












# Dienstleistungsroboter im Handel – Einsatzmöglichkeiten und verantwortungsbewusster Einsatz

Ruth Stock-Homburg , Merlind Knof , Jérôme Kirchhoff ,  
Judith S. Heinisch , Andreas Ebert , Philip Busch , Klaus David ,  
Janine Wendt , Indra Spiecker gen. Döhmann , Oskar von Stryk   
und Martin Hannig 

## Inhaltsverzeichnis

- 1 Einleitung
- 2 Forschungsstand zur verantwortungsbewussten MRI
  - 2.1 Ethik in der MRI
  - 2.2 RoboCompliance – Sicherheit in der MRI
  - 2.3 Privatsphäre und Datenschutz in der MRI
  - 2.4 Transparenz in der MRI
  - 2.5 Bewertung der verantwortungsbewussten MRI

Der vorliegende Beitrag ist ein Ergebnis des Pilotprojektes „RoboTrust“ des Zentrums verantwortungsbewusste Digitalisierung (ZEVEDI).

Dieser Beitrag beruht im Wesentlichen auf einer Studie der „RoboTrust“ Projektgruppe aus dem Jahr 2020 sowie aus den folgenden englischsprachigen Konferenzbeiträgen der AutorInnen, die bei der 55. Hawaii International Conference on System Sciences eingereicht und im Rahmen der Conference Proceedings im Januar 2022 veröffentlicht wurden: Stock-Homburg et al. (2022) und Knof et al. (2022).

---

R. Stock-Homburg (✉) · M. Knof · J. Kirchhoff · P. Busch · J. Wendt · O. von Stryk ·  
M. Hannig  
Technische Universität Darmstadt, Darmstadt, Deutschland  
E-Mail: [rsh@bwl.tu-darmstadt.de](mailto:rsh@bwl.tu-darmstadt.de)

M. Knof  
E-Mail: [merlind.knof@bwl.tu-darmstadt.de](mailto:merlind.knof@bwl.tu-darmstadt.de)

- 3 Interdisziplinärer Bezugsrahmen
- 3.1 Verantwortungsbewusstes physisches Design anthropomorpher Dienstleistungsroboter
- 3.2 Verantwortungsbewusste Programmierung
- 3.3 Verantwortungsbewusster Einsatz anthropomorpher Dienstleistungsroboter
- 4 Implikationen und Gestaltungsempfehlungen
- Literatur

## Zusammenfassung

Anthropomorphe Dienstleistungsroboter gewinnen immer mehr an Popularität. Je leistungsfähiger sie werden und je stärker sie in unserem Alltag integriert sind, desto wichtiger wird die verantwortungsbewusste Gestaltung der Mensch-Roboter-Interaktion (MRI). Hierbei sind Menschenwürde, Transparenz, Privatsphäre, Datenschutz und Compliance im verantwortungsbewussten Einsatz anthropomorpher Dienstleistungsroboter von zentraler Bedeutung. Dieser Beitrag nennt Tätigkeiten von Dienstleistungsrobotern im Handel und bietet einen interdisziplinären Überblick über den aktuellen Forschungsstand zur verantwortungsbewussten Gestaltung der MRI unter besonderer Berücksichtigung der vier Disziplinen Ethik, Recht, Psychologie

J. Kirchhoff

E-Mail: [kirchhoff@sim.tu-darmstadt.de](mailto:kirchhoff@sim.tu-darmstadt.de)

P. Busch

E-Mail: [busch@jus.tu-darmstadt.de](mailto:busch@jus.tu-darmstadt.de)

J. Wendt

E-Mail: [wendt@jus.tu-darmstadt.de](mailto:wendt@jus.tu-darmstadt.de)

O. von Stryk

E-Mail: [stryk@sim.tu-darmstadt.de](mailto:stryk@sim.tu-darmstadt.de)

M. Hannig

E-Mail: [martin.hannig@bwl.tu-darmstadt.de](mailto:martin.hannig@bwl.tu-darmstadt.de)

J. S. Heinisch · K. David

Universität Kassel, Kassel, Deutschland

E-Mail: [judith.heinisch@uni-kassel.de](mailto:judith.heinisch@uni-kassel.de)

K. David

E-Mail: [david@uni-kassel.de](mailto:david@uni-kassel.de)

A. Ebert · I. S. g. Döhmman

Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland

E-Mail: [ebert@jur.uni-frankfurt.de](mailto:ebert@jur.uni-frankfurt.de)

I. S. g. Döhmman

E-Mail: [spiecker@jur.uni-frankfurt.de](mailto:spiecker@jur.uni-frankfurt.de)

und Technik. Zudem wird ein interdisziplinärer Bezugsrahmen für die Gestaltung einer verantwortungsbewussten MRI mit anthropomorphen Dienstleistungsrobotern entwickelt und präsentiert. Abschließend werden wissenschaftliche Implikationen abgeleitet und weitere Forschungsfelder hinsichtlich einer verantwortungsbewussten Gestaltung der MRI mit anthropomorphen Dienstleistungsrobotern formuliert.

## 1 Einleitung

In der vergangenen Dekade haben sowohl anthropomorphe Dienstleistungsroboter mit Extremitäten, wie Armen, Beinen oder einem Kopf (Mara & Appel, 2015b), als auch autonomes Verhalten und der Einsatz von künstlicher Intelligenz (KI) bei der Erbringung von Dienstleistungen vielversprechende Chancen und nützliche Perspektiven geboten (siehe Rawal & Stock-Homburg, 2021; Stock & Merkle, 2017 für Übersichten). Hierbei sind Geldautomaten, Selbstbedienungskassen und Touchscreen-Bestellungen für Fast Food bereits etablierte Technologien und Vorläufer von Dienstleistungsrobotern. Obwohl die Robotik sich noch nicht in allen Bereichen etabliert hat, sind „human-humanoid encounters in the marketplace [...] not as futuristic as they might seem, and they represent a primary arena for innovation in services“ (Mende et al., 2019, S. 535). Zukünftig werden Dienstleistungsroboter auf allen wirtschaftlichen Ebenen (d. h. Mikro-, Meso- und Makroebene) eine wesentliche Rolle spielen (Wirtz et al., 2018). So werden Roboter den Menschen in Zukunft verstärkt unterstützen, indem sie verschiedene Dienstleistungen anbieten. Hierzu zählen z. B. das Bereitstellen von Informationen und Beratungsleistungen in verschiedenen Bereichen, wie dem Bankwesen, dem Einzelhandel, dem Gastgewerbe, dem Gesundheitswesen und dem Bildungswesen (Hofmann, 2020; Ivanov et al., 2017; Merkle, 2019).

Insbesondere für den Einzelhandel bietet der Einsatz von Robotern vielversprechende Perspektiven. Die Ansprüche und Bedürfnisse von KundInnen im Kaufprozess verändern sich zunehmend (Gieselmann & Gremmer, 2018). Durch den Einsatz von KI im Einzelhandel kann beispielsweise der gestiegenen Komplexität im Handel begegnet werden, die unter anderem durch die wachsenden Anforderungen der KundInnen bedingt ist (Gläß, 2018). So stellt der Einsatz von Robotern im stationären Einzelhandel eine wesentliche Komponente dar, um die HändlerInnen-KundInnen-Beziehung neu zu definieren und so der rasant steigenden Entwicklung von online getätigten Einkäufen begegnen zu können

(Gieselmann & Gremmer, 2018). Das mögliche Aufgabenfeld von Dienstleistungsrobotern ist vielfältig und erstreckt sich im stationären Einzelhandel von der Begrüßung von KundInnen über die Vorbereitungs- bis hin zur Empfehlungs- und Informationsfunktion (Gieselmann & Gremmer, 2018). Abb. 1 liefert einen Überblick über die Ergebnisse einer Online-Befragung von NutzerInnen in Deutschland und in den USA über die möglichen Tätigkeiten von Dienstleistungsrobotern im Handel. Bemerkenswert ist zunächst, dass die Befragten beider Länder sich zahlreiche Tätigkeiten von Dienstleistungsrobotern vorstellen konnten. Darüber hinaus erachtet der Großteil der Befragten in den USA<sup>1</sup> (66,7 %) und in Deutschland<sup>2</sup> (59,6 %) einen verstärkten Einsatz von Dienstleistungsrobotern aufgrund der COVID-19-Pandemie als sinnvoll.

Roboter „with human-like appearance features such as eyes, hands, or faces“ (Phillips et al., 2018, S. 105) werden als *anthropomorphe Roboter* bezeichnet (Phillips et al., 2018). Im Gegensatz zu digitalen Assistenten oder Automaten können anthropomorphe Roboter emotionsähnliche Signale durch Gestik, Mimik und Haptik (d. h. Handkontakt und Berührungen) ausdrücken. Dies ermöglicht es Menschen, mit anthropomorphen Robotern auch über nicht-verbale Signale zu interagieren (Rawal & Stock-Homburg, 2021) und sogar Beziehungen mit ihnen aufzubauen (Schweitzer et al., 2019). Die physische Präsenz anthropomorpher Dienstleistungsroboter ist ein wesentlicher Erfolgsfaktor der Mensch-Roboter-Interaktion (MRI) (Leyzberg et al., 2012), da diese die Akzeptanz der Menschen gegenüber dieser Roboter bedingt (Gieselmann & Gremmer, 2018).

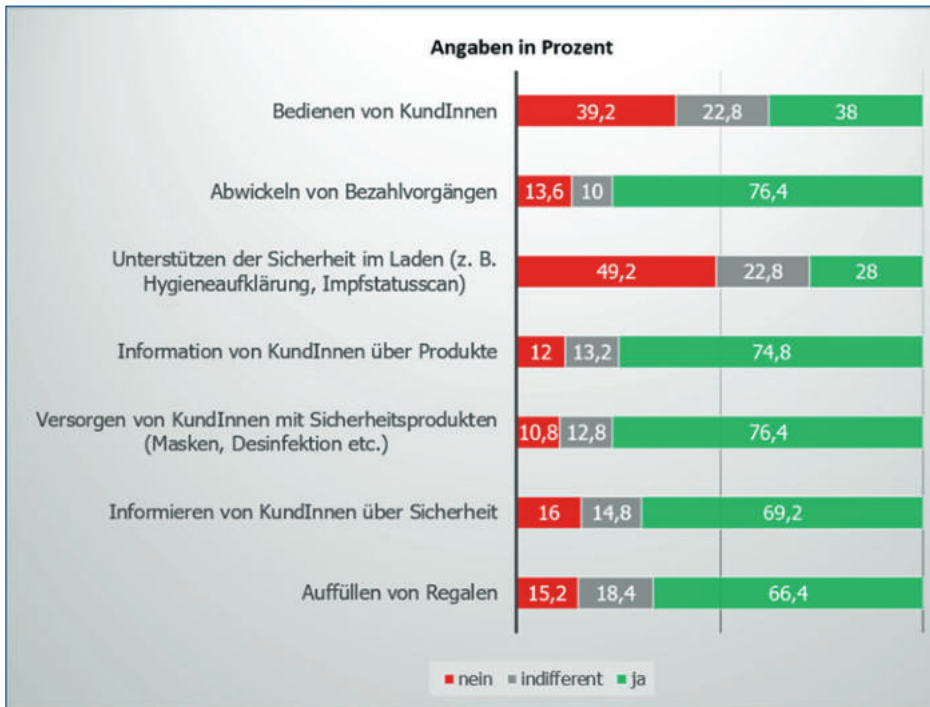
In diesem Beitrag wird MRI definiert als „the study of the humans, robots, and the ways they influence each other“ (Fong et al., 2003, S. 256). Um Roboter in unser Leben zu integrieren, ist es wichtig, die MRI verantwortungsbewusst zu gestalten und ethische Fragestellungen sowie die Wirkungsmechanismen und Regulierungen der MRI zu berücksichtigen. „If we do not address futuristic ethical issues now, it may be too late once technological innovations force them upon us“ (Belk, 2020, S. 2).

Unser interdisziplinäres Forschungsprojekt „RoboTrust“ verfolgt den wissenschaftlichen Leitgedanken einer verantwortungsbewussten Gestaltung der MRI mit anthropomorphen Dienstleistungsrobotern. Hierfür wurden fünf Prämissen zugrunde gelegt.

<sup>1</sup>Befragte: n=475 private NutzerInnen via Amazon Mechanical Turk; Alter: 22 % zwischen 21 und 30 Jahre, 39 % zwischen 31 und 40 Jahre, 18 % zwischen 41 und 50 Jahre, 21 % über 50 Jahre; Geschlecht: 57 % männlich, 42 % weiblich, 1 % divers.

<sup>2</sup>Befragte: n=252 private NutzerInnen via Amazon Mechanical Turk; Alter: 9 % jünger als 20 Jahre, 34 % zwischen 21 und 30 Jahre, 30 % zwischen 31 und 40 Jahre, 17 % zwischen 41 und 50 Jahre, 10 % über 50 Jahre; Geschlecht: 70 % männlich, 29 % weiblich, 1 % divers.

## Mögliche Tätigkeiten von Dienstleistungsrobotern im Einzelhandel



## Weitere Tätigkeiten für Dienstleistungsroboter im Handel

- Zählen von Personen in Geschäften
- Sicherstellen von Abstandswahrung
- Automatisiertes Bezahlen
- Reinigung von Böden
- Kommissionierung von Lebensmittelgeschäften
- Scannen von Verkäufen, um die Interaktion von Mensch zu Mensch zu vermeiden
- Desinfizieren von Regalen
- Einkaufsassistenz
- Entladen von Lieferwagen
- Desinfizieren von Einkaufswägen
- Einsammeln von Einkaufswägen
- Ein- und Auschecken von KundInnen
- Einladen / Einpacken von Lebensmitteln

**Abb. 1** Mögliche Tätigkeiten von Dienstleistungsrobotern im Handel (unveröffentlichte Ergebnisse der „RoboTrust-Studie 2020“)

Anthropomorphe Dienstleistungsroboter sollen

1. den Menschen in Bereichen unterstützen (und nicht ersetzen), die der Mensch aufgrund von Ressourcenbeschränkungen nicht bedienen kann (Abschn. 2.1);
2. die Würde des Menschen achten und sich verantwortungsbewusst verhalten (Abschn. 2.1 und 2.5);
3. in ihrem Interaktionsverhalten transparent und vorhersehbar sein (Abschn. 2.4);
4. personenbezogene und vertrauliche Daten in transparenter Weise und in Übereinstimmung mit den geltenden Datenschutzgesetzen erheben und verarbeiten (Abschn. 2.3) und
5. einer klar definierten Haftungsregelung unterliegen (Abschn. 2.2).

Entsprechend diesen Anforderungen definieren wir den Begriff des *verantwortungsbewussten Designs* als das Design der MRI zwischen anthropomorphen Dienstleistungsrobotern und NutzerInnen unter besonderer Berücksichtigung der menschlichen Würde, der emotionalen Bedürfnisse, der Privatsphäre und des Datenschutzes, der Transparenz sowie der Sicherheit und des Schutzes. Dieser Grundgedanke von „RoboTrust“ ermöglicht die Entwicklung anthropomorpher Dienstleistungsroboter, die zu einer für Menschen authentischen und akzeptablen MRI führen. Daraus ergibt sich unter anderem die Notwendigkeit, psychologische Bedürfnisse innerhalb der MRI im Rahmen von Studien sowie in der täglichen Interaktion zu evaluieren (siehe Knof et al., 2022 für einen Überblick).

Die MRI wird von verschiedenen Fachrichtungen beeinflusst, welche den interdisziplinären Charakter dieses Forschungsgebietes begründen. Die vier Disziplinen Ethik, Recht, Psychologie und Technik sind hierbei von zentraler Bedeutung (Stock-Homburg, 2021). Nach aktuellem Wissensstand der AutorInnen ist der derzeitige Stand der Forschung zur MRI mit anthropomorphen Dienstleistungsrobotern jedoch unvollständig, da er auf einer abstrakten Ebene bleibt oder die Anforderungen nur teilweise bzw. nicht ganzheitlich abdeckt (Ahmad et al., 2017; Baxter et al., 2016; Lutz et al., 2019; Wullenkord & Eyssel, 2020; Zech, 2021). Dies gilt insbesondere dann, wenn die Interdisziplinarität der Gestaltung der MRI berücksichtigt wird. Insofern ist zunächst ein Literaturüberblick zur interdisziplinären Forschung zur verantwortungsbewussten Gestaltung von MRI mit anthropomorphen Dienstleistungsrobotern notwendig (Kap. 2). Zudem haben wir einen interdisziplinären Bezugsrahmen entwickelt und bieten damit Leitlinien für die verantwortungsbewusste Gestaltung von MRI mit anthropomorphen Dienstleistungsrobotern an (Kap. 3). Darüber hinaus werden die Interdependenzen unterschiedlicher Forschungsdisziplinen (insbesondere Ethik, Recht, Psychologie und Technik) betrachtet. Abschließend werden auf der Grundlage der gewonnenen Erkenntnisse künftige Forschungsbereiche identifiziert und weitere Forschungsfelder abgeleitet (Kap. 4).

## 2 Forschungsstand zur verantwortungsbewussten MRI

Mehrere Literaturübersichten über MRI betonen die Bedeutung eines interdisziplinären Ansatzes für eine verantwortungsbewusste MRI (Lutz et al., 2019; Stock-Homburg, 2021; Stock-Homburg et al., 2022; Wirtz et al., 2018; Wullenkord & Eyssel, 2020; Zech, 2021). Um diesen interdisziplinären Rahmen zu schaffen, wird in diesem Kapitel zunächst der allgemeine Forschungsstand aus Sicht verschiedener relevanter Fachdisziplinen wiedergegeben.

### 2.1 Ethik in der MRI

Im Bereich der KI und Servicerobotik werden in der Fachliteratur diverse ethische Herausforderungen thematisiert (Belk, 2020). Dazu zählen die allgegenwärtige Überwachung und die Speicherung persönlicher und verhaltensbezogener Daten (Sodemann et al., 2012). Digitale Assistenten, wie Apples Siri, Amazons Alexa, Microsofts Cortana und Google Home, „not only listen to what we say, they also record it and send it back to Amazon, Apple, or Google“ (Belk, 2020, S. 3). Dies wirft zum einen Fragen der Compliance (Abschn. 2.2) und der Privatsphäre und des Datenschutzes (Abschn. 2.3) auf. Des Weiteren sind jedoch auch Aspekte der Menschenwürde betroffen, da technische Systeme zunehmend nach vorgegeben Kriterien autonom handeln und unklar bleibt, wie weit diese Entscheidungen aus ethischer Sicht gehen dürfen. Studien haben gezeigt, dass NutzerInnen gegenüber diesen digitalen Agenten durchaus verhalten reagieren (Oremus, 2016; Schweitzer et al., 2019). Vor diesem Hintergrund sind ethische Überlegungen in Verbindung mit MRI in akademische und praktische Diskussionen verschiedener Disziplinen eingeflossen, wie z. B. in der Maschinenethik (Anderson & Anderson, 2020), der Computerethik (Crnkovic & Çürüklü, 2012) und der Roboterethik (Veruggio & Operto, 2007). Im Jahr 2019 wurde ein ergänzendes „Ethical Aligned Design“ für autonome und intelligente Systeme vorgeschlagen (The IEEE Global Initiative on Ethics of Autonomous und Intelligent Systems, 2019). Bis heute gibt es eine anhaltende Diskussion darüber, ob Roboter, die ethischen Grundsätzen folgen, überhaupt möglich oder gar wünschenswert sind.

Der Begriff *Roboterethik*, oft auch als „RoboEthics“ (Veruggio & Operto, 2007) bezeichnet, hat mindestens drei verschiedene Bedeutungen (Abney, 2012): a) die Berufsethik der Robotik, b) der Moralkodex, der in einem Roboter selbst einprogrammiert ist, und c) „the self-conscious ability to do ethical reasoning by robots—to a robot’s own, self-chosen moral code“ (Abney, 2012, S. 35). Das Projekt „RoboTrust“ konzentriert sich auf die verantwortungsbewusste Gestaltung von MRI mit anthropomorphen Dienstleistungsrobotern. Daher definieren wir *Roboterethik* als „the moral code programmed into the robots themselves with a special attention to protect human dignity, transparency, privacy, and security“ (Stock-Homburg et al., 2022, S. 2066).

Für die Strukturierung der bestehenden Forschung im Bereich der Roboterethik sind drei Disziplinen von besonderer Bedeutung (Bringsjord & Taylor, 2012): a) die Metaethik, b) die angewandte Ethik und c) die normative Ethik. Die *Metaethik* ist eher abstrakt und konzentriert sich auf die Fragen, ob ein Roboter gut oder schlecht handeln (DeBaets, 2014; Floridi & Sanders, 2004) oder für seine Handlungen verantwortlich gemacht werden kann (Parthemore & Whitby, 2013). Die *angewandte Ethik* konzentriert sich dagegen auf moralische Leitlinien für bestimmte Bereiche und untersucht, welche spezifischen moralischen Dilemmata sich aus dem Einsatz von Robotern ergeben können (Malle, 2016). Eine wichtige Frage ist hierbei, wie EntwicklerInnen einen Roboter für eine konkrete Situation konzipieren, einsetzen und behandeln sollten (Veruggio et al., 2011). Die *normative Ethik* reflektiert schließlich Aspekte des Robotereinsatzes, die obligatorisch, erlaubt oder verboten sind. Aus dieser Perspektive argumentiert die *Moral Mediation Theory* der Technologie (Verbeek, 2011, 2014), dass aufgrund der zunehmenden Nähe zwischen Menschen die Technologie vermehrt als Mediator moralischer Wahrnehmungen und Entscheidungen der Menschen fungiert. Darüber hinaus werden in der Forschung zur *Service-Dominant Logik* (S-D-Logik; (Vargo & Lusch, 2008)) mehrere allgemeine ethische Grundsätze genannt, denen KI-Technologien folgen sollten, wie z. B. Transparenz, Zuverlässigkeit und Fairness (vgl. unter anderem Bostrom & Yudkowsky, 2014). Diese theoretischen Erkenntnisse sprechen für eine bewusste Integration ethischer Werte und Prinzipien in Technologien (Du & Xie, 2021; Friedman & Hendry, 2019; Riva et al., 2012).

Die Moralphilosophie unterscheidet sich von der Moralpsychologie insofern, als die Psychologie versucht, das Zusammenspiel zwischen Emotionen (Abschn. 2.5) und Vernunft als Ursache für individuelle moralische Entscheidungen zu verstehen. Die philosophische Perspektive konzentriert sich hingegen eher auf das „Soll“ als auf das „Ist“ (Wallach, 2010). Der vorliegende Beitrag verfolgt einen umfassenden Ansatz zur angewandten Roboterethik. Dabei sollen Perspektiven der Roboterethik und der Roboterpsychologie (Stock & Nguyen, 2019a; Stock-Homburg, 2021) integriert werden, um eine verantwortungsbewusste Gestaltung anthropomorpher Dienstleistungsroboter zu ermöglichen.

## 2.2 RoboCompliance – Sicherheit in der MRI

Mit der Erweiterung der Einsatzmöglichkeiten von Dienstleistungsrobotern durch den Einsatz von KI steigt auch deren Potenzial, Schaden zu verursachen (Matsuzaki & Lindemann, 2016; Zech, 2021). Durch *Robot Compliance* soll dieses Gefahrenpotenzial minimiert werden (im Folgenden „RoboCompliance“ genannt).

Unter *RoboCompliance* verstehen wir – in Anlehnung an die allgemein anerkannte Definition von Produkt-Compliance als „a product conforming to rules, such as specifications, policies, standards, or laws“ (Loznen et al., 2017, S. 3) – einen Roboter, der regelkonform (van den Hoven van Genderen, 2017) mit den notwendigerweise erhobenen Daten umgeht und kein unangemessen hohes Risiko für eine Schädigung Dritter befürchten lässt (Barfield, 2018). Beim Einsatz von KI müssen demnach vernünftige



Entwicklungs- und Kontrollabläufe in allen Lebensphasen des Roboters gewährleistet werden. Folglich definieren wir *RoboCompliance* als die „compliance with all regulations specifically relevant to intelligent anthropomorphic social robots, including those in data protection and liability law“ (Stock-Homburg et al., 2022, S. 2067).

Transparente und faire Haftungsregeln für Schäden, die durch Roboter verursacht werden, spielen eine zentrale Rolle für die Akzeptanz ihres Einsatzes. Sie sorgen für eine Berechenbarkeit der von NutzerInnen eventuell befürchteten Worst-Case-Szenarien und schaffen damit erst die Voraussetzung, dass sich Roboter etablieren können. Daher ist es entscheidend, sich bereits im Entwicklungsprozess mit Worst-Case-Szenarien von durch Roboter verursachten Haftungsfällen auseinanderzusetzen. Dadurch wird sichergestellt, dass Autonomie und Sicherheit sinnvoll gegeneinander abgewogen werden können (Matsuzaki & Lindemann, 2016).

Wie nachvollziehbare und gerechte Haftungsregeln konkret ausgestaltet werden sollen, ist in Zusammenarbeit mit anderen Disziplinen, darunter insbesondere der Roboterethik (Abschn. 2.1), zu erarbeiten. Deren Beitrag ist für die Umsetzung einer allgemein akzeptierten Haftungsverteilung unerlässlich. Entscheidend ist auch die technologische Transparenz (Abschn. 2.4), die die Grundlage für Kausalitätsnachweise bildet. Die korrekte Datenerhebung und -verarbeitung (Abschn. 2.3) ist für eine sichere autonome Bewegung in der Nähe von und in der Interaktion mit Menschen unerlässlich. Verstöße gegen die Datenschutz-Grundverordnung (DS-GVO) durch den Einsatz eines Roboters können zu empfindlichen Bußgeldern für Unternehmen und Personen führen, die Roboter einsetzen. Ein Gleichgewicht zwischen Innovationsförderung, Regulierung und Verbraucherschutz zu finden, ist daher eine anspruchsvolle Aufgabe, die interdisziplinär bearbeitet werden muss (Hubbard, 2014).

Das erhöhte Schadenspotenzial anthropomorpher Roboter ist insbesondere auf den Einsatz von KI zurückzuführen. Aus deren Einsatz resultiert ein nicht immer vorhersehbares Verhalten (Abschn. 2.4), was neben der Unvorhersehbarkeit robotischer Handlungen im Haftungsfall auch zu Problemen beim Nachweis der Kausalität ex post führt (Lior, 2020; Zech, 2021).

Im Rahmen bestehender Regelungen wäre eine Gefährdungshaftung, also eine verschuldensunabhängige Haftung für Fehlverhalten, eine sinnvolle Option. Neben einigen Problemen in der Anwendung, die sich auch aus der geringen Vorhersehbarkeit komplexer Software ergeben (Erdélyi & Erdélyi, 2021), würde eine verschuldensunabhängige Haftung auch einen deutschen und österreichischen Sonderweg bedeuten, da europaweit nur wenige Staaten dieses rechtliche Konstrukt kennen (Abbott, 2020).

Schließlich sollte bei allen bestehenden Risiken beim Einsatz anthropomorpher Dienstleistungsroboter das große Ganze nicht aus den Augen verloren werden: Auch wenn einzelne Risiken und Schäden nicht außer Acht gelassen werden dürfen, wird die Sicherheit durch den Einsatz von KI in der Summe höchstwahrscheinlich erhöht. Es ist davon auszugehen, dass die Automatisierung mehr Unfälle verhindert als sie verursacht (Abbott, 2020; Wagner, 2019). Daher ist nach Lösungen zu suchen, die weder die Akzeptanz der Innovation noch diese selbst behindern. Hier könnte eine Fonds- (Erdélyi & Erdélyi, 2021)

oder Versicherungslösung (Expert Group on Liability and New Technologies – New Technologies Formation, 2019) sinnvoll sein. Die Europäische Kommission schlägt aktuell einen risikobasierten Regulierungsansatz für die Regulierung von KI vor (European Commission, 2021): Besonders riskante KI-Systeme sollen verboten oder als „high-risk“ streng reguliert werden. Neben Dokumentations- und Transparenzpflichten sind Risikomanagementsysteme einzusetzen. KI muss zudem robust, sicher und genau sein. Damit sind Voraussetzungen geschaffen, die auch den Nachweis der Kausalität erleichtern. Wir stimmen dem risikobasierten Regulierungsansatz zu. Es bedarf aber einer interdisziplinären Zusammenarbeit, wie sie von „RoboTrust“ angestrebt wird, um das spezifische Risiko beim Einsatz von KI realistisch einschätzen zu können.

### 2.3 Privatsphäre und Datenschutz in der MRI

Die weitreichende Verbreitung von Robotern in den verschiedensten sozialen Kontexten wirft die Frage auf, ob und wie diese durch die Verwendung personenbezogener Informationen womöglich die Entfaltung der Persönlichkeit und soziale Dynamiken (Rueben et al., 2018) gefährden. Dies betrifft beispielsweise Daten in Form von Kamerabildern und anderen Sensordaten, die für die korrekte Funktion des Roboters erforderlich sein können, aber gleichzeitig unerwünscht vielfältige Erkenntnisse über die umgebenden und nutzenden Personen ermöglichen. Für die verantwortungsbewusste Gestaltung von MRI ist folglich die Einführung angemessener Maßnahmen für den Schutz personenbezogener Daten und der Privatsphäre unerlässlich.

In der Rechtswissenschaft ist der Begriff *Privacy* (im Deutschen und in Europa nur begrenzt erfasst durch den Begriff der „Privatsphäre“, dort eher informationelle Selbstbestimmung) seit jeher Gegenstand wissenschaftlicher Überlegungen. Trotz scheinbar identischer Terminologie wird das Konzept in der EU und den USA jedoch sehr unterschiedlich betrachtet. In den USA ist der Begriff *Privacy* eng mit dem physischen Eindringen in private Bereiche und dem Umstand verbunden, dass es kein in der Verfassung verankertes, unmittelbares Recht auf eine solche Privatsphäre gibt. *Privacy* wird hauptsächlich als „the claim of individuals [...] when, how, and to what extent information about them is communicated to others“ (Westin, 1968, S. 7) angesehen.

In der EU entspricht die Regulierung von *Privacy* der Regulierung von Datenschutz (Hustinx, 2014). Die Verarbeitung personenbezogener Daten unterliegt einer strengen Kontrolle, weil sie Auswirkungen für die Persönlichkeitsrechte des Einzelnen, aber auch für Rechtspositionen der Gesellschaft haben kann (Lynskey, 2014). Die DS-GVO der EU bietet sich seit ihrem Inkrafttreten im Jahr 2018 insofern als offensichtlicher Maßstab für die Bewertung eines die Privatsphäre schützenden Roboters an. Dies gilt erst recht, weil sich die verschiedenen Ansätze aus den USA und der EU in Zukunft annähern könnten, da die DS-GVO bereits als Vorbild für Regulierung in mehreren anderen Ländern wie Brasilien und Japan diente, vor allem aber auch für den kalifornischen *Consumer Privacy*

Act (Schwartz, 2019). Ein Roboter, der die Bestimmungen der DS-GVO einhält, wahrt damit für die Zwecke dieses Beitrags auch in aller Regel die Anforderungen des US-amerikanischen Privatsphärenschutzes, da die Bestimmungen der Verordnung typischerweise anspruchsvoller sind als jene, die im US-amerikanischen Konzept der Privacy formuliert werden (Kuner et al., 2020).

Trotz der Bedeutung findet sich bislang nur wenig Literatur zu den Anforderungen DS-GVO-konformer Robotik. Die vorhandenen Arbeiten spiegeln häufig den neuen Rechtsrahmen nicht wider, der durch die DS-GVO eingeführt wurde (Dreier & Spiecker gen. Döhmman, 2012) oder konzentrieren sich auf bestimmte Bereiche, wie z. B. auf den Einsatz von Robotern im Gesundheitswesen (Fosch-Villaronga et al., 2018) oder im Rahmen von Lieferdiensten (Hoffmann & Prause, 2018). Andere befassen sich mit den Besonderheiten bestimmter Plattformen, z. B. mit Robotern, die Cloud Computing nutzen (Fosch-Villaronga et al., 2018; Fosch-Villaronga und Millard, 2018). Wiederum andere Beiträge analysieren bestimmte Grundsätze und Instrumente der DS-GVO in spezifischen Kontexten, z. B. Transparenz (Felzmann et al., 2019).

Dagegen liegt Forschung zur datenschutzkonformen Verwendung von Sensordaten vor, die bereits als solche Bedenken hervorrufen (Lutz & Tamò, 2015). Diese Bedenken beziehen sich allerdings auf anthropomorphe Dienstleistungsroboter, die mit Menschen interagieren, und weniger auf mechanische, nicht-soziale Roboter (Lutz & Tamò, 2015). In einer Studie sorgten sich Befragte am meisten um ihre informationelle Selbstbestimmung, wenn HerstellerInnen oder BetreiberInnen von sozialen Robotern ihre Daten verwendeten (Lutz & Tamò-Larrieux, 2020). Auch Hacking und Stalking rufen erhebliche Beunruhigung hervor, wohingegen das physische Eindringen in die Privatsphäre auf weniger Bedenken trifft (Lutz & Tamò-Larrieux, 2020). In diesem Zusammenhang finden sich verschiedene Vorarbeiten für eine verantwortungsbewusste, DS-GVO-konforme Verarbeitung von Videodaten, die auch Analysen möglicher technischer Lösungen umfassen (Barnovicu et al., 2019; Šidlauskas, 2019).

Obwohl verschiedene Interessengruppen sich für einen stärkeren Datenschutz einsetzen, deuten mehrere Studien darauf hin, dass Individuen sich oft kaum mit datenschutzbezogenen Fragen auseinandersetzen bzw. nicht entsprechend ihrer Präferenzen agieren. Obwohl viele NutzerInnen von Technik Bedenken hinsichtlich eines hinreichenden Schutzes ihrer Privatsphäre haben, geben sie häufig dennoch persönliche Daten preis; ein Umstand, der auch als Privacy Paradox bezeichnet wird (Barnes, 2006; Norberg et al., 2007). Erste Studien deuten darauf hin, dass dieses Privacy Paradox auch in der realen MRI auftreten kann und bei androiden Robotern sogar noch stärker ausgeprägt ist als bei humanoiden Robotern (Stock-Homburg & Hannig, 2020). Humanoide Roboter haben in der Regel menschliche Extremitäten, jedoch noch ein ganzheitliches mechanisches Aussehen, wohingegen androide Roboter den Menschen so realistisch wie möglich abbilden (Mara & Appel, 2015a). Abb. 2 zeigt den humanoiden Roboter TIAGo++ sowie die androide Roboterfrau Elenoide in einer Einzelhandelsumgebung.



**Abb. 2** Humanoider Roboter TIAGo++ (links) und androide Roboterfrau Elenoide (rechts) in einer Einzelhandelsumgebung

## 2.4 Transparenz in der MRI

Mit dem anhaltenden technologischen Fortschritt von Robotern, insbesondere von sozialen Robotern und Dienstleistungsrobotern, nimmt auch die Anzahl der möglichen Anwendungen für solche intelligenten und autonomen Systeme immer weiter zu. Die Entwicklung dieser Roboter wird mit besonderen Herausforderungen konfrontiert, wenn für die erfolgreiche und effektive gemeinsame Aufgabenerfüllung die Interaktion mit dem Menschen essenziell ist. Das Konzept der Transparenz soll dazu beitragen, diese Herausforderungen zu überwinden. In der Literatur wird Transparenz unter anderem als das Bereitstellen von Erläuterungen über durchgeführte Aktionen (Kim & Hinds, 2006), als die Menge an Informationen, die das System den BenutzerInnen über interne Vorgänge vermittelt (Sanders et al., 2014), oder als eine Designanforderung beschrieben, um BenutzerInnen nicht auf trügerische Weise auszunutzen (Wortham et al., 2016). Darüber hinaus kann Transparenz ein Ansatz sein, um Vertrauen in automatisierte Systeme oder Roboter zu schaffen (Lyons et al., 2017; Nettet et al., 2021), wobei ihr positiver Einfluss auf die menschliche Leistungsfähigkeit von der Zuverlässigkeit des spezifischen Roboters abhängen kann (Wright et al., 2020).

Unter Betrachtung von 1) autonomen Agenten (z. B. Roboter) und 2) MRI (Chen et al., 2014; Lyons, 2013) wurden zwei mittlerweile häufig verwendete Transparenz-

modelle entwickelt. Das auf Situationsbewusstsein basierende Agententransparenzmodell (SAT) (Chen et al., 2014) beruht auf den Arbeiten von Endsley (1995), Lee & See (2004) und Rao und Georgeff (1995) und besteht aus drei Ebenen, die verschiedene Transparenzaspekte (Aufgabenparameter, Ausführungslogik und erwartete Ergebnisse) repräsentieren. Es zielt darauf ab, eine Grundlage für transparente Benutzerschnittstellen zu schaffen, die es den BedienerInnen ermöglichen, fundierte Entscheidungen über ihr Eingreifen zu treffen. Das zweite Modell (Lyons, 2013) unterscheidet zwischen der „Roboter-zu-Mensch“- und „Roboter-von-Mensch“-Transparenz. Diese Unterscheidung gruppiert die Transparenzfaktoren in Informationen über den Roboter (Absicht, Aufgabe, Analytik, Umgebung) und über die Wahrnehmung des Menschen durch den Roboter (Teamarbeit, Zustand des Menschen). Das SAT-Modell und die „Roboter-zu-Mensch“-Transparenz weisen einige Gemeinsamkeiten auf.

In den Arbeiten von Bhaskara et al. (2020) und Rajabiyazdi und Jamieson (2020) wurden die Auswirkungen von Transparenz, basierend auf Veröffentlichungen empirischer Studien, untersucht. Diese zeigen, dass sich die Literatur bisher hauptsächlich auf das SAT-Modell fokussiert, die Ergebnisse dazu jedoch unvollständig oder inkonsistent sind, z. B. in Bezug auf das wünschenswerte Maß an Transparenz. Darüber hinaus schlagen sie vor, in zukünftigen Forschungsfragen Erkenntnisse aus der kognitiven Psychologie zu berücksichtigen sowie zu untersuchen, welche Transparenzniveaus in Bezug auf den Kontext angemessen sind und warum Transparenz die BedienerInnen unterstützt.

In der bisher referenzierten Literatur lag der Forschungsfokus in erster Linie auf den Auswirkungen von Transparenz in Bezug auf eine/n BedienerIn oder eine/n TeamkollegIn. Zwar sind in diesen Bereichen noch viele Fragestellungen ungeklärt, jedoch sind insbesondere die ethischen und gesellschaftlichen Aspekte meist nicht abgedeckt. Im Gegensatz dazu zeigen Felzmann et al. (2019) die Transparenzerwartungen verschiedener Stakeholder in Anlehnung an die Arbeit von Weller (2017) auf und betonen, dass Transparenz eine interdisziplinäre Zusammenarbeit von Recht, Sozialwissenschaft und Technik erfordert. Darüber hinaus wurde eine Checkliste erarbeitet, die RoboterentwicklerInnen bei der Umsetzung der Transparenzanforderungen der DS-GVO anleiten soll und die insbesondere für Serviceroboter, die personenbezogene Daten verarbeiten, relevant ist (Abschn. 2.3).

Um ein Maß für Transparenz zu schaffen, ist die Norm „P7001 – Transparency of Autonomous Systems“ (IEEE Standards Association, 2020) der „IEEE Standards Association“ erarbeitet worden und Anfang 2022 erschienen. In Übereinstimmung mit Felzmann et al. (2019) werden auch hier verschiedene Interaktions-Stakeholder berücksichtigt, um den Grad der Normerfüllung eines autonomen Systems objektiv zu bestimmen. Zudem werden technische Lösungen gefordert, um die Ursachen für Fehlfunktionen oder Zwischenfälle zu ermitteln und die für die Post-Mortem-Analyse erforderlichen Daten effizient und mit hoher Integrität zu speichern (Thomas, 2010). Die Entwicklung eines Blackbox-Recorders für autonomes und KI-gesteuertes Roboterverhalten, welcher diese und darüber hinaus rechtliche Anforderungen integriert,

würde helfen, die Ursache einer Fehlfunktion auf die Hardware, die Software oder die BenutzerInnen zurückzuführen (Abschn. 2.2).

Transparenz spielt in der MRI eine wichtige Rolle und die Forschungsbemühungen dazu stellen eine solide Basis dar, weisen teilweise jedoch noch inkonsistente Ergebnisse und offene Fragenstellungen auf. Weiterhin fehlt ein evidenzbasierter interdisziplinärer Leitfaden, wie Transparenz in spezifischen Kontexten realisiert werden kann. Dies gilt auch für den Fall, dass Transparenz potenziell nicht erwünscht ist, z. B. bei sozialen Robotern (Fischer, 2018).

## 2.5 Bewertung der verantwortungsbewussten MRI

Evaluierung ist „[t]he process of determining the merit or worth or value of something“ (Scriven, 1981, S. 53). Für die MRI bietet sie einen systematischen Ansatz, um die Zielerreichung einer verantwortungsbewussten MRI zu bewerten. Die bisherige Forschung zur Messung von MRI lässt sich in drei Kategorien einteilen: 1) (subjektive) Wahrnehmungsmessungen (d. h. Selbst- und Fremdwahrnehmungen), 2) (objektive) Leistungsmessungen und 3) sensorbasierte Messungen (z. B. psychophysiologische Messungen, Gesichtserkennungsmessungen) (siehe Bethel et al., 2007; Bethel & Murphy, 2010; Homburg, 2018; Tiberio et al., 2013 für einen Überblick).

Subjektive Wahrnehmungsmessungen können aus der Innen- oder Außenperspektive und anhand von verhaltensbezogenen Indikatoren oder Selbsteinschätzungen durchgeführt werden. Selbstwahrnehmungen beruhen auf subjektiven Beobachtungen oder Emotionen einer Person. Oft sind sie nicht direkt beobachtbar, stattdessen durch Indikatoren messbar oder anhand von Selbsteinschätzungen der menschlichen Teilnehmenden innerhalb der MRI ableitbar (de Jong et al., 2020; Tiberio et al., 2013). Psychometrische Skalen, Umfragen und Fragebögen (Bethel et al., 2007; Steinfeld et al., 2006) sind geeignete Instrumente, um Vertrauen (Li et al., 2015; Schaefer, 2016), Akzeptanz (de Jong et al., 2020; Schaefer, 2016) und verschiedene Charakteristika der Teilnehmenden (Homburg, 2018) zu bewerten. Im Gegensatz dazu untersucht die Außenperspektive, wie BeobachterInnen Verhaltensaspekte der direkt an der MRI beteiligten Parteien bewerten (Coyne et al., 2020; Li et al., 2015). Um emotionale und/oder verhaltensbezogene Reaktionen von Teilnehmenden zu beurteilen, werden Beobachtungen (durch Audio- und Videoaufnahmen) aufgezeichnet und anschließend ausgewertet (Bethel & Murphy, 2010).

Objektive Messungen der Leistung quantifizieren das Ergebnis der MRI (Admoni et al., 2016; Murphy & Schreckenghost, 2013) und geben beispielsweise an, wie gut ein anthropomorpher Roboter einen Menschen bei der Erfüllung einer Aufgabe unterstützt (Bethel & Murphy, 2010). Beispielhafte Indikatoren sind die Effizienz und die Geschwindigkeit einer ausgeführten Aufgabe oder die Anzahl der dabei auftretenden Fehler (Admoni et al., 2016; Murphy & Schreckenghost, 2013).

Sensordaten können empirischer oder analytischer Natur sein. Beispiele für empirische Sensordaten sind Mikrofone und Kameras; diese erlauben es, das Verhalten

menschlicher InteraktionspartnerInnen während der MRI aufzuzeichnen (Stiefelhagen et al., 2004). In der MRI ist zudem auch die Wahrnehmung der Bewegungen der menschlichen InteraktionspartnerInnen ein wesentlicher Aspekt, da daraus angemessene Reaktionen abgeleitet werden können (Bütepage et al., 2018; Jarrassé et al., 2008). Durch empirische Sensordaten kann der emotionale Zustand der menschlichen InteraktionspartnerInnen über Audio (Rázuri et al., 2015), Video (Filintisis et al., 2020; Rawal & Stock-Homburg, 2021) oder beide Formate (Avots et al., 2019; Kansizoglou et al., 2019) erfasst werden.

Emotionen sind „an episode of interrelated synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism“ (Scherer, 2005, S. 697). Sie sind ein wichtiger Indikator für das Wohlbefinden von Menschen während der MRI (siehe Stock-Homburg, 2021 für einen Überblick). Zudem hat die Roboterpsychologie gezeigt, dass Menschen künstliche Emotionen von anthropomorphen Robotern (Stock & Merkle, 2017) im Rahmen eines emotionalen Ansteckungsprozesses wahrnehmen können (siehe Stock & Nguyen, 2019b für einen Überblick). Ungeachtet der emotionalen Expression von Robotern können menschliche Emotionen durch Sensoren erkannt und anhand trainierter Modelle kann auf die Emotionen der Person zurückgeschlossen werden.

Diese Fähigkeit, „to encode an ensemble of sensory stimuli providing information about the emotional state of another individual“ (Ferretti & Papaleo, 2019, S. 1), wird als Emotionserkennung bezeichnet. Die menschliche Mimik ist eine essenzielle Quelle für die Erkennung von Emotionen und zieht seit einigen Jahren großes Forschungsinteresse auf sich (Canedo & Neves, 2019; Li & Deng, 2020). In letzter Zeit sind Deep-Learning-Algorithmen populär geworden. Das vorverarbeitete Bild wird direkt in tiefe neuronale Netze eingespeist, um ein Ergebnis vorherzusagen (siehe Rawal & Stock-Homburg, 2021 für einen Überblick). Dieses bezieht sich hauptsächlich auf die sechs allgemein bekannten Emotionen Glück, Traurigkeit, Ekel, Wut, Angst und Überraschung und wird manchmal durch einen neutralen Ausdruck erweitert (Li & Deng, 2020).

Ein weiteres wichtiges Mittel zur Erfassung von Informationen über menschliche Emotionen während der MRI sind psychophysiologische Daten. Zu diesen zählen elektrodermale Aktivitäten (EDA), Elektroenzephalographie (EEG), Elektrokardiogramm (EKG), Elektromyographie (EMG) und Respiration (RESP) (Bethel et al., 2007). EDA, EEG und EKG sind hierbei die am häufigsten verwendeten Indikatoren (Greene et al., 2016). In einer Stresssituation nimmt beispielsweise die Transpiration zu, was die Leitfähigkeit der Haut erhöht und das EDA-Signal ansteigen lässt. Emotionen stehen mit dem vegetativen Nervensystem in Zusammenhang, und werden durch dieses von Veränderungen der physiologischen Signale begleitet (Kreibig, 2010; Levenson, 1988). Folglich können physiologische Reaktionen der Teilnehmenden als Indikatoren für die Erregung von Menschen verwendet werden (Homburg, 2018). Geeignete Armbänder können beispielsweise physiologische Daten, wie die Herzfrequenz, aufzeichnen, wodurch Rückschlüsse auf das Stressniveau der Teilnehmenden möglich werden (Ollander et al., 2016). Physiologische Signale lassen sich kaum, wenn nicht sogar gar nicht kontrollieren (Gunes & Pantic, 2010). Die Verwendung analytischer Sensordaten

führt daher zu einer Objektivierung der Daten, da menschliche Teilnehmende im Allgemeinen kaum in der Lage sind, ihre physiologischen Reaktionen zu manipulieren (Bethel et al., 2007). Beobachtungen und analytische Sensoren können verwendet werden, um die Emotionen einer Person anhand von akustischen (Yoon et al., 2018), physiologischen (Shu et al., 2018) oder visuellen Signalen (Do et al., 2021; Rawal & Stock-Homburg, 2021) zu messen.

Eine Literaturübersicht über die Verwendung physiologischer Signale in der MRI zur Erkennung von Emotionen weist darauf hin, dass der Referenzwert der physiologischen Signale erfasst werden sollte, bevor der Einfluss der Emotionen erkannt wird (Liu et al., 2006; Val-Calvo et al., 2020). Kidd und Breazeal (2005) weisen darauf hin, dass die erste Begegnung mit einem Roboter vor Beginn des Experiments stattfinden sollte. Dies ist sicherlich richtig, wenn die Auswirkungen verschiedener Verhaltensweisen des Roboters als Emotionsauslöser betrachtet werden sollen und nicht die Reaktionen auf die erste Begegnung des Menschen mit dem Roboter. In den betrachteten Studien (Kagawa et al., 2018; Kulic & Croft, 2005; Mohammad & Nishida, 2010; Rani & Sarkar, 2005a, 2005b; Suzuki et al., 2020; Val-Calvo et al., 2020; Zanchettin et al., 2013) wird der Versuchsaufbau während der MRI in kleine Teilabschnitte unterteilt. Diese stellen einzelne Bewegungen des Roboters oder Aufgaben der menschlichen Teilnehmenden in der Interaktion mit dem Roboter dar. Innerhalb dieser Abschnitte wird eine bestimmte emotionale Reaktion der menschlichen Teilnehmenden erwartet und zu einem späteren Zeitpunkt ausgewertet.

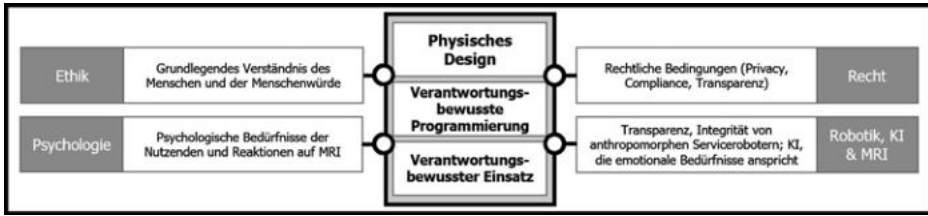
Aus konzeptioneller Sicht verwendet die Forschung zu Emotionen in der MRI Emotionsmodelle, die Emotionen der Versuchsperson in unterschiedlichem Maße abstrahieren (siehe Stock-Homburg, 2021 für einen Überblick). Das populärste Modell ist das von James Russel entwickelte *Circumplex Model of Emotion* (Russell, 1980). Die „Ground Truth“ beschreibt die tatsächlich empfundenen Emotionen der Versuchsperson in den einzelnen Abschnitten und wird durch die Klassifizierung der physiologischen Signale, durch Fragebögen oder durch die zusätzliche Nutzung weiterer Emotionserkennungsmethoden (z. B. der Gesichtserkennung) sichergestellt.

Die bisherige Forschung bietet wertvolle Bewertungskriterien für die menschliche Reaktion auf die Gestaltung von Robotern. Diese sind jedoch überwiegend allgemeiner Natur, sodass es kaum spezifische Kriterien für ein verantwortungsbewusstes Design der MRI gibt. Hierbei sind insbesondere Kriterien betroffen, die ethische (Abschn. 2.1) und rechtliche Aspekte, wie die Privatsphäre des Menschen (Abschn. 2.3), Sicherheit (Abschn. 2.2) und Transparenz (Abschn. 2.4) der Gestaltung von Robotern (Kap. 3) berücksichtigen.

### **3 Interdisziplinärer Bezugsrahmen**

Basierend auf dem Literaturüberblick und den dort beschriebenen Anforderungen und interdisziplinären Wechselwirkungen (Kap. 2) haben wir den Bedarf eines interdisziplinären Bezugsrahmens identifiziert, der eine Leitlinie für die verantwortungsbewusste Gestaltung der MRI mit anthropomorphen Dienstleistungsrobotern bietet. In





**Abb. 3** Interdisziplinärer Bezugsrahmen für die verantwortungsbewusste MRI mit anthropomorphen Dienstleistungsrobotern

diesem Kapitel präsentieren wir einen solchen interdisziplinären Bezugsrahmen. Die Grundlage des interdisziplinären Bezugsrahmens sind die ethischen, psychologischen, rechtlichen und technischen Anforderungen der NutzerInnen.

Der entwickelte interdisziplinäre Bezugsrahmen (siehe Abb. 3) besteht aus drei Kernschritten, welche Interdependenzen mit den vier Disziplinen Ethik, Recht, Psychologie und Technologie aufweisen. Bei der Gestaltung von anthropomorphen Dienstleistungsrobotern sowie bei der Durchführung von experimentellen Studien sollten diese drei Kernschritte berücksichtigt werden, um die MRI verantwortungsbewusst zu gestalten. Dadurch wird eine MRI ermöglicht, die authentisch und akzeptabel für menschliche NutzerInnen ist. In den Abschn. 3.1 bis Abschn. 3.3 werden die einzelnen Kernschritte erläutert.

### 3.1 Verantwortungsbewusstes physisches Design anthropomorpher Dienstleistungsroboter

Um anthropomorphe Roboter zu erschaffen, wurden in der Vergangenheit zahlreiche Anstrengungen unternommen (Mori et al., 2012). Die Überlegungen stützten sich auf verschiedene theoretische Paradigmen.

Das *Computers-as-Social-Actors-Paradigma* (CASA) zeigt, „that individuals mindlessly apply social rules and expectations to computers“ (Nass & Moon, 2000, S. 81). Abhängig von verschiedenen Faktoren, einschließlich des Grades der sozialen Ausprägungen von Technologien (z. B. Roboter), neigen Menschen dazu, diese Technologien wie Menschen zu behandeln, indem sie soziale Heuristiken von Mensch-Mensch-Interaktionen auf Mensch-Computer-Interaktionen übertragen (Gambino et al., 2020; Kim & Sundar, 2012).

Die Tendenz von Menschen, nicht-menschlichen Wesen menschliche Eigenschaften und Merkmale zuzuschreiben, wird allgemein als Anthropomorphismus bezeichnet (Damiano & Dumouchel, 2018; Fink, 2012). In der Robotik wird ein anthropomorphes Design verwendet, um MRI zu begünstigen und so die Akzeptanz der Roboter durch die NutzerInnen sowie die Effektivität der MRI zu steigern (z. B. durch die Verwendung

von Gesichtsausdrücken oder ein menschenähnliches physisches Design des Roboters) (Damiano & Dumouchel, 2018; Fink, 2012). Trotz dieses Trends in der Roboterforschung, Roboter (z. B. Dienstleistungsroboter) zu entwerfen und zu entwickeln, die menschenähnlich sind, gehen Studien davon aus, dass die Zufriedenheit der AnwenderInnen bei einer reinen Mensch-Mensch-Interaktion höher als bei einer MRI ist (Merkle, 2019; Mori et al., 2012).

Das *Uncanny-Valley-Paradigma* sagt einen positiven Zusammenhang zwischen der Menschenähnlichkeit eines Roboters und der Sympathie bzw. Akzeptanz des Menschen gegenüber dem Roboter voraus (Ho & MacDorman, 2017; Mara & Appel, 2015a; Mori et al., 2012). Dieser Zusammenhang erfolgt bei steigender Menschenähnlichkeit jedoch nicht streng monoton steigend, denn im Falle eines fast, aber nicht vollkommen realistischen menschlichen Aussehens eines Roboters berichten menschliche AnwenderInnen von negativen Gefühlen gegenüber solchen Robotern (Ho & MacDorman, 2017; Mara & Appel, 2015a; Mori et al., 2012; Seyama & Nagayama, 2007).

Gieselmann und Gremmer (2018) beschreiben zudem, dass die Erwartungshaltung der KundInnen hinsichtlich der Fähigkeiten und Funktionen anthropomorpher Dienstleistungsroboter durch ihre Menschenähnlichkeit bedingt ist. HändlerInnen im Einzelhandel sollten daher berücksichtigen, dass KundInnen „von einem sehr menschenähnlichen Roboter eine höhere Intelligenz, Komfort und Leistung“ (Gieselmann & Gremmer, 2018, S. 439) erwarten.

In einer Online-Studie haben wir mittels Amazon Mechanical Turk (MTurk) Daten von OnlinearbeiterInnen erhoben. Die Teilnehmenden haben verschiedene Service-Szenarien hinsichtlich ihrer Wichtigkeit und der Wahrscheinlichkeit, dass diese in naher Zukunft eintreten, bewertet.<sup>3</sup> Bei der Untersuchung des Uncanny-Valley-Paradigmas haben wir analysiert, welche/r DienstleistungsmitarbeiterIn, d. h. ein Mensch, ein humanoider Dienstleistungsroboter oder ein androider Dienstleistungsroboter, von den Teilnehmenden bevorzugt wurde. Die Teilnehmenden wurden gebeten, die drei möglichen DienstleistungsmitarbeiterInnen nach ihrer Präferenz einzustufen, wenn sie mit diesen in einem Einzelhandelsszenario interagieren würden. Die Studie bestätigte die Vorhersagen des Uncanny-Valley-Paradigmas, wonach humanoide und androide Roboter im Vergleich zu Menschen eine geringere Akzeptanz erfahren. Es überrascht nicht, dass die Teilnehmenden an erster Stelle den Menschen als bevorzugten Interaktionspartner

<sup>3</sup>Die Umfrage mit Amazon Mechanical Turk (MTurk) war für OnlinearbeiterInnen mit einer historischen Akzeptanzrate von über 95 % und mit einem Wohnsitz in den USA sichtbar. Dadurch konnte sichergestellt werden, dass nur solche OnlinearbeiterInnen an der Studie teilnahmen, die in der Vergangenheit ihre Aufgaben zu mindestens 95 % sorgfältig durchgeführt hatten. In einem Between-Subject-Design haben wir vier verschiedene Szenarien miteinander verglichen: 1) Handel, 2) Gesundheitswesen, 3) Reisebüro und 4) Restaurant. Nach Bereinigung der Daten konnten wir Daten von 392 US-BürgerInnen auswerten (60 % Männer, 39 % Frauen, 1 % divers, Durchschnittsalter=40 Jahre, Standardabweichung=13). Der primäre berufliche Status war angestellt (67 %), gefolgt von selbstständig (10 %), nichterwerbstätig (8 %) und sonstige (14 %).

nannten (73 %), gefolgt von humanoiden Robotern (23 %) und androiden Robotern (3 %). Andere Studien zeigen jedoch ein differenzierteres Bild im Verhalten gegenüber androiden und humanoiden Robotern. So gaben MitarbeiterInnen während eines Experimentes in einem Unternehmen bei komplexen Aufgabenstellungen mehr persönliche Informationen gegenüber androiden Robotern als gegenüber humanoiden Robotern preis (Stock-Homburg & Hannig, 2020).

Folglich hat das physische Design anthropomorpher Dienstleistungsroboter einen wichtigen Einfluss auf die Wahrnehmung von AnwenderInnen. Bei der Auswahl des physischen Designs ist dies neben dem Einsatzzweck und Einsatzort der anthropomorphen Dienstleistungsroboter zu berücksichtigen (siehe Abb. 3, physisches Design).

## **3.2 Verantwortungsbewusste Programmierung**

### **Transparenz**

Mit der zunehmenden Komplexität anthropomorpher Dienstleistungsroboter nehmen auch deren künstliche Denkfähigkeit und Anwendungsbereiche zu. Dies ermöglicht die Integration von Systemen in reale, dynamische Umgebungen mit einer Vielzahl von Interaktionskontexten und -partnern. Hierbei kann Transparenz ein wichtiger Ansatz sein, um ein verantwortungsbewusstes Roboterdesign zu erreichen und damit Vertrauen und soziale Akzeptanz von anthropomorphen Dienstleistungsrobotern zu etablieren (Abschn. 2.4).

Transparenz wurde häufig im Zusammenhang mit der Interaktion von Robotern und NutzerInnen bzw. Teammitgliedern betrachtet (Bhaskara et al., 2020; Rajabiyazdi & Jamieson, 2020), um die Entscheidungen des Roboters zu verstehen und nachzuvollziehen. Im Fall von anthropomorphen Dienstleistungsrobotern müssen jedoch auch rechtliche und gesellschaftliche Anforderungen berücksichtigt werden. Dazu gehört mitunter die Untersuchung von Zwischenfällen, die durch unerwünschtes Roboterverhalten in dynamischen Interaktionsszenarien verursacht werden. Transparenz hat je nach MRI-Akteuren eine unterschiedliche Bedeutung und Bedarf daher gesonderter Anforderungsanalysen und -bewertungen (Felzmann et al., 2019; IEEE Standards Association, 2020). So werden sich voraussichtlich die Informationen voneinander unterscheiden, die EndnutzerInnen und WartungstechnikerInnen zur Verfügung gestellt werden sollten, um das Verhalten des Dienstleistungsroboters nachvollziehen zu können.

Ein Blackbox-Recorder (IEEE Standards Association, 2020; Winfield & Jirotko, 2017) könnte die nachträgliche Untersuchung von Zwischenfällen und getroffenen Roboterentscheidungen für verschiedene Akteure im Nachhinein erleichtern. Hierfür sind eine effiziente Speicherung und ein semantischer Zugriff (Niemueller et al., 2012) auf die aus verschiedenen Quellen stammenden Daten erforderlich, ebenso sind die Datensicherheit und die Einhaltung der rechtlichen Vorgaben für personenbezogene Daten (Vasylkovskyi et al., 2020) zu gewährleisten. Darüber hinaus müssen die Authentizität und Integrität der Aufnahmen sichergestellt sein, um die juristische Haftung bei Unfällen durch Anwälte oder Sachverständige zu klären. Eine technische Lösung für

den benötigten Blackbox-Recorder mit geeigneten Tools zur Untersuchung von Fehlfunktionen oder Vorfallsursachen ist für eine verantwortungsbewusste MRI Voraussetzung. Damit soll die technische Lücke geschlossen werden, um die Realisierung akteurabhängiger Transparenzanforderungen zu erleichtern und gleichzeitig einen verantwortungsbewussten Umgang mit den gesammelten Daten zu gewährleisten.

Anthropomorphe Dienstleistungsroboter haben hinsichtlich der transparenten Gestaltung von Roboterentscheidungen und anderen Interaktionsinformationen vorteilhafte MRI-Fähigkeiten, die es für die Interaktion nicht erfordern, dass EndnutzerInnen ein tiefes Verständnis über die internen Mechanismen des Roboters besitzen (Park et al., 2011; Salem et al., 2013). Da diese Roboter in der Lage sind, Informationen durch eine Kombination aus Gestik, Mimik und Sprache auszudrücken, kann die Intuition ausgeführter und zukünftiger Aktionen in einer für NutzerInnen gewohnten Weise ausgedrückt werden. Neben der Übertragung von Gesten zwischen unterschiedlichen anthropomorphen oder humanoiden Robotern (Van de Perre et al., 2015) besteht eine besondere Herausforderung darin, diese Körpersprache in Abhängigkeit von der aktuellen Situation in einer generischen Weise zu modellieren und zu programmieren, welche auch eine Online-Anpassung der Gesten erlaubt. Dies würde es ermöglichen, Bewegungen von einem spezifischen Roboter zu abstrahieren und allgemeine Interaktionsstrategien für eine transparente und kontextsensitive Interaktion zu entwickeln.

### **Emotional Responsiveness**

Emotionen spielen grundsätzlich bei Interaktionen eine wesentliche Rolle, sowohl in der Interaktion zwischen Menschen als auch in der Interaktion zwischen Menschen und anthropomorphen Robotern (Rawal & Stock-Homburg, 2021; Stock-Homburg, 2021). Insbesondere bei der Interaktion mit anthropomorphen Robotern erwarten viele Menschen, dass der Roboter Emotionen ausdrückt und sich empathisch verhält (Fussell et al., 2008). Nach der in diesem Beitrag verwendeten Definition für Emotionen (Abschn. 2.5) sind Emotionen zeitlich begrenzt (etwa zwischen 0,5 und 4 s (Levenson, 1988)) und werden durch physische oder psychische Reize (z. B. einen sich nähernden Roboter) hervorgerufen. Das vegetative Nervensystem ist ein Beispiel eines organismischen Teilsystems. Es ist für die Systemregulierung verantwortlich und beeinflusst daher die Physiologie einer Person. Durch Sensoren können Veränderungen in der Physiologie erkannt und mithilfe gelernter Modelle kann auf die Emotionen der Person geschlossen werden (Abschn. 2.5).

Auf dem Gebiet der Emotionserkennung (Abschn. 2.5) hat sich in Bezug auf Transparenz und Ethik für Versuchsaufbauten viel verändert. Spätestens mit der Arbeit von Picard et al. (2001), in der fünf Faktoren genannt werden, welche die Datenerfassung in der Emotionserkennung beeinflussen, muss sich die Scientific Community mit diesem Themenbereich auseinandersetzen. Einer der von den Picard et al. (2001) identifizierten Faktoren wirft die Frage auf, ob sich eine Versuchsperson über den Zweck des Experiments bewusst ist oder nicht. Picard et al. (2001) weisen darauf hin, dass es aus Gründen des Datenschutzes und der Ethik praktisch unmöglich ist, eine Datenerfassung umzusetzen, bei der sich Versuchspersonen weder über die Datenerfassung selbst noch über den

Zweck des Experiments bewusst sind. Larradet et al. (2020) zeigen allerdings, dass sich ethische Bedenken durch die Durchführung von Feldstudien ändern können. Zum Beispiel ist das Erfassen negativer Emotionen (wie z. B. Wut, Traurigkeit) mit hoher Intensität aus ethischer Sicht weniger problematisch als in einer Laborstudie. Dies begründet sich daraus, dass diese Emotionen in einer Feldstudie nicht vorsätzlich hervorgerufen werden.

Im Einzelhandel ist es essenziell, dass Dienstleistungsroboter optimal programmiert werden, damit sie ihre unterstützende Funktion bestmöglich ausführen können (Gieselmann & Gremmer, 2018). Damit wird die verantwortungsbewusste Programmierung ein wesentlicher Bestandteil bei der Gestaltung zur Erzielung einer für den Menschen authentischen und akzeptablen MRI (siehe Abb. 3).

### **3.3 Verantwortungsbewusster Einsatz anthropomorpher Dienstleistungsroboter**

Die Verbreitung von Dienstleistungsrobotern nimmt stetig zu, sodass auch menschliche Reaktionen auf diese Roboter zu einer zunehmend relevanten Thematik geworden sind (Mara & Appel, 2015a). In diesem Beitrag wird betrachtet, wie der verantwortungsbewusste Einsatz anthropomorpher Dienstleistungsroboter die Akzeptanz dieser Roboter in der MRI steigern kann. Anthropomorphe Dienstleistungsroboter und Menschen koexistieren, wobei anthropomorphe Dienstleistungsroboter sowohl eine Vielzahl von Aufgaben in verschiedenen Bereichen ausführen als auch mit dem Menschen interagieren und kommunizieren. Bei der Interaktion mit Menschen zeigen diese Roboter einige grundlegende intelligente Verhaltensweisen, um zugewiesene Aufgaben autonom auszuführen (Forlizzi & DiSalvo, 2006; Haidegger et al., 2013; Holland et al., 2021; Mara & Appel, 2015a).

Das anthropomorphe Design und die immer fortschrittlichere Verkörperung von Robotern haben einen entscheidenden Einfluss auf das Wohlbefinden und die Akzeptanz von anthropomorphen Dienstleistungsrobotern durch die NutzerInnen (Fusté-Forné & Jamal, 2021; Mori et al., 2012) (Abschn. 3.1). Für eine verantwortungsbewusste Gestaltung der MRI sind die Akzeptanz von Robotern durch die NutzerInnen und die Verringerung des menschlichen Unbehagens von grundlegender Bedeutung. Das Ausmaß, in dem Menschen bereit sind, mit Robotern zu interagieren, hängt jedoch nicht nur von den Eigenschaften des Roboters (Mara & Appel, 2015a; Mori et al., 2012) (Abschn. 3.1 und 3.2) oder von technischen, sicherheitsrelevanten und rechtlichen Aspekten (Savela et al., 2018) (Abschn. 3.2) ab. Vielmehr hat auch das ausgewählte Szenario der MRI (d. h. der Ort, an dem die MRI stattfindet, sowie die dem Roboter zugewiesenen Aufgaben) eine erhebliche Auswirkung (Webster & Ivanov, 2020).

Aus der Perspektive der NutzerInnen bestehen Präferenzen hinsichtlich der Aktivitäten, Dienstleistungen und Aufgaben, die von anthropomorphen Dienstleistungsrobotern (nicht) durchgeführt werden sollten. Hierzu wurden bereits mehrere Studien in den Bereichen Reisen, Tourismus, Gastgewerbe und Gesundheitswesen durchgeführt, die die Einstellung der NutzerInnen zu Dienstleistungsrobotern untersuchen (Fusté-Forné & Jamal, 2021;

Ivanov et al., 2018; Smarr et al., 2012; Webster & Ivanov, 2020). Diese Studien weisen darauf hin, dass sich die AnwenderInnen wohl fühlen, wenn Roboter repetitive Aufgaben übernehmen, wie beispielsweise die Bereitstellung von Informationen oder die Durchführung von Haushaltstätigkeiten (z. B. Auslieferung von Gegenständen und Reinigung). Demgegenüber neigen NutzerInnen dazu, Tätigkeiten von Dienstleistungsrobotern eher nicht zu akzeptieren, bei denen der eigene Körper einem Dienstleistungsroboter untergeordnet werden muss (z. B. beim Babysitting, bei Massagen oder bei Friseurarbeiten) (Ivanov et al., 2018; Ivanov & Webster, 2019; Webster & Ivanov, 2020).

In der von uns durchgeführten Online-Studie (Abschn. 3.1) haben wir die Teilnehmenden vor dem Hintergrund der COVID-19-Pandemie anhand einer 7-stufigen Likert-Skala gebeten einzuschätzen, als wie realistisch sie den Einsatz von Dienstleistungsrobotern in den einzelnen Bereichen (d. h. Handel, Gesundheitswesen, Reisebüro oder Restaurant) aus Sicht des Unternehmens einschätzen. Die Teilnehmenden bewerteten den Einsatz von Dienstleistungsrobotern in jedem Bereich überwiegend als wahrscheinliches Zukunftsszenario, in einem Möbelhaus zu 74 %, einer Apotheke zu 63 %, einem Reisebüro zu 66 % und einem Restaurant zu 70 %. Die Ergebnisse zeigten, dass der Einzelhandel von den US-Teilnehmenden als der wahrscheinlichste und relevanteste Bereich angesehen wurde. Dieses Ergebnis wurde durch die offenen Bemerkungen der Teilnehmenden in dem Fragebogen unterstrichen.

Um den verantwortungsbewussten Einsatz anthropomorpher Dienstleistungsroboter zu bestimmen, sollten die psychologischen Bedürfnisse und Reaktionen der NutzerInnen der MRI angemessen berücksichtigt werden. Daher sollte das zu untersuchende Szenario vorab kritisch hinterfragt werden (siehe Abb. 3).

## **4 Implikationen und Gestaltungsempfehlungen**

Der Einsatz von anthropomorphen Dienstleistungsrobotern im stationären Einzelhandel ermöglicht es den HändlerInnen, dem zunehmenden Trend zum Online-Handel zu begegnen. So können anthropomorphe Dienstleistungsroboter ihre menschlichen KollegInnen nicht nur bei der Beratung von KundInnen unterstützen, ihr Einsatz kann vielmehr zu einem signifikant größeren Einkaufserlebnis der KundInnen führen (Gieselmann & Gremmer, 2018). Essenzielle Grundlage hierfür ist jedoch eine verantwortungsbewusste Gestaltung der MRI.

Der vorliegende Beitrag bietet eine Übersicht über den aktuellen Forschungsstand zur verantwortungsbewussten Gestaltung von MRI mit anthropomorphen Dienstleistungsrobotern in den Disziplinen Ethik, Recht, Psychologie und Technik. Des Weiteren haben wir einen interdisziplinären Bezugsrahmen für die Gestaltung einer verantwortungsbewussten MRI mit anthropomorphen Dienstleistungsrobotern entwickelt. Dieser bildet in drei Kernschritten die Gestaltung einer verantwortungsbewussten und demnach menschengerechten Gestaltung der MRI ab. Dabei fließen ethische, psychologische, rechtliche und technische Erkenntnisse ein.

Basierend auf unseren bisher durchgeführten Studien leiten wir den Einzelhandel als angemessenes Szenario ab (Abschn. 3.3), um die verantwortungsbewusste Gestaltung der MRI mit anthropomorphen Dienstleistungsrobotern weiter zu untersuchen. Die Bandbreite an Tätigkeiten, die diese Roboter im stationären Einzelhandel übernehmen können, ist groß (Kap. 1). Neben der Begrüßung von KundInnen können Informations- und Beratungsleistungen, aber auch Unterstützungsleistungen, etwa bei der Bezahlung, durch anthropomorphe Dienstleistungsroboter übernommen werden. Bei der Gestaltung einer verantwortungsbewussten MRI sind Menschenwürde, Transparenz, Privatsphäre, Datenschutz und Compliance von zentraler Bedeutung.

In der Vergangenheit wurden bereits erhebliche Fortschritte bei einer verantwortungsbewussten Gestaltung der MRI erzielt. In diesem Beitrag demonstrieren wir jedoch, dass aus Sicht der einzelnen Disziplinen einige Einschränkungen der bisherigen empirischen Forschung vorliegen. Um die zukünftige Forschung zur verantwortungsbewussten Gestaltung von MRI voranzutreiben, formulieren wir fünf wesentliche Gestaltungsempfehlungen:

- 1. Ausweitung der interdisziplinären Perspektive auf die verantwortungsbewusste MRI:** Dieser Beitrag zeigt, wie wichtig ein interdisziplinärer Ansatz ist, um eine verantwortungsbewusste MRI mit anthropomorphen Dienstleistungsrobotern zu gestalten (Kap. 2 und 3). Des Weiteren ist ein interdisziplinärer Ansatz für die Durchführung und Auswertung verantwortungsbewusster MRI-Experimente unerlässlich. Daher sind weitere Forschungsarbeiten erforderlich, die verschiedene Disziplinen integrieren, um die MRI so zu gestalten, dass Menschenwürde, Datenschutz, Privatsphäre, Compliance und Transparenz eine besondere Berücksichtigung erfahren.
- 2. Entwicklung einheitlicher Standards zur Bewertung der verantwortungsbewussten Gestaltung von MRI:** In den meisten Veröffentlichungen, die in diesem Beitrag zitiert wurden, wurde die menschliche Reaktion auf MRI auf Grundlage von Daten bewertet, die mittels Selbst- oder Fremdeinschätzungen oder mittels empirischer und analytischer Sensoren erhoben wurden (Abschn. 2.5). Zukünftige Forschungen sollten sich stärker auf Kriterien konzentrieren, die Transparenz, Privatsphäre und Sicherheit oder menschliche Reaktionen auf diese Gestaltungsparameter, wie z. B. Vertrauen, erfassen.
- 3. Anwendung des entwickelten interdisziplinären Bezugsrahmens zur Gestaltung von verantwortungsbewussten MRI-Studien:** Im Rahmen des Projekts „RoboTrust“ haben wir einen interdisziplinären Bezugsrahmen (Kap. 3) für die Gestaltung einer verantwortungsbewussten MRI mit anthropomorphen Dienstleistungsrobotern entwickelt. Hierbei haben wir die Disziplinen Ethik, Psychologie, Recht und Technik als Rahmengerüst für die drei Kernschritte des Bezugsrahmens verstanden. Wir empfehlen die Anwendung unseres interdisziplinären Bezugsrahmens für die Gestaltung von verantwortungsbewussten MRI-Studien mit anthropomorphen Dienstleistungsrobotern.

4. **Analyse von Langzeiteffekten einer verantwortungsbewussten MRI:** Anstelle des Fokus auf Querschnittstudien sollte die zukünftige Forschung auch die Langzeiteffekte einer verantwortungsbewussten Gestaltung von Robotern untersuchen. Durch die Nutzung von Längsschnittstudien können die menschlichen Reaktionen auf MRI über ihren ersten Eindruck hinaus untersucht werden, sodass auch langfristige Reaktionen der Menschen auf anthropomorphe Dienstleistungsroboter untersucht werden können.
5. **Nutzung realer Umgebungen, um die Auswirkungen einer verantwortungsbewussten Gestaltung von MRI zu testen:** Bislang wurden die meisten Studien über MRI unter Laborbedingungen durchgeführt (Abschn. 2.5). Die Durchführung von Feldstudien würde zu einem tieferen Verständnis und einer realistischen und damit „authentischen“ MRI beitragen. Daher empfehlen wir die Durchführung von Experimenten mit anthropomorphen Dienstleistungsrobotern in realen Einzelhandlungsumgebungen auf Basis unserer Leitlinien (Kap. 3), um weitere Erkenntnisse für eine verantwortungsbewusste MRI abzuleiten.

## Literatur

- Abbott, R. (2020). *The reasonable robot: artificial intelligence and the law*. Cambridge Law Review.
- Abney, K. (2012). Robotics, ethical theory, and metaethics: a guide for the perplexed. In P. Lin, K. Abney, & G. A. Bekey (Hrsg.), *IEEE Xplore digital library. Robot ethics: the ethical and social implications of robotics* (S. 35–52). The MIT Press.
- Admoni, H., Weng, T., Hayes, B., & Scassellati, B. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction*, 7–10 March 2016.
- Ahmad, M., Mubin, O., & Orlando, J. (2017). A systematic review of adaptivity in human-robot interaction. *Multimodal Technologies and Interaction*, 1(3), 14.
- Anderson, M., & Anderson, S. L. (2020). Machine ethics: creating an ethical intelligent agent. In P. Asaro & W. Wallbach (Hrsg.), *Machine ethics and robot ethics* (S. 237–248). Routledge.
- Avots, E., Sapiński, T., Bachmann, M., & Kamińska, D. (2019). Audiovisual emotion recognition in wild. *Machine Vision and Applications*, 30(5), 975–985.
- Barfield, W. (2018). Liability for autonomous and artificially intelligent Robots. *Paladyn, Journal of Behavioral Robotics*, 9(1), 193–203.
- Barnes, S. B. (2006). A privacy paradox: social networking in the United States. *First Monday*, 11(9). <https://doi.org/10.5210/fm.v11i9.1394>.
- Barnoviciu, E., Ghenescu, V., Carata, S.-V., Ghenescu, M., Mihaescu, R., & Chindea, M. (2019). GDPR Compliance in video surveillance and video processing application. In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Timisoara, Romania, 10–12 October 2019.
- Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., & Belpaeme, T. (2016). From characterising three years of HRI to methodology and reporting recommendations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction*, 7–10 March 2016.
- Belk, R. (2020). Ethical issues in service robotics and artificial intelligence. *The Service Industries Journal*, 41(13–14), 1–17.



- Bethel, C. L., & Murphy, R. R. (2010). Review of human studies methods in HRI and recommendations. *International Journal of Social Robotics*, 2(4), 347–359.
- Bethel, C. L., Salomon, K., Murphy, R. R., & Burke, J. L. (2007). Survey of psychophysiology measurements applied to human-robot interaction. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 26–29 August 2007.
- Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent transparency: a review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, 50(3), 215–224.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Hrsg.), *The cambridge handbook of artificial intelligence* (Bd. 1, S. 316–334). Cambridge University Press.
- Bringsjord, S., & Taylor, J. (2012). Introducing divine-command approach to robot ethics. In P. Lin, K. Abney, & G. A. Bekey (Hrsg.), *IEEE Xplore digital library. Robot ethics: the ethical and social implications of robotics* (S. 85–108). The MIT Press.
- Bütepage, J., Kjellström, H., & Kragic, D. (2018). Anticipating many futures: online human motion prediction and generation for human-robot interaction. In *2018 IEEE international Conference on Robotics and Automation (ICRA)*. Symposium conducted at the meeting of IEEE, 21–25 May 2018.
- Canedo, D., & Neves, A. (2019). Facial expression recognition using computer vision: a systematic review. *Applied Sciences*, 9(21), 4678. <https://doi.org/10.3390/app9214678>
- Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (ARL-TR-6905). Aberdeen Proving Ground (MD): Army Research Laboratory (US).
- Coyne, A. K., Murtagh, A., & McGinn, C. (2020). Using the geneva emotion wheel to measure perceived affect in human-robot interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 23–26 March 2020.
- Crnkovic, G. D., & Çürüklü, B. (2012). Robots: ethical by design. *Ethics and Information Technology*, 14(1), 61–71.
- Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human-robot co-evolution. *Frontiers in Psychology*, 9, 468.
- de Jong, C., Kühne, R., Peter, J., van Straten, C. L., & Barco, A. (2020). Intentional acceptance of social robots: development and validation of a self-report measure for children. *International Journal of Human-Computer Studies*, 139, 102426.
- DeBaets, A. M. (2014). Can a robot pursue the good? Exploring artificial moral agency. *Journal of Evolution and Technology*, 24(3), 76–86.
- Do, L.-N., Yang, H.-J., Nguyen, H.-D., Kim, S.-H., Lee, G.-S., & Na, I.-S. (2021). Deep neural network-based fusion model for emotion recognition using visual data. *The Journal of Supercomputing*, 77(3), 10773–10790.
- Dreier, T., & Spiecker gen. Döhmman, I. (2012). Legal Aspects of Service Robotics. *Poiesis & Praxis*, 9(3-4), 201–217.
- Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: ethical challenges and opportunities. *Journal of Business Research*, 129, 961–974.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Erdélyi, O. J., & Erdélyi, G. (2021). The AI liability puzzle and a fund-based work-around. *Journal of Artificial Intelligence Research*, 70, 1309–1334.
- European Commission. (2021). *Proposal for a regulation of the european parliament and of the council: laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*.
- Expert Group on Liability and New Technologies – New Technologies Formation. (2019). *Liability for artificial intelligence and other emerging digital technologies*. European Commission.

- Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and transparency: the multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics & Automation Magazine*, 26(2), 71–78.
- Ferretti, V., & Papaleo, F. (2019). Understanding others: emotion recognition in humans and other animals. *Genes, Brain and Behavior*, 18(1), e12544.
- Filntisis, P. P., Efthymiou, N., Potamianos, G., & Maragos, P. (2020). Emotion understanding in videos through body, context, and visual-semantic embedding loss. In *European Conference on Computer Vision*. Symposium conducted at the meeting of Springer, 23–28 August 2020.
- Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*. Symposium conducted at the meeting of Springer, 29–31 October 2012.
- Fischer, K. (2018). When transparent does not mean explainable. In *Papers of HRI Workshop on Explainable Robotic Systems*, Chicago, 5–8 March 2018.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379.
- Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, Dialogue, Human-Robot Interaction. In R. A. Jarvis & A. Zelinsky (Hrsg.), *Robotics Research* (S. 255–266). Springer.
- Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2–3 March 2006.
- Fosch-Villaronga, E., Felzmann, H., Mahler, T., & Ramos Montero, M. (2018). Cloud services for robotic nurses? Assessing legal and ethical issues in the use of cloud services for healthcare robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1–5 October 2018.
- Fosch-Villaronga, E., & Millard, C. (2018). Cloud Robotics Law and Regulation. *Queen Mary School of Law Legal Studies Research Paper*, (295).
- Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: shaping technology with moral imagination*. The MIT Press.
- Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the 3rd international conference on Human robot interaction – HRI '08*, 12–15 March 2008.
- Fusté-Forné, F., & Jamal, T. (2021). Co-creating new directions for service robots in hospitality and tourism. *Tourism and Hospitality*, 2(1), 43–61.
- Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: extending the computers are social actors paradigm. *Human-Machine Communication*, 1(1), 71–86.
- Gieselmann, C., & Gremmer, E. (2018). Wie Digitale Innovationen den Stationären Kaufprozess Revolutionieren – Mögliche Antworten auf den Online-Trend. In F. Keuper, M. Schomann, & L. I. Sikora (Hrsg.), *Homo Connectus* (S. 431–452). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-19133-7\\_18](https://doi.org/10.1007/978-3-658-19133-7_18).
- Gläß, R. (2018). *Künstliche Intelligenz im Handel 1 – Überblick*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-23803-2>
- Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A survey of affective computing for stress detection: evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, 5(4), 44–56.
- Gunes, H., & Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, 1(1), 68–99.
- Haidegger, T., Barreto, M., Gonçalves, P., Habib, M. K., Ragavan, S. K. V., Li, H., Vaccarella, A., Perrone, R., & Prestes, E. (2013). Applied ontologies and standards for service robots. *Robotics and Autonomous Systems*, 61(11), 1215–1223.

- Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1), 129–139.
- Hoffmann, T., & Prause, G. (2018). On the regulatory framework for last-mile Robots. *Machines*, 6(3), 33.
- Hofmann, J. (2020). Ausgewählte Technologische Grundlagen. In L. Fend & J. Hofmann (Hrsg.), *Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen* (S. 3–40). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-26964-7\\_1](https://doi.org/10.1007/978-3-658-26964-7_1).
- Holland, J., Kingston, L., McCarthy, C., Armstrong, E., O'Dwyer, P., Merz, F., & McConnell, M. (2021). Service robots in the healthcare sector. *Robotics*, 10(1), 47.
- Homburg, N. (2018). How to include humanoid robots into experimental research: a multi-step approach. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, Hilton Waikoloa Village, 3–6 January 2018.
- Hubbard, F. P. (2014). Sophisticated robots: balancing liability, regulation, and innovation. *Florida Law Review*, 66(5), 1803–1872.
- Hustinx, P. (2014). The reform of EU data protection: towards more effective and more consistent data protection across the EU: 4. In N. Witzleb, D. Lindsay, M. Paterson, & S. Rodrick (Hrsg.), *Emerging Challenges in Privacy Law* (S. 62–72). Cambridge University Press.
- The IEEE Global Initiative on Ethics of Autonomous, & Intelligent Systems (2019), Ethically aligned design – a vision for prioritizing human well-being with autonomous and intelligent systems, 1–294.
- IEEE Standards Association (2020, Juni). IEEE draft standard for transparency of autonomous systems. *IEEE P7001/D1*, 1–70.
- Ivanov, S., & Webster, C. (2019). What should robots do? A comparative analysis of industry professionals, educators and tourists. In J. Personen & J. Neidhardt (Hrsg.), *Information and Communication Technologies in Tourism 2019: Proceedings of the International Conference* (S. 249–262). Springer.
- Ivanov, S., Webster, C., & Berezina, K. (2017). Adoption of robots and service automation by tourism and hospitality companies. *Revista Turismo & Desenvolvimento*, 27(28), 1501–1517.
- Ivanov, S., Webster, C., & Seyyedi, P. (2018). Consumers' attitudes towards the introduction of robots in accommodation establishments. *Tourism: An International Interdisciplinary Journal*, 66(3), 302–317.
- Jarrassé, N., Paik, J., Pasqui, V., & Morel, G. (2008). How can human motion prediction increase transparency? In *2008 IEEE International Conference on Robotics and Automation*. Symposium conducted at the meeting of IEEE, 19–23 May 2008.
- Kagawa, R., Nobuto, M., Someya, Y., Yoshida, R., & Sugaya, M. (2018). Affect evaluation of biological information approached by a nursing/care robot. In *Proceedings of the Asia Pacific Conference on Robot IoT System Development and Platform*, Phuket, 30 October–2 November 2018.
- Kansizoglou, I., Bampis, L., & Gasteratos, A. (2019). An active learning paradigm for online audio-visual emotion recognition. *IEEE Transactions on Affective Computing*, 1, <https://doi.org/10.1109/TAFFC.2019.2961089>
- Kidd, C. D., & Breazeal, C. (2005). Human-robot interaction experiments: lessons learned. In *Proceeding of AISB*, University of Hertfordshire, 12–15 April 2005.
- Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 6–8 September 2006.
- Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, 28(1), 241–250.
- Knof, M., Heinisch, J., Kirchhoff, J., Rawal, N., David, K., Stryk, O. von, & Stock-Homburg, R. (2022). Implications from responsible human-robot interaction with anthropomorphic service robots for design science. In *Proceedings of the 55th Hawaii international conference on system sciences*.

- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biological Psychology*, 84(3), 394–421.
- Kulic, D., & Croft, E. (2005). Anxiety detection during human-robot interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2–6 August 2005.
- Kuner, C., Bygrave, L. A., Docksey, C., & Drechsler, L. (2020). *The EU general data protection regulation (GDPR): a commentary*. Oxford University Press.
- Larradet, F., Niewiadomski, R., Barresi, G., Caldwell, D. G., & Mattos, L. S. (2020). Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection. *Frontiers in Psychology*, 11, 1111. <https://doi.org/10.3389/fpsyg.2020.01111>
- Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Levenson, R. W. (1988). Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In H. L. Wagner (Hrsg.), *Social psychophysiology and emotion: theory and clinical applications* (S. 17–42). Wiley.
- Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 1–4 August 2012.
- Li, J., Ju, W., & Nass, C. (2015). Observer perception of dominance and mirroring behavior in human-robot relationships. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction*, 2–5 March 2015.
- Li, S., & Deng, W. (2020). Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing*, 1–20. <https://doi.org/10.1109/TAFFC.2020.2981446>
- Lior, A. (2020). AI entities as AI agents: artificial intelligence liability and the AI respondeat superior analogy. *Mitchell Hamline Law Review*, 46(5), 1043–1102.
- Liu, C., Rani, P., & Sarkar, N. (2006). Human-robot interaction using affective cues. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, University of Hertfordshire, 6–8 September 2006.
- Loznen, S., Bolinteanu, C., & Swart, J. (2017). *Electrical Product Compliance and Safety Engineering*. Norwood: Artech House.
- Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: scoping review and expert interviews. *Mobile Media & Communication*, 7(3), 412–434.
- Lutz, C., & Tamò, A. (2015). RoboCode-ethicists: privacy-friendly robots, an ethical responsibility of engineers? In *Proceedings of the ACM Web Science Conference*, 1 July 2015.
- Lutz, C., & Tamó-Larriex, A. (2020). The robot privacy paradox: understanding how privacy concerns shape intentions to use social robots. *Human-Machine Communication Journal*, 1(1), 87–111.
- Lynskey, O. (2014). Deconstructing data protection: the ‘Added-Value’ of a right to data protection in the EU legal order. *International and Comparative Law Quarterly*, 63(4), 569–597.
- Lyons, J. (2013). *Being Transparent about Transparency: A Model for Human-Robot Interaction* (S. 13–17). AAAI Spring Symposium Series.
- Lyons, J. B., Sadler, G. G., Koltai, K., Battiste, H., Ho, N. T., Hoffmann, L. C., Smith, D., Johnson, W., & Shively, R. (2017). Shaping trust through transparent design: theoretical and experimental guidelines. In *Advances in human factors in robots and unmanned systems* (S. 127–136). Springer International Publishing Switzerland.
- Malle, B. F. (2016). Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*, 18(4), 243–256.
- Mara, M., & Appel, M. (2015a). Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior*, 44, 326–334.

- Mara, M., & Appel, M. (2015b). Science fiction reduces the eeriness of android robots: a field experiment. *Computers in Human Behavior*, 48, 156–162.
- Matsuzaki, H., & Lindemann, G. (2016). The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. *AI & Society*, 31(4), 501–517.
- Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: how humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 56(4), 535–556.
- Merkle, M. (2019). Customer responses to service robots-comparing human-robot interaction with human-human interaction. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 8–11 January 2019.
- Mohammad, Y., & Nishida, T. (2010). Using physiological signals to detect natural interactive behavior. *Applied Intelligence*, 33(1), 79–92.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the Field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100.
- Murphy, R. R., & Schreckenghost, D. (2013). Survey of metrics for human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 3–6 March 2013.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues*, 56(1), 81–103.
- Nesbet, B., Robb, D. A., Lopes, J., & Hastie, H. (2021). Transparency in HRI: trust and decision making in the face of robot errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 9–11 March 2021.
- Niemueller, T., Lakemeyer, G., & Srinivasa, S. S. (2012). A generic robot database and its application in fault analysis and performance evaluation. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 7–12 October 2012.
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: personal information disclosure intentions versus behaviors. *The Journal of Consumer Affairs*, 41(1), 100–126.
- Ollander, S., Godin, C., Campagne, A., & Charbonnier, S. (2016). A comparison of wearable and stationary sensors for stress detection. In *2016 IEEE International Conference on Systems, Man, and Cybernetics*, 9–12 October 2016.
- Oremus, W. (2016). *Terrifyingly convenient: AI assistants can give you the news, order you a pizza, and tell you a joke. All you have to do is trust them—completely*. Slate.
- Park, E., Kim, K. J., & Del Pobil, A. P. (2011). The effects of robot's body gesture and gender in human-robot interaction. *Human-Computer Interaction*, 6, 91–96.
- Parthemore, J., & Whitby, B. (2013). What makes any agent a moral agent? Reflections on machine consciousness and moral agency. *International Journal of Machine Consciousness*, 5(02), 105–129.
- Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like? decomposing robots' human-like appearance using the anthropomorphic RoBOT (ABOT) database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 5–8 March 2018.
- Picard, R., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175–1191.
- Rajabiyazdi, F., & Jamieson, G. A. (2020). A review of transparency (seeing-into) models. In *2020 IEEE International Conference on Systems, Man, and Cybernetics*, 11–14 October 2020.
- Rani, P., & Sarkar, N. (2005a). Making robots emotion-sensitive – preliminary experiments and results. In *ROMAN 2005a. IEEE International Workshop on Robot and Human Interactive Communication*, 13–15 August 2005a.
- Rani, P., & Sarkar, N. (2005b). Operator engagement detection and robot behavior adaptation in human-robot interaction. In *Proceedings of the 2005b IEEE International Conference on Robotics and Automation*, 18–22 April 2005b.

- Rao, A. S., & Georgeff, M. P. (1995). BDI agents: from theory to practice. In *Proceedings of the First International Conference on Multiagent Systems*, 12–14 June 1995.
- Rawal, N., & Stock-Homburg, R. M. (2021). Facial emotion expressions in human-robot interaction: a survey. *International Journal of Social Robotics*. In Press.
- Rázuri, J. G., Sundgren, D., Rahmani, R., Moran, A., Bonet, I., & Larsson, A. (2015). Speech emotion recognition in emotional feedback for human-robot interaction. *International Journal of Advanced Research in Artificial Intelligence*, 4(2), 20–27.
- Riva, G., Banos, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 69–77.
- Rueben, M., Aroyo, A., Lutz, C., Schmolz, J., van Cleynenbreugel, P., Corti, A., Agrawal, S., & Smart, W. (2018). Themes and research directions in privacy-sensitive robotics. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts*, 27–29 September 2018.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.
- Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joubin, F. (2013). To err is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, 5(3), 313–323.
- Sanders, T. L., Wixon, T., Schafer, K. E., Chen, J. Y. C., & Hancock, P. A. (2014). The influence of modality and transparency on trust in human-robot interaction. In *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, 3–6 March 2014.
- Savela, N., Turja, T., & Oksanen, A. (2018). Social acceptance of robots in different occupational fields: a systematic literature review. *International Journal of Social Robotics*, 10(4), 493–502.
- Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the “Trust Perception Scale-HRI”: 10. *Robust Intelligence and Trust in Autonomous Systems* (S. 191–218). Springer.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729.
- Schwartz, P. M. (2019). Global data privacy: the EU way. *New York University Law Review*, 94(1), 771–818.
- Schweitzer, F., Belk, R., Jordan, W., & Ortner, M. (2019). Servant, friend or master? The relationships users build with voice-controlled smart devices. *Journal of Marketing Management*, 35(7–8), 693–715.
- Scriven, M. (1981). *Evaluation thesaurus*. Sage.
- Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4), 337–351.
- Shu, L., Xie, J., Yang, M., Li, Z [Ziyi], Li, Z [Zhenqi], Liao, D., Xu, X., & Yang, X. (2018). A Review of Emotion Recognition Using Physiological Signals. *Sensors*, 18(7), 2074. <https://doi.org/10.3390/s18072074>.
- Šidlauskas, A. (2019). Video Surveillance and the GDPR. In *Social Transformations in Contemporary Society 2019* (7), 55–65.
- Smarr, C.-A., Prakash, A., Beer, J. M., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2012). Older adults’ preferences for and acceptance of robot assistance for everyday living tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 22–26 October 2012.
- Sodemann, A. A., Ross, M. P., & Borghetti, B. J. (2012). A review of anomaly detection in automated surveillance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1257–1272. <https://doi.org/10.1109/TSMCC.2012.2215319>.

- Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common Metrics for Human-Robot Interaction. In M. A. Goodrich, A. C. Schultz, & D. J. Brummer (Hrsg.), *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction – HRI '06* (S. 33–40). ACM Press, 2–3 March 2006. <https://doi.org/10.1145/1121241.1121249>.
- Stiefelhagen, R., Fugen, C., Gieselmann, R., Holzapfel, H., Nickel, K., & Waibel, A. (2004). Natural human-robot interaction using speech, head pose and gestures. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)* (Bd. 3, S. 2422–2427). IEEE, 28 September–2 October 2004. <https://doi.org/10.1109/IROS.2004.1389771>.
- Stock, R., & Nguyen, M. A. (2019a). Robotic psychology. What do We know about human-robot interaction and what do we still need to learn? In T. Bui (Hrsg.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 52nd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 8–11 January 2019a. <https://doi.org/10.24251/HICSS.2019a.234>.
- Stock, R., & Nguyen, M. A. (2019b). Robotic psychology. What do we know about human-robot interaction and what do we still need to learn? In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 8–11 January 2019b.
- Stock, R. M., & Merkle, M. (2017). A service robot acceptance model: user acceptance of humanoid robots during service encounters. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops* (S. 339–344). IEEE, 13–17 March 2017. <https://doi.org/10.1109/PERCOMW.2017.7917585>.
- Stock-Homburg, R. (2021). Survey of emotions in human-robot interactions: perspectives from robotic psychology on 20 years of research. *International Journal of Social Robotics*, 1–23, <https://doi.org/10.1007/s12369-021-00778-6>
- Stock-Homburg, R., & Hannig, M. (2020). Is there a privacy paradox in the workplace? In *Proceedings of the Forty-First International Conference on Information Systems*, 13–16 December 2020.
- Stock-Homburg, R., Kirchhoff, J., Heinisch, J., Ebert, A., Busch, P., Rawal, N., David, K., Wendt, J., Spiecker gen. Döhmman, I., Stryk, O. von, Hannig, M., & Knof, M. (2022). Responsible human-robot interaction with anthropomorphic service robots: state of the art of an interdisciplinary research challenge. In *Proceedings of the 55th Hawaii international Conference on System Sciences*.
- Suzuki, S., Anuardi, MNAM., Sripiyan, P., Matsuhira, N., & Sugaya, M. (2020). Multi-user robot impression with a virtual agent and features modification according to real-time emotion from physiological signals. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 31 August–4 September 2020. <https://doi.org/10.1109/RO-MAN47096.2020.9223585>.
- Thomas, D. (2010). *Middleware for efficient programming of autonomous mobile robots*. Dissertation. Technische Universität Darmstadt.
- Tiberio, L., Cesta, A., & Olivetti Belardinelli, M. (2013). Psychophysiological methods to evaluate user's response in human robot interaction: a review and feasibility study. *Robotics*, 2(2), 92–121.
- Val-Calvo, M., Alvarez-Sanchez, J. R., Ferrandez-Vicente, J. M., & Fernandez, E. (2020). Affective robot story-telling human-robot interaction: exploratory real-time emotion estimation analysis using facial expressions and physiological signals. *IEEE Access*, 8, 134051–134066. <https://doi.org/10.1109/ACCESS.2020.3007109>
- Van de Perre, G., Van Damme, M., Lefeber, D., & Vanderborght, B. (2015). Development of a generic method to generate upper-body emotional expressions for different social robots. *Advanced Robotics*, 29(9), 597–609.

- van den Hoven van Genderen, R. (2017). Privacy and data protection in the age of pervasive technologies in AI and robotics. *European Data Protection Law Review*, 3(3), 338–352.
- Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science*, 36(1), 1–10. <https://doi.org/10.1007/s11747-007-0069-6>
- Vasylykovskiy, V., Guerreiro, S., & Sequeira, J. S. (2020). BlockRobot: Increasing Privacy in Human Robot Interaction by Using Blockchain. In *2020 IEEE International Conference on Blockchain*, 2–6 November 2020.
- Verbeek, P.-P. (2011). *Moralizing technology: understanding and designing the morality of things*. University of Chicago Press.
- Verbeek, P.-P. (2014). Some misunderstandings about the moral significance of technology. In P. Kroes & P.-P. Verbeek (Hrsg.), *Philosophy of engineering and technology. The moral status of technical artefacts*, Bd. 17 (S. 75–88). Springer Netherlands. [https://doi.org/10.1007/978-94-007-7914-3\\_5](https://doi.org/10.1007/978-94-007-7914-3_5).
- Veruggio, G., & Operto, F. (2007). The debate on roboethics. *Philosophy & Engineering*, 23.
- Veruggio, G., Solis, J., & van der Loos, M. (2011). Roboethics: ethics applied to robotics [From the Guest Editors]. *IEEE Robotics & Automation Magazine*, 18(1), 21–22. <https://doi.org/10.1109/MRA.2010.940149>
- Wagner, G. (2019). Robot Liability. In S. Lohsse, R. Schulze, & D. Staudenmeyer (Hrsg.), *Liability for artificial intelligence and the internet of things* (S. 27–62). Nomos.
- Wallach, W. (2010). Robot minds and human ethics: the need for a comprehensive model of moral decision making. *Ethics and Information Technology*, 12(3), 243–250. <https://doi.org/10.1007/s10676-010-9232-8>
- Webster, C., & Ivanov, S. (2020). *Robots in travel, tourism and hospitality: key findings from a global study*. Zangador.
- Weller, A. (2017). Transparency: motivations and challenges. In *Proceedings of the 2017 ICML Workshop on Human Interpretability in Machine Learning*, 10 August 2017.
- Westin, A. F. (1968). *Privacy and freedom*. Atheneum.
- Winfield, A. F. T., & Jirotko, M. (2017). The case for an ethical black box. In *Annual Conference Towards Autonomous Robotic Systems*, 19–21 July 2017.
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management*, 29(5), 907–931. <https://doi.org/10.1108/JOSM-04-2018-0119>
- Worham, R. H., Theodorou, A., & Bryson, J. J. (2016). What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In *Proceedings of the IJCAI Workshop on Ethics for Artificial Intelligence*, 9 July 2016.
- Wright, J. L., Chen, J. Y. C., & Lakhmani, S. G. (2020). Agent transparency and reliability in human-robot interaction: the influence on user confidence and perceived reliability. *IEEE Transactions on Human-Machine Systems*, 50(3), 254–263.
- Wullenkord, R., & Eyssel, F. (2020). Societal and ethical issues in HRI. *Current Robotics Reports*, 1(3), 85–96. <https://doi.org/10.1007/s43154-020-00010-9>
- Yoon, S., Byun, S., & Jung, K. (2018). Multimodal speech emotion recognition using audio and text. In *2018 IEEE Spoken Language Technology Workshop*. Symposium conducted at the meeting of IEEE, 18–21 December 2018.
- Zanchettin, A. M., Bascetta, L., & Rocco, P. (2013). Acceptability of robotic manipulators in shared working environments through human-like redundancy resolution. *Applied Ergonomics*, 44(6), 982–989. <https://doi.org/10.1016/j.apergo.2013.03.028>
- Zech, H. (2021). Liability for AI: public policy considerations. *ERA Forum*, 22(1), 147–158. <https://doi.org/10.1007/s12027-020-00648-0>



**Prof. Dr. Dr. Ruth Stock-Homburg**, promovierte Wirtschaftswissenschaftlerin und Psychologin, ist Leiterin des Fachgebiets Marketing & Personalmanagement an der Technischen Universität Darmstadt. Sie forschte als Gastprofessorin an der Sloan School of Management des MIT Cambridge. Für ihre Forschungsleistungen und Veröffentlichungen in renommierten Fachzeitschriften wie dem Journal of the Academy of Marketing Science (JAMS) und dem Journal of Product Innovation Management (JPIM) wurde sie 2005, 2009 und 2014 als forschungsstärkste deutsche Professorin im Bereich der Betriebswirtschaftslehre ausgezeichnet. Darüber hinaus ist Ruth Stock-Homburg Autorin zahlreicher Bücher und wurde mehrfach mit dem Best Paper Award der American Marketing Association ausgezeichnet. Sie ist die Gründerin der leap in time GmbH, eines Forschungsinstituts, das sich der Untersuchung zukünftiger Arbeitswelten gewidmet hat. Ihre Forschungsinteressen umfassen insbesondere die Themenfelder Arbeitswelt 4.0, Mensch-Roboter-Interaktion in Büroarbeitswelten, Innovationsmanagement und Leadership.

**Merlind Knof** ist wissenschaftliche Mitarbeiterin am Fachgebiet Marketing & Personalmanagement der Technischen Universität Darmstadt. Sie hat Masterabschlüsse in Quantitative Finance and Risk Management und International Area Studies. Von 2018 bis 2021 arbeitete sie bei der KPMG AG Wirtschaftsprüfungsgesellschaft in den Bereichen Audit, Corporate Governance und Deal Advisory Valuation. Ihre Forschungsinteressen umfassen insbesondere die Themenfelder Personalmanagement, Erfolgsmessungen und die Bewertung der Mensch-Roboter-Interaktion.

**Dr. Jérôme Kirchhoff** erhielt den Bachelor- und Masterabschluss in Informatik (Bereich Robotik) in den Jahren 2008 und 2011 sowie den Dokortitel im Jahr 2017 an der Technischen Universität Darmstadt, Deutschland. Derzeit ist er Postdoktorand im Fachgebiet Simulation, Systemoptimierung und Robotik der Technischen Universität Darmstadt. Neben seiner Forschungstätigkeit arbeitete er von 2011 bis 2014 im Start-up Bionic Robotics GmbH, Darmstadt, an der Entwicklung der Kernsoftware (Simulation, Steuerung, Dynamik, Kinematik, Sicherheit) für einen elastischen, sehnengetriebenen Ultraleichtroboter, der für die sichere Mensch-Roboter-Interaktion konzipiert wurde, und unterstützte hier zudem die Hardwareweiterentwicklung. Seine Forschungsinteressen umfassen Themen der menschenzentrierten Robotik, die den Menschen in seinem täglichen Leben oder seiner Arbeit unterstützen, unter anderem durch den Einsatz von Dienstleistungsrobotern, intelligenten Orthesen und Manipulatoren.

**Judith Simone Heinisch** ist Doktorandin am Fachgebiet für Kommunikationstechnik an der Universität Kassel. Sie erhielt ihren B. Sc. und M. Sc. in Informatik mit den Schwerpunkten Aktivitätserkennung und Data Mining von der Universität Kassel in 2014 bzw. 2017. Ihre Forschungsschwerpunkte sind Pervasive Computing, Maschinelles Lernen, Data Mining, Emotions- und Aktivitätserkennung.

**Andreas Ebert** ist wissenschaftlicher Mitarbeiter und Doktorand am Fachgebiet Rechtswissenschaft an der Goethe-Universität Frankfurt am Main sowie wissenschaftlicher Mitarbeiter im Kompetenzzentrum für Angewandte Sicherheitstechnologie (KASTEL) am Karlsruher Institut für Technologie. Nach seinem Studium der Rechtswissenschaft an der Ruprecht-Karls-Universität Heidelberg absolvierte er die Erste Juristische Prüfung des Landes Baden-Württemberg. Im Anschluss arbeitete er als wissenschaftlicher Mitarbeiter für eine internationale Kanzlei in Düsseldorf im Bereich IT- und Datenschutzrecht. Seine Forschungsinteressen umfassen das Datenschutzrecht, das IT-Sicherheitsrecht sowie die Regulierung von künstlicher Intelligenz und Robotik.

**Philip Busch** studierte der Rechtswissenschaften an der Goethe-Universität Frankfurt am Main und absolvierte sein Rechtsreferendariat am Landgericht Darmstadt. Neben dem Referendariat arbeitete er ab 2017 als wissenschaftlicher Mitarbeiter am Lehrstuhl von Prof. Dr. Janine Wendt an der Technischen Universität Darmstadt. Seit 2020 betreut er das Pilotprojekt RoboTrust des Zentrums für verantwortungsbewusste Digitalisierung. Seine Forschungsschwerpunkte sind die Regulierung künstlicher Intelligenz und das Technikrecht.

**Prof. Dr. Klaus David** ist seit 1998 Professor und seit 2000 Inhaber des Lehrstuhls für Kommunikationstechnik an der Universität Kassel, Deutschland. Seine Forschungsinteressen umfassen mobile Netzwerke, Anwendungen und context awareness. Er hat mehr als 200 wissenschaftliche Artikel veröffentlicht, darunter drei Bücher, und hat mehr als zehn Patente angemeldet. Er war Editor-in-Chief des IEEE Vehicular Technology Magazine (2015–2018) und Mitglied des IEEE Vehicular Technology Society Board of Governors (2015 bis 2017).

**Prof. Dr. iur. Janine Wendt** studierte Rechtswissenschaften und promovierte an der Universität Wien. Seit 2011 ist sie Inhaberin des Lehrstuhls für Bürgerliches Recht und Unternehmensrecht an der Technischen Universität Darmstadt. Janine Wendt ist Herausgeberin mehrerer Kommentare und stellvertretende Leiterin des Pilotprojekts RoboTrust des Zentrums für verantwortungsbewusste Digitalisierung. Ihre Forschungsschwerpunkte sind das Unternehmensrecht, das Finanzmarktrecht, das Recht der Digitalisierung sowie die Regulierung künstlicher Intelligenz.

**Prof. Dr. iur. Indra Spiecker genannt Döhmann, LL.M.** (Georgetown Univ.), studierte von 1990 bis 1996 u. a. Rechtswissenschaften an den Universitäten Bonn, Mainz, Heidelberg und der Georgetown Universität. 2000 erfolgte die Promotion an der Universität Bonn und 2007 die Habilitation an der Universität Osnabrück. Von 1994 bis 2008 war Indra Spiecker Wissenschaftliche Mitarbeiterin an den Universitäten Bonn und Heidelberg sowie am MPI Gemeinschaftsgüter, Bonn; 2006 bis 2008 nahm sie diverse Lehrstuhlvertretungen in Freiburg, Konstanz und Karlsruhe wahr. Seit 2008 nahm sie eine ordentliche W3-Professur für Öffentliches Recht, Telekommunikationsrecht, Datenschutzrecht wahr und war gleichzeitig Direktorin des Instituts für Informationsrecht am Karlsruher Institut für Technologie. Es erfolgten Rufe an die Universitäten Berlin, Mannheim, Trier und Hannover. Seit 2013 ist Indra Spiecker W3-Professorin für Öffentliches Recht, Informationsrecht, Umweltrecht und Verwaltungswissenschaft an der Goethe Universität Frankfurt a. M. Dort leitet sie außerdem als Direktorin die Forschungsstelle Datenschutz (geschäftsführend), das Institut für Europäisches Sozialrecht und Gesundheitspolitik (ineges) (geschäftsführend) sowie die Forschungsstelle Umweltrecht. Indra Spiecker füllt diverse Beirats- und Kuratoriumspositionen aus. Sie ist Mit-Herausgeberin diverser Law Journals sowie des größten DSGVO-Kommentars. Als erste Juristin

wurde sie 2016 in die in die Akademie der Technikwissenschaften berufen; für die Leopoldina ist sie u. a. in der Arbeitsgruppe Gesellschaft und Digitalisierung tätig und wirkte u. a. an der Stellungnahme zu „Digitalisierung und Öffentlichkeit“ mit. Ihre Expertise wird national wie international geschätzt; sie berät Unternehmen, NGOs, Behörden, Parlamente und die Regierung.

**Prof. Dr. Oskar von Stryk** ist Informatik-Professor an der TU Darmstadt, Vizepräsident der internationalen RoboCup-Federation, Founding Trustee der Association for the Understanding of Artificial Intelligence und stellvertretender Vorstandsvorsitzender des von ihm mitgegründeten Deutschen Rettungsrobotik-Zentrums e.V. Er hat mehr als 200 internationale Fachpublikationen veröffentlicht und mehrere Robotik-Start-ups wie Energy Robotics und Freemotion Systems mitgegründet. Oskar von Stryk ist als Gutachter für nationale und internationale Forschungsprojekte tätig. Seine Forschung zu kooperierenden autonomen mobilen Robotern, bio-inspirierten Robotern sowie Optimierungs- und Steuerungsmethoden wurde in mehr als 30 Projekten durch nationale, europäische und US-Forschungsagenturen gefördert. Die Forschungsergebnisse wurden mehrfach ausgezeichnet, unter anderem mit dem ersten Preis des European Robotics Technology Transfer Award, dem Hessischen Kooperationspreis sowie dem Louis Vuitton Best Humanoid Award und führten zum Gewinnen vieler internationaler Wettbewerbe für autonome mobile Roboter (unter anderem ARGOS Challenge, ENRICH, RoboCup, World Robot Summit). Er wird von Academic Influence als einer der 25 weltweit einflussreichsten Informatikerinnen und Informatikern angesehen.

**Martin Hannig** hat einen Masterabschluss in Betriebswirtschaftslehre und ist wissenschaftlicher Mitarbeiter und Doktorand am Fachgebiet Marketing & Personalmanagement der Technischen Universität Darmstadt. Seine Forschungsinteressen umfassen insbesondere die Themenfelder Personalmanagement, Mensch-Roboter-Interaktion in Unternehmen und Datenschutz. In diesen Themenfeldern hat er mehrere Publikationen veröffentlicht und einen Best Paper Award der International Conference on Information Systems 2020 erhalten.