitung, Hypothesengenerierung und Normenkontextualisierung eingesetzt werden. Die Simulationsprozesse der Mentalisierung sollten jedoch in einer separaten, transparenten, normsensitiven Inferenz- und Planungsarchitektur untergebracht werden, die LLM-Informationen nutzt. Sicher ist, dass es Prüfungs- und Korrekturmechanismen geben muss, etwa Human-in-the-Loop⁶.

Neuere Arbeiten berichten, dass LLMs funktionale »Emotions«-Korrelate aufweisen, die das LLM-Verhalten kausal beeinflussen⁶⁴. Diese unterscheiden sich jedoch strukturell von dem, was KET erfordert: LLM-Emotions-Vektoren sind implizit in Modellaktivierungen verteilt — keine explizit benannten, persistenten, personenspezifischen Zustandsrepräsentationen mit prüfbarer Prozessarchitektur, wie sie KET für transparente, erklärbare Co-Regulation braucht^{65,66}.

LLMs sind prinzipiell nicht als Ersatz für eine prozessorientierte Mentalisierung geeignet. Sie generieren statistisch plausible Aussagen über mentale Zustände, verfügen aber über keine inneren Regulationsprozesse und Zustandsrepräsentationen, wie sie KET erfordert

Grundsätzlich bleibt die Frage, welche Anteile der Mentalisierung technisch simulierbar sind. Yirmiya und Fonagy⁶⁷ argumentieren, dass KI-Systemen reziproke Mentalisierung fehlt: Sie können nicht

63 *Sayed Hossain et al.*, AutoPsyC: Automatic Recognition of Psychodynamic Conflicts from Semi-structured Interviews with Large Language Models. arXiv preprint, 2025.

64 *Anthropic Interpretability Team*, Emotion-like representations in large language models, <http://anthropic.com/research/emotion-concepts-function>, 2025 (zuletzt aufgerufen am 24. April 2026).

65 *Michal Kosinski*, Evaluating large language models in theory of mind tasks, *Proceedings of National Academy of Science. U.S.A.* 121 (45), 2024.

66 *Jennifer Hu, Felix Sosa und Thomer Ullman*, Re-evaluating Theory of Mind evaluation in large language models. *Philosophical Transactions of the Royal Society B*, 2025.

67 *Karen Yirmiya und Peter Fonagy*, Mentalizing Without a Mind: Psychotherapeutic Potential of Generative AI, *Journal of Medical Internet Research* 27:e79156, 2025.

erfahren, verstanden zu werden — eine Asymmetrie, die sie als »*the sound of one hand clapping*« bezeichnen. Diese Sicht markiert eine prinzipielle Grenze, die KET explizit anerkennt: KET *simuliert* Mentalisierungsprozesse ohne eigene Erlebnisfähigkeit (Kapitel 2) und ist kein therapeutischer Agent, sondern ein Assistenzsystem. Die fehlende Reziprozität wird institutionell getragen: Durch das Human-in-the-Loop-Prinzip bleibt die Pflegekraft Trägerin der relationalen Präsenz und der epistemischen Vertrauensbildung; KET strukturiert und erweitert ihre Mentalisierungskapazität, ersetzt sie jedoch nicht.

4. Von CASA zu relationaler KET

Menschen behandeln Medien und Computer wie reale Personen. Diese Beobachtung markiert den Schritt vom Werkzeuggebrauch zur sozialen Interaktion mit technischen Systemen und verändert damit auch die normativen und psychologischen Bezugspunkte der Gestaltung⁶⁸.

Könnten Menschen unbewusst ähnliche Beziehungen zu SIAs aufbauen wie zu anderen Menschen — mit allen normativen und psychologischen Konsequenzen⁶⁹? Das Paradigma *Computers As Social Actors (CASA)*⁷⁰ und dessen Verallgemeinerung im Buch *Media Equation*⁷¹ gehen davon aus, dass Menschen Medien und Computer wie reale Personen behandeln und dabei unbewusst soziale Skripte auf die Interaktion mit sozialen Technologien übertragen. Nass und Reeves belegen dies empirisch: Ein Computer wird besser bewertet, wenn er von einem anderen Computer gelobt wird, als wenn er

68 Anna Strasser, *From Tool Use to Social Interactions, Social Robotics and The Good Life*, 2022.

69 Jean Piaget, *Der Aufbau der Wirklichkeit beim Kinde*. Klett-Cotta, 1974.

70 Clifford Nass, Jonathan Steuer und Ellen R. Tauber, *Computer are social actors*, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, 72–78, 1994.

71 Bryon Reeves und Clifford Nass, *The Media Equation*, Cambridge, UK, 1996.

sich selbst lobt⁷²; Menschen helfen eher einem Computer, der ihnen zuvor geholfen hat, als einem, der ihnen nicht geholfen hat⁷³.

Das CASA-Paradigma berücksichtigt innere Prozesse wie soziale Regeln, Normen und Erwartungen, liefert jedoch keine Leitlinien dazu, wie diese zusammenhängen und sich gegenseitig beeinflussen. Eine solche Strukturierung ist notwendig, um interaktive Systeme individuell anpassen zu können.

Erleben, Kognition, Emotionen und soziales Verhalten sind zentrale Bausteine des menschlichen Erlebens und des damit verbundenen Konzepts der menschlichen Empathiefähigkeitv (Kapitel 3.1).

5. KET-Rahmenwerk – Regulation, Individuum, Beziehung, Interaktion

Dieser Abschnitt erläutert die interne Organisation von KET: Welche Prozesse und Repräsentationen erforderlich sind, damit das System sozial agieren und sich selbst regulieren kann. Ausgangspunkt sind die menschlichen Mentalisierungsfähigkeiten (Kapitel 3.2); daraus werden die zentralen technischen Komponenten abgeleitet.

Das Rahmenwerk ermöglicht es, Anforderungen, Bedingungen und Richtlinien für die Realisierung konkreter KET-Systeme abzuleiten — insbesondere im Kontext der Pflege. Das Rahmenwerk adressiert die interdisziplinäre Verknüpfung verschiedener Disziplinen bei der KET-Entwicklung, um Anforderungen und Bedingungen unterschiedlicher Perspektiven sowie deren Implikationen sichtbarer und greifbarer zu machen und den notwendigen Forschungsbedarf zu verdeutlichen.

Das KET-Rahmenwerk behandelt die Dyade *System–Patient* als Kern: KET ist als autonom operierendes Assistenzsystem konzipiert, das unter institutioneller Human-in-the-Loop-Aufsicht direkt mit der pflegebedürftigen Person interagiert. Die menschliche Pflegekraft tritt dabei nicht als direkter Interaktionspartner in Erschei-

72 Clifford Nass und Jonathan Steuer, Voices, boxes, and sources of messages: Computers and social actors, Human communication research, 19.4: 504–527, 1993.

73 Clifford Nass und Youngme Moon, Machines and mindlessness: Social responses to computers, Journal of social issues 56.1: 81–103, 2000.

nung, sondern ist strukturell als Governance-Instanz eingebunden. Zu berücksichtigen ist, dass Pflegebedürftige wahrscheinlich Attribute ihrer primären Bezugsperson auf das System projizieren — ein aus der Bindungs- und Mentalisierungstheorie erwartbares Phänomen. Das Beziehungsmodell (Abschnitt 5.4) trägt dem Rechnung, indem es solche projizierten Erwartungen, Vertrauensmuster und Rollenbilder explizit repräsentiert⁷⁴. Dieses Kapitel fokussiert bewusst auf die Dyade *Pflegebedürftige/r*–*KET*. Eine konzeptuell naheliegende Erweiterung bestünde darin, jede weitere dyadische Beziehung, z. B. *Pflegekraft*–*KET* oder *Angehörige*–*KET* als eigenständig nach demselben Ansatz zu modellieren. Die Simulation der Beziehungen zu diesen externen Personen wäre ein weiterführendes Forschungsvorhaben, das den Rahmen dieses Kapitels übersteigt.

Kommunikationsmodelle bilden einen wichtigen Ausgangspunkt. Einen ersten Formalisierungsschritt lieferte das Übertragungsmodell von Shannon und Weaver⁷⁵, das trotz seiner ursprünglich nachrichtentechnischen Intention als Ausgangspunkt für Kommunikationsmodelle rezipiert wurde. Weiterentwicklungen davon berücksichtigen die Adaption der Ausgabewerte — also was ein System verbal, nonverbal, visuell oder haptisch in einer Interaktion von sich gibt^{76,77}; haptische Ausgaben kommen hinzu.

Transaktionsmodelle fokussieren auf parallel stattfindenden Austausch, der durch individuelle und kulturelle Werte gefärbt ist und Rollen sowie verschiedene Modalitäten umfasst⁷⁸. Weitere Arbeiten beschreiben, dass Kontextinformationen wie Umgebung, Situation

74 Auf Repräsentationsebene ist KETs Architektur prinzipiell auf $n > 2$ Entitäten erweiterbar, indem für weitere relevante Akteure eigene Zustands- und Beziehungsmodelle geführt werden. Die damit verbundenen Implikationen für Ziel-funktions-Aggregation und Wertekonflikt-Resolution gehen jedoch über den Rahmen dieses Kapitels hinaus.

75 Claude E. Shannon und Warren Weaver. *The mathematical theory of communication*, University of Illinois press, 1998.

76 Norbert Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, MIT press, 2019.

77 Steve Yohanan und Karon E. MacLean, *The Role of Affective Touch in Human-Robot Interaction: Human Intent and Expectations in Touching the Haptic Creature*. *Int J of Soc Robotics* 4, 163–180, 2012.

78 Dean Barnlund, *A Transactional Model of Communication*, *Communication Theory*, Routledge, 47–57, 2017.

und Beziehung eine wichtige Rolle in der Kommunikation spielen⁷⁹. Alle Ansätze haben gemeinsam, dass interne Regulationsprozesse nicht berücksichtigt wurden.

Für die Erweiterung dient die Theorie der Mentalisierung (Kapitel 3.2) als Ausgangsbasis. Der hier verfolgte Ansatz erweitert klassische ToM-basierte kognitive Modelle für Nutzende⁸⁰ um zusätzliche Prozesse und Repräsentationen. KET beschreibt erstmals folgende Aspekte prozessorientiert:

- 1) *Regulation: Integration von Emotion und Kognition*. Sie dient als zentrale mentale organisierende Einheit.
- 2) *Innere Prozesse und Repräsentationen des Individuums* mit interagierenden unbewussten Regulationsprozessen zur Wahrnehmung, Selbstregulation, Regulation des gezeigten Verhaltens sowie notwendigen Repräsentationen und abstrakten (vor)bewussten emotionalen Informationen.
- 3) *Beziehungen und Begegnungen mit Menschen und SIA* — mit beteiligten Prozessen und deren Interaktionen sowie den notwendigen Repräsentationen, einschließlich der Beziehung der Interaktionspartner aus Sicht des Individuums.
- 4) *Soziale Interaktion*, die auf den vorangegangenen Bausteinen (1–3) aufbaut und die Komplexität innerer Abhängigkeiten illustriert. Bevor das zentrale Konzept des Regulierungsprozesses mit den relevanten Komponenten erläutert wird, wird kurz auf die technischen Anforderungen eingegangen. Darauf aufbauend werden mentale Prozesse und die notwendigen Repräsentationen modelliert, die für die (Co-)Regulation — also Selbststabilisierung und soziale Interaktion — erforderlich sind.

79 Steven A. Beebe, Susan J. Beebe und Diana K. Ivy, *Communication: Principles for a lifetime*, Boston, MA: Allyn & Bacon, 2010.

80 Christelle Langley et al., *Theory of Mind and Preference Learning at the Interface of Cognitive Science, Neuroscience, and AI: A Review*, *Frontiers in Artificial Intelligence* 5: 778852, 2022.

5.1 Anforderungen an Sensorik, Erfassung und Interpretation

KET braucht eine robuste Fusion aus Sprache, Mimik, Gestik, Physiologie und Kontext^{81,82}. Nicht nur zur Erkennung, sondern auch zur Mentalisierung im Rollen- und Beziehungsrahmen. Technisch anspruchsvoll sind (a) Domänentransfer und Personalisierung bei variablen Signalen, (b) Abschätzung von Unsicherheiten in prädiktiven Modellen, (c) simulative Folgenabschätzung hinsichtlich Sicherheit, Autonomie und Selbstachtung, einschließlich Mindestgrenzen, sowie (d) Erklärbarkeit von KI-Modellen auf Interaktionsebene mit Bezug zu Alternativen und Werten. Passende algorithmische Ansätze für Predictive Processing oder Active Inference, Kostenfunktionen und Werteschwellen müssen mit klaren Fehlermodi und Reparaturpfaden implementiert werden. Diese Punkte werden in den folgenden Kapiteln konzeptuell genauer adressiert.

5.2 Regulation – Integration von Emotion und Kognition

Wenn eine Pflegekraft morgens eine vertraute pflegebedürftige Person trifft, weiß sie oft schon nach einem ersten Blick, ob es ein guter oder schwieriger Tag wird — noch bevor ein Wort fällt. Dieses Vorwissen entsteht nicht zufällig: Das Gehirn gleicht ständig ab, was es erwartet, mit dem, was es tatsächlich wahrnimmt (Kapitel 3.2). Weichen die Signale ab, reagiert es sofort. Genau dieses Prinzip der Vorhersage und Fehlerkorrektur wird im KET-Rahmenwerk genutzt.

Um interne Prozesse zu organisieren, die soziales und selbststabilisierendes Handeln ermöglichen, greift KET auf das Paradigma der Regulations- und Vorhersagesysteme zurück.

81 Helen Schneider et al., Datasets for Valence and Arousal Inference: A Survey, Proceedings of the Computer Vision and Pattern Recognition Conference, 2025.

82 Ronak Kosti et al., Context Based Emotion Recognition using EMOTIC Dataset, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42.11: 2755–2766, 2019.

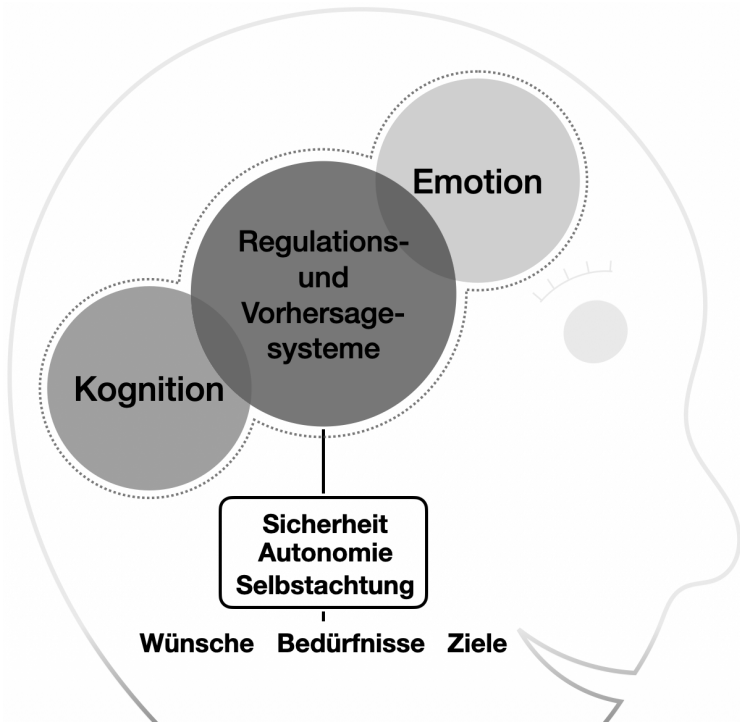


Abbildung 2: Integration von Emotion und Kognition in Regulations- und Vorhersagesysteme.

Die Abbildung 2 illustriert, wie Emotion und Kognition in Regulations- und Vorhersagesystemen zusammenwirken. Emotionen — verstanden unter anderem nach Moser und von Zeppelin als mentale Informationen über eigenes Handeln, Beziehungen, (soziale) Situationen und Selbstempfinden — fließen fortlaufend in die *Regulation*⁸³ ein. Wie dieses Prozessmuster zur Modellierung der eigenen Psyche, sozialer Interaktion und Beziehung verwendet werden kann, erläutert das Folgekapitel.

Aus Sicht des *Predictive Processing* (prädiiktive Verarbeitung) werden diese emotionalen Hinweise mit kognitiven Erwartungen ab-

83 Ulrich Moser und Ilka von Zeppelin. Die Entwicklung des Affektsystems, *Psyche* 50.1: 32–84, 1996.

geglichen; Abweichungen (Vorhersagefehler) beeinflussen Wahrnehmung, Entscheidungen und Handeln⁸⁴. Formale *Active-Inference-Modelle* präzisieren, wie aus inneren (interozeptiven) und äußeren (exterozeptiven) Hinweisen aktuelle emotionale Informationen abgeschätzt, daraus gelernt und Handlungen ausgewählt werden können⁸⁵. Warum Vorhersage nötig ist, verdeutlicht das Tennisspiel: Die menschliche Sensorik mit der weiteren mentalen Verarbeitung ist nicht schnell⁸⁶ genug, um Handlungen auf Spielereignisse hin reaktiv zu planen und auszuführen; kinematische und kontextuelle Hinweise ermöglichen jedoch oft eine erfolgreiche antizipative Vorbereitung der Bewegung⁸⁷. In der Pflege ist die Vorhersageaufgabe deutlich komplexer: Eine pflegebedürftige Person sagt »*Ich bin in Ordnung*«, hält aber den Arm schützend vor den Körper. KET muss beide Signale gleichzeitig verarbeiten, den Widerspruch im Beziehungskontext einordnen und antizipieren, wie Handlungen — berühren, zurücktreten, nachfragen — erlebt werden könnten. Mehrere unsichere Vorhersagen müssen dabei simultan abgewogen werden.

Die Modellierung der Regulation hat zum Ziel, zentrale psychisch organisierende Informationen wie *Sicherheit*, *Autonomie* und *Selbstachtung* unter Berücksichtigung von *Wünschen*, *Bedürfnissen* und *Zielen* zu optimieren. Diese Ausrichtung erfolgt nach den Arbeiten der Psychologen Moser und von Zeppelin in ihrer Arbeit *Die Entwicklung des Affektsystems*⁸⁰, die zudem Arbeiten aus dem Gebiet der Entwicklungs- und Neuropsychologie einbezieht⁵².

Die Wahl dieses Modells ist theoretisch begründet: Seine dreifache Differenzierung von strukturellen, situativen und kommunikativen Emotionen sowie die Verbindung von Regulationsprozessen mit sozialen Signalen erweisen sich als kompatibel mit Predictive-Processing-Ansätzen und konstruktivistischen Emotionstheorien⁴¹; eine ausführliche Begründung findet sich in⁵⁶. Die Abfolge der dabei be-

84 Clark, Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science, *Behavioral and Brain Sciences* 36.3, 2013.

85 Ryan Smith, Thomas Parr und Karl J. Friston, Simulating Emotions: An Active Inference Model of Emotional State Inference and Emotion Concept Learning. *Front. Psychol.*, 2019.

86 Ein gespielter Ball kann Geschwindigkeiten von 100km/h und höher annehmen.

87 Jaeho Shim et. al, The Use of Anticipatory Visual Cues by Highly Skilled Tennis Players. *Journal of Motivational Behavior*, 2005.

teiligten Prozesse ist in sechs Schritte gegliedert und kann als Computermodell(e) simuliert werden (Beispiel im Kapitel 6): 1) Wahrnehmen, 2) Vorhersagen, 3) Bewerten (Emotion) und Abgleichen (Kognition), 4) Handeln, 5) Rückmelden, 6) Lernen. Die Schritte 2) und 4) bis 6) sind nicht durch das Regulations-/Vorhersagemodell erklärt und erfordern zusätzliche Strukturen, die in den folgenden Kapiteln erläutert werden.

5.3 Innere Prozesse und Repräsentationen des Individuums

Die auf soziale und selbststabilisierende Fähigkeiten fokussierte Organisation interner Prozesse beginnt mit der menschlichen *Wahrnehmung* (Abbildung 3, blaue 1, fein gepunktete Linie).

Der Mensch nimmt zugleich die Außenwelt, den eigenen Körper und Aspekte seiner Innenwelt unbewusst und teilweise auch bewusst wahr. Aus exterozeptiven (Sehen, Hören, ...) sowie interozeptiv-/propriozeptiven Signalen (Körperzustände, Bewegung) entstehen laufend mentale Repräsentationen, die miteinander abgeglichen und durch mentale Prozesse Bedeutung zugewiesen werden^{88,89}. Dabei bilden sich: 1) eine Körperrepräsentation als Modell aktueller Bedürfnisse, Möglichkeiten und Grenzen, 2) eine *Repräsentation der Situation (Sit)* der Außenwelt mit Objekten und Ereignissen (Abbildung 3, unten, Teile markiert durch blaue 1, fein gepunktete Linie), die räumlich-zeitliche, aufgaben- und zielbezogene Kontexte integriert, sowie 3) eine *Repräsentation des Sozialen (Soz)*, die unter anderem Rollen⁹⁰ und Normen⁹¹ umfasst. Im Folgenden stehen 2) und 3) im Vordergrund. *Soz* umfasst nicht nur situative Rollenerwartungen, sondern auch soziale Normen, die für KET-Systeme explizit modelliert und geprüft werden müssen⁹².

88 Simon Grondin, *Psychology of Perception*, Springer Cham, 2016.

89 George N. Kenyon und Kabir C. Sen, *The Perception Process*. in *The Perception of Quality*, Springer, London, 2015.

90 Susan T. Fiske und Shelly E. Taylor, *Social cognition* (2nd ed.). McGraw-Hill Book Company, 1991.

91 Michael E. Hogg und Scott A. Reid, *Social Identity, Self-Categorization, and the Communication of Group Norms*. *Communication Theory*, 16: 7–30, 2006.

92 Anna Strasser, *Social Norms for Artificial Systems, Culturally Sustainable Social Robotics*, IOS Press, 295–304, 2020.

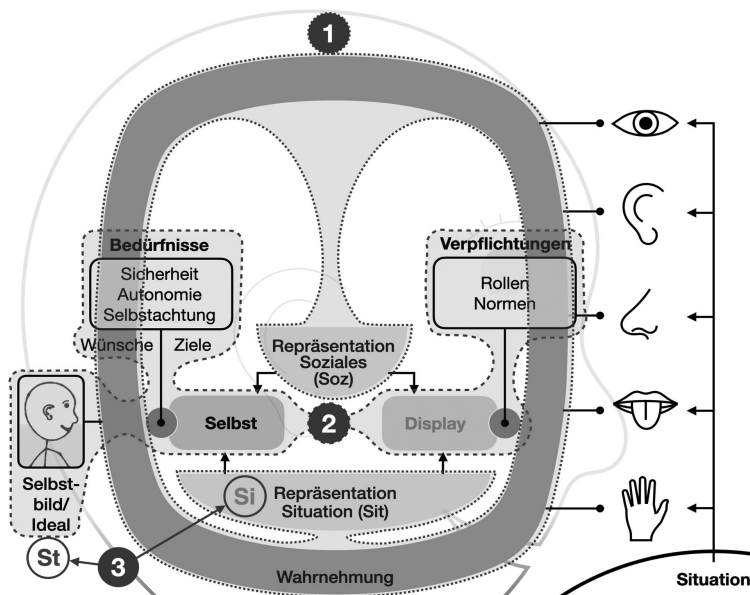


Abbildung 3: Struktur eines funktionalen Benutzermodells mit Wahrnehmungsprozess (1) und Regulationsprozessen (2, 3).

Die nächste Organisationseinheit sind die nach der Wahrnehmung nachgeschalteten Regulationsprozesse: *Selbst* und *Display*⁹³ (Abbildung 3, blaue 2, gestrichelte Linie). Die Selbst-Regulation hat die Aufgabe, die eigene Psyche hinsichtlich der Bedürfnisse, Wünsche und Ziele zu stabilisieren (Kapitel 5.2). Dieser Prozess verläuft weitgehend im Unbewussten. Die Display-Regulation optimiert das über den Körper (Pose, Gesten, Mimik, Sprache, ...) Ausgedrückte hinsichtlich der in der Situation geltenden sozialen Verpflichtungen, Rollenwerte und Normen. Anzumerken ist, dass es sich um verinnerlichte individuelle Werte handelt, die mit der Selbst-Regulation hinsichtlich der Bedürfnisse, Wünsche und Ziele abgestimmt sind⁹⁴. Auch die Display-Regulation erfolgt weitgehend unbewusst. Jedoch

93 Paul Ekman, und Wallace V. Friesen, The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding, Semiotica, 1.1: 49–98, 1969.

94 Amy H. Mezulis et al., Is There a Universal Positivity Bias in Attributions? A Meta-analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias, Psychological Bulletin 130.5:711, 2004.

können Menschen im Allgemeinen sich leichter darüber bewusst werden als über die Selbst-Regulation⁷. Dieser Modellierungsansatz respektiert, dass jeder Mensch ein individuelles inneres Moralsystem hat. Die Selbstregulation konstruiert ein *Selbstbild*: ein relativ stabiles, aber veränderbares Modell der eigenen Eigenschaften und ein *Idealbild*: eine normativ aufgeladene, angestrebte Version des Selbst. Beide entstehen aus der Integration aktueller Wahrnehmungen mit gespeicherten Erfahrungen, sozialen Vergleichen und kulturellen Erwartungen und werden durch Rückmeldungen des sozialen Umfelds kontinuierlich angepasst. Als Teil der Innenwelt wirken Selbst- und Idealbild wiederum rückkoppelnd auf Wahrnehmung und Interpretation: Sie lenken Aufmerksamkeit, bewerten Eingänge und beeinflussen Handlungsentscheidungen^{95,96}.

Der Theorie von Moser und von Zeppelin⁸⁰ folgend, generieren der Wahrnehmungsprozess und beide Regulationsprozesse die *strukturelle emotionale Information* (St) und die *situative emotionale Information* (Si) (Abbildung 3, blaue 3 und Kreise). Wie bewusst Menschen dieser Information sind, ist individuell unterschiedlich⁹⁷. St steht für die Bewertung des eigenen Ichs (Aussehen, Handeln, Rollenpassung, Werte, ...). Beispielsweise kann St in einem Bewerbungsgespräch die emotionale Information Scham annehmen, wenn das eigene Verhalten als unzureichend interpretiert wurde und das Gegenüber dies entsprechend nach kulturellen Normen kommuniziert hat⁹⁸; Si steht für die Bewertung der Situation, wird im Gedächtnis abgespeichert und dient zukünftigen ähnlichen Situationen als Referenzinformation. In dem Fall einer misslungenen Bewerbungssituation könnte dies auch die Information Scham sein, die dann bei einer zukünftigen Bewerbungssituation zunächst als emotionale situative Information reguliert werden muss. Diese differen-

95 Bareera Saeed et al., Shaping Self-Perception: The Intricate Relationship Between Self-Concept, Self-Image, and Body Dysmorphic Disorder, *Journal of Positive School Psychology*, 7.5:1043–1059, 2023.

96 George N. Kenyon und Kabir C. Sen, *The Perception of Quality*, Londra, Springer-Verlag, 2015.

97 Svenja Taubner et al., Internal Structure of the Reflective Functioning Scale, *Psychological Assessment*, 25.1 127, 2013.

98 Tanja Schneeberger et al., Can Social Agents Elicit Shame as Humans do?, 8th International Conference on Affective Computing and Intelligent Interaction, IEEE, 2019.

zierte Modellierung mentaler und emotionaler Information wurde erstmals durch MARSSI⁵⁷ technisch umgesetzt.

Der hier vorgestellte Ansatz erweitert klassische bewertungsbaasierte Emotionsmodelle^{99,100} und deren Realisierung als Softwareagent^{101,102,103}. Er basiert auf einer care-ethisch fundierten, rechnerisch expliziten Zielfunktion für Sicherheit, Autonomie und Selbstachtung mit Mindestgrenzen und adaptiven Gewichten (Kapitel 5.2). Daran gekoppelt sind Selbst- und Display-Regulation als zwei getrennte, jedoch hinsichtlich der Rollen und Normen abgestimmte Regelmechanismen. Selbst- und implizite Display-Regulation ist bereits durch verschiedene technische Ansätze realisiert, z. B. durch das MARSSI⁵⁷ oder verfeinert durch das DEEP-Modell⁵⁸.

Zudem erlaubt die hier erstmals vorgestellte technisch realisierbare Modellierung der Trennung in Selbst- und Display-Regulation, regulative Assistenzaktionen transparent zu bewerten (z. B. als Abweichung von Soll-Werten, Beispiel in Kapitel 6) und zugleich normative Grenzen zu erzwingen, was grundlegend für Pflege- oder Klinikszenerien ist. Technisch visiert das Modell die Kombination aus generativer aktiver Inferenz, gewünschten zukünftigen Beobachtungen (Prior-Preferences)¹⁰⁴ zur Zustands- und Präzisionsschätzung sowie sicherer Policyauswahl und Reinforcement-Learning¹⁰⁵ an. Dazu wird eine mehrdimensionale Personalisierung anhand von *Selbst- und Idealbild* sowie deren Unterschieden benötigt. Der konkrete Mehrwert ist dreifach: Ethische Werte werden rechnerisch

-
- 99 Andrew Ortony, Gerald L. Clore, and Allan Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, 2022.
 - 100 Klaus R. Scherer, *Appraisal Considered as a Process of Multilevel Sequential Checking*, *Appraisal Processes in Emotion: Theory, methods, Research*, 92.120:57, 2001.
 - 101 Stacy C. Marsella und Jonathan Gratch, *EMA: A Process Model of Appraisal Dynamics*, *Cognitive Systems Research* 10.1:70–90, 2009.
 - 102 Christian Becker-Asano, *WASABI: Affect Simulation for Agents with Believable Interactivity*, Vol. 319, IOS Press, 2008.
 - 103 Joost Broekens *et al.*, *Fine-grained Affective Processing Capabilities Emerging from Large Language Models.*, *Proceedings of 11th International Conference on Affective Computing and Intelligent Interaction*, IEEE, 2023.
 - 104 Karl Friston *et al.*, *Active Inference: A Process Theory*, *Neural Computation*, 29.1:1–49, 2017.
 - 105 Javier García und Fernando Fernández, *A Comprehensive Survey on Safe Reinforcement Learning*, *Journal of Machine Learning Research*, 16.1:1437–1480, 2015.

operationalisiert und sind damit prüfbar repräsentiert; Entscheidungen lassen sich durch Beitragszerlegung verändern und erklären; die Automatisierung erfolgt schrittweise und sicher unter menschlicher Aufsicht (Human-in-the-Loop).

5.4 Beziehung

Menschen sind soziale Wesen und leben in einer Welt, die maßgeblich von sozialen Beziehungen geprägt ist¹⁰⁶. Ein grundlegendes Bedürfnis des Menschen ist das Bedürfnis nach Zugehörigkeit. Die vorgestellte Modellierung berücksichtigt den von Asendorpf und Kollegen beschriebenen Paradigmenwechsel in der Psychologie, zur Beschreibung menschlicher Eigenschaften von der persönlichkeitsorientierten Psychologie zur beziehungsorientierten Psychologie. Da Beziehungen für Menschen zentral sind, sollten sie die Ausgangsposition bei der Modellierung und Simulation individueller Eigenschaften und sozialer Interaktionen bilden. Wie lässt sich das mit dem vorgestellten funktionalen Organisationsmodell von Individuen repräsentieren?

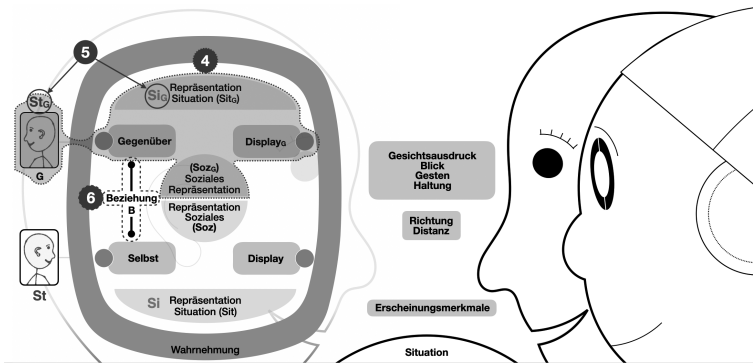


Abbildung 4: Erweiterung des Benutzermodells durch die Simulationsprozesse (4, 5) des Gegenübers sowie die dazugehörige Beziehung (6).

¹⁰⁶ Jens B. Asendorpf, Rainer Banse, und Franz J. Neyer, Psychologie der Beziehung. Hogrefe, 2017.

Bei einer Begegnung mit anderen Menschen wird, ausgehend von psychologischen Theorien, unter anderem von Moser und Piaget, zunächst ein passendes, erfahrungsbasiertes Wissens- und Erwartungsmuster über eine Rolle oder Situation (*Schema*)⁶⁶ aktiviert. Diese Schemata werden durch Eltern und andere in der Kind- und Jugendzeit angelegt und können im Lauf des Lebens verändert werden¹⁰⁷. Im Kontext dieser Arbeit wird davon ausgegangen, dass die Begegnung mit sozial-interaktiven Agenten auch Schemata aktiviert. Eine Hypothese, für die konvergierende Indizien sprechen: PARO-Studien belegen Schemaaktivierung und Beziehungserleben auch bei kognitiv eingeschränkten Personen^{29,115}; Billie zeigt Akzeptanz sozialer Interaktion in dieser Zielgruppe¹⁷. Für traumatisierte Personen kann die wahrgenommene Nicht-Menschlichkeit des Systems sogar einen Vorteil darstellen — geringere soziale Bedrohung und weniger Scham erleichtern das Anvertrauen, wie DeVault und Kollegen mit dem SimSensei-Kiosk bei Soldaten mit PTSD zeigen¹⁰⁸. Systematische psychologische Studien für die genannten Zielgruppen fehlen jedoch und sind notwendig.

Das aktivierte Schema steuert die Aufmerksamkeit darauf, wie soziale Signale und Handlungen interpretiert und als passend empfunden werden. Ein entscheidender Punkt ist, dass das SIA-Verhalten der Aufgabe *nicht* anhand statistischer Aussagen, sondern *individuell angepasst* modelliert werden sollte. Die Signale des Gegenübers (Abbildung 4, Mitte) werden als Instanzen bekannter Rollen- und Handlungsmuster (*Skripte*) interpretiert (Kapitel 9.2). Dazu gehört eine konsistente Repräsentation des *Selbst* (Abbildung 4, links unten) sowie des Personenmodells des Gegenübers *G* (Abbildung 4, links oben). Zum Gegenüber gehören Erfahrungsepisoden (Szenen) mit Informationen über Rollen und Ziele sowie über die Intentionen der Beteiligten, mit emotionalen Informationen und Handlungsmustern, die durch Interaktionen entstanden sind. Eine computerbasierte Modellierung dieser Aspekte in einem integrierten Modell gibt es derzeit nicht.

107 Susan T. Fiske und Shelley E. Taylor, Social Cognition: From Brains to Culture, 2020, Kapitel 4.

108 David DeVault et al., SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support, Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2014.

Das hier vorgestellte Modell sieht vor, dass das Gegenüber zunächst durch das Kopieren von 1) Selbst- und Display-Regulations-Prozessen und 2) der Repräsentation der Situation (Sit_G) und des Sozialen (Soz_G) simuliert wird (Abbildung 4, blaue 4, fein gepunktete Linie). Der tiefgestellte Index $_G$ steht im Folgenden durchgängig für das *Gegenüber*. In diesem Fall, wie das Gegenüber seine Situation (Sit_G) und sein soziales Umfeld (Soz_G) wahrnimmt. Diese Prozesse und Repräsentationen erzeugen ein Objektmodell des Gegenübers G ; nach denselben Prinzipien, wie auch das Selbstbild erzeugt wird (Kapitel 5.3).

Nach Fonagy sind die Inhalte von Sit_G und Soz_G , die vom Gegenüber angenommen werden, mit den eigenen (unbewussten) Fähigkeiten zur Mentalisierung (Kapitel 3.2) gekoppelt. Sit_G repräsentiert die Annahmen des Individuums über die Wahrnehmung der Außenwelt, einschließlich Objekten und Ereignissen des Gegenübers. Die Anwesenheit dieser Repräsentation ist besonders relevant, wenn das Gegenüber körperliche oder psychische Einschränkungen hat, die den Annahmen entsprechen. Soz_G repräsentiert die Annahmen des Individuums über die sozialen Werte (Rolle, Kompetenz, Wohlwollen, erwartete Aufgaben) des Gegenübers. Dies können auch angenommene oder visionierte Werte sein, die mit dem beschriebenen Wirkmechanismus der Beziehung dazu führen, dass die eigene Selbst-Regulation gestärkt wird. Das heißt, dass darin auch angenommene oder gewünschte Rolle(n), Aufgaben, Historien usw. enthalten sind. Inhalte sind zu finden, die letztendlich dazu dienen, die mit der Selbst-Regulierung verbundenen Werte zu stärken. Eine große Herausforderung in der Pflegepraxis besteht darin, diese phantasierten Werte zu erkennen und die Interaktion entsprechend zu moderieren. Eine weitaus größere ist es, dies mittels KET-Systemen festzustellen, Abweichungen zwischen angenommenem und tatsächlichem Erleben zu erkennen und auszugleichen.

Ansätze dazu, wie das Personenmodell des Gegenübers G strukturiert ist und welche Informationen es enthält, sind nicht Teil dieser Übersicht. Konkrete Ansätze könnten Arbeiten zu Archetypen¹⁰⁹ liefern, die in der datengetriebenen Modellierung von Nutzenden

109 Carl Gustav Jung, *Four Archetypes*, Routledge, 2014.

(Personas)¹¹⁰ eingesetzt werden. Eine Verbindung zu den hier vorgestellten Prozessen und Repräsentationen besteht derzeit nicht und bedarf zukünftiger Forschung.

Parallel dazu werden situative (Si_G) und strukturelle (St_G) emotionale Informationen des Gegenübers simuliert (Abbildung 4, blaue 5 und Kreise). Informationen, die einfacher und bewusster zugänglich sind als die konkreten Inhalte von Soz_G und Sit_G , die für deren Simulation verwendet werden. Diese Werte stehen in Beziehung zu den Theorien über Emotionen sowie zu kognitiver und emotionaler Empathie (Kapitel 3.2). Für eine technische Realisierung dieser emotionalen Werte können bestehende Ansätze benutzt werden (Ende Kapitel 5.3).

Wie erwähnt, wird ein Gegenüber durch Kopieren der Selbst- und Display-Regulation sowie der notwendigen Repräsentation der Situation und des Sozialen simuliert.

Beziehungen werden als partnerspezifische, relativ stabile Repräsentationen modelliert, deren Wahrnehmung, Bewertung und Handlungsbereitschaft systematisch reziprok moduliert werden. Im Sinne des Beziehungsparadigmas¹⁰³ enthält die Beziehungsrepräsentation B unter anderem geteilte Handlungshistorien und Routinen, Rollen (z. B. Pflegekraft–Pflegebedürftige/r), Vertrauen und Verlässlichkeit, Nähe und Distanz, Macht- und Abhängigkeitsasymmetrien sowie normative Erwartungen. Die unterschiedlichen Repräsentationen der Situation speichern dies gegliedert (z. B. Ort, Zeit, Aufgabe, Risiken, Routinen) für beide Individuen.

Dieselben sozialen Signale werden durch die simulierte Display-Regulation des Gegenübers interpretiert und für die Gegenüber-Regulationssimulation genutzt. Über die Beziehung B entfalten sie je nach Kontext unterschiedliche Wirkungen auf die Selbst-Regulation einer Person und die Effizienz von Co-Regulation.

Die Selbst-Regulation und die damit gekoppelten inneren Werte der Sicherheit, Autonomie und Selbstachtung werden durch (a) eigene Strategien, (b) situative Bedingungen und (c) Handlungen des Gegenübers beeinflusst. Die Beziehung (B) (Abbildung 4, blaue gestrichelte Linie) bestimmt, unter welchen Bedingungen Co-Regulation wirksam werden kann: Eine tragfähige, geklärte Beziehung er-

110 Bernard J. Jansen et al., Data-Driven Personas, Springer Nature, 2022.

möglicht und verstärkt beziehungsnahe Co-Regulation (z. B. strukturierendes Führen); unklare Einwilligung, Rollenunklarheit oder erlebte Verletzungen dämpfen oder blockieren sie. Da Selbst-Regulation — wie in Kapitel 5.3 dargestellt — der primäre innere Prozess zur Optimierung von Sicherheit, Autonomie und Selbstachtung ist, tritt Co-Regulation nicht an ihre Stelle, sondern wird zugelassen oder eingefordert, wenn sie benötigt wird und die Beziehung sie trägt. *B* liefert damit die Repräsentation, auf deren Grundlage der Einfluss der Co-Regulation in Kapitel 7.2.2 modelliert wird.

Co-Regulation (durch Handlungen anderer, beispielsweise einer Pflegeperson) wird akzeptiert oder gefordert, wenn sie zur Rolle der anderen Person passt (repräsentiert in *Soz_G*) und als angemessen erwartet wird — beziehungsweise wenn eigene innere Normen (repräsentiert in *Soz*) gewahrt bleiben. Co-Regulation kann im Konflikt mit allgemeinen Normen stehen — aber auch mit inneren Bindungsmustern, die die Annahme von Hilfe erschweren, selbst wenn äußere Bedingungen erfüllt sind. Das erfordert, dass jede regulative Assistenzaktion vorab auf Passung zu *Sit*, *Soz* und *Sit_G* sowie zu *Soz_G* geprüft wird. Dieselbe Handlung kann je nach Situation und den gegenseitigen Zuschreibungen ganz unterschiedlich wirken.

5.5 Soziale Interaktion

Die Organisation sozialer Interaktion ist für KET zentral. Dafür ist ein Prozessmodell notwendig, das die zuvor vorgestellten Prozesse und Repräsentationen miteinander verknüpft, Abhängigkeiten modelliert und die Komplexität verdeutlicht: co-regulierende soziale Interaktion, innere Regulationsprozesse sowie beteiligte Repräsentationen wie Normen und Rollen, Situationen, Beziehungen, Selbst- und Personenmodelle. Für Erschaffende von KET-Systemen dient das Vorgestellte dazu, die psychischen Prozesse und internen emotionalen Informationen von Nutzenden durch ein Computermodell zu simulieren und darauf aufbauend Handlungsoptionen passgenauer zum Individuum und zu kulturellen Werten zu modellieren.

Im Pflegekontext sind regulative Assistenzaktionen stets an gängige Vorgehensmodelle und care-ethische Standards gebunden (Kapitel 1). Anders als Menschen sollen sich KET-Systeme die Inhalte von *Soz_G* und *Sit_G* nicht frei wählen oder ihre eigenen kopieren.

Je nach Definition und Einschätzung des Gegenübers (z. B. hinsichtlich Rolle, Aufgabe, Historie) müssen diese Werte spezifisch gesucht, überprüft und später in der Interaktion angepasst (Kapitel 2) werden. Diese Schritte müssen sowohl in der Konzeption als auch in der Realisierung berücksichtigt werden.

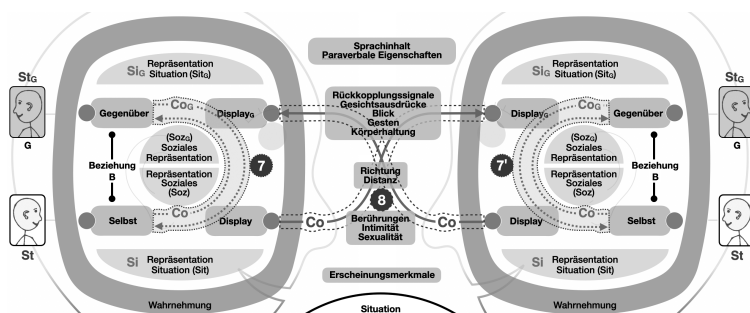


Abbildung 5: Regulation und Co-Regulation mit innerer (7, 7') und äußerer Interaktion (8).

Das Prozessmodell der sozialen Interaktion in der Dyade (Abbildung 5) ergänzt die vorgestellten Modelle um *innere Kommunikationspfade* (Abbildung 5, 7 und 7') und *äußere Kommunikationspfade* (Abbildung 5, 8), die die innere regulative Prozesse des Einzelnen und einer Dyade miteinander verknüpfen. Auf allen Pfaden, inneren wie äußeren, werden die in Abbildung 5, Mitte, genannten Modalitäten genutzt, um Erlebtes sowie Selbst- und Beziehungssignale mittels *kommunikativer Emotionen* (Co) in körperlichen Handlungen reziprok auszutauschen. Kommunikative Emotionen ergänzen situative Emotionen Si und strukturelle Emotionen St und können, da sie dargestellt werden, direkt beobachtet werden. Diese Differenzierung basiert auf dem Affektmodell von Moser und von Zeppelin⁸⁰ und wurde erstmals technisch mittels MARSSI⁵⁷ realisiert (Kapitel 3.1 und 3.2).

Die Dyadenkopplung geschieht über die co-regulative reziproke soziale Interaktion. Letztere ist beobachtbar. Die damit zusammenhängenden inneren regulativen Prozesse sind nicht direkt beobachtbar. Hinweise aus neurologischen Studien zur Hirnsynchronizität deuten jedoch darauf hin, dass soziale Interaktionen generell eine

Kopplung auf Gehirnebene hervorrufen¹¹¹. Diese Sicht ist jedoch zu grob, um die oben genannten inneren Abläufe zu beschreiben, zu erklären, oder KET-Systeme daraufhin zu sensibilisieren. Eine Beschreibung aller Aspekte der inneren Vorgänge bei der reziproken sozialen Interaktion in der Dyade mittels des vorgestellten Modells kann hier nicht geleistet werden und ist Gegenstand zukünftiger Arbeiten. Eine grobe strukturgebende Beschreibung der Motivation und der Arbeitsweise wird jedoch gegeben.

Zentral für die Arbeitsweise sind das Personenmodell *G* des Gegenübers sowie die damit verbundenen Prozesse und Repräsentationen (Kapitel 5.4). Das Personenmodell bestimmt, wie die Beziehung *B* angelegt ist. Eine intime Partnerschaft wirkt, im Sinne der Bindungstheorie, als »sichere Basis«. Die antizipierte oder tatsächliche Verfügbarkeit der Partnerperson dämpft nachweislich die Aktivierung von Bindungs- und Stresssystemen, fördert die eigene Emotionsregulation und erhöht so das innere Sicherheitsgefühl (»felt security«, cf. Zielfunktion Sicherheit, insbesondere Kapitel 5.2). Asendorpf und Kollegen¹⁰³ beschreiben diese Sicherheits- und Stützfunktion als zentrale Beziehungsfunktion und zeigen, dass Verlässlichkeit, Nähe und wechselseitige Unterstützung die Partnerschaftsqualität und damit die gefühlte Sicherheit stärken. Die Repräsentation eines responsiven Partners reduziert die Bewertung von Bedrohungen und erleichtert die Exploration. Somit werden auch Autonomie und Selbstachtung gestärkt. Umgekehrt unterminieren kontrollierende, abwertende oder inkonsistente Interaktionen das Autonomieerleben und beeinträchtigen die innere Emotionsregulation, was den Wert der Autonomie und damit die Selbstachtung verringert.

5.5.1 Innere Kommunikationspfade

Die *innere Kommunikation*, ein strukturgebender, organisierender »Denkprozess«, dient der Antizipation oder Prädiktion sozialer Darstellungen und Handlungen des Gegenübers. Generell gibt es Hinweise darauf, dass Antizipation oder Prädiktion eine zentrale

111 Lyle Kingsbury und Weizhe Hong, A multi-brain framework for social interaction, Trends in neurosciences, 43.9: 651–666, 2020.

Arbeitsweise des Gehirns ist (Kapitel 5.2). Hier wird sie auf soziale Interaktion angewendet.

Ziel der inneren Kommunikation ist stets, einen Gewinn für sich selbst und im besten Fall für das Gegenüber durch wertegerechte sozial-interaktive Handlungen zu erreichen. Der Gewinn ist mit einer Erhöhung der inneren Werte von Sicherheit, Autonomie und Selbstachtung verbunden, die die eigenen Wünsche und Ziele berücksichtigen. Natürlich ist das nicht immer erreichbar, beispielsweise bei ärztlichen Diagnosen, bei Ablehnungen in Beziehungen oder bei Kündigungen. Diese Fälle stellen eine enorme Herausforderung hinsichtlich der inneren Stabilisierung mittels des *Selbst*-Regulationsprozess dar. Diese können durch die Beziehung und den *Gegenüber*-Regulationsprozess in der äußeren Kommunikation (Kapitel 5.5.2) unterstützt werden (Kapitel 7.2.2). Eine genaue Beschreibung der Wirkweisen ist Gegenstand zukünftiger Arbeiten.

In dem Fall der sozialen Interaktion wird der unbewusste Prozess der inneren Kommunikation durch das tatsächliche oder imaginierte Vorhandensein des Gegenübers und dessen Handlungen ausgelöst. Bevor nun selbst außen beobachtbar gehandelt wird, werden, wie oben beschrieben, Erlebtes und Selbst- und Beziehungssignale mittels kommunikativer Emotionen *Co* in virtuellen körperlichen Handlungen reziprok zwischen den inneren Regulationsprozessen *Selbst* und *Gegenüber* ausgetauscht. Die Gegenübersimulation basiert auf 1) den Annahmen über Soziales *Soz_G* und Situation *Sit_G* des Gegenübers sowie der eigenen *Soz*, *Sit*, und deren Verarbeitung durch die Regulationsprozesse *Display_G* und *Display*, die bei der inneren Kommunikation virtuell aktiv sind. Die Repräsentationen des Gegenübers werden je nachdem, wie sie durch dessen Handeln bestätigt werden, angepasst. Die eigenen Repräsentationen des Sozialen *Soz* und der Situation *Sit* unterliegen Schutzmechanismen, die sicherstellen, dass keine Minderung der zentralen Werte Sicherheit, Autonomie und Selbstachtung erfolgt¹¹². Das kann dazu führen, dass die inneren Repräsentationen des Sozialen und der Situation von denen des Gegenübers abweichen. Eine Realisierung davon ist Gegenstand zukünftiger Arbeiten.

112 Laura Parolin et al., The Interplay of Mentalization and Epistemic Trust: A Protective Mechanism Against Emotional Dysregulation in Adolescent Internalizing Aymptoms, Research in Psychotherapy: Psychopathology, Process, and Outcome 26.3:707, 2024.

Die innere Kommunikation ist ein weitgehend unbewusst ablaufender Prozess und vor dem Gegenüber verborgen — auch vor der Technologie. Menschen können den Prozess jedoch beeinflussen und ihr Handeln entsprechend kontinuierlich anpassen (Kapitel 3.2). KET-Systeme benötigen dies auch. So wird es möglich, transparente, erklärbare, individuelle und wertgerechte regulative Assistenzaktionen zu generieren. Deren Realisierung unterliegt den Rahmenbedingungen, die in den Kapiteln 8 und 9 beschrieben werden.

5.5.2 Äußere Kommunikationspfade

Die *äußere Kommunikation* ist beobachtbar. In einer Dyade lässt sich beobachten, dass der gesamte Körper einer Person reziprok auf die Signale der anderen reagiert. Alle Signale sind individuell und von den persönlichen Erlebnissen sowie den kulturellen Faktoren der jeweiligen Person geprägt¹¹³.

Das Modell (Abbildung 5, Mitte) illustriert, dass dargestellte non-verbale und verbale Signale einer anderen Person vom Display-Regulation *Display_G* — der Simulation dessen, wie eine Person sich das Verhalten der anderen Person vorstellt — empfangen und interpretiert werden. Die Interpretation hängt von den angenommenen Werten für Soziales *Soz_G* und Situation *Sit_G* des Gegenübers ab. Abstrakter zusammengefasst, lässt sich sagen, dass die Interpretation in Abhängigkeit vom Personenmodell *G* sowie allen damit verbundenen Prozessen und Repräsentationen erfolgt. Passen die Signale somit nicht zusammen, muss nachjustiert werden (Kapitel 9.2). Wie bereits durch andere Kommunikationsmodelle grob beschrieben, lässt sich die Interpretation sozialer Signale nun feiner darstellen. Missverständene Signale können auf der Ebene des *Display_G*-Regulationsprozesses liegen oder ihre Ursache im *Gegenüber*-Regulationsprozess beziehungsweise im damit verbundenen Personenmodell *G* sowie in der Beziehung haben. Zukünftige Forschung hat die Aufgabe, dies genauer zu identifizieren.

Die *äußere Kommunikation* findet parallel zur *inneren* statt. Die beschriebenen Turn-Taking-Mechanismen zur Organisation der so-

113 Hedda Lausberg. Der Körper in der Psychotherapie. Kohlhammer Verlag, 2022.

zialen Kommunikation^{114,115} physischer Handlungen und anderer Prozesse sind Mittel zur Synchronisation beider Prozesse. Zukünftige Forschung kann Verbindungen zwischen den Kommunikationssignalen beider Personen einer Dyade, insbesondere deren Form, räumliche und zeitliche Ausprägung sowie Reziprozität, mit den vorgestellten internen Prozessen und Repräsentationen untersuchen. Im Vordergrund stehen Fragen wie die Synchronizität der Kommunikationssignale und die Korrelationen in der Beziehung, was für den Aufbau therapeutischer Allianzen — insbesondere in Pflege- und Behandlungsbeziehungen — zentral ist. Dabei sind die Implikationen der Regulationsprozesse und die beteiligten Repräsentationen stets zu berücksichtigen.

Durch dieses Modell der äußeren und inneren Kommunikation wird auch deutlich, dass die Erkennung von »Emotionen« als Teil der äußeren Kommunikation, wie es aus computertechnischer Sicht zurzeit möglich scheint (Kapitel 3.1), bei Weitem nicht ausreicht, um auf irgendwelche internen Vorgänge oder Zustände zu schließen.

6. Berechnungsorientiertes Beispiel der Selbst-Regulation

Das folgende Beispiel macht das abstrakte Modell aus Kapitel 5 greifbar: Es zeigt, wie KET die drei Zielwerte Sicherheit, Autonomie und Selbstachtung der Selbst-Regulation (Kapitel 5.3) in einer realen Pflegesituation berechnet und gegeneinander abwägt — und welche Assistenzaktion daraus folgt. Der Anteil des Modells, der das Gegenüber und die innere Kommunikation einbezieht (Kapitel 5.4 und 5.5), wird in Kapitel 7 durch ein weiteres Beispiel veranschaulicht.

Die in dieser Beispielsituation verwendeten Parameterwerte, Sollwerte und Gewichte werden im tatsächlichen Einsatz nicht willkürlich festgelegt, sondern durch einen strukturierten Initialisierungsprozess bestimmt. Als Ausgangsbasis dienen normbasierte Referenz-

114 Julia A. Goldberg, Interrupting the Discourse on Interruptions: An Analysis in Terms of Relationally Neutral, Power-and Rapport-Oriented Acts, *Journal of Pragmatics*, 14.6: 883–903, 1990.

115 Han Z. Li, Cooperative and Intrusive Interruptions in Inter-and Intracultural Dyadic Discourse, *Journal of Language and Social Psychology* 20.3: 259–284, 2001.

werte, die aus pflegefachlichen Standards und care-ethischen Leitlinien abgeleitet und im System explizit repräsentiert werden. Diese werden durch ein strukturiertes Erstgespräch und/oder die Übernahme relevanter Angaben aus der Pflegeakte individualisiert. Abweichungen von den normbasierten Werten sind stets zu beschreiben und, soweit möglich, gemeinsam von der Pflegekraft, der pflegebedürftigen Person und der Person mit Vorsorgevollmacht zu bestätigen. Dieses Vorgehen gewährleistet, dass die Personalisierung gezielt und nachvollziehbar erfolgt (Kapitel 8.1).

Eine pflegebedürftige Person will unmittelbar nach dem Aufstehen selbstständig zur Toilette gehen. Sie meldet über die Notfallsprechanlage leichten Schwindel; KET erfasst zudem den unsicheren Stand sensorisch. Ausgehend vom beschriebenen Modell können die Zielvariablen Sicherheit S , Autonomie A und Selbstachtung SA mit den Sollwerten S_s , A_s und SA_s sowie den Minimalwerten $S_{\min}$, $A_{\min}$ und $SA_{\min}$ reguliert werden. Im Folgenden gilt: Index $_s$ = individueller Sollwert, $_{\min}$ = Minimalwert, ho = Vorhersage für eine konkrete Handlungsoption (z. B. ho_1 , ho_2), S' = aktueller Schätzwert der Sicherheit. Die für die jeweilige Person ermittelten Werte müssen in einem vorgelagerten Prozess bestimmt werden. In Computermodellen werden dafür traditionell numerische Werte im Bereich von 0,0 bis 1,0 verwendet, z. B. $S = 0,3$ (geringe Sicherheit), $S_s = 0,8$ (hohe Soll-Sicherheit).

Aus internen und externen Hinweisen (1) entstehen laufend Schätzungen für die Zielvariablen S' , A' und SA' sowie für jede regulative Assistenzaktion ho (z. B. ho_1 =allein gehen, ho_2 =begleitet gehen, ho_3 =kurz abwarten) und Vorhersagen (2) S_{ho} , A_{ho} und SA_{ho} .

Im Anschluss vergleicht die *Kognition* (3) die Soll- und Vorhersage und berechnet den Vorhersagefehler; die *Emotion* (4) bestimmt, wie stark der Fehler wiegt, mit den Gewichten w für jede Zielvariable w_s , w_A und w_{SA} . Die *emotionale Bewertung* passt diese Wichtigkeit *situativ* an: Bei belastenden inneren Zuständen wie Unruhe, erlebtem Schwindel oder Alarmgefühl steigt das Sicherheitsgewicht w_s — kleine Sicherheitsverluste haben damit einen großen Einfluss. Bei ruhigen, zuversichtlichen Zuständen sinkt w_s entsprechend. Das hängt von individuellen Erfahrungen ab, und das Modell muss entsprechend individualisiert werden. Die Kognition hat die Aufgaben die

$$\text{Kosten}(ho) = w_S \times |S_s - S_{ho}| + w_A \times |A_s - A_{ho}| + w_{SA} \times |SA_s - SA_{ho}|$$

zu minimieren unter der Bedingung, dass die Minimalgrenzen nicht unterschritten werden, z. B. S_{ho} muss größer als der individuelle Minimalwert für die Sicherheit S_{min} sein.

Zurück zum Beispiel: Diese Hinweise senken S' und erhöhen w_S . Für ho_1 wäre S_{ho} voraussichtlich unter S_{min} ; für ho_3 stabilisiert S_{ho} zwar, senkt aber auch A und SA . Damit minimiert ho_2 die gewichtete Abweichung der Kosten (ho) am stärksten: Die Sicherheit steigt ausreichend, Autonomie und Selbstachtung bleiben etwa gleich (z. B. verbalisiert durch die Pflegekraft: »*Sie gehen, ich sichere nur*«) (4, Handeln). Rückmeldesignale (5) während der Ausführung aktualisieren S' , A' , SA' und deren Gewichte w . Stabilisiert sich der Stand, sinkt w_S wieder, sodass bei der nächsten Gelegenheit (6, Lernen) eine autonomiestärkere Option (ho_2 mit Aufsicht einer Pflegekraft) wahrscheinlicher wird.

Insgesamt ist die modellbasierte Priorisierung transparent: Akute Unsicherheit → temporär w_S hoch und Sicherheitsgrenze aktiv; nach Stabilisierung Rückkehr zur Autonomie. Jede solche Parameteranpassung wird mit Zeitstempel und Begründung dokumentiert und ist auf Anfrage adressatengerecht erklärbar (Kapitel 8.1). Bei sicherheitskritischen Abweichungen greift eine zusätzliche Kontrollebene, die die physische Anwesenheit der Pflegekraft voraussetzt (Kapitel 8.1).

Inwieweit ein KET-System dies durch konkrete Handlungen umsetzen könnte, bleibt an dieser Stelle unbeantwortet. Es lässt sich jedoch sagen, dass sich diese prozessorientierte und berechnende Vorgehensweise für Simulationen und Trainingszwecke im Kontext der Ausbildung nutzen lässt, was per se einen Mehrwert bietet. Im unterstützenden Betrieb würde KET der Pflegekraft keine Handlung vorgeben, sondern die Abwägung transparent machen: Welche Option sichert unter den aktuellen Bedingungen Sicherheit, Autonomie und Selbstachtung am besten? Die Entscheidung verbleibt bei der Pflegekraft — KET strukturiert und begründet, die Pflegekraft urteilt und handelt. Das stellt keine Infragestellung der professionellen Kompetenz dar, sondern deren Entlastung und Erweiterung: Implizite Abwägungen werden explizit und nachvollziehbar — z. B. dass begleitetes Gehen die Sicherheit ausreichend erhöht, ohne Autonomie wesentlich zu mindern – und damit auch lernbar.

7. Einsatzfelder in der Pflege

7.1 Co-Regulative Assistenz

Der Einsatz von künstlich empathischer Technologie (KET) bietet dort Mehrwert, wo sozial-kommunikative Brücken, Co-Regulation, also externe Unterstützung bei emotional schwierigen Situationen, oder Begleitung kontinuierlich benötigt werden. Kapitel 9 informiert über die allgemeinen Risiken des Einsatzes. Die folgenden exemplarischen Einsatzfelder sind 1) nach dem Ziel, das durch den 2) KET-Beitrag erreicht werden soll, sowie 3) nach den dazu notwendigen Bedingungen gegliedert:

- 1) *Adhärenz und Alltagsstruktur in der Häuslichkeit.* Das Ziel ist dabei die Assistenz für eine sichere, selbstbestimmte Routine, z. B. Medikation, Flüssigkeitsaufnahme oder Mobilität. KET trägt mit personalisierten Erinnerungs- und Motivationsstrategien dazu bei. Diese sind situativ und individuell angepasst. Sie berücksichtigen beispielsweise die Auswahl geeigneter Tageszeitpunkte, folgen ritualisierten Schemata und passen sich in der Interaktion an kulturelle Normen an. Wichtig ist die Einhaltung der Bedingungen Transparenz, modifizierbare Einwilligung und Einbindung von Bezugspersonen.
- 2) *Kommunikationsbrücken in der Langzeitpflege.* Das Ziel ist, trotz kognitiver, sensorischer oder sprachlicher Barrieren Interaktion zu ermöglichen. KET trägt dazu bei, mit adaptiven, multimodalen Ausdruck (Sprache, Bild, Gestik), der an individuelle Tagesformen und Präferenzen (Normen) gekoppelt ist. Die Wahrung der Autonomie und die Vermeidung von Überforderung sind dabei wichtige Bedingungen, die eingehalten werden müssen.
- 3) *Agitationsmanagement bei Demenz.* Das Ziel ist, Anspannungen und motorische Unruhe zu senken und eine allgemeine Überforderung zu vermeiden. KET trägt dazu bei, durch 1) eine kontinuierliche sensorische Erfassung und Interpretation von Anzeichen aufkommender Unruhe, durch 2) eine darauf basierende Simulation möglicher Interventionen und durch 3) Assistenz durch individualisierte psychologische Maßnahmen wie Ansprachen, Reizreduktionsstrategien und anschlussfähige Inhalte. Zentrale Bedingungen sind dabei eine enge menschliche Supervision, kei-

ne Täuschung über Fähigkeiten sowie definierte, klare Abbruchkriterien. Erste Evidenz für dieses Einsatzpotenzial liefert Pu und Kollegen¹¹⁶: Menschen mit Demenz berichten durch die Interaktion mit PARO von verbessertem emotionalem Wohlbefinden und Schmerzlinderung, äußern jedoch den Wunsch nach breiteren Interaktionsmodalitäten — insbesondere nach sprachlicher Kommunikation. Genau diese Lücke adressiert KET durch adaptiven, multimodalen Ausdruck (Sprache, Bild, Gestik).

7.2 Bezug zur Pflege

7.2.1 Erster Eindruck und Beziehungsmanagement

Auch bei der Pflege ist der erste Eindruck wichtig¹¹⁷. Dazu gehören soziale Aspekte, wie das Auftreten, insbesondere eine ruhige Begrüßung durch Pflegekräfte aber auch situative Aspekte, wie eine einheitliche, gepflegte Kleidung mit Namensschild¹¹⁸ und eine aufgeräumte, helle Station¹¹⁹. Damit schlägt die Literatur eine Brücke zu KETs Beziehungsmodell (Kapitel 5.4): Kompetentes und warmes Verhalten¹²⁰ beim ersten Eindruck stärkt das Vertrauen und erhöht die inneren Werte der Sicherheit und der Autonomie. Systeme der künstlich empathischen Technologie (KET) nutzen das Vorgestellte, um den ersten Eindruck entsprechend den erwünschten Werten für

-
- 116 Lihui Pu, Wendy Moyle und Cindy Jones, How People with Dementia Perceive a Therapeutic Robot called PARO in Relation to Their Pain and Mood: A Qualitative Study, *Journal of Clinical Nursing*, 29.3 – 4: 437–446, 2020.
 - 117 Andrea Eikelenboom-Boskamp et al., Preferences for healthcare worker attire among nursing home residents and residents' preferences as perceived by workers: A cross-sectional study, *International Journal of Nursing Studies Advances* 4:100112, 2022.
 - 118 Nancy M. Albert et al., Impact of Nurses' Uniforms on Patient and Family Perceptions of Nurse Professionalism, *Applied Nursing Research* 21.4: 181–190, 2008.
 - 119 John Reiling, Ronda G. Hughes und Mike R. Murphy, The Impact of Facility Design on Patient Safety, Patient Safety and Quality: An Evidence-Based Handbook for Nurses, Chapter 28, Agency for Healthcare Research and Quality, 2008.
 - 120 Lauren C. Howe, Kari A. Leibowitz und Alia J. Crum, When your Doctor »gets it« and »gets you«, The Critical Role of Competence and Warmth in the Patient–Provider Interaction., *Frontiers in Psychiatry* 10: 475, 2019.

die situationalen und sozialen Repräsentationen sowie das Modell des Gegenübers G bei Pflegenden so zu gestalten, dass über die Beziehung hinweg die individuell bestmögliche Co-Regulation möglich wird. Dazu können bekannte Vorgehensweisen als Ausgangsbasis herangezogen werden (siehe vorherigen Abschnitt). Dennoch sind weitere Grundlagenforschungen notwendig — insbesondere über die Modellentstehung des Gegenübers, wenn das Gegenüber ein technisches System ist und wenn unter anderem verschiedene individuelle Charakteristiken und Krankheitsbilder, Einschränkungen beachtet werden.

Eine konsistente Handlungs- oder generell Verhaltensmodellierung von KET-Systemen ist ebenso wichtig (Kapitel 5.5). Die regulativen Assistenzaktionen sind stets an gängigen Vorgehens-Modellen und care-ethischen Standards gebunden (Kapitel 1). Anders als Menschen sollen sich KET-Systeme die Inhalte von Soz_G und Sit_G nicht frei wählen oder ihre eigenen kopieren. Je nach Definition und Einschätzung des Gegenübers (z. B. hinsichtlich Rolle, Aufgabe, Historie) müssen diese Werte spezifisch ausgewählt oder von Pflegekräften definiert, überprüft und später in der Interaktion angepasst (Kapitel 2) werden. Diese Schritte müssen sowohl bei der Konzeption als auch bei der Realisierung berücksichtigt werden. Auch Strategien zur Beziehungsreparatur müssen berücksichtigt werden, die unter anderem auf eine andere Rollen- oder Aufgabenannahme der Pflegenden im KET-System beruhen. Das wird in den nächsten Kapiteln näher erläutert.

7.2.2 Modellierung der inneren und äußeren Kommunikation

Warum sind die vorgestellten Kommunikationspfade (Kapitel 5.5) für die Pflege wichtig? Ein Beispiel: Ein technisches System ermöglicht einer Pflegebedürftigen mehr Autonomie — etwa, allein wieder auf die Toilette zu gehen. In einer Modellierung nach KET hätte das eine Steigerung der Sicherheit, der Autonomie und der Selbstachtung zur Folge. Wird dieses Versprechen jedoch nicht erfüllt — die Pflegebedürftige stürzt, braucht doch Hilfe oder traut sich beim nächsten Mal nicht mehr — sind diese inneren Werte beschädigt: Das Vertrauen sinkt, die Selbstachtung leidet, die Beziehung zum System ist belastet.

Ein technisches System ohne KET erfasst diese relationale Dimension nicht. Es registriert allenfalls die körperliche Situation, aber kein beschädigtes Vertrauen, keine gesunkene Selbstachtung, keine Angst vor Wiederholung. Beim nächsten Interaktionsversuch bietet es dieselbe Autonomieoption an — weil es keine Repräsentation des Misserfolgs und seiner inneren Folgen hat. Damit reaktiviert es genau die Zustände, die es eigentlich unterstützen sollte: Sicherheit und Selbstachtung sinken erneut, die Pflegebedürftige zieht sich zurück, und das System kann dies mangels entsprechender Repräsentationen nicht interpretieren.

KET hingegen modelliert diese inneren Zustände und deren Veränderungen durch Interaktionen. Nach einer Nichterfüllung sind Reparaturschritte notwendig: schrittweiser Wiederaufbau des Vertrauens, angepasste Angebote, Anerkennung des Erlebten. Diese Fähigkeit — Kommunikationspfade auch nach Brüchen zu reparieren — unterscheidet KET von konventionellen Assistenzsystemen.

Pflegetätigkeiten sind per se im Spannungsfeld von Bedürfnissen, Erwartungen und Annahmen angesiedelt. Pflegebedürftige und Angehörige neigen unter Unsicherheit dazu, ihre eigenen Fähigkeiten und die des Personals zu überschätzen^{121,122} um die inneren Werte der Sicherheit und Autonomie zu erhöhen (Kapitel 5.2). Zum Beispiel sagt ein Pflegebedürftiger nach der Visite: »*Ins Bett schaffe ich es alleine*«. Diese Annahmen können im Pflegekontext leicht zu Fehleinschätzungen hinsichtlich der Risiken und des Unterstützungsbedarfs führen. Ein professioneller Umgang umfasst immer den Abgleich mit den tatsächlichen Fähigkeiten. Pflegekräfte ergänzen Selbstauskünfte systematisch durch kurze, leistungsbasierte Assessments; durch den TUG-Test¹²³ kann im genannten Beispiel das Verstehen gesichert werden (z. B. »*Bitte zeigen Sie mir, wie Sie aufstehen und losgehen*«) und erreichbare, personenzentrierte Ziele verhandeln im Sinne eines beziehungsstärkenden Vorgehens (z. B. »*Bis morgen nehmen wir den Rollator und die Rufklingel, dann prüfen wir*

121 Ellen J. Langer, The Illusion of Control, Journal of Personality and Social Psychology 32.2: 311, 1975.

122 Martin Hagger et al., The Common Sense Model of Self-Regulation: Meta-Analysis and Test of a Process Model, Psychological bulletin 143.11: 1117, 2017.

123 Diane Podsiadlo und Sandra Richardson, The Timed »Up & Go«: A Test of Basic Functional Mobility for Frail Elderly Persons. Journal of the American Geriatric Society, 39(2):142–8, 1991.

erneut«). So bleibt das innere Sicherheitsgefühl erhalten, ohne die tatsächliche Sicherheit zu gefährden.

Das folgende Beispiel setzt voraus, dass das Pflegeteam gemeinsam mit dem KET-System vorab festgelegt hat, nach welchen Leitlinien es in Situationen wie Mentalisierungsabbrüchen agiert (hier exemplarisch nach MbT-Prinzipien¹²⁶). KET kann ebenso nach anderen pflegfachlichen Ansätzen konfiguriert werden; die Wahl obliegt dem Team. Nach entsprechender Supervision kann KET solche Situationen auch autonom bearbeiten (Kapitel 2, Kapitel 8.1).

Fonagys Mentalisierungstheorie und die darauf aufbauende mentalisierungsbasierte Therapie (MbT) beschreibt Kompetenzen, die auch für KET relevant sind: das aufmerksame, zugewandte Arbeiten bei ruhiger eigener Haltung — also präsent bleiben, ohne von der Situation überwältigt zu werden — sowie das Erkennen und Reparieren von Momenten, in denen das gegenseitige Verstehen abbricht^{124,125}.

Mentalisierungsabbrüche spielen in der Pflege eine große Rolle¹²⁶, wie folgendes Beispiel illustriert: Eine gepflegte Person mit starker Anspannung ruft laut, als die Pflegekraft vorbeigeht: *»Sie ignorieren mich absichtlich! Ihnen ist egal, dass ich Schmerzen habe!«*. Erste Schritte der Reparatur wären (adaptiert von MbT): 1) Verlangsamung, Sicherheit erhöhen: Pflegekraft bleibt stehen, senkt Stimme und Tempo. *»Einen Moment — ich möchte erst verstehen, was gerade los ist.«* 2) Markiertes Spiegeln — das Wahrgenommene benennen, ohne es zu übernehmen: *»Es wirkt, als fühlten Sie sich gerade allein gelassen und nicht ernst genommen; das ist sehr belastend.«* 3) Gemeinsame Hypothesenbildung: *»Ich weiß noch nicht genau, wie es für Sie ist. Könnte es sein, dass das Warten das Gefühl verstärkt hat, unwichtig zu sein?«* 4) Gemeinsamer Realitätscheck: *»In Ihrer Kurve steht, dass ich Ihr Schmerzmittel in 10 Minuten geben darf. Bis dahin: Was hilft Ihnen eher – kalter Waschlappen oder Atemübung? Ich bleibe jetzt 2 Minuten hier und komme in 10 Minuten mit dem*

124 Klaus Wieglerling, Exposition einer Theorie der Widerständigkeit, *Filozofija i društvo*, 32.4:641–661, 2021.

125 Anthony Bateman und Peter Fonagy, *Mentalization-Based Treatment for Personality Disorders: A Practical Guide*, Oxford: Oxford University Press., 2015.

126 Gieke Free et al., Empathy and Mentalizing of Mental Health Nurses: A Cross-Sectional Correlational Study, *International Journal of Mental Health Nursing*, 2025.

Medikament zurück.« 5) Metakommunikation — Rückblick auf das eigene Handeln mit Reparatur: *»Vorhin habe ich schnell erklärt, statt erst zu verstehen. Das konnte so wirken, als hätte ich Sie abgewertet — das tut mir leid«*, 6) Mini-Plan & Bindungsrahmen: *»Ich notiere den Plan im Team und komme wie besprochen. Wenn es vorher zu viel wird, drücken Sie bitte die Klingel — ich reagiere«*.

Taubner und Volkert¹²⁷ haben diesen Ansatz im deutschsprachigen Raum maßgeblich für die Behandlung von Erwachsenen ausgearbeitet und zeigen mit Hauschild die Wirksamkeit von MBT und skizzieren Veränderungen im Mentalisieren als spezifischen Wirkpfad¹²⁸. Für die Pflegepraxis bedeutet das: eine neugierig-validierende Grundhaltung, fein dosierte Co-Regulation, konsequente Fokussierung auf wahrnehmbare innere Zustände sowie strukturierende Routinen und kurze Reflexionsschleifen im Team- und Familienkontakt — allesamt Bausteine eines mentalisierungsförderlichen Pflegekontexts, der Vertrauen, Adhärenz und Beziehungssicherheit stärkt.

7.2.3 Beispiel innere Kommunikation bei Medikamentengabe

Die innere Kommunikation simuliert, welchen Einfluss (später gezeigte) eigene Handlungen auf das Gegenübermodell G haben und insbesondere welche Handlungen des Gegenübers daraufhin die Folge sein könnten. Ziel ist, durch die Beziehung B zum Selbst einen Gewinn für die eigenen Bedürfnisse zu erzielen. Ein KET-System kann diese inneren Dialoge simulieren und mögliche Gewinne antizipieren.

Wie das Gegenüber tatsächlich handelt und wie dies zu dem Gegenübermodell und den antizipierten Handlungen passt, wird anhand des folgenden Beispiels erläutert, wie eine Pflegekraft (PK) im Kontext einer sicheren Medikamentenverabreichung möglicherweise handelt.

Die pflegebedürftige Person (PP) ist abends erschöpft und unsicher, ob sie die Medikamente jetzt einnehmen möchte. Die PK

127 Svenja Taubner und Jana Volkert. Mentalisierungsbasierte Therapie für Adoleszente (MBT-A). Vandenhoeck & Ruprecht, 2016.

128 Jana Volkert, Sophie Hauschild und Svenja Taubner, Mentalization-Based Treatment for Personality Disorders: Efficacy, Effectiveness, and New Developments. Current Psychiatry Reports, 2019.

kommt und beginnt mit der Verabreichung. Sie geht davon aus, dass sie bei PP grundsätzlich als strikt, zeitgebunden und fair wahrgenommen wird (Gegenübermodell: G_{PK} von PP). Unbewusst laufen bei PP auf den inneren Kommunikationspfaden mehrere simulierte Dialoge (Handlungen) ab:

- (1) »Wenn ich jetzt zustimme, ist es schnell erledigt — kann ich machen«; die Folgen sind Autonomie leicht eingeschränkt, Sicherheit hoch, Selbstachtung gleich.
- (2) »Wenn ich zögere, werde ich gedrängt — gefällt mir nicht«; die Folgen sind: Autonomie verringert, Sicherheit erhöht, Selbstachtung verringert.
- (3) »Wenn ich klar wähle, wird die Wahl akzeptiert — wenig Spielraum«; die Folgen sind: Autonomie bleibt gleich; Sicherheit und Selbstachtung sind abhängig von der kommunizierten Klarheit. Letzteres hängt von den aktuellen mentalen Fähigkeiten sowie von individuellen und kulturellen Normen (Display) ab.

Optimale Beispielreaktion: Die PK (die denselben Aufgaben hätte ein KET-System) schätzt anhand vorangegangener ähnlicher Situationen ab, welche dieser Varianten dominant ist, und simuliert die Folgen möglicher äußerer Handlungen hinsichtlich Sicherheit, Autonomie und Selbstachtung bei PP in einer sicherheitsgebenden Beziehung B zwischen PK und PP. Ziel ist es, die G_{PK} unverändert zu lassen. Ist es Variante (3), empfiehlt die PK eine äußere Formulierung, die mit ihrem G_{PK} kompatibel ist und kleine Wahl im engen Rahmen ermöglicht: »Ich habe hier auf Station 10 Minuten zu tun. Lieber jetzt in Ruhe oder in 15 Minuten? Beides ist möglich; ich komme in jedem Fall wieder«. So bleibt die Autonomie von PP gleich, die Selbstachtung wird geschützt (respektvoller Ton) und die Sicherheit wird leicht erhöht (vgl. Planbarkeit). Die Beziehung B wird positiv aktualisiert.

Unpassende Beispielreaktionen: Handlungen der PK, die dem G_{PK} widersprechen oder konterkarieren: (1) Zu vage: »Später irgendwann, wir schauen mal...«; Passt nicht zum G_{PK} (wegen Zeitgebundenheit). In der Folge wird bei PP die Sicherheit verringert. In der Folge dürfte es zu vermehrten inneren Dialogen kommen, die im schlimmsten Fall G_{PK} anpassen (Kapitel 9.2) und Variante (2, oben) triggern, was zusätzlich die Selbstachtung reduzieren würde; (2) Zu

hart: »Wir machen das jetzt ohne Diskussion«; Strikt ist zwar kompatibel mit G_{PK} , fair aber nicht, was am Ende G_{PK} verletzt. In der Folge werden Autonomie und Selbstachtung verringert, was bei PP innere Dialoge und Handlungssimulationen auslöst, die ein erhöhtes Risiko für Widerstand beziehungsweise für die Schwächung der Beziehung mit sich bringen. (3) Zu flexibel: »Es wird garantiert ganz entspannt, ganz wie Sie wollen«; widerspricht G_{PK} (zu viel Flexibilität); in der Folge erzeugt der Erwartungsbruch eine Minderung der Selbstachtung und löst bei PP innere Dialoge/Handlungssimulationen aus, die G_{PK} anpassen und in der Folge eine Schwächung der Beziehung mit sich bringen können.

Reparaturaufforderungen von PP: verzögerte oder ausweichende Antworten (»Schon gut...«), Arousal-Anstieg (Tempo, Lautstärke) oder rhetorische Rückfragen (»Muss das jetzt sein?«). Durch zum G_{PK} passende Reparaturen kann eine Anpassung des G_{PK} vermieden werden. Beispiel: »Verstanden. Ich bin heute bis 19:30 auf Ihrer Station. Zwei Möglichkeiten: Jetzt gemeinsam in Ruhe oder um 19:15. Ich komme sicher wieder. Was passt Ihnen besser?« (vgl. optimale Beispielreaktion (1, oben). Autonomie (eng, aber echt), Sicherheit (klare Zeit) und Selbstachtung (respektvoll) werden gleichzeitig bedient, G_{PK} muss nicht angepasst werden und B wird erhalten.

Eine Evaluation, gerade wenn derartige Aufgaben von einem KET-System erledigt werden, ist nicht einfach. Expertinnen und Experten können unter anderem die Passfähigkeit von PK-Handlungen zum G_{PK} , die wahrgenommene Reaktionsfähigkeit sowie die empfundene Sicherheit (Felt-Security-Fragebogen) messen. Objektive Indikatoren aus den Systemlogs könnten (multimodale) Synchronizität¹²⁹ sowie Veränderungen im Sprechtempo und in der Unterbrechungsrate von Dialogbeiträgen oder allgemein von Handlungen markieren. Das Ziel sollte eine hohe Passung sein.

129 Fabian Ramseyer und Wolfgang Tschacher, Nonverbal Synchrony in Psychotherapy: Coordinated Body Movement Reflects Relationship Quality and Outcome, *Journal of Consulting and Clinical Psychology* 79.3: 284, 2011.

8. Evaluation und Governance festlegen

Die bisherigen Systeme emotional adaptiver Technologie greifen zentrale Konzepte von Emotion und Empathie verkürzt auf. Emotion wird häufig als statisches oder numerisches Label statt als situations- und rollenabhängigen Bewertungsprozess behandelt. Das Konzept der Empathie reduziert sich meist auf sprachliche Validierung statt auf Mentalisierung und dyadische Co-Regulation mit innerer Kommunikation. Es fehlen zentrale Repräsentationen zu Beziehung, zeitlicher Dynamik, Erlebenskontext (situative, strukturelle und kommunizierte emotionale Informationen) sowie eine wertegebundene Zielfunktion für Sicherheit, Autonomie und Selbstachtung. KET adressiert diese Lücken. Bis zu belastbaren, skalierten Lösungen besteht jedoch Forschungs- und Entwicklungsbedarf. Zunächst werden Methoden zur Messung von Transparenz, Autonomie, Sicherheitserleben und Handover-Angemessenheit vorgeschlagen. Dabei sind kurze Frageitems, einfache Checks (Teach-Back¹³⁰/Checkliste) kombiniert. Die kurzen Erhebungen (weniger als 5 Min.) sind so ausgelegt, dass sie in Routinesequenzen durchgeführt werden. Zudem empfiehlt sich die Integration von MEESTAR¹⁰ als regelmäßige Prüfroutine (halbjährlich) mit einer Schnittstelle zum Pflege-Quality-Management.

Tabelle 2 Messung von einigen relevanten Werten für KET.

Wert	Motivation	Messmethode
Transparenz	Versteht die Person, warum diese Option gewählt wurde?	Teach-Back-Accuracy, Erklär-Checkliste
Autonomie	Erlebte Wahlfreiheit bzw. Nein-Sagen-Können	Kurzskala zu Wahlmöglichkeiten: »Ich hatte echte Wahlmöglichkeiten.«, »Ich konnte Nein sagen, ohne Nachteile.«
Sicherheitserleben	Subjektives Sicherheitsgefühl	3 Kurzitems (z. B. »Ich wusste, was als Nächstes passiert.«, »Ich fühlte mich

130 Kathryn M. Anderson et al., The 5Ts for Teach Back: An Operational Definition for Teach-Back Training, HLRP: Health Literacy Research and Practice, 4.2e94-e103, 2020.

Wert	Motivation	Messmethode
Handover-Angemessenheit	Übergabe erfolgt, wenn Unsicherheit dies verlangt?	sicher.«,»Ich fühlte mich geschützt. Kontinuierliche norm- und sicherheitsbasierte Auslösung durch System mit tatsächlichem Handover und Fachkraft-Begutachtung

Die in Tabelle 2 vorgeschlagenen Werte und Methoden sind keinesfalls ausreichend für eine umfassende Einschätzung eines KET-Systems; sie sollen als Ausgangspunkt für zukünftige Arbeiten dienen. Die Teach-Back-Methode wird im Gesundheitswesen angewendet, um Pflegebedürftige in eigenen Worten erklären zu lassen, was sie verstanden haben und was sie als Nächstes tun sollen.

8.1 Datenschutz, Datensparsamkeit und Governance

Da KET intime Signale erfasst, sind Privacy-by-Design, minimale Datenerhebung (Data-Protection-by-Design), zweckgebundene Speicherung, Protokollierung und Human-in-the-Loop zwingend. Das gilt insbesondere in sensiblen Pflegesettings. Bereits adressiert wurden Transparenz, Einwilligung, Protokolle, Erklärbarkeit sowie regelmäßige Wirksamkeits- und Risikochecks. Technisch bedeutet das konkret, dass eine On-Device-/Premise-Verarbeitung erfolgen muss, solange keine geeigneten Verfahren zur Anonymisierung zur Verfügung stehen. Alle individuellen Daten müssen in verschlüsselter Form gespeichert werden.

Neben der sicheren Speicherung erfordert Transparenz eine differenzierte Darstellung der internen Parameterwerte: Das System führt nicht nur aktuelle Annahmen zu Sicherheit, Autonomie und Selbstachtung der Person, sondern auch die Ausgangswerte der initialen Erfassung sowie die Differenz zwischen beiden. Verändert das System seine Annahmen aufgrund von Interaktionsbeobachtungen, wird dies mit Zeitstempel, Auslöser und Begründung protokolliert.

Auf Nachfrage muss das System versuchen, diese Veränderungen in adressatengerechter Sprache zu erklären. Der pflegebedürftigen Person in einfacher Sprache, wobei die betroffenen Werte und ihre

Veränderungsrichtung benannt werden (z. B.: *»Ich habe bemerkt, dass Sie beim Aufstehen öfter gezögert haben. Das hat meine Einschätzung verändert: Ich gehe jetzt davon aus, dass Ihr Sicherheitsgefühl etwas geringer ist als zu Beginn — und dass Sie sich weniger selbstständig fühlen als früher. Ich schlage deshalb vor, mehr Unterstützung anzubieten. Kann ich Ihnen das genauer erklären oder haben Sie Fragen dazu?«*). Der Pflegekraft auf Fachebene mit explizitem Bezug zur Parameterentwicklung (z. B.: *»Meine Einschätzung zum Sicherheitserleben der Person hat sich merklich verringert (Sicherheit: 0,8 auf 0,6) und auch die wahrgenommene Autonomie ist gesunken (Autonomie: 0,9 auf 0,7).« Grundlage sind drei der letzten fünf Morgeninteraktionen, bei denen sensorisch erhöhte Bewegungsögerung und sprachlich geäußerte Unsicherheit erfasst wurden. Die Abweichung vom Ausgangswert beträgt jeweils $-0,2$.«*). Personen mit rechtlich erteilter Vorsorgevollmacht erhalten Zugang zu denselben Informationen im Rahmen ihrer Befugnisse. Alle Anfragen zu diesen Werten — wer, wann, welcher Parameter — werden revisionssicher protokolliert.

Die professionelle Einschätzung der Pflegekraft hat gegenüber dem Systemmodell stets Vorrang: Parameter können jederzeit überprüft werden; alle Änderungen werden protokolliert. Bei mittleren Abweichungen zwischen Systemannahme und Pflegekraft-Einschätzung empfiehlt das System eine Intervention — eine strukturierte Peer-Reflexion, bei der das KET-System Beobachtungsdaten und den Parameterverlauf als Gesprächsgrundlage bereitstellt, ohne über die Pflegekraft zu urteilen. Bei großen Abweichungen oder wenn sicherheitsrelevante Minimalwerte betroffen sind, greift eine Kontrollebene: Der Schutz der pflegebedürftigen Person hat dabei absoluten Vorrang. Die physische Anwesenheit der Pflegekraft ist in diesen Fällen stets Pflicht — sie führt die Änderung durch und trägt die Verantwortung. Ist eine zweite Pflegeperson verfügbar, muss die Änderung zusätzlich im Vier-Augen-Prinzip bestätigt werden. Ist dies nicht möglich (z. B. Nachtschicht, Unterbesetzung), genügt die physische Anwesenheit; der Fall wird jedoch automatisch für eine nachträgliche Supervision markiert. Ein Remote-Override auf sicherheitskritischer Stufe ist nicht möglich.

8.2 Limitierte Erlebnissimulation versus echte Erlebnisfähigkeit

KET erlebt nicht selbst. Sie rekonstruiert relevante innere Prozesse durch Simulation und wählt Handlungen wertegebunden aus. Auf technischer Ebene ist eine klare Grenze zwischen simulierter Empathie und menschlicher Erlebnisfähigkeit festgelegt. So werden Täuschungen im Kontext, in dem die Maschine »erlebt« oder »fühlt«, offensichtlich. Erklärungen und die Validierung der Simulation müssen sich auf nachvollziehbare Wirkungsgrößen stützen: wie verständlich das System seine Entscheidungen kommuniziert, ob das Erleben von Autonomie gestärkt oder geschwächt wurde und wie sicher sich die Person subjektiv fühlt (Tabelle 2).

9. Risiken und Grenzen beleuchten

9.1 Risiken und ethisch-anthropologische Herausforderungen

Emotional adaptive Technologie und KET gehen mit Risiken einher. Diese sind unter anderem erleichterte Manipulationsmöglichkeiten durch Anthropomorphisierung, emotionale Abhängigkeit, technische Fehlinterpretationen sensibler Signale sowie Verzerrungen durch unausgewogene Trainingsdaten^{131,132}. Risiken müssen durch Leitplanken, wie sie beispielsweise der EU AI Act¹³³ und andere, reguliert sein: 1) Transparenz darüber, dass es sich um Technologie handelt; 2) informierte Einwilligung und besondere Schutzregeln bei eingeschränkter Entscheidungsfähigkeit; 3) Human-in-the-Loop insbesondere für sensible Assistenz; 4) Protokollierung, Erklärbarkeit und minimale Datenerhebung; 5) Risikoklassifizierung (z. B. Hochrisiko im Pflegekontext) und daraus folgende Konformitäts- und Qualitätsmanagementpflichten (auch Kapitel 8.1, 6) regelmäßige Wirksamkeits- und Risiko-Überprüfungen mit Hand-Over im laufenden Betrieb.

Der Gewinn kann nicht maßgeblich an der Erkennungsgenauigkeit von Beobachtbarem und der Simulationsfähigkeit innerer

131 Andrew McStay, *Emotional AI: The Rise of Empathic Media*, 1–248, 2018.

132 Saif M. Mohammad, *Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis*, *Computational Linguistics*, 48.2: 239–278, 2022.

133 <https://artificialintelligenceact.eu/de/> (zuletzt aufgerufen 12. September 2025).

Prozesse gemessen werden, sondern an pflegespezifischen Werten: Autonomie-Wahrung, Angemessenheit der Regulierung (z. B. de-eskalierend statt kontrollierend), subjektive Erlebensverbesserung, Beziehungsqualität, Transparenz und Verständlichkeit, Sicherheitsergebnisse (Reduktion) sowie arbeitspraktische Effekte (z. B. wahrgenommene Entlastung, Zeitgewinne ohne Qualitätsverlust).

Täuschungsrisiken. Sozial wirkende Systeme führen leicht zu einer falschen Zuschreibung von Verständnis und Fürsorge: Simulierte Empathie, Höflichkeit und flüssiges interaktives Verhalten können als »echtes (Mit)Fühlen« missverstanden werden. Über die wohl per se stattfindende Anthropomorphisierung kann es zu einer Schwächung der individuellen Autonomie kommen. Damit kann eine Verschiebung der Verantwortung einhergehen. KET muss daher Fähigkeiten und Grenzen transparent und überprüfbar kommunizieren, Alternativen benennen und Erklärungen mit Wertebezug liefern. Besondere Aufmerksamkeit gilt der Tendenz, simulierte Empathie unkritisch als Abbild der Wirklichkeit zu akzeptieren, statt sie als überprüfbare Hypothese zu behandeln¹³⁴. KET begegnet dem durch explizite Unsicherheitskalibrierung und Transparenz hinsichtlich der Modellgrundlage.

Abhängigkeitsrisiken. In Pflege- und Hilfekontexten drohen funktionale und emotionale Abhängigkeit. Durch Bequemlichkeit, soziale Kompensation und ständige Verfügbarkeit können Selbstwirksamkeit und Beziehungsnetzwerke erodieren. Durch institutionelle Asymmetrien verschärft sich das Machtgefälle. Die entstehenden emotionalen Bindungen sind nicht nur psychologisch, sondern auch normativ zu bewerten — etwa im Hinblick auf Selbstwirksamkeit, Verantwortungszuschreibungen und Vorstellungen vom guten Leben¹³⁵. KET braucht nicht zwingende Defaults, Widerrufbarkeit, Human-in-the-Loop, sowie Autonomie- und Selbstachtungs-Checks durch regelmäßige Prüfung durch externe Experten.

Auswirkungen auf Menschenbild und Würde. Werden Menschen primär als Datenprofile modelliert, droht eine Verengung des Men-

134 Karen Yirmiya und Peter Fonagy, Mentalizing Without a Mind: Psychotherapeutic Potential of Generative AI, *Journal of Medical Internet Research* 27:e79156, 2025.

135 Janina Loh und Wulf Loh, eds. *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots*, 1st ed, transcript Verlag, 2023.

schenbildes auf Vorhersagbarkeit und Steuerbarkeit. Würde wird berührt, wenn intime Signale zweckentfremdet werden, Rollen entgrenzt oder Interaktionen ohne transparente Aushandlung und Erklärung gestaltet werden. KET setzt dem ein relationales Menschenbild entgegen. Es nimmt Beziehungen statt Etiketten als Maßstab. Dazu zählen zentrale menschliche Werte wie Sicherheit, Autonomie, Selbstachtung, Reparaturpfade bei Fehlpassung sowie Datensparsamkeit. Empathie wird als verantwortete, erklärbare Co-Regulation umgesetzt.

9.2 Das Unheimliche Tal: Wenn Erwartungen brechen

Durch soziale Interaktionen werden Beziehungen gepflegt und abgestimmt; sie sind für uns von entscheidender Bedeutung. Passen soziale Interaktionssignale nicht zum Personenmodell des Gegenübers *G* und den dazugehörigen Repräsentationen (Abbildung 4), führt dies zu einer Korrektur der Erwartungen und zu einem Umbau, was Piaget als Akkommodation bezeichnet.

Daraus lässt sich die Hypothese ableiten, dass dieser Prozess zum Uncanny-Phänomen im Kontext künstlicher menschenähnlicher Figuren¹³⁶ beiträgt. Masahiro Mori beschrieb 1970, dass stark menschenähnliche Androiden und realistische Prothesen plötzlich Unbehagen auslösen können. Dieses Phänomen ist nicht mit dem Uncanny-Valley-Graphen zu verwechseln. Der Graph beschreibt als technisches Modell den Zusammenhang zwischen Menschenähnlichkeit und Vertrautheit, wobei ein Einbruch als »Tal« im Graphen dargestellt wird. Das Uncanny-Phänomen bezeichnet das psychologische Erleben dieses Einbruchs: Plötzliche Fremdheit oder Unbehagen trotz hoher Menschenähnlichkeit. Das Konzept des Unheimlichen und seine Verwendung im Kontext menschenähnlicher Automaten wurden 1906 von Jentsch eingeführt¹³⁷.

Die durch den Umbau verursachte Rekategorisierung beziehungsweise Dehumanisierung löst Ekel oder Angst aus, wie einige Erklä-

136 Masahiro Mori, Karl F. MacDorman und Norri Kageki, The Uncanny Valley [from the field], IEEE Robotics & Automation Magazine 19.2: 98–100, 2012.

137 Ernst Jentsch, Zur Psychologie des Unheimlichen, 1906.

rungen¹³⁸ bemerken. Diese gehen davon aus, dass die Ursache der Unheimlichkeit im Prozess liegt, der von einer automatisierten Personalisierung ausgeht, wodurch das Unmenschliche enttarnt wird und ablehnende Emotionen ausgelöst werden. Eine tiefergehende modellbasierte Erklärung der beteiligten Prozesse (Vorhersageverletzung mit Einfluss auf die innere Sicherheit) wird nicht gegeben; das gesamte Phänomen ist jedoch für eine konsistente, vertrauensvolle Erstellung von KET-Systemen sehr wichtig und muss mit weiterer Grundlagenforschung ergründet werden, etwa durch den Ansatz von Wessler und Kollegen¹³⁹.

9.3 Kritische Reflexionen zum Menschenbild

Der vorgestellte Ansatz versteht Empathie als modellierte, erklärbare Co-Regulation statt Erlebnisfähigkeit. Damit knüpft er an eine relationale Anthropologie an, die den Menschen weder auf Datenpunkte reduziert noch romantisiert.

Die weitere Reflexion folgt klaren Leitprinzipien: Verkörperung und Handlungsfähigkeit. Entlastende Assistenz unter Regeln. Würde- und Beziehungspraxis. Transparenz und Rechenschaft. Eingebettet in Pluralität, Verletzlichkeit und Datengerechtigkeit.

Die von Plessner benannte *Doppelaspektivität* beschreibt den Menschen als natürlich-verkörpert und frei-handelnd¹⁴⁰. Für technische Systeme bedeutet das, dass sie funktional ergänzen, aber nicht substituieren dürfen, was personale Freiheit, Verantwortung und Sinnsetzung ausmacht. Simulation sozialer Funktionen (Kapitel 5, insbesondere 5.5) ist nicht gleichzusetzen mit dem Besitz der Qualitäten, auf die sich diese Funktionen beziehen (Würde, Verantwortung, Intentionalität). Technische Simulationen von Empathie dürfen soziale Kommunikationshinweise aufnehmen und relativ

138 Shensheng Wang, Scott O. Lilienfeld und Philippe Rochat, The Uncanny Valley: Existence and Explanations, *Review of General Psychology* 19: 393 – 407, 2015.

139 Janet Wessler, Patrick Gebhard und Sigal Zilcha-Mano, Investigating Movement Synchrony in Therapeutic Settings using Socially Interactive Agents: An Experimental Toolkit, *Frontiers in Psychiatry* 15: 1330158, 2024.

140 Helmut Plessner, *Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie*, Walter de Gruyter GmbH & Co KG, 2019 (Nachdruck von 1928).

unterstützen, aber sie können persönliches, frei verantwortetes Handeln nicht ersetzen. Daraus ergeben sich Design- und Governance-Prinzipien, die Autonomie und Verantwortlichkeit sichern und im KET-Rahmenwerk berücksichtigt werden.

Gehlen beschreibt den Menschen als Mängelwesen — biologisch wenig spezialisiert, dafür handlungs- und kulturfähig¹⁴¹. Menschen bleiben handlungsfähig, weil klare Spielregeln, Rollen und verlässliche Abläufe sowie Werkzeuge und Technologie sie im Alltag entlasten. Diese Entlastung schafft Raum für das Wesentliche: Selbstbestimmung, Beziehung, verantwortungsvolles Handeln. Technische Systeme sind nur dann legitim, wenn sie Entlastung als Ermöglichung von Selbst- und Mit-Regulation verstehen, ohne Autonomie und Verantwortlichkeit auszuhöhlen. Genau diese Bedingungen adressiert das KET-Rahmenwerk durch Komplementarität, Human-in-the-Loop, Erklärbarkeit und Verantwortungsklarheit. KET ergänzt menschliche Urteilskraft und Beziehungsgestaltung, erhöht die Handlungsfähigkeit — und überschreitet nie die Grenze, an der KET den Menschen als verantwortliches Subjekt ersetzt.

Eine medizinethische Position stellt die Würde des Menschen ins Zentrum¹⁴². Entscheidungen sollen im Rahmen einer professionellen Beziehung zwischen Fachkräften und Pflegebedürftigen im Dialog getroffen werden. Die Basis dafür ist situatives und individuelles Kontextwissen sowie eine verantwortungsvolle Abwägung. Technologie kann unterstützen, darf jedoch nicht die professionelle Urteilskraft ersetzen. Denn sonst kann es passieren, dass Werte, Individualität und situative Gegebenheiten hinter Metriken verschwinden. Dieses Leitbild wird in KET strukturell umgesetzt und durch begleitende Rahmenbedingungen abgesichert.

Digitaler Humanismus¹⁴³ und die Leitlinien für Trustworthy AI¹⁴⁴ verlangen, dass Technologie Menschen stärkt, nicht steuert. KET übersetzt dies in prüfbare Betriebsauflagen: transparente Kennzeichnung inklusive Fähigkeiten und Grenzen, Human-in-the-Loop bei

141 Arnold Gehlen, *Der Mensch. Seine Natur und seine Stellung in der Welt*, 1940.

142 Giovanni Maio, *Werte für die Medizin: Warum die Heilberufe ihre eigene Identität verteidigen müssen*, Kösel-Verlag, 2018.

143 Julian Nida-Rümelin, *Digitaler Humanismus: Eine Ethik für das Zeitalter der Künstlichen Intelligenz*, 2018.

144 Michel Cannarsa, *Ethics Guidelines for Trustworthy AI*, The Cambridge Handbook of Lawyering in the Digital Age, 283–297, 2021.

werterelevanten Entscheidungen, Privacy-by-Design sowie revisionssichere Protokollierung und regelmäßige Audits. MEESTAR¹⁰ dient als wiederkehrende Governance-Routine und ist mit der Pflege-QM sowie den relevanten Messmethoden (Kapitel 8) verzahnt.

Autonomie ist stets sozial in Beziehungen, Rollen und Institutionen eingebettet¹⁴⁵ — KET trägt dem Rechnung, indem es relationale Autonomie modelliert und individuelle Werte wie Sicherheit und Selbstachtung in anpassbaren Lernschleifen verankert.

Mehrwert entsteht dort, wo KET Kommunikationshindernisse überbrückt, Co-Regulation stabilisiert, das Sicherheitserleben erhöht und Teilhabe und Autonomie stärkt. Die Bewertung richtet sich daher weniger nach Klassifikationsmetriken als nach relationalen Outcomes (z. B. Transparenz, Autonomie, Sicherheitsgefühl). Gleichzeitig verlangen Täuschungs- und Abhängigkeitsrisiken strikte Leitplanken: keine Suggestion von »Fühlen«, klare Fähigkeits- und Grenzkommunikation, Widerrufbarkeit und Übergabe an Menschen, faire Defaults sowie Datensparsamkeit. Praktisch heißt das: partizipative Modellierung beziehungsweise Co-Design, wertegebundene Zielfunktion mit Schwellen und Ausnahmen, Simulation mit Unsicherheitskalibrierung, Erklärbarkeit und Protokoll für Entscheidungen und Alternativen, Evaluation von Beziehungs- und Werteindikatoren sowie kontinuierliche Governance (siehe oben).

10. Zusammenfassung und Ausblick

Dieses Buchkapitel beschreibt *künstlich empathische Technologie* (KET) als wertegebundene Interaktionsarchitektur. Empathie ist dort nicht als Erlebnisfähigkeit, sondern als modellierte, erklärbare Co-Regulation gestaltet, die explizit individuelle innere Prozesse und Repräsentationen einbezieht. Entlang der Prozesskette (1) Wahrnehmung, (2) Interpretation, (3) Mentalisierung und Regulation, (4) Ausdruck integriert KET die eigenen Kommunikationssignale und die des Gegenübers, simuliert Alternativhypothesen für Bedürfnisse und Ziele und betrachtet dabei den Einfluss der Beziehungsreprä-

145 Catriona Mackenzie und Natalie Stoljar, *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, Oxford University Press, 2000.

sensation, wägt Handlungen simulativ gegen die Zielwerte Sicherheit, Autonomie und Selbstachtung ab. Entscheidungen müssen empfängergerecht erklärt werden. Damit übersetzt KET ein relationales, würde- und autonomiesensibles Menschenbild in konkrete Designentscheidungen — mit Human-in-the-Loop, Werteschwellen, Widerrufbarkeit und Datensparsamkeit als verbindliche Leitplanken.

Im Vergleich zur emotional adaptiven Technologie verschiebt KET den Maßstab. Entscheidend ist nicht primär die Güte der Erkennung, sondern die Angemessenheit regulativer Entscheidungen im Hinblick auf Sicherheit, Autonomie und Selbstachtung. Diese Indikatoren ermöglichen eine relationale Evaluation jenseits rein technischer Klassifikationsmetriken. Täuschungs- und Abhängigkeitsrisiken werden nicht nur benannt, sondern durch Transparenz über Nicht-Fühlbarkeit, faire Defaults, die Übergabe an Menschen sowie kontinuierliche Risiko- und Wirkungsprüfungen aktiv bearbeitet.

Um KET vom Konzept in die Praxis zu überführen, bedarf es einiger Schritte:

- 1) Domänenabhängige partizipative Modellierung mit Zielgruppen in realen Settings;
- 2) Zentrale Realisierung von Zielfunktionen für Sicherheit, Autonomie und Selbstachtung mit begründeten Gewichten, Mindestgrenzen und Ausnahmefällen;
- 3) Simulative Entscheidungs- und Handlungsmodellierung mit Unsicherheits-/Risikokalibrierung und expliziten Alternativen basierend auf 2);
- 4) Empfängerorientierte Erklärbarkeit sowie Protokollierung auf Interaktionsebene;
- 5) Evaluation über relationale Metriken und MEESTAR-kompatible¹⁰ Prüffragen;
- 6) Realisierung von Governance-Methoden im Betrieb: Human-in-the-Loop, Widerruf, Datenminimierung, regelmäßige Audits und (Re)Kalibrierung.

KET entwirft Technologie als Stütze menschlicher Handlungs- und Beziehungskompetenz. Wo Kommunikation stockt, Routinen brüchig sind oder Sicherheit leidet, bietet KET strukturierte Co-Regulation, ohne menschliche Urteilskraft zu ersetzen. Der nächste Schritt besteht darin, skalierbare Prototypen mit klaren Werte- und

Evidenznachweisen in Pflege-, Bildungs- und Assistenzkontexten zu erproben — sodass Menschenbild, Würde und Autonomie nicht nur gewahrt, sondern auch gestärkt werden.

Glossar

<i>Affektiv / kognitiv</i>	Affektiv bezieht sich auf Gefühle und emotionale Reaktionen; kognitiv auf Denkprozesse, Überzeugungen und Wahrnehmung. Empathie und Regulation verbinden stets beide Aspekte.
<i>CASA-Paradigma</i> (Computers As Social Actors)	Befund, dass Menschen Computern und Robotern unbewusst soziale Eigenschaften zuschreiben und entsprechend darauf reagieren — auch wenn sie wissen, dass es sich um Technik handelt.
<i>Co-Regulation</i>	gegenseitige emotionale Stabilisierung zwischen zwei Personen in sozialer Interaktion; z. B., wenn eine ruhige Pflegekraft einer aufgewühlten Person hilft, sich zu beruhigen.
<i>Display-Regulation</i>	Steuerung des nach außen sichtbaren emotionalen Ausdrucks (Mimik, Sprache, Gestik), unabhängig vom tatsächlichen inneren Erleben.
<i>Emotionsregulation</i>	innere Prozesse, mit denen Menschen ihre Gefühle wahrnehmen, bewerten und anpassen — bewusst oder unbewusst.
<i>Mentalisierung</i>	Fähigkeit, eigenes und fremdes Verhalten als durch innere Zustände (Gedanken, Gefühle, Absichten, Werte) verursacht zu verstehen und zu regulieren. Grundlage sozialer Verständigung und Empathie.
<i>Mentalisierungs- basierte Therapie</i> (MbT)	therapeutischer Ansatz zur Stärkung und Wiederherstellung von Mentalisierungsfähig-

	keiten, insbesondere nach Interaktionsabbrüchen oder in Belastungssituationen.
<i>Personenmodell G</i>	KET-interne Repräsentation der angenommenen Wahrnehmung, Werte und Erwartungen des Gegenübers; Grundlage für die Simulation, wie eine Handlung die andere Person berührt.
<i>Predictive Processing</i>	Grundprinzip, nach dem das Gehirn (und KET) ständig Vorhersagen über die Außenwelt trifft und diese bei Abweichungen korrigiert. Erlaubt vorausschauendes, nicht nur reaktives Handeln.
<i>Regulative Assistenzaktionen</i>	KET-Handlungen, die gezielt darauf ausgerichtet sind, Sicherheit, Autonomie und Selbstachtung der pflegebedürftigen Person zu stärken — von einem aufmunternden Satz bis zu einem konkreten Hilfsangebot.
<i>Strukturelle / situative / kommunikative Emotionen</i>	Moser & von Zeppelin: Unterscheidung zwischen inneren emotionalen Informationen, die sich auf sich selbst (strukturell), auf die Situation (situativ) und auf das außen im Verhalten Gezeigte (kommunikativ) beziehen.
<i>Theory of Mind (ToM)</i>	Fähigkeit, sich vorzustellen, was andere Menschen denken, fühlen oder beabsichtigen; Voraussetzung für soziales Verstehen und Empathie.

