



The trustworthy and acceptable HRI checklist (TA-HRI): questions and design recommendations to support a trust-worthy and acceptable design of human-robot interaction

Johannes Kraus¹ · Franziska Babel¹ · Philipp Hock¹ · Katrin Hauber¹ · Martin Baumann¹

Accepted: 12 July 2022 / Published online: 4 August 2022
© The Author(s) 2022

Abstract

This contribution to the journal *Gruppe. Interaktion. Organisation. (GIO)* presents a checklist of questions and design recommendations for designing acceptable and trustworthy human-robot interaction (HRI). In order to extend the application scope of robots towards more complex contexts in the public domain and in private households, robots have to fulfill requirements regarding social interaction between humans and robots in addition to safety and efficiency. In particular, this results in recommendations for the design of the appearance, behavior, and interaction strategies of robots that can contribute to acceptance and appropriate trust. The presented checklist was derived from existing guidelines of associated fields of application, the current state of research on HRI, and the results of the BMBF-funded project RobotKoop. The trustworthy and acceptable HRI checklist (TA-HRI) contains 60 design topics with questions and design recommendations for the development and design of acceptable and trustworthy robots. The TA-HRI Checklist provides a basis for discussion of the design of service robots for use in public and private environments and will be continuously refined based on feedback from the community.

Keywords Human-robot interaction · Social robots · Trustworthy design · Trust in robots · Trust in technology · Technology acceptance

The authors J. Kraus and F. Babel contributed equally to this work.

✉ Dr. Johannes Kraus
johannes.kraus@uni-ulm.de

¹ Department of Human Factors, Ulm University, Ulm, Germany

Die Trustworthy and Acceptable HRI Checklist (TA-HRI): Fragen und Gestaltungsempfehlungen zur Unterstützung eines akzeptablen und vertrauenswürdigen Designs der Mensch-Roboter-Interaktion

Zusammenfassung

Dieser Beitrag der Zeitschrift *Gruppe. Interaktion. Organisation. (GIO)* stellt eine Checkliste mit Fragen und Designempfehlungen zur Gestaltung einer akzeptablen und vertrauenswürdigen Mensch-Roboter-Interaktion (MRI) vor. Um den Anwendungsbereich von Robotern auf komplexere Kontexte im öffentlichen Bereich und im privaten Haushalt auszuweiten, müssen Roboter neben der Sicherheit und Effizienz auch Anforderungen hinsichtlich eines sozialen Miteinanders zwischen Menschen und Robotern erfüllen. Insbesondere ergeben sich hieraus Empfehlungen für das Design des Äußeren, des Verhaltens und der Interaktionsstrategien von Robotern, die zu Akzeptanz und einem angemessenen Vertrauen beitragen können. Die vorgestellte Checkliste wurde aus vorhandenen Richtlinien aus assoziierten Anwendungs- und Themenfeldern, dem aktuellen Forschungsstand zur MRI und den Ergebnissen des BMBF-geförderten Projektes RobotKoop abgeleitet. Die Trustworthy and Acceptable HRI Checklist (TA-HRI) enthält 60 Designthemen mit Fragen und Designempfehlungen für die Entwicklung und Gestaltung von akzeptablen und vertrauenswürdigen Robotern. Die TA-HRI-Checklist ist Grundlage für die Diskussion der Gestaltung von Servicerobotern zum Einsatz im öffentlichen und privaten Umfeld und wird auf Grundlage von Feedback aus der Community kontinuierlich weiterentwickelt.

Schlüsselwörter Mensch-Roboter-Interaktion · Soziale Roboter · Vertrauenswürdiges Design · Vertrauen in Roboter · Vertrauen in Technik · Technikakzeptanz

1 Motivation

In recent years, robots have started to be used in more and more everyday as well as professional contexts with service robots increasingly performing their tasks in direct interaction with humans. Here, service robots are conceived as robots that perform useful tasks for humans in a partially or fully autonomous manner, whereby the industrial application is excluded from this scope (International Organization for Standardization 2021). The area of application of such service robots thus includes both public and private spaces. In the public sector, for example, service robots are already used as information assistants, but more complex tasks such as autonomous cleaning or deliveries can already be realized. Also, in the private context, service robots already take over simple cleaning tasks today. In the near future, it is conceivable that robots will be used as personal assistants to help people with a wide range of everyday tasks. In this context, robots will take on new roles in the social system (Enz et al. 2011). In the course of this development, requirements for the design of the interaction of robots with humans in their task environment are changing (e.g., Bartneck and Forlizzi 2004). This does not only concern the interaction with people who own the robots or who directly perform a task together with them, but increasingly also people who happen to be in the task environment of the robots (e.g. passers-by in public spaces). In this sense, service robots are turning more and more into social robots that can be considered part of a mixed society of humans and robots, in which robots, among other things, interact and communicate socially with humans, thereby learning and acquiring experiences (e.g., Hegel et al. 2009; Fong

et al. 2003). Accordingly, a service robot which performs its task in the context of social interaction can be described as a social (service) robot. This work refers to all types of social service robots (hereafter also referred to simplified as “service robots” or “robots”).

In the context of social interaction (both in the private and in the public domain), the appearance and behavior of service robots, as well as the interaction concepts they are equipped with, must fulfil several basic requirements. On the one hand, they must allow efficient task performance, and on the other hand, they must be suitable for the social structure in which the robot operates and above this be acceptable on an individual level. On this basis, the RobotKoop project funded by the BMBF (2018–2021) pursued the vision of cooperative, intelligent service robots which operate in dynamic social settings, act in a trustworthy and acceptable manner, and thereby negotiate and coordinate their actions with the people around them. One major step towards this vision is a context-sensitive, cooperative human-robot interaction strategy. The advantages of robot use in areas such as household, care, communication and service can only fully play out if the employed robots are equipped with interaction strategies that are context- and need-sensitive. These strategies should allow situation- and goal-oriented communication with the environment in which the robot is used. Such a human-robot interaction (HRI) needs to be perceived as transparent, acceptable and trustworthy on the part of the users and the persons in the environment of task execution. In particular, the promotion of a minimum level of acceptance and an appropriate level of trust are fundamental prerequisites for efficient, safe and pleasant togetherness of humans and robots.

Against this background, in the following a checklist is presented that is intended to support a human-centered design of robots, their behavior and communication in both science and practice. By providing 60 questions on design topics related to acceptance and trust and making design recommendations based on these questions, the checklist is intended to contribute to an increase in subjective trustworthiness and acceptance of HRI design. The questions and design recommendations in the checklist are intended to serve as an orientation aid, source of inspiration, and working tool for practitioners (e.g., in engineering, computer science, and product development) and researchers, and provide a starting point and a basis for discussion to enhance human-centered HRI design in specific HRI projects.

2 Theoretical background

To answer the question of how robots and their interaction with humans can be optimally designed with respect to acceptance and trust, it is not feasible to formulate a generalizable strategy due to the wide range of applications, tasks, and design possibilities of robots. With respect to the current state of HRI research, in regard to many specific design decisions partly contradictory results from specific studies and various individual collections of guidelines, recommendations, and requirements from different application areas can be found.

The aim of the presented checklist is to integrate and complement the existing results and approaches with the findings from studies and expert discussions of the RobotKoop project. Through its application in science and practice, the investigation and practical implementation of trustworthy and acceptable HRI might be advanced and promoted.

In the following, a theoretical introduction to the underlying concepts cooperation, trust, and acceptability is provided. Furthermore, a preliminary overview of existing collections of recommendations and requirements from related fields is given and their relevance for the present work is discussed.

2.1 Cooperation between humans and robots

In recent years, the distribution of tasks between humans and robots in the socio-technical system of joint task execution has been changing from a situation in which robots and humans mainly co-exist (the tasks of robots and humans are independent of each other) to the possibility of cooperative teamwork between humans and robots. In this respect, the concept of human-technology cooperation describes a type of working together between humans and technical systems in which both parties pursue a common

goal as team players, coordinating, aligning, and complementing their task performance with each other (e.g., Hoc 2001; Christoffersen and Woods 2002; Klein et al. 2004).

In this context, different types of cooperative interaction between humans and robots can be distinguished, of which, for example, two types are differentiated in the work of Onnasch et al. (2016). In the first type of cooperation (referred to as human-robot cooperation in Onnasch's work), humans and robots work together towards a common goal, whereby there is a clear division of the tasks between humans and robots and their actions are not directly dependent on each other. The second type of joint working is referred to as human-robot collaboration, which goes beyond cooperation by describing a working relationship between the human and the robot, in which sub goals are also worked on together simultaneously (and direct physical contact may also occur under certain circumstances). In these collaborative scenarios, the roles of humans and robots change. Humans and robots enter into a social exchange with each other—naturally on the side of the robot within its restrictions in regard to subjectivism and intentionality—and coordinate dynamic solutions to problems, taking into account their respective capabilities. In the following, the term human-robot cooperation is used to refer to both types of cooperation, since both situations are difficult to distinguish in practice or merge into each other, or sometimes this distinction does not seem meaningful (Onnasch describes cooperation as a type of collaboration). Ultimately, both types of collaboration between humans and robots require similar basic prerequisites on the part of the design of robot behavior and the user interface.

As compared to a mere coexistence, this more complex and communication-intensive cooperation between humans and robots creates new requirements for the design of robots and HRI (e.g., Walch et al. 2017; Babel et al. 2021). First, in order to ensure successful human-robot cooperation, the interface should foster a shared situational awareness between humans and robots (e.g., human-robot awareness; Yanco and Drury 2004; Drury et al. 2004). In this sense, each partner should be well informed about the current status of subtasks, as well as the current activities, and plans of the other. In this way, a certain predictability of the actions of the robotic or human counterpart can be established.

Furthermore, some degree of controllability of the actions of the cooperation partner seems necessary (e.g., Christoffersen and Woods 2002). For example, human users should be able to dynamically adapt the task scope of the robot or, to a certain extent, to control the way in which the robot performs tasks. Similarly, in many scenarios it seems desirable that the robot informs the human partner about upcoming actions or need for support. In many application contexts, it might be a necessity for the robot to prevent the human from carrying out potentially

risky activities or to point out potential errors. These and other scenarios that can occur in cooperative collaboration between humans and robots require a certain degree of acceptance and trust in the robotic team partner or the robotic household helper. Without this, it could be unpleasant for the user to grant the robot own competencies and decision-making scope. This, in turn, could have negative psychological consequences, such as anxiety or stress, which—both in work and private contexts—could have even more serious long-term consequences for the user.

In addition to the cooperative and collaborative teamwork of humans and robots, also a mere co-existence of humans and robots requires a coordination of the respective independent goals and interests. Consequently, also in this case, the robots should be equipped with a cooperative interaction interface that enables an effective and acceptable coordination between humans and robots.

Against this background, this work aims to provide a collection of design topics for informing the design of the appearance and interface of robots in both the public and private domains. This is intended to support optimized cooperation between humans and robots and support the formation of an appropriate level of acceptance and trust on the part of the users or people present in the area of the robots' tasks.

2.2 Acceptance of robots

The understanding and the prediction of acceptance of robots and robot behavior is a central research topic in HRI (de Graaf and Allouch 2013), as acceptance is a basic prerequisite for the use of automated technology (see, e.g., Technology Acceptance Model; TAM, Ghazizadeh et al. 2012). In this way, for example, a minimum level of acceptance constitutes a subjective prerequisite for the use of a technical system, e.g., a robot. Due to the large number of publications on the topic of acceptance, there is currently a wide variety of acceptance definitions, some of which contradict each other (e.g., Arndt 2011). This is summarized by Königstorfer and Gröppel-Klein (2009): “In acceptance [...] research, there is now a consensus that acceptance moves along a continuum that ranges from attitude [...] to action (purchase) and regular use of technological innovations” (p. 849, German original translated by the authors). In this sense, acceptance is defined in this paper as the intention to use—a subjective evaluation that influences the extent to which a robot is used (e.g., Naneva et al. 2020).

Since a minimum level of acceptance is a necessary prerequisite for the use of robots, the factors that influence users' acceptance are of particular interest for HRI design. In line with the general framework of the TAM, Ghazizadeh et al. (2012) identified perceived usefulness and ease of

use as subjective factors influencing robot acceptance. In addition, robot-specific models for predicting acceptance include emotional processes on the part of users. For example, the USUS Evaluation Framework postulates that robot acceptance is significantly influenced by users' attitudes towards robots and their emotional attachment to robots, in addition to expectations regarding robot performance and efficiency (Weiss et al. 2009). Studies also identified additional robot characteristics influencing acceptance. For example, numerous studies have shown that transparency increases robot acceptance (Alonso and de la Puente 2018; Cramer et al. 2008; Ososky et al. 2014). Similarly, appropriate social behavior of the robot has been found to promote acceptance (politeness, social distance, communication behavior; de Graaf et al. 2015; Babel et al. 2021, 2022a, b). Furthermore, some studies have shown that the degree of human-likeness affects robot acceptance. In this regard, some studies report higher acceptance of human-like robot design (e.g., Barnes et al. 2017; Eyssel et al. 2012; Louie et al. 2014).

However, it can be noted that the influence of specific design and interaction features can vary depending on the studied robot types, tasks, subject groups, etc. This is supported by the findings of a recently published meta-analysis by Naneva et al. (2020). Overall, there was a wide variation in the average acceptance levels across the included studies. Namely, the average acceptance of robots tended to be in the negative range in 42% of the 26 included studies. The authors identify several factors in the study design that influence the level of acceptance. On the one hand, studies in which robots were either directly interacted with or not interacted with at all showed a higher average acceptance compared to studies in which robots were indirectly presented (picture/video of a robot). On the other hand, robots were better accepted in studies with a setting in the educational field than in studies in the health and care field or in studies in which the application field was not explained in detail. In contrast, age, gender, and publication year had no effect on the acceptance of social robots (Naneva et al. 2020).

2.3 Trust in robots

A second fundamental subjective prerequisite for enjoyable, efficient and safe use of and interaction with service robots is an appropriate level of trust in these robots. While trust has been researched in Psychology for many decades in the context of interpersonal relationships (e.g., Rempel et al. 1985; Holmes and Rempel 1989) the area of automation trust, i.e., trust in automated technical systems, represents a comparatively new research direction. Since the late 1980s, this research has been increasing, and initially the focus was on trust processes in the monitoring and op-

eration of professional, automated industrial systems (e.g., Muir and Moray 1996; Lee and Moray 1992). Over the past two decades, the number of research papers on trust processes in automated vehicles (e.g., Hergeth et al. 2017; Kraus et al. 2019, 2020; Beggiato and Krems 2013, 2015) and robots (Miller et al. 2021; Babel et al. 2021, 2022a, b; Kraus et al. 2018) increased. Fundamentally, at the psychological level, one can distinguish between several layers of trust (e.g., Marsh and Dibben 2003). Here, a basic distinction needs to be made between the personality tendency to trust automated technology in general (propensity to automation trust), and a learned attitude with respect to a specific technical system (learned trust).

The general dispositional tendency to trust automated technology has been defined as an overarching individual predisposition to trust automated technology across different contexts, systems, and tasks (e.g., Hoff and Bashir 2015; Kraus 2020). It describes a user's individual personality tendency to trust a broad set of automatized technology across a range of situations. It is hypothesized that this individual predisposition to trust automated technology arises from a combination of the individual user's personality and the experiences they have with technology over the course of their learning history (e.g., Kraus 2020). This individual predisposition to trust a technical device thus represents an individual psychological basis for the formation of learned trust in a specific, new technical system. In line with this, Miller et al. (2021) found the propensity to trust to predict learned trust in the assistance robot Tiago (PAL Robotics) in a laboratory study. This supports previous findings by Kraus et al. (2021) in the domain of automated driving. Additionally, in terms of personality variables mainly a positive association between extraversion and trust in robots has been reported (e.g., Haring et al. 2013; Alarcon et al. 2021). From this, it can be concluded that when considering and optimizing trust processes in interaction or cooperation with robots, differences in the personality and experiences of users should also be taken into account.

Furthermore, according to Lee and See (2004), learned trust in automation is commonly defined as "an attitude that an agent will help an individual achieve a goal in a situation of uncertainty and vulnerability" (p. 51). In this respect, trust is a dynamic psychological attitude related to a specific technical (automated) system that develops in the course of getting to know and building a relationship with this technical system (e.g., Miller et al. 2021). The level of trust is influenced by available information—so-called trust cues (Thielmann and Hilbig 2015)—on the basis of which it is calibrated over time. Both, information available before the actual interaction with the system and information available during the use of the system, are considered in this process of trust calibration. The optimal result of such a trust calibration process is a calibrated level of trust—a situation

of adequate trust that is characterized by the fact that users trust a technical system exactly to a degree that corresponds to the capabilities and reliability of the system (e.g., Forster et al. 2018; Lee and See 2004).

The psychological variable trust has particular relevance for behavior in situations in which the trust-giving agent is exposed to a particular degree of uncertainty, risk, and vulnerability (e.g., Thielmann and Hilbig 2015). This applies to interaction or collaboration with novel robots in both private and public settings. In the recent meta-analysis by Naneva et al. (2020), a total of 30 studies that investigated trust in social robots were analyzed. Overall, a wide variability of the average trust was found. Above this, the authors report various factors on the side of the systems and the study setup that seem to affect the level of trust. The presented checklist aims at optimizing calibration of trust in robots by stimulating considerations and an informed design of robots' appearance and interaction concepts as well as a systematic communication of information to the users about the robot (e.g. in the form of trainings, user manuals or tutorials). On the one hand, this should promote the development of a sufficient degree of trust on the part of the user, so that he or she wants to use the robot or accepts its task execution at all. On the other hand, however, this is also intended to prevent overtrust in the robot, which could lead to the robot being assigned with tasks that it is not designed to perform or being used in situations that go beyond its scope of application (e.g., Parasuraman and Riley 1997). Similarly, overtrust could also lead to not keeping a necessary safety distance or not sufficiently considering the needs of vulnerable groups of people (e.g. children, elderly people). All these possible consequences of overtrust represent potential dangers of the use of robots in the public and private sector. In this way, supporting a calibrated level of trust in robot design and in the design of human-robot interfaces are likely to considerably foster a pleasant, safe, and stress-free interaction with the robot.

Based on these theoretical foundations and considerations, the presented checklist was developed with the goal of promoting an acceptable and trustworthy interaction between humans and robots. Thereby, the level of trustworthiness that is aimed at in the design process should reflect the actual capabilities and reliability of the robot—and not exceed it—to prevent overtrust and associated, potentially harmful interaction decisions. To level out and calibrate the trustworthiness of robot design is an important responsibility of robot and HRI designers. The development and structure of the checklist are described in the following.

3 Development and structure of the checklist

3.1 Development process

The checklist was created in an iterative process in partnership with interdisciplinary experts. Here, existing collections of criteria from different disciplines and technical domains, with different focuses were to be integrated and expanded. A multi-stage procedure was used, which is outlined in the following (Fig. 1).

1. The development process started with a broad literature search on existing ethical, safety, privacy, and interaction guidelines for the use of robots in private and/or public spaces (e.g. European Parliament 2017; Gelin 2017; Kreis 2018; Salvini et al. 2010). Search platforms used here were Google Scholar, Science Direct, and Scopus. Keywords used included: Robot requirements; Roboter Anforderungen; robot guidelines; human-machine interaction ethic*; Mensch-Roboter-Interaktion; Interaktion Mensch-Roboter; Sicherheit Anforderung; Serviceroboter Anforderungen; service robot ethic* & safety & guideline; collaborative robot guideline; personal robot data safety guideline; Datenschutz Richtlinie Roboter.

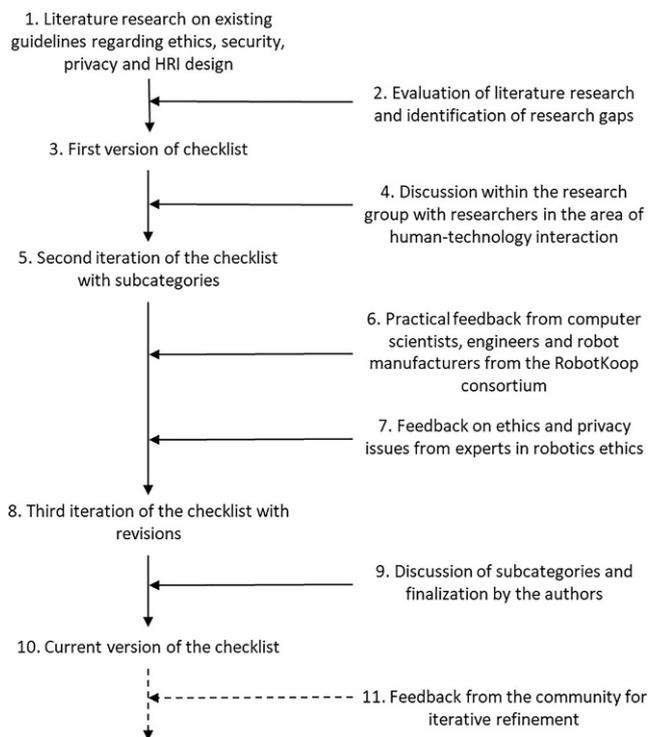


Fig. 1 Flow diagram of the development process of the Trustworthy and Acceptable HRI Checklist (TA-HRI)

2. In addition, key factors that can influence acceptance and trust towards robots were identified on the basis of existing literature (e.g. Hancock et al. 2021; de Graaf and Allouch 2013).
3. The current state of the literature was then assessed and evaluated. It was found that there is additional potential to more strongly stress the individual perspective of robot users and of people who are in the environment of robots in terms of trust and acceptance in HRI design.
4. Based on the literature review, a first version of the checklist (design topics with associated questions and design recommendations) was created, which summarized the preliminary results. The different identified questions and recommendations were grouped into categories. This list was further updated on the basis of additional literature.
5. The compiled results were further expanded and adapted through an expert survey and discussion with HRI experts (both from a research and a practice and application perspective). On this basis, subcategories were introduced into the structure of the checklist. The questionnaire included eleven open questions and mainly referred to the experts' views on several topics in HRI (e.g. "In your experience: What general requirements must a cooperative robot fulfil?"). The questionnaire was sent to 12 experts. Five completed questionnaires were returned by the expert groups (jointly completed answers). The questionnaire and the preliminary criteria collection then served as a basis for discussion of further requirements and design recommendations.
6. On this basis a second version of the checklist was developed incorporating subcategories of design topics.
7. On this basis, the design topics, questions and design recommendations of the checklist were further developed in discussions with additional interdisciplinary experts (psychologists, computer scientists, engineers and robot manufacturers) within the RobotKoop project. This enabled the integration of further practical feedback into the recommendations.
8. In addition, experts on the subject of ethics and data protection in the domain were consulted to provide feedback.
9. The interdisciplinary feedback was integrated to a third iteration step of the checklist.
10. The resulting version was again discussed (especially in regard to the organization and naming of the subcategories) within the author team.
11. This resulted in the current version of the checklist.
12. The current version of the checklist does not claim to be a final, complete version, but is to be further developed on the basis of feedback from the community.

3.2 Included factors influencing robot trust and acceptance in the checklist

Trust and acceptance are important foundations of successful HRI and integration of robots in everyday life. Study results indicate that amongst others user characteristics (e.g., individual attitudes towards robots, expectations), environmental and task factors (e.g., team collaboration, task characteristics), and robot characteristics can influence the trust towards robots (cf. Hancock et al. 2021; de Graaf and Allouch 2013). On the robot side, Hancock et al. (2021) identified in a meta-analysis that especially robot performance (e.g., low error rate, high reliability) as well as properties of appearance and robot behavior (e.g., anthropomorphism, physical proximity) as relevant for trust. In this regard, especially, transparency of the robot's plans, processes and actions are important to establish a realistic expectation towards it, which in turn builds an essential basis for calibrated trust (e.g., Kraus 2020; Kraus et al. 2019). In this sense, communication, bilateral understanding and task coordination are essential for trust. In line with this, the acceptance of a robot can be promoted if it is perceived as useful, adaptable and controllable, as well as a sociable companion by its design (de Graaf and Allouch 2013).

Against the background of the discussed state of research (see 2.2 and 2.3), the presented checklist integrates a large number of the aforementioned factors influencing the acceptance of and trust in robots in the entailed design topics, questions and recommendations. In particular, transparency, understandability, and a trustworthy design of both robot appearance and interaction are considered. Furthermore, in order to pay respect to the discussed individual differences between users, throughout the checklist possibilities for customizability and individualization are suggested.

Additionally, characteristics of the situation, in which the interaction between humans and robots takes place, are commonly viewed as essential for the formation of trust (e.g., Lee and See 2004; Kraus 2020; Hancock et al. 2021) and acceptance (Abrams et al. 2021; Turja et al. 2020; de Graaf et al. 2019). For example, ethical and legal concerns regarding the use of a robot can negatively impact trust (Alaiad and Zhou 2014). Consequently, a design of robots that promotes trust calibration and acceptance should also consider ethical, safety, and privacy aspects as these seem to establish a framework in which trust and acceptance can prosper.

Therefore, in order to establish a functioning, efficient as well as from a subjective viewpoint enjoyable integration of robots into existing social systems, an adaptive, norm congruent and appropriate social behavior of robots is an essential design goal to foster both trust and acceptance of robots. Therefore, additionally to design and interaction considerations, the presented checklist integrates legal and

societal framework conditions which are mainly based on previous work as discussed in the following.

3.3 Included ethical, safety and legal requirements in the checklist

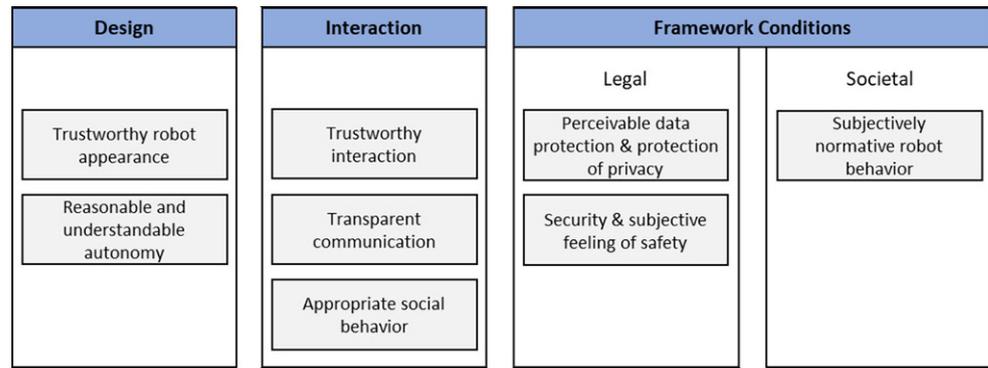
The introduction of artificial intelligence (AI) technologies into society poses potential risks to physical safety, data protection, human rights, and fundamental freedoms (Yeung 2018). For this reason and as technological developments progressed, in recent years a large number of ethical guidelines and recommendation lists have been published. A recent prominent example are the ethical guidelines for trustworthy AI which were developed by an independent group of experts on behalf of the EU Commission (European Commission 2019). These guidelines describe seven ethical prerequisites, which should be examined before a system enters the market. These are (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) societal and environmental well-being, and (7) accountability.

In addition, there are long-established guidelines regulating safety for people working with industrial robots. The Machinery Directives 2006/42/EC set out basic safety requirements for machinery, such as maintaining a minimum distance from people, fitting guards and an emergency stop switch. In Germany, these requirements have been transposed into national law by the Product Safety Act. For collaborative robot systems, the safety requirements under the DIN ISO/TS 15066 (Deutsches Institut für Normung e. V. 2017) standard for industrial robots apply in particular. For personal assistance robots, DIN EN ISO 13482 (Deutsches Institut für Normung e. V. 2014) specifies requirements for safe design, protective measures and user information.

Furthermore, the EU's General Data Protection Regulation (GDPR) applies to robots that store the personal data of users. This means that individuals must give their consent to data processing and that this consent can be withdrawn at any time. This applies in particular to robots with sensors that process audiovisual data in order to interact with their environment. However, the GDPR does not yet contain any explicit specifications in regard to robots, which was pointed out by the EU Parliament in a resolution on civil law regulations in the area of robotics. The recommendations of the latter include data protection-friendly default settings and transparent control procedures for affected persons (European Parliament 2017).

The development of this checklist incorporates this and other work (see footnotes in the checklist).

Fig. 2 Structure of the Trustworthy and Acceptable HRI Checklist (TA-HRI): the checklist is established by 60 design topics with respective questions and design recommendations in four areas and eight categories



3.4 Structure of the checklist

The checklist includes 60 design topics at different levels of HRI design (see Table 1 for the English version and Table 2 for the German version). A distinction is made here between private service robots in private households and robots that perform work in public spaces. Based on the preceding theoretical considerations, the design topics are assigned to the four areas: 1) design, 2) interaction, 3) legal, and 4) societal environment, which are in turn subdivided into eight categories (Fig. 2). For each design topic questions are provided in the checklist to favor acceptable and trustworthy HRI design. Based on each question exemplary design recommendations are listed that can help to optimize the acceptance and trustworthiness of robots.

The area of ‘design’ includes categories that relate to the design and development of the robot. A *trustworthy robot appearance* of the, as well as a *reasonable and understandable autonomy*, should be designed in a way that does not cause uncertainty for the user (Kreis 2018; Rosenstrauch and Kruger 2017; Salvini et al. 2010) and ultimately fosters a calibrated level of trust.

The area of ‘interaction’ describes processes that relate to the direct exchange between humans and robots. During interaction, users gain experience in dealing with the robot and, as a result, calibrate their trust in it. For this, on the one hand, the robot must *interact in a trustworthy manner* during the interaction and *communicate its actions and states transparently* (Gelin 2017; Hancock et al. 2011). On the other hand, the expectations towards an *appropriate social behavior* of the robot must be fulfilled.

In the areas of ‘legal and societal framework conditions’, a distinction is made between the following three categories: 1. *perceivable data protection & protection of privacy*, 2. *security & subjective feeling of safety*, and 3. *subjectively normative robot behavior*. The questions of the categories regarding data protection and security refer, among other things, to the legally regulated demands placed on the technical system in order to protect the users (Jacobs 2013; Müller 2014). The questions on the societal frame-

work conditions refer to the subjective compliance with normative and moral principles (e.g., European Parliament 2017; Gelin 2017).

4 Discussion

4.1 Application, considerations and scope

This work presented the TA-HRI Checklist incorporating design topics to support trustworthy and acceptable HRI design. The design topics, questions and design recommendations covered in the TA-HRI Checklist pursue the goal to stimulate a consideration of potential design approaches that can contribute positively to the trustworthiness and acceptance of robots. Thereby, the questions and recommendations address different levels of design aspects. In addition to the physical appearance and the interaction, also the integration of the robots into the legal and social context is addressed.

The checklist can be used in robot development as a heuristic framework for optimizing the interaction design at an early stage. The questions on the respective design topics can serve as a starting point for discussions and the evaluation of the design of the HRI in the individual design or research project. They are intended to help examining design ideas with regard to possible trust- or acceptance-promoting (or calibrating) aspects and, if necessary, they might contribute to optimize the design in this regard. Thereby, the evaluation should always be carried out under consideration of the system, its specific task and the users interacting with this system. In this sense, successful implementation of the design topics should be evaluated in the specific application context.

The listed recommendations are not to be understood as indisputable design rules. Some of the recommendations listed may not be appropriate for all contexts, robotic types, tasks, and user groups. In this sense, the appropriateness and expected gain of each recommendation should be evaluated in the context of the intended robot task and operational

context. Also, in practice, it may well be that the design criterion of trustworthiness or acceptability must be subordinated to other criteria (such as effectiveness or efficiency of task execution). This may be the case, for example, in security and emergency scenarios, where the effectiveness of the interaction (e.g., evacuation of a building) may be considered more important than its acceptability (Babel et al. 2022b). Also, increasing trust to a maximum that exceeds the actual capabilities and reliability of a robot (overtrust) can lead to dangerous interaction decisions (misuse). Therefore, the ultimate design goal in terms of trust should always be a calibrated instead of a maximum level of trust.

Furthermore, it can sometimes be useful to not implement the single recommendations as mandatory features, but to offer them to the user as an option and individually customizable features. In this sense, the checklist also offers a set of suggestions to implement personalization in the interaction with robots (e.g., recommendations regarding transparency).

The checklist neither has the claim to represent the requirements for a trustworthy HRI in an all-encompassing nor final way and is intended to form a basis for an interdisciplinary iterative development process. For this reason, feedback and comments on the checklist are explicitly welcome and serve as a basis to continuously update and enhance the checklist in an ongoing discussion within the community.

4.2 Strengths, limitations and further development

The presented checklist implements several areas of factors affecting trust and acceptance of social robots and integrates them with recommendations from a safety and ethical viewpoint. It is a result of an iterative process combining several methods—besides others literature search and expert discussions with interdisciplinary experts from the fields of engineering, computer science, psychology and ethics—to arrive at the presented design topics. With this, it is intended as a practical tool both for practitioners and researchers to support the consideration of acceptance and trust in a human-centered robot and HRI design process emphasizing the perspective of users. Thereby, in its current form the checklist aims at providing design impulses rather than a fixed list of rules.

At this stage of development, the checklist still has some limitations, which might be addressed in future iteration steps and further research. While existing models from the domain of acceptance and trust were used as a basis for the checklist, there was no explicit integrative theoretical model underlying the checklist development. Based on the areas and categories entailed in the checklist a model might be derived and empirically tested. This in turn would be a valuable starting point for the derivation of metrics and a sys-

tematic validation of the checklist. Derived metrics could be used as an extension of the scope of the checklist for evaluating a specific robot design against the different areas and categories of the checklist. With this the checklist would constitute a helpful evaluation tool for trustworthy and acceptable robot design for usability studies and A/B testing. A further limitation is the mainly European-centered perspective of the checklist, which might be extended in future iteration steps to include additional perspectives. Also, with ongoing technological progress, amongst others in the field of AI, new technological solutions for the covered challenges for acceptable and trustworthy robot design might be available. Therefore, continuous updating of the checklist to integrate technological developments seems necessary. In the same manner, in the next years with robots being more and more present in different domains of daily life, legislation and social norms will be adapted and refined, which should also be considered in coming iterations of the checklist.

To conclude, the checklist is intended as a practical tool to enhance the consideration of trust and acceptance in HRI design both in practice and research. The authors look forward to feedback and directions for continuous updating of the checklist from the community.

Table 1 The Trustworthy and Acceptable HRI Checklist (*TA-HRI*)—English Version

Questions on Design	Design Recommendation	Use Context
Trustworthy robot appearance		
1 <i>Is the robot designed to look human-like for no compelling reason?</i>	Human likeness of the robot is purpose-built and appropriate. The robot remains recognizable as a machine ^a	Both
2 <i>Does the robot appear threatening?</i>	The robot's dimensions are chosen to allow for optimal task completion with minimal threat. ^b The robot's face is designed to be neutral to friendly to support a basic level of trust ^c	Both
3 <i>Is the appearance of the robot meaningful?</i>	All design features of the robot are linked to functions and do not create false expectations in users ^d	Both
4 <i>Is the degree of humanlikeness appropriate to the task, meaningful, and adapted to the user group?</i>	Human-likeness of robot behavior is reasonably balanced and fits the users/operators, task, and situation ^e	Both
Reasonable and understandable autonomy		
1 <i>Does the robot have an appropriate degree of autonomy/decision-making power?</i>	It should be considered where it makes sense to grant robots autonomy and decision-making power. Higher acceptance by users can be expected if a monitoring/override function is implemented. The user/operator remains responsible for robot use ^{f,g}	Both
2 <i>Can the scope and range of the robot's autonomous task execution be coordinated with the user/operator?</i>	The robot only performs actions to be performed within the scope of the task assigned to it. The robot requires permission for each function/task from the user/operator before the robot can perform it ^f	Both
3 <i>Does the robot signal autonomy?</i>	If the robot is working in a fully autonomous mode, this is communicated in the interface (e.g., icon, lights) ^h ; depending on the area of application and task this option can be implemented in a way that it can be deselected	Both
4 <i>Does the robot act autonomously to an appropriate degree within the scope of the tasks assigned to it and does it communicate efficiently?</i>	If tasks are given to the robot, they are performed as effectively and efficiently as possible. This includes the reduction of task-related queries and information to a user-adaptive minimum ⁱ	Both
5 <i>Is an optional successive increase in the level of autonomy implemented for standard tasks?</i>	If desired by the users the robot can become successively more autonomous in the execution of standard tasks by learning from past interactions and reducing the extent of queries ⁱ	Private
6 <i>Is the level of autonomy and proactivity of task execution adaptable to the task, area of use, and user group?</i>	The level of autonomy and proactivity can be adjusted to the task according to user preferences ^j	Private
Trustworthy interaction		
1 <i>Does the robot support a calibrated level of trust?</i>	The robot and its interaction are designed in a way to support dynamic and situation-specific formation of a calibrated level of trust for each subtask. For this, the design recommendations of the category "Transparent communication" are fundamental prerequisites ^k	Both
2 <i>Does the robot adapt immediately to user input?</i>	The robot adapts its task execution directly after receiving an input. The task scope of the robot can be extended and restricted if necessary ^f	Both
3 <i>Are the robot's current reliability and probability of error communicated?</i>	The robot's design and interaction mechanisms allow for maximum reliability in task execution. The robot communicates errors and limitations to its reliability dynamically and in a timely manner ^{l,m,n}	Both
4 <i>Does the robot coordinate the task execution with users to an appropriate degree?</i>	The robot reassures itself to an appropriate degree with the users prior to action execution ^g	Both
5 <i>Is a user-adaptive level of reassurance by the robot implemented?</i>	The extent to which the robot coordinates its task execution with the user/operator can be configured by the user (e.g., all actions vs. unusual actions)	Both
6 <i>Does the robot use intuitive interaction mechanisms resembling social, interpersonal communication?</i>	The robot uses intuitive mechanisms of interpersonal communication appropriately (without excessive anthropomorphising or an inappropriate degree of attachment) ^o	Both
7 <i>Are deviations from expected objects or situations communicated?</i>	If an object or task anomaly is detected by the robot, this is communicated to the user and clarification is attempted	Private

Table 1 (Continued)

Questions on Design	Design Recommendation	Use Context
Transparent communication		
1 Does the robot have the ability to show what movements it will perform (both locomotion and manipulator)?	The robot communicates the planned path and/or occupied movement space (e.g., projection on the floor) ^p	Both
2 Does the robot have the ability to communicate its current state and plans?	Robot states (e.g. battery status, errors), plans (e.g. schedule, remaining sub tasks) and degree of autonomy are communicated transparently and can be checked at any time ^h	Both
3 Does the robot show whether and which people and objects it has detected?	The robot makes the object/person recognition transparent and comprehensible for users, and thereby allowing for the identification of errors in the person recognition ^q	Both
4 Is the robot's communication modality adapted to the environment?	The interaction of the robot is adapted to the task environment and is (in the optimal case) implemented in a multimodal design to ensure universal usability. ^f If applicable in private households, a voice dialog is recommended In public spaces and noisy environments, warning sounds and visual interaction are often more beneficial	Both Private Public
5 Are system boundaries transparent and comprehensible?	The robot communicates situations for which system limitations exist, explains their consequences and warns about possible errors ^f	Both
6 Is the robot able to draw attention to itself?	The robot's interaction concept is designed in a way to allow to attract attention to the robot when necessary ^p	Both
7 Does the robot communicate unnecessary information?	The robot by default limits the communicated information to what is necessary for task execution, unless its task is communication ^s	Both
8 Does the robot give feedback on faulty operation/mistreatment?	The robot provides feedback when operation by the users is not in accordance with the task or could cause damage to the robot's hardware or software ^t	Both
9 Is the robot equipped with a possibility of announcing its entry into a room?	To prevent startling by sudden, unexpected entry, the robot signals its entry beforehand. To avoid excessive disturbance, this option can be switched off in accordance with the situation	Both
10 Is there an adaptive level of coordination with users?	Users can adjust the robot's frequency of the coordination with the robot and the autonomy level for individual tasks ^j	Private
11 Does the robot demonstrate critical tasks before it first executes these?	The robot demonstrates critical tasks to users first, before final permission to perform these tasks in the future is given (e.g., demo mode or tutorial)	Private
Appropriate social behavior		
1 Does the robot adapt to the environment and its interaction partners when performing its tasks?	When people enter the robot's movement space, the robot adjusts its movement sequences in a way that people can move around undisturbed ^{h,u}	Both
2 Is the robot as inconspicuous, discreet, and non-disruptive as possible?	The robot performs its task discreetly, unobtrusively and with a minimum level of interference. Both noise generation of the task and communication are reduced to the minimum required for the task execution ^v	Both
3 Does the robot have an suitable and culturally appropriate level of politeness?	The robot adheres to social norms and communicates in a culturally compliant, friendly and polite manner that at the same time allows efficient task completion ^w	Both
4 Does the robot respect the personal distance zone?	The robot does not violate the human's personal space (a minimum distance of 1.5 m is recommended). Physical contact with humans is acceptable if it is relevant to the task and permission has been granted by the user. ^{x,y} The robot does not violate the human's personal space. A minimum distance of 1.5 m is recommended ^x	Private Public
5 Does the robot react appropriately to inattentive persons?	The robot recognizes when people in its environment are inattentive and adjusts its movements and actions accordingly ^s	Both
6 Does the robot assert itself only within defined limits (e.g. emergencies)?	The situations in which assertive behaviour by the robot is allowed are to be coordinated with the users. It should be possible for the user to stop the assertive action at any time. ^{z,aa}	Both

Table 1 (Continued)

Questions on Design	Design Recommendation	Use Context
Perceptible data protection and protection of privacy		
1 <i>Have the data protection regulations/laws of the respective country and the corresponding situation at the robot's operating location been considered in design?</i>	Depending on the applicable law or regulation, the robot requires explicit consent for the use of cameras/microphones and the further processing of the collected data. The implemented data protection measures are communicated transparently to the users ^j	Both
2 <i>The processing and storage of data is limited only to the personal data needed for the robot to perform the task?</i>	The robot does not process and store any specific identification features of the surrounding persons beyond those required for task completion	Both
3 <i>Does the robot provide transparency as to when and what personal data is collected for what purpose and under what conditions it is deleted?</i>	Data recording by the robot is recognizable to users and the scope and extent is comprehensible. If applicable, the purpose of data collection is communicated ^f Appropriate procedures for (automated or user-initiated) data deletion are implemented and communicated transparently ^g	Both
4 <i>Is personal identification by the robot without user consent avoided?</i>	The robot protects peoples' privacy and personally identifies people only after their consent ^f The robot protects privacy and avoids personal identification. If the robot needs to distinguish users, it does so in pseudonymous form whenever possible. ^{ab}	Private Public
5 <i>Is the data transmission encrypted in a comprehensible way?</i>	All data transferred between the robot and other parties is encrypted in a way that guarantees data security. This is communicated to the users in a comprehensible and transparent manner ^{ab}	Both
6 <i>Is the robot secured against hacking and misuse?</i>	The robot's hardware and software are secured against unauthorized access (e.g. hacking, illegitimate access to user data). The users are reassured in this respect (actively or on request) ^g	Both
7 <i>Does the robot respect privacy in the home?</i>	The possibility of coordinating the area of use and reducing robot activity to the agreed rooms and task areas can increase acceptance. For particularly private rooms (e.g., bathrooms and bedrooms), the robot has an individualizable time- and situation-based coordination concept. ^{ac}	Private
Security & subjective feeling of safety		
1 <i>Can the robot be switched off at any time?</i>	The robot has a clearly marked and easily accessible emergency stop switch ^{ad,ae,af}	Both
2 <i>Is the physical force of the robot limited to a maximum level that does not exceed the maximum necessary for the task?</i>	Functionality, force application and speeds are limited to the maximum required for successful task completion. Accordingly, in this regard, realistic expectations of users are fostered by appearance and instruction ^{ad,af}	Both
3 <i>Are the robot and its components (e.g. manipulators) designed to be minimally hazardous?</i>	The robot is designed and built e.g. as a lightweight construction, without clamping points and with soft/flexible surfaces ^{ad,ae,af}	Both
5 <i>Does the robot handle sensitive and dangerous objects with care?</i>	The robot recognizes critical and dangerous objects and interacts with them with limited force and speed and without endangering its environment ^{ad,af}	Both
6 <i>Does the robot avoid collisions and warns of them in a timely manner?</i>	The robot is equipped with sensor technology that monitors distances to people in the immediate environment and has automatic emergency braking as well as a perceivable, preventive collision avoidance system ^{h,ag,ah}	Both
7 <i>Does the robot keep a safe distance to people?</i>	The robot detects persons and acts with a perceivable minimum distance ^{ad,ae}	Both
Subjectively normative robot behavior		
1 <i>Does the robot respect the dignity and rights of humans?</i>	Actions of the robot do not affect human rights and respect human dignity ^{f,g}	Both
2 <i>Does the robot coordinate moral decisions with a human?</i>	To increase acceptance and trustworthiness of the robot, decisions involving a moral component are not made by the robot, but by a human ^g	Both
3 <i>Does the robot follow generally applicable legislation?</i>	Robot behavior and robot interaction do not cross any legal boundaries ^{f,g}	Both
4 <i>Is discrimination of groups of people by the robot ruled out?</i>	The robot does not discriminate (e.g., based on gender, age, ethnicity) ^g . User-adaptive interaction concepts build on factual requirements of users and not on stereotyped assumptions	Both
5 <i>Does the robot allow for universal usability and inclusion of vulnerable and impaired people in the interaction?</i>	The interaction of the robot is internationally understandable and includes people with disabilities, for example, through multimodality. The robot can adapt its behavior to the needs of vulnerable and impaired persons. ^{ai}	Public
6 <i>Does the robot help to provide relief for humans?</i>	The robot takes over monotonous, repetitive or stressful tasks. The robot is not used in competition with humans ^d	Public

Table 1 (Continued)

	Questions on Design	Design Recommendation	Use Context
7	<i>Does the implementation of robots allow for complete tasks for humans?</i>	The robot takes over tasks in the socio-technical system in a way that allows the design of complete tasks for humans as well as an experience of competence and self-efficacy ^g . As far as possible, the human does not come into the position of a mere supervisor of the task execution of the robot	Public
8	<i>Are the activities of the robot retrospectively reconstructible?</i>	The robot has a black box that keeps an activity log to reconstruct task execution (e.g., after accidents); this recording is done in accordance with data protection regulations ^g	Both
9	<i>Does the robot not replace interpersonal, social contacts, but complement these?</i>	The robot does not simulate a human being. It encourages users to have real social contact with other people ^{i,aj}	Private
10	<i>Does the robot avoid emotional attachment of the users beyond a healthy level?</i>	The robot is designed to prevent excessive emotional attachment of the users to it. In this regard, decisions regarding humanization, robot personality, and communication style of the robot are made in an informed manner ^{ah}	Private

^aKreis (2018)^bSong and Luximon (2020)^cHiroi and Ito (2008)^dHaring et al. (2018)^eGoetz et al. (2003)^fNevejans (2016)^gEuropean Commission (2019)^hElkmann (2013)ⁱGoodrich and Schultz (2007)^jGelin (2017)^kDe Visser et al. (2020)^lChen et al. (2018)^mBeller et al. (2013)ⁿEder et al. (2014)^oKirchner et al. (2015)^pJanowski et al. (2018)^qBansal et al. (2014)^rKardos et al. (2018)^sDevin and Alami (2016)^tDirective 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) (2006)^uSchenk and Elkmann (2012)^vBendel (2021)^wSalem et al. (2014)^xDautenhahn (2007)^yRuijten and Cuijpers (2020)^zBabel et al. (2022a)^{aa}Babel et al. (2022b)^{ab}Regulation (EU) 2016/679 of The European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016)^{ac}Lutz et al. (2019)^{ad}Deutsches Institut für Normung e. V. (2017), DIN ISO/TS 15066:2017-04^{ae}Deutsches Institut für Normung e. V. (2012), DIN EN ISO 10218-1:2012-01^{af}Deutsches Institut für Normung e. V. (2014), DIN EN ISO 13482:2014-11^{ag}Jacobs (2013)^{ah}Rosenstrauch and Kruger (2017)^{ai}Kildal et al. (2019)^{aj}Kornwachs (2019)

Table 2 The Trustworthy and Acceptable HRI Checklist (TA-HRI)—German Version

Fragen zum Design	Gestaltungsempfehlung	Einsatzort
Vertrauenswürdige Erscheinungsbild		
1 <i>Ist der Roboter ohne zwingenden Grund menschenähnlich gestaltet?</i>	Menschenähnlichkeit des Roboters ist zweckgebunden und wird in einem angemessenen Grad umgesetzt. Der Roboter bleibt als Maschine erkennbar ^a	Beide
2 <i>Wirkt der Roboter bedrohlich?</i>	Der Roboter ist in seinen Abmaßen auf eine optimale Aufgabenerledigung bei minimaler Bedrohlichkeit ausgelegt. ^b Das Gesicht des Roboters ist neutral bis freundlich gestaltet, um ein grundlegendes Maß an Vertrauen zu begünstigen ^c	Beide
3 <i>Ist die Erscheinung des Roboters zweckgebunden?</i>	Alle Gestaltungsmerkmale des Roboters sind mit Funktionen verknüpft und erwecken keine falschen Erwartungen bei den Nutzenden ^d	Beide
4 <i>Ist der Grad an Menschlichkeit aufgabenangemessen, sinnvoll und an die Nutzergruppe angepasst?</i>	Menschenähnliches Verhalten ist in einem sinnvollen Maß im Einklang mit Nutzer:in/Operator:in, Aufgabe und Situation umgesetzt ^e	Beide
Angemessene und nachvollziehbare Autonomie		
1 <i>Besitzt der Roboter ein angemessenes Maß an Autonomie/Entscheidungsgewalt?</i>	Es sollte abgewogen werden wo es sinnvoll ist, Robotern Autonomie und Entscheidungsgewalt einzuräumen. Es ist eine höhere Akzeptanz bei Nutzer:innen zu erwarten, wenn eine menschliche Überwachungs-/Override-Funktion implementiert ist. Der/die Nutzer:in/Operator:in bleibt verantwortlich für Robotereinsatz. ^{f,g}	Beide
2 <i>Kann Umfang und Bereich der autonomen Aufgabenausführung des Roboters mit dem/der Nutzer:in/Operator:in abgestimmt werden?</i>	Der Roboter führt nur Aktionen aus, die im Rahmen der ihm übertragenen Aufgabe ausgeführt werden müssen. Der Roboter benötigt die Erlaubnis für einzelne Funktionen durch den/die Nutzer:in/Operator:in, bevor der Roboter diese ausführen kann ^f	Beide
3 <i>Signalisiert der Roboter Autonomie?</i>	Wenn der Roboter im vollautonomen Zustand arbeitet, wird dies im Interface kommuniziert (z. B. Symbol, Leuchten) ^h ; diese Option kann je nach Einsatzbereich und Aufgabe abwählbar implementiert werden	Beide
4 <i>Handelt der Roboter im Rahmen der ihm übertragenen Aufgaben angemessen eigenständig und kommuniziert er effizient?</i>	Werden dem Roboter Aufgaben erteilt, werden diese möglichst effektiv und effizient erfüllt. Dies beinhaltet die Reduktion von aufgabenbezogenen Rückfragen und Informationen auf ein nutzeradaptives Minimum ⁱ	Beide
5 <i>Ist eine optionale sukzessive Zunahme des Autonomiegrades bei Standardaufgaben implementiert?</i>	Der Roboter kann (wenn von den Nutzenden erwünscht) in der Ausführung von Standardfunktionen sukzessiv autonomer werden, indem er aus vergangenen Interaktionen lernt und das Ausmaß an Rücksprachen reduziert ⁱ	Privat
6 <i>Ist das Ausmaß an Autonomie und Proaktivität der Aufgabenausführung an Aufgabe, Einsatzbereich und Nutzergruppe anpassbar?</i>	Grad an Autonomie und Proaktivität können nach den Wünschen von Nutzenden entsprechend der Aufgabenangemessenheit angepasst werden ^j	Privat
Vertrauenswürdige Interaktion		
1 <i>Begünstigt der Roboter ein kalibriertes Maß an Vertrauen?</i>	Der Roboter und seine Interaktion sind auf eine Weise gestaltet, dass dynamisch und situational für jede Teilaufgabe ein angemessenes Maß an Vertrauen erzeugt wird. Hierfür sind die Gestaltungsempfehlungen in der Kategorie „Transparente Kommunikation“ grundlegende Voraussetzung ^k	Beide
2 <i>Passt sich der Roboter unmittelbar an Nutzereingaben an?</i>	Der Roboter passt seine Aufgabenausführung direkt nach Erhalt einer Eingabe an. Der Aufgabenumfang des Roboters ist ggf. erweiter- und einschränkbar ^f	Beide
3 <i>Werden die aktuelle Zuverlässigkeit und Fehlerwahrscheinlichkeit des Roboters kommuniziert?</i>	Die Gestaltung und Interaktionsmechanismen des Roboters erlauben ein Maximum an Zuverlässigkeit in der Aufgabenausführung. Der Roboter kommuniziert Fehler und Einschränkungen in der Zuverlässigkeit dynamisch und unmittelbar. ^{l,m,n}	Beide
4 <i>Spricht der Roboter die Aufgabenausführung in angemessenem Maß mit den Nutzenden ab?</i>	Der Roboter versichert sich vor der Handlungsdurchführung in einem angemessenen Maß bei den Nutzenden zurück ^g	Beide
5 <i>Ist ein nutzeradaptives Ausmaß an Rückversicherungen durch den Roboter implementiert?</i>	Das Ausmaß, in dem der Roboter seine Aufgabenerledigung mit dem/der Nutzer:in/Operator:in abstimmt, kann durch den/die Nutzer:in konfiguriert werden (z. B. alle vs. ungewöhnliche Aktionen)	Beide
6 <i>Nutzt der Roboter intuitive Interaktionsmechanismen, die an soziale, zwischenmenschliche Kommunikation angelehnt sind?</i>	Der Roboter bedient sich intuitiver Mechanismen zwischenmenschlicher Kommunikation in angemessener Weise (ohne übermäßige Vermenschlichung und einem zu hohen Maß an Bindung) ^o	Beide
7 <i>Werden Abweichungen von erwarteten Menschen oder Objekten kommuniziert?</i>	Wenn eine Objekt- oder Aufgabenanomalie vom Roboter festgestellt wird, wird dies dem/der Nutzer:in mitgeteilt und es wird eine Abklärung angestrebt	Privat

Table 2 (Continued)

Fragen zum Design	Gestaltungsempfehlung	Einsatzort
Transparente Kommunikation		
1 <i>Hat der Roboter die Möglichkeit zu zeigen, welche Bewegungen er ausführen wird (sowohl Fortbewegung als auch Manipulation)?</i>	Der Roboter kommuniziert den geplanten Weg und/oder vereinnahmten Bewegungsraum (z. B. Projektion auf den Boden) ^p	Beide
2 <i>Hat der Roboter die Möglichkeit, seinen aktuellen Zustand und seine Pläne zu kommunizieren?</i>	Zustände (z. B. Batteriestatus, Fehler), Pläne (z. B. Zeitplan, verbleibende Teilschritte) und Autonomiegrad des Roboters werden transparent kommuniziert und sind jederzeit überprüfbar ^b	Beide
3 <i>Zeigt der Roboter, ob und wenn ja welche Personen und Objekte er erkannt hat?</i>	Der Roboter macht die Objekt-/Personenerkennung für Nutzende transparent und nachvollziehbar, wodurch Fehler in der Personenerkennung identifiziert werden können ^q	Beide
4 <i>Ist die Kommunikationsmodalität des Roboters an die Umgebung angepasst?</i>	Die Interaktionsform des Roboters ist an die Umwelt der Aufgabenerledigung angepasst und erfolgt bestenfalls zur Sicherstellung einer universellen Usability multimodal ^r Im Privathaushalt ist ggf. ein Sprachdialog empfehlenswert Im öffentlichen Raum und lauten Umgebungen sind häufig Warntöne und visuelle Interaktion vorteilhafter	Beide Privat Öffentl.
5 <i>Sind Systemgrenzen transparent und nachvollziehbar?</i>	Der Roboter kommuniziert die Situationen, in denen Systemgrenzen vorliegen, erläutert deren Konsequenzen und warnt vor möglichen Fehlern ^f	Beide
6 <i>Ist der Roboter in der Lage, auf sich aufmerksam zu machen?</i>	Das Interaktionskonzept des Roboters ist so gestaltet, dass es die Aufmerksamkeit auf den Roboter gelenkt werden kann, wenn dies notwendig ist ^p	Beide
7 <i>Kommuniziert der Roboter unnötige Informationen?</i>	Der Roboter beschränkt die standardmäßig kommunizierten Informationen auf das zur Aufgabenausführung notwendige Maß, es sei denn seine Aufgabe besteht in Kommunikation ^s	Beide
8 <i>Gibt der Roboter Rückmeldung zu Fehlbedienung/Misshandlung?</i>	Der Roboter gibt Feedback, wenn Bedienung durch die Nutzenden nicht im Einklang mit der Aufgabe steht oder Schaden an der Hard- oder Software des Roboters herbeiführen (könnte) ^t	Beide
9 <i>Hat der Roboter eine Möglichkeit zur Ankündigung seines Eintretens in einen Raum?</i>	Um Erschrecken durch plötzliches, unerwartetes Eintreten zu verhindern, signalisiert der Roboter sein Eintreten vorher. Um übermäßige Störungen zu vermeiden, kann diese Option situationsangemessen abschaltbar sein	Beide
10 <i>Besteht ein adaptives Ausmaß an Abstimmung mit Nutzenden?</i>	Nutzende können die Abstimmungshäufigkeit und den Autonomiegrad des Roboters für einzelne Aufgaben anpassen ^l	Privat
11 <i>Demonstriert der Roboter kritische Aufgaben vor der ersten Ausführung?</i>	Der Roboter demonstriert kritische Aufgaben den Nutzenden zunächst, bevor die endgültige Erlaubnis zur künftigen Ausführung dieser Aufgaben erteilt wird (z. B. Demomodus oder Tutorial)	Privat
Angemessenes Sozialverhalten		
1 <i>Passt sich der Roboter bei seiner Aufgabenerfüllung adaptiv der Umgebung und seinen Interaktionspartner:innen an?</i>	Wenn Menschen in den Bewegungsraum des Roboters treten, passt dieser seine Bewegungsabläufe so an, dass Menschen sich ungestört bewegen können ^{h,ai}	Beide
2 <i>Ist der Roboter möglichst unauffällig, diskret und wenig störend?</i>	Der Roboter erledigt seine Aufgabe diskret, unauffällig und störungsfrei. Sowohl Geräusentwicklung der Arbeitsaufgabe als auch Kommunikation wird auf das für die Aufgabenausführung benötigte Minimum reduziert ^v	Beide
3 <i>Verfügt der Roboter über ein angemessenes und kulturkonformes Maß an Höflichkeit?</i>	Der Roboter hält soziale Normen ein und kommuniziert in kulturell konformer Weise freundlich und höflich in dem Ausmaß, das zugleich eine effiziente Aufgabenerledigung ermöglicht ^w	Beide
4 <i>Respektiert der Roboter die persönliche Distanzzone?</i>	Der Roboter verletzt den persönlichen Bereich des Menschen nicht (hierbei ist ein Mindestabstand von 1,5 m empfehlenswert). Physischer Kontakt mit dem Menschen ist akzeptabel, wenn er aufgabenrelevant ist und von dem/der Nutzenden eine Erlaubnis erteilt wurde ^{x,y} Der Roboter verletzt den persönlichen Bereich des Menschen nicht. Hierbei ist ein Mindestabstand von 1,5 m empfehlenswert ^x	Privat Öffentl.
5 <i>Reagiert der Roboter angemessen auf unachtsame Personen?</i>	Der Roboter erkennt, wenn Personen in seiner Umgebung unaufmerksam sind, und passt seine Bewegungen und Handlungen dementsprechend an ^s	Beide
6 <i>Setzt sich der Roboter nur innerhalb definierter Grenzen (z. B. Notfällen) durch?</i>	Die Situationen, in denen eine Durchsetzung des Roboters erlaubt sind, werden mit den Nutzenden abgestimmt. Die Durchsetzung sollte jederzeit durch den/die Nutzer:in abgebrochen werden können ^{z,aa}	Beide

Table 2 (Continued)

Fragen zum Design	Gestaltungsempfehlung	Einsatzort
Wahrnehmbarer Datenschutz und Achtung der Privatsphäre		
1 <i>Sind die datenschutzrechtlichen Vorschriften/Gesetze des jeweiligen Landes und der entsprechen Situation am Einsatzort des Roboters berücksichtigt?</i>	Je nach geltendem Gesetz bzw. Vorschrift benötigt der Roboter die eindeutige Zustimmung für die Nutzung von Kameras/Mikrofonen und der Weiterverarbeitung der erhobenen Daten. Die implementierten Datenschutzmaßnahmen werden den Nutzenden transparent kommuniziert ^j	Beide
2 <i>Die Verarbeitung und Speicherung von Daten beschränkt sich nur auf die für die Aufgabenerledigung des Roboters benötigten persönliche Daten?</i>	Der Roboter verarbeitet und speichert keine spezifischen Identifikationsmerkmale der umgebenden Personen über diejenigen hinaus, die zur Aufgabenerledigung notwendig sind	Beide
3 <i>Macht der Roboter transparent, wann und welche personenbezogenen Daten zu welchem Zweck gesammelt werden und unter welchen Bedingungen deren Löschung erfolgt?</i>	Datenaufzeichnung durch den Roboter ist für Nutzende erkennbar und der Umfang nachvollziehbar. Ggf. wird der Zweck der Datenerfassung kommuniziert ^f Entsprechende Prozeduren zur (automatisierten oder nutzerinitiierten) Datenlöschung sind implementiert und werden transparent kommuniziert ^g	Beide
4 <i>Wird eine persönliche Identifikation durch den Roboter ohne Einwilligung der Nutzenden vermieden?</i>	Der Roboter wahrt den Schutz der Privatsphäre und identifiziert Personen nur nach Zustimmung persönlich ^f	Privat
5 <i>Ist die Datenübertragung nachvollziehbar verschlüsselt?</i>	Der Roboter wahrt den Schutz der Privatsphäre und eine persönliche Identifikation wird vermieden. Sollte der Roboter Nutzende unterscheiden müssen, so tut er dies möglichst in pseudonymisierter Form ^{ab}	Öffentl.
6 <i>Ist die Datenübertragung nachvollziehbar verschlüsselt?</i>	Alle Daten, die zwischen dem Roboter und anderen Stellen übertragen werden, werden auf eine Weise verschlüsselt, die Datensicherheit gewährleistet. Dies wird den Nutzer:innen nachvollziehbar und transparent kommuniziert ^{ab}	Beide
7 <i>Ist der Roboter gegen Hacking und Missbrauch abgesichert?</i>	Die Hard- und Software des Roboters ist gegen unautorisierten Zugriff abgesichert (z. B. Hacking, unberechtigter Zugriff auf Nutzerdaten). Die Nutzer:innen werden in dieser Hinsicht (aktiv oder auf Nachfrage) rückversichert ^g	Beide
8 <i>Respektiert der Roboter die Privatsphäre im Haushalt?</i>	Eine Möglichkeit zur Abstimmung des Einsatzbereiches und die Reduktion der Roboteraktivität auf die vereinbarten Räumen und Aufgabenbereichen kann die Akzeptanz erhöhen. Für besonders private Räume (z. B. Bade- und Schlafzimmer) verfügt der Roboter über ein individualisierbares, zeit- und situationsbasiertes Abstimmungskonzept ^{ac}	Privat
Sicherheit & Subjektives Sicherheitsgefühl		
1 <i>Ist der Roboter jederzeit abschaltbar?</i>	Der Roboter besitzt einen deutlich markierten und leicht erreichbaren Not-Ausschalter ^{ad,ae,af}	Beide
2 <i>Ist die physikalische Wirkmächtigkeit des Roboters auf das aufgabenangemessene Maß beschränkt?</i>	Funktionalität, Krafteinwirkung und Geschwindigkeiten werden bei der Roboterentwicklung auf das für die Aufgabenerledigung erforderliche Maß limitiert. Entsprechend werden in dieser Hinsicht bei den Nutzenden realistische Erwartungen durch Erscheinungsbild und Instruktion begünstigt ^{ad,af}	Beide
3 <i>Sind der Roboter und seine Anbauteile (z. B. Manipulatoren) in ihrer Bauweise minimal gefährdend gestaltet?</i>	Der Roboter ist bspw. als Leichtbau, ohne Klemmstellen und mit weichen/nachgiebigen Oberflächen konstruiert und gebaut ^{ad,ae,af}	Beide
5 <i>Geht der Roboter mit empfindlichen und gefährlichen Gegenständen vorsichtig um?</i>	Der Roboter erkennt kritische und gefährliche Objekte und interagiert mit ihnen mit limitierter Krafteinwirkung und Geschwindigkeit und unter Ausschluss von Gefährdung des Umfelds ^{ad,af}	Beide
6 <i>Vermeidet der Roboter Zusammenstöße und warnt rechtzeitig vor diesen?</i>	Der Roboter besitzt Sensorik, die Abstand zu Personen im unmittelbaren Umfeld überwacht und verfügt über eine automatische Notbremsung und eine wahrnehmbare, präventive Kollisionsvermeidung ^{h,ag,ah}	Beide
7 <i>Hält der Roboter einen Sicherheitsabstand zu Menschen ein?</i>	Der Roboter erkennt Personen und agiert mit einem wahrnehmbaren Mindestabstand ^{ad,ae}	Beide

Table 2 (Continued)

Fragen zum Design	Gestaltungsempfehlung	Einsatzort
Subjektiv normatives Roboterverhalten		
1 <i>Respektiert der Roboter die Würde und Rechte von Menschen?</i>	Handlungen des Roboters beeinträchtigen nicht die Menschenrechte und er beachtet die Menschenwürde ^{f,g}	Beide
2 <i>Stimmt der Roboter moralische Entscheidungen mit einem Menschen ab?</i>	Um Akzeptanz und Vertrauenswürdigkeit des Roboters zu erhöhen, werden Entscheidungen, die eine moralische Komponente beinhalten, nicht vom Roboter, sondern vom Menschen getroffen ^g	Beide
3 <i>Befolgt der Roboter allgemein geltende Gesetzgebung?</i>	Das Roboterverhalten und die Roboterinteraktion überschreiten keine Gesetzesgrenzen ^{f,g}	Beide
4 <i>Ist eine Diskriminierung von Menschengruppen durch den Roboter ausgeschlossen?</i>	Der Roboter diskriminiert nicht (z. B. aufgrund von Geschlecht, Alter, ethnischer Herkunft) ^g . Nutzeradaptive Interaktionskonzepte bauen auf faktischen Anforderungen von Nutzenden und nicht auf stereotypisierten Annahmen auf	Beide
5 <i>Beachtet der Roboter universelle Usability und Inklusion von verletzlichen und eingeschränkten Personen in der Interaktion?</i>	Die Interaktion des Roboters ist international verständlich und inkludiert beispielsweise durch Multimodalität Personen mit Handicaps. Der Roboter kann sein Verhalten adaptiv an Bedürfnisse von besonders verletzlichen und eingeschränkten Personen anpassen ^{ai}	Öffentl.
6 <i>Dient der Roboter zur Entlastung des Menschen?</i>	Der Roboter übernimmt stupide, repetitive oder stressige Aufgaben. Der Roboter wird nicht in Konkurrenz zum Menschen eingesetzt ^a	Öffentl.
7 <i>Ermöglicht der Robotereinsatz vollständige Aufgaben für den Menschen?</i>	Der Roboter übernimmt Aufgaben im sozio-technischen System auf eine Weise, welche eine Gestaltung vollständiger Aufgaben für die Menschen sowie ein Erleben von Kompetenz und Selbstwirksamkeit erlaubt ^g . Der Mensch kommt möglichst nicht in die Position eines/einer reinen Überwacher:in der Aufgabenausführung des Roboters	Öffentl.

Table 2 (Continued)

	Fragen zum Design	Gestaltungsempfehlung	Einsatzort
8	<i>Sind die Aktivitäten des Roboters rückwirkend nachvollziehbar?</i>	Der Roboter besitzt eine Blackbox, die ein Aktivitätsprotokoll führt, um die Aufgabenausführung zu rekonstruieren (z. B. nach Unfällen); diese Aufzeichnung erfolgt im Einklang mit Datenschutzrichtlinien ^g	Beide
9	<i>Ersetzt der Roboter keine zwischenmenschlichen, sozialen Kontakte, sondern ergänzt diese?</i>	Der Roboter simuliert keinen Menschen und regt Nutzende zu realem sozialem Kontakt mit anderen Menschen an ^{i,aj}	Privat
10	<i>Vermeidet der Roboter eine emotionale Bindung der Nutzenden über ein gesundes Maß hinaus?</i>	Der Roboter ist so gestaltet, dass eine übermäßige emotionale Bindung der Nutzenden an den Roboter verhindert wird. In dieser Hinsicht werden Entscheidungen bezüglich Vermenschlichung, Roboterpersönlichkeit und Kommunikationsstil des Roboters informiert getroffen ^{aj}	Privat

^aKreis (2018)^bSong und Luximon (2020)^cHiroi und Ito (2008)^dHaring et al. (2018)^eGoetz et al. (2003)^fNevejans (2016)^gEuropean Commission (2019)^hElkmann (2013)ⁱGoodrich und Schultz (2007)^jGelin (2017)^kDe Visser et al. (2020)^lChen et al. (2018)^mBeller et al. (2013)ⁿEder et al. (2014)^oKirchner et al. (2015)^pJanowski et al. (2018)^qBansal et al. (2014)^rKardos et al. (2018)^sDevin und Alami (2016)^tRichtlinie 2006/42/EG des Europäischen Parlaments und des Rates vom 17. Mai 2006 über Maschinen und zur Änderung der Richtlinie 95/16/EG (Neufassung) (2006)^uSchenk und Elkmann (2012)^vBendel (2021)^wSalem et al. (2014)^xDautenhahn (2007)^yRuijten und Cuijpers (2020)^zBabel et al. (2022a)^{aa}Babel et al. (2022b)^{ab}Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) (2016)^{ac}Lutz et al. (2019)^{ad}Deutsches Institut für Normung e. V. (2017), DIN ISO/TS 15066:2017-04^{ae}Deutsches Institut für Normung e. V. (2012), DIN EN ISO 10218-1:2012-01^{af}Deutsches Institut für Normung e. V. (2014), DIN EN ISO 13482:2014-11^{ag}Jacobs (2013)^{ah}Rosenstrauch und Kruger (2017)^{ai}Kildal et al. (2019)^{aj}Kornwachs (2019)

Funding This research was funded by the German Federal Ministry of Education and Research in the RobotKoop project (grant number 16SV7967).

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrams, A. M., Dautzenberg, P. S., Jakobowsky, C., Ladwig, S., & Rosenthal-von der Pütten, A. M. (2021). A theoretical and empirical reflection on technology acceptance models for autonomous delivery robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 272–280).
- Alaiad, A., & Zhou, L. (2014). The determinants of home healthcare robots adoption: an empirical investigation. *International Journal of Medical Informatics*, 83(11), 825–840. <https://doi.org/10.1016/j.ijmedinf.2014.07.003>.
- Alarcon, G. M., Capiola, A., & Pfahler, M. D. (2021). The role of human personality on trust in human-robot interaction. In *Trust in human-robot interaction* (pp. 159–178). Academic Press.
- Alonso, V., & de la Puente, P. (2018). System transparency in shared autonomy: a mini review. *Frontiers in Neurobotics*, 12, 83. <https://doi.org/10.3389/fnbot.2018.00083>.
- Arndt, S. (2011). *Evaluierung der Akzeptanz von Fahrerassistenzsystemen*. Wiesbaden: VS.
- Babel, F., Kraus, J., Müller, L., Kraus, M., Wagner, N., Minker, W., & Baumann, M. (2021). Small talk with a robot? The impact of dialog content, talk initiative, and gaze behavior of a social robot on trust, acceptance, and proximity. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-020-00730-0>.
- Babel, F., Hock, P., Kraus, J., & Baumann, M. (2022a). It Will not take long! Longitudinal effects of robot conflict resolution strategies on compliance, acceptance and trust. In *Proceedings of the 2022 ACM/IEEE international conference on human-robot interaction* (pp. 225–235).
- Babel, F., Vogt, A., Hock, P., Kraus, J., Angerer, F., Seufert, T., & Baumann, M. (2022b). Step aside! VR-based evaluation of adaptive robot conflict resolution strategies for domestic service robots. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-021-00858-7>.
- Bansal, A., Farhadi, A., & Parikh, D. (2014). Towards transparent systems: semantic characterization of failure modes. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – ECCV 2014* (Vol. 8694, pp. 366–381). Springer. https://doi.org/10.1007/978-3-319-10599-4_24.
- Barnes, J., FakhrHosseini, M., Jeon, M., Park, C.-H., & Howard, A. (2017). The influence of robot design on acceptance of social robots. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. Maison Glad Jeju, June 28–July 1, 2017. (pp. 51–55). Piscataway: IEEE. <https://doi.org/10.1109/URAI.2017.7992883>.
- Bartneck, C., & Forlizzi, J. (2004). *A design-centred framework for social human-robot interaction*. Proceedings of the Ro-Man2004, Kurashiki. (pp. 591–594). <https://doi.org/10.1109/ROMAN.2004.1374827>.
- Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust, and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour*, 18, 47–57. <https://doi.org/10.1016/j.trf.2012.12.006>.
- Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, 75–84. <https://doi.org/10.1016/j.trf.2015.10.005>.
- Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver–automation interaction: an approach using automation uncertainty. *Human Factors*, 55(6), 1130–1141. <https://doi.org/10.1177/0018720813482327>.
- Bendel, O. (2021). *Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen*. Springer. <https://doi.org/10.1007/978-3-658-31114-8>.
- Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3), 259–282. <https://doi.org/10.1080/1463922X.2017.1315750>.
- Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in Human Performance and Cognitive Engineering Research*, 2, 1–12. [https://doi.org/10.1016/S1479-3601\(02\)02003-9](https://doi.org/10.1016/S1479-3601(02)02003-9).
- Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5), 455–496. <https://doi.org/10.1007/s11257-008-9051-3>.
- Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1480), 679–704. <https://doi.org/10.1098/rstb.2006.2004>.
- de Graaf, M. M. A., & Allouch, B. S. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61(12), 1476–1486. <https://doi.org/10.1016/j.robot.2013.07.007>.
- de Graaf, M. M. A., Allouch, S. B., & Klamer, T. (2015). Sharing a life with Harvey: exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior*, 43, 1–14. <https://doi.org/10.1016/j.chb.2014.10.030>.
- de Graaf, M. M. A., Allouch, B. S., & van Dijk, J. A. G. M. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human-Computer Interaction*, 34(2), 115–173. <https://doi.org/10.1080/07370024.2017.1312406>.
- De Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerinx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, 12(2), 459–478. <https://doi.org/10.1007/s12369-019-00596-x>.
- Deutsches Institut für Normung e. V. (2012). *Industrieroboter – Sicherheitsanforderungen – Teil 1: Roboter*. (Norm, DIN EN ISO 10218-1:2012-01). Berlin: Beuth Verlag GmbH.
- Deutsches Institut für Normung e. V. (2014). *Roboter und Robotikgeräte – Sicherheitsanforderungen für persönliche Assistenzroboter*. (Norm, DIN EN ISO 13482:2014-11). Berlin: Beuth Verlag GmbH.
- Deutsches Institut für Normung e. V. (2017). *Roboter und Robotikgeräte – Kollaborierende Roboter*. (Norm, DIN ISO/TS 15066:2017-04). Berlin: Beuth Verlag GmbH.
- Devin, S., & Alami, R. (2016). An implemented theory of mind to improve human-robot shared plans execution. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 319–326). <https://doi.org/10.1109/HRI.2016.7451768>.
- Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) (2006).
- Drury, J. L., Scholtz, J., & Yanco, H. A. (2004). *Awareness in human-robot interactions* (pp. 912–918). <https://doi.org/10.1109/icsmc.2003.1243931>.
- Eder, K., Harper, C., & Leonards, U. (2014). Towards the safety of human-in-the-loop robotics: challenges and opportunities for safety assurance of robotic co-workers. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 660–665). <https://doi.org/10.1109/ROMAN.2014.6926328>.
- Elkmann, N. (2013). Sichere Mensch-Roboter-Kooperation: Normenlage, Forschungsfelder und neue Technologien. *Zeitschrift Für Arbeitswissenschaft*, 67(3), 143–149. <https://doi.org/10.1007/BF03374401>.
- Enz, S., Diruf, M., Spielhagen, C., Zoll, C., & Vargas, P. A. (2011). The social role of robots in the future—explorative measurement

- of hopes and fears. *International Journal of Social Robotics*, 3(3), 263–271.
- European Commission (2019). Ethics guidelines for trustworthy AI. European commission. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>. Accessed September 24, 2021. <https://doi.org/10.2759/346720>
- European Parliament (2017). *Civil regulations in the field of robotics: European parliament resolution of 16 February 2017 with recommendations to the commission on civil law rules on robotics (2015/2103(INL))*
- Eyssel, F., Kuchenbrandt, D., Bobinger, S., de Ruitter, L., & Hegel, F. (2012). ‘if you sound like me, you must be more human. In H. Yanco (Ed.), *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (p. 125). New York: ACM. <https://doi.org/10.1145/2157689.2157717>.
- Fong, T. W., Nourbakhsh, I., & Dautenhahn, I. K. (2003). A survey of socially interactive robots: concepts, design, and applications. *Robotics and Autonomous Systems*, 42(3–4), 142–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X).
- Forster, Y., Kraus, J., Feinauer, S., & Baumann, M. (2018). *Calibration of trust expectancies in conditionally automated driving by brand, reliability information and introductory videos: An online study*. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. (pp. 118–128). <https://doi.org/10.1145/3239060.3239070>.
- Gelin, R. (2017). The Domestic Robot: Ethical and Technical Concerns. In M. I. A. Ferreira, J. S. Sequeira, M. O. Tokhi, E. E. Kadar & G. S. Virk (Eds.), *A world with robots: International Conference on Robot Ethics: ICRE 2015* (Vol. 84, pp. 207–216). Cham: Springer. https://doi.org/10.1007/978-3-319-46667-5_16.
- Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology and Work*, 14(1), 39–49. <https://doi.org/10.1007/s10111-011-0194-3>.
- Goetz, J., Kiesler, S., & Powers, A. (2003). *Matching robot appearance and behavior to tasks to improve human-robot cooperation*. The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. (pp. 55–60). <https://doi.org/10.1109/ROMAN.2003.1251796>.
- Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: a survey. *FNT in Human-Computer Interaction (foundations and Trends in Human-Computer Interaction)*, 1(3), 203–275. <https://doi.org/10.1561/1100000005>.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517–527. <https://doi.org/10.1177/0018720811417254>.
- Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2021). Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses. *Human Factors*, 63(7), 1196–1229. <https://doi.org/10.1177/0018720820922080>.
- Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. In *Proceedings of the world congress on engineering and computer science* (Vol. 1, pp. 425–430).
- Haring, K. S., Watanabe, K., Velonaki, M., Tossell, C. C., & Finomore, V. (2018). FFAB—The form function attribution bias in human–robot interaction. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4), 843–851. <https://doi.org/10.1109/TCDS.2018.2851569>.
- Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., & Sagerer, G. (2009). *Understanding social robots*. 2009 Second International Conferences on Advances in Computer-Human Interactions, Cancun. (pp. 169–174). <https://doi.org/10.1109/ACHI.2009.51>.
- Hergeth, S., Lorenz, L., & Krems, J. (2017). Prior familiarization with takeover requests affects drivers’ takeover performance and automation trust. *Human Factors*, 59(3), 457–470. <https://doi.org/10.1177/0018720816678714>.
- Hiroi, Y., & Ito, A. (2008). Are bigger robots scary?—The relationship between robot size and psychological threat. In *2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics* (pp. 546–551). <https://doi.org/10.1109/AIM.2008.4601719>.
- Hoc, J.-M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies*, 54(4), 509–540. <https://doi.org/10.1006/ijhc.2000.0454>.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>.
- Holmes, J. G., & Rempel, J. K. (1989). Trust in close relationships. In C. Hendrick (Ed.), *Close relationships*. Review of personality and social psychology, (Vol. 10, pp. 187–220). SAGE.
- International Organization for Standardization (2021). *Robotics—Vocabulary (ISO/DIS Standard No. 8373:2021)*. <https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-3:v1:en>. Accessed September 24, 2021.
- Jacobs, T. (2013). *Validierung der funktionalen Sicherheit bei der mobilen Manipulation mit Servicerobotern: Anwenderleitfaden*. Stuttgart.
- Janowski, K., Ritschel, H., Lugin, B., & André, E. (2018). Sozial interagierende Roboter in der Pflege. In O. Bendel (Ed.), *Pflegroboter* (pp. 63–87). Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-22698-5_4.
- Kardos, C., Kemény, Z., Kovács, A., Pataki, B. E., & Váncza, J. (2018). Context-dependent multimodal communication in human-robot collaboration. *Procedia CIRP*, 72, 15–20. <https://doi.org/10.1016/j.procir.2018.03.027>.
- Kildal, J., Martín, M., Ipiña, I., & Maurtua, I. (2019). Empowering assembly workers with cognitive disabilities by working with collaborative robots: A study to capture design requirements. *Procedia CIRP*, 81, 797–802. <https://doi.org/10.1016/j.procir.2019.03.202>.
- Kirchner, E. A., de Gea Fernandez, J., Kampmann, P., Schröer, M., Metzen, J. H., & Kirchner, F. (2015). Intuitive interaction with robots—Technical approaches and challenges. In R. Drechsler & U. Kühne (Eds.), *Formal modeling and verification of Cyber-physical systems: 1st international summer school on methods and tools for the design of digital systems*. Bremen, 09.2015. (pp. 224–248). Springer. https://doi.org/10.1007/978-3-658-09994-7_8.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltoch, P. J. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91–95. <https://doi.org/10.1109/mis.2004.74>.
- Königstorfer, J., & Gröppel-Klein, A. (2009). Projektive Verfahren zur Ermittlung der Akzeptanz technologischer Innovationen. In R. Buber & H. H. Holz Müller (Eds.), *Qualitative Marktforschung*. Gabler. https://doi.org/10.1007/978-3-8349-9441-7_51.
- Kornwachs, K. (2019). Smart robots—smart ethics? *Datenschutz Und Datensicherheit – DuD*, 43(6), 332–341. <https://doi.org/10.1007/s11623-019-1118-2>.
- Kraus, J. M. (2020). *Psychological processes in the formation and calibration of trust in automation*. Open Access Repository. Dissertation. Ulm: Universität Ulm. <https://doi.org/10.18725/OPARU-32583>.
- Kraus, M., Kraus, J., Baumann, M., & Minker, W. (2018). Effects of gender stereotypes on trust and likability in spoken human-robot interaction. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. <https://www.aclweb.org/anthology/L18-1018>. Accessed September 24, 2021.
- Kraus, J. M., Forster, Y., Hergeth, S., & Baumann, M. (2019). Two routes to trust calibration: effects of reliability and brand informa-

- tion on trust in automation. *International Journal of Mobile Human Computer Interaction*, 11(3), 1–17. <https://doi.org/10.4018/IJMHCI.2019070101>.
- Kraus, J., Scholz, D., Messner, E.-M., Messner, M., & Baumann, M. (2020). Scared to trust?—predicting trust in highly automated driving by depressiveness, negative self-evaluations and state anxiety. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2019.02917>.
- Kraus, J., Scholz, D., & Baumann, M. (2021). What's driving me? Exploration and validation of a hierarchical personality model for trust in automated driving. *Human factors*, 63(6), 1076–1105. <https://doi.org/10.1177/0018720820922653>.
- Kreis, J. (2018). Umsorgen, überwachen, unterhalten – sind Pflegeroboter ethisch vertretbar? In O. Bendel (Ed.), *Pflegeroboter* (pp. 213–228). Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-22698-5_12.
- Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270. <https://doi.org/10.1080/00140139208967392>.
- Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.
- Louie, W.-Y. G., McColl, D., & Nejat, G. (2014). Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assistive Technology: The Official Journal of RESNA*, 26(3), 140–150. <https://doi.org/10.1080/10400435.2013.869703>.
- 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. <https://doi.org/10.1177/2050157919843961>.
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37(1), 465–498. <https://doi.org/10.1002/aris.1440370111>.
- Miller, L., Kraus, J., Babel, F., & Baumann, M. (2021). More than a feeling—interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2021.592711>.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429–460. <https://doi.org/10.1080/00140139608964474>.
- Müller, M. F. (2014). Roboter und Recht. *Aktuelle Juristische Praxis (AJP/PJA)*, (5), 595–608. http://www.robotics.tu-berlin.de/fileadmin/fg170/Publikationen_pdf/01_Aufsatz_MelindaMueller.pdf. Accessed September 24, 2021.
- Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-020-00659-4>.
- Nevejans, N. (2016). European civil law in robotics. [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf). Accessed September 24, 2021.
- Onnasch, L., Maier, X., & Jürgensohn, T. (2016). *Mensch-Roboter-Interaktion – Eine Taxonomie für alle Anwendungsfälle* (1st edn.). (pp. 1–12). baa, Fokus, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. <https://doi.org/10.21934/baua:fokus20160630>.
- Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. Y. C. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In R. E. Karlson, D. W. Gage, C. M. Shoemaker & G. R. Gerhart (Eds.), *SPIE proceedings, unmanned systems technology XVI (90840E)*. SPIE. <https://doi.org/10.1117/12.2050622>.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, 39, 230–253.
- Regulation (EU) 2016/679 of The European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016)
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95. <https://doi.org/10.1037/0022-3514.49.1.95>.
- Rosenstrauch, M. J., & Kruger, J. (2017). Safe human-robot-collaboration-introduction and experiment using ISO/TS 15066. In *2017 3rd International Conference on Control, Automation and Robotics—ICCAR 2017*. Nagoya, 22 Apr.–24 Apr., 2017. (pp. 740–744). Piscataway: IEEE. <https://doi.org/10.1109/ICCAR.2017.7942795>.
- Ruijten, P. A. M., & Cuijpers, R. H. (2020). Do not let the robot get too close: Investigating the shape and size of shared interaction space for two people in a conversation. *Information*, 11(3), 147. <https://doi.org/10.3390/info11030147>.
- Salem, M., Ziadee, M., & Sakr, M. (2014). Marhaba, how may I help you? Effects of politeness and culture on robot acceptance and anthropomorphism. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 74–81). <https://doi.org/10.1145/2559636.2559683>.
- Salvini, P., Laschi, C., & Dario, P. (2010). Design for acceptability: improving robots' coexistence in human society. *International Journal of Social Robotics*, 2(4), 451–460. <https://doi.org/10.1007/s12369-010-0079-2>.
- Schenk, M., & Elkmann, N. (2012). Sichere Mensch-Roboter-Interaktion: Anforderungen, Voraussetzungen, Szenarien und Lösungsansätze. In E. Müller (Ed.), *Demographischer Wandel: Herausforderung für die Arbeits- und Betriebsorganisation der Zukunft. Tagungsband zum 25. HAB-Forschungsseminar. Schriftenreihe der Hochschulgruppe für Arbeits- und Betriebsorganisation e. V. (HAB)*. (pp. 109–122). Berlin: GITO.
- Song, Y., & Luximon, Y. (2020). Trust in AI agent: a systematic review of facial anthropomorphic trustworthiness for social robot design. *Sensors*, 20(18), 5087. <https://doi.org/10.3390/s20185087>.
- Thielmann, I., & Hilbig, B. E. (2015). Trust: an integrative review from a person-situation perspective. *Review of General Psychology*, 19(3), 249–277. <https://doi.org/10.1037/gpr0000046>.
- Turja, T., Aaltonen, I., Taipale, S., & Oksanen, A. (2020). Robot acceptance model for care (RAM-care): A principled approach to the intention to use care robots. *Information & Management*, 57(5), 103220. <https://doi.org/10.1016/j.im.2019.103220>.
- Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M., & Weber, M. (2017). From car-driver-handovers to cooperative interfaces: Visions for driver-vehicle interaction in automated driving. In G. Meixner & C. Müller (Eds.), *Automotive user interfaces: creating interactive experiences in the car* (pp. 273–294). Springer. https://doi.org/10.1007/978-3-319-49448-7_10.
- Weiss, A., Bernhaupt, R., Lankes, M., & Tscheligi, M. (2009). *The USUS evaluation framework for human-robot interaction*. Adaptive and Emergent Behaviour and Complex Systems—Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009. (pp. 158–165).
- Yanco, H. A., & Drury, J. (2004). *Classifying human-robot interaction: an updated taxonomy*. IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), The Hague. Vol. 3 (pp. 2841–2846). <https://doi.org/10.1109/ICSMC.2004.1400763>.
- Yeung, K. (2018). *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework* (MSI-AUT 05). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027. Accessed September 24, 2021.



Dr. Johannes Kraus is a post-doctoral researcher at the Human Factors department of Ulm University and head of the subject area “human-robot interaction”. His research interests lie in the decision processes in the interaction with automated systems, especially automated vehicles and robots. His research focuses on trust processes and the role of user personality and attitudes.



Katrin Hauber is a student assistant in the Department of Human Factors at Ulm University. Her research interests are human-robot interaction and automated driving.



Franziska Babel is a PhD student at the Department of Human Factors at Ulm University. She has a background in engineering psychology and human factors. Her research interests are human-robot interaction (technology acceptance, persuasive robotics), human-machine interaction, and persuasive technology. In her thesis, she investigates conflict resolution between humans and robots.



Prof. Dr. Martin Baumann is head of the Department of Human Factors at Ulm University. His main research interests are the psychological basis of human-machine interaction in different domains, mainly traffic, human-robot interaction, interaction with intelligent systems, and the development and validation of concepts of cooperative human-machine systems.

Dr. Philipp Hock is a post-doctoral researcher at the Department of Human Factors at Ulm University. He has a background in computer science and his research interests are human-machine interaction, persuasive technology and automated driving.