Implications from Responsible Human-Robot Interaction with Anthropomorphic Service Robots for Design Science

Merlind Knof[1], Judith S. Heinisch[2], Jérôme Kirchhoff[4], Niyati Rawal[1],
Klaus David[2], Oskar von Stryk[4], and Ruth Stock-Homburg[1]

[1]Department of Law and Economics, Technische Universität Darmstadt,
Email: {rsh,merlind.knof,niyati.rawal}@bwl.tu-darmstadt.de
[2]Department of Electrical Engineering and Computer Science, Universität Kassel,
Email: {judith.heinisch,david}@uni-kassel.de
[4]Department of Computer Science, Technische Universität Darmstadt, Email:{kirchhoff,stryk}@sim.tu-darmstadt.de

Abstract

Accelerated by the COVID-19 pandemic, anthropomorphic service robots are continuously penetrating various domains of our daily lives. With this development, the urge for an interdisciplinary approach to responsibly design human-robot interaction (HRI), with particular attention to human dignity, privacy, compliance, and transparency, increases. This paper contributes to design science, in developing a new artifact, i.e., an interdisciplinary framework for designing responsible HRI with anthropomorphic service robots, which covers the three design science research cycles. Furthermore, we propose a multi-method approach by applying this interdisciplinary framework. Thereby, our finding offer implications for designing HRI in a responsible manner.

1. Introduction

The COVID-19 pandemic significantly influenced the observed growing importance of the service sector within the last decades, leading to an acceleration of technological developments and hence, the increased use of autonomous and intelligent service robots [1–4]. This can be experienced in everybody's daily life, where technologies have already penetrated several areas [3–5], e.g., customers interact with airport self-service kiosks, internet-based services, and automated teller machines [6]. Especially, service robots offer useful perspectives and can assist humans by providing several advising and informing services for customers in numerous areas, e.g., banking, education, health, hospitality, and retail [4, 6]. Service robots will play an essential role in the future on all economic levels (i.e., micro, meso, and macro level) [7]. This paper proposes an interdisciplinary framework to provide comprehensible recommendations and suggestions regarding the design activities for the responsible investigation of human-robot interaction (HRI). This framework is one result of the interdisciplinary research project "RoboTrust". We understand HRI as "the study of the humans, robots, and the ways they influence each other" [8, p. 256], with regard to anthropomorphic service robots.

Several literature reviews on HRI emphasize the importance to interdisciplinary approach responsible HRI [7, 9–13] involving the following four disciplines: (1) ethics, (2) law, (3) psychology, and (4) technology. This leads to our definition of responsible design of HRI (see also [13]) as the design of HRI between users and anthropomorphic service robots with particular attention to humans' dignity, emotional needs, data & privacy protection, security & safety, and transparency. This fundamental notion identified in "RoboTrust" captures the interdisciplinary character of HRI. It allows a responsible design which enables HRI in a human-oriented and trustful manner. Thus, the anthropomorphic service robot serves as assistant that (a) supports (and does not replace) humans in domains that are not perceivable by humans due to resource constraints, and (b) acts in a responsible manner with particular attention to human dignity. Further, (c) the anthropomorphic service robot's interaction behavior is predictable and transparent. (d) All data collected by the anthropomorphic service robot are processed transparently and in conformity with the applicable data protection laws. (e) The liability of the robotic assistant's behavior is clearly established.

HĬCSS

From a design science perspective it is common practice to involve different disciplines within information and communication technologies research [14–16]. Value-sensitive design, for instance, is a methodology and research approach which seeks to integrate moral and social values into the design and development of technologies [14]. In contrast, our approach emphasizes interdisciplinary research relevant for the responsible design of HRI with anthropomorphic service robots, considering the interdependencies of the individual disciplines that shape each other.

In design science, innovative and new artifacts (e.g., constructs, methods or models) aim to solve problems or achieve improvements, instead of explaining the existing reality [4, 17, 18]. Thus, the proposed interdisciplinary framework, as a method that provides guidance on how to design HRI with anthropomorphic service robots in a responsible manner, represents a new artifact drawing on design science.

In order to create this interdisciplinary framework, we made an overall state of the art study (see [13] for an overview) and identified the overarching research objectives for this paper:

Objective 1: Development of a concept for the responsible, human-oriented application and placement of anthropomorphic service robots based on ethical, psychological, legal, and technical requirements of human users for authenticity and acceptability of HRI (see Section 2, *responsible placement*).

Objective 2: Development of an exemplary socio-technical system of HRI for anthropomorphic service robots that takes corresponding ethical, psychological, legal, and technical aspects into account, that positively symbiotically enrich the actions of a human (see Section 3, *responsible programming*).

Objective 3: Experimental investigation of the socio-technical system of HRI for a retailing scenario using two types of anthropomorphic (i.e., humanoid and android) service robots (see Section 3, *physical design, responsible placement*).

Based on these objectives, the interdisciplinary framework consisting of three steps was developed as shown in Figure 1.

The paper is organized as follows. Firstly, the interdisciplinary framework is introduced (see Figure 1 and Section 2). Secondly, a multi-method approach is proposed to which the framework is applied in Section 3. Finally, Section 4 outlines implications for design science research focusing on HRI.

2. Interdisciplinary Framework

Inspired by the design science *Information Systems Research Framework* [17], our interdisciplinary framework and its application (see Section 3) involve the three common cycles (i.e., Relevance, Design, and Rigor Cycle) [19]. The described three core framework steps in this Section reside in the *Design Cycle* of our research activity to realize responsible HRI with anthropomorphic service robots. The identified four disciplines determine the relevant *Environment* and *Knowledge Base* [19].
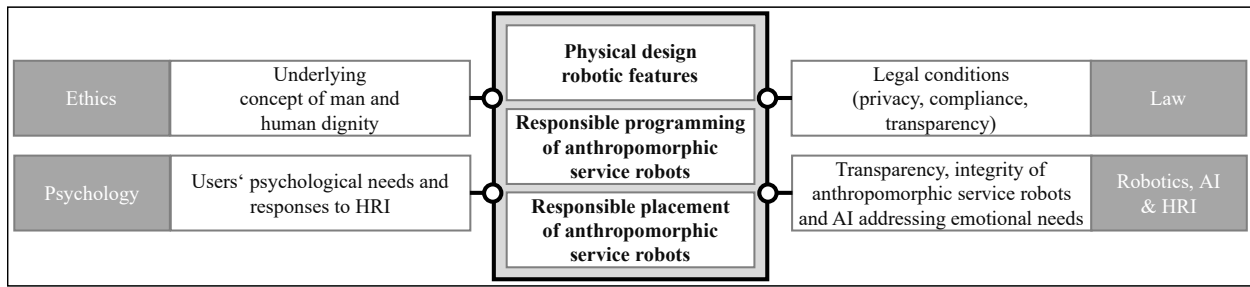
2.1. Physical Design

Various efforts have been made to create human-like robots [20]. These robots "with human-like appearance features such as eyes, hands, or faces" [21, p. 105], are classified as anthropomorphic robots [21]. Typically, robots which are constructed to behave or look like humans, are referred to as androids or humanoids [3]. "Whereas humanoid robots often come with extremities ... but still have an overall mechanical look, android robots are intended to mimic human beings as realistically as possible" [3, p. 327].

The *Computers as Social Actors* (CASA) paradigm demonstrates "that individuals mindlessly apply social rules and expectations to computers" [22, p. 81]. Hence, depending on several factors, including the degree of social characteristics of technologies (e.g., robots), humans tend to treat these technologies like humans by adopting social heuristics of human-human interactions to human-computer interactions [23, 24].

The tendency of humans to assign human characteristics and traits to non-human entities is generally understood as *anthropomorphism* [25, 26]. In robotics anthropomorphic design is used to facilitate HRI, since user acceptance of robots, as well as the effectiveness of HRI, are influenced by attributing social behavior and presence to robots (e.g., through the use of facial expressions or a human-like physical design of the robot) [25, 26]. Yet, despite this trend in robotic research to design and develop robots (e.g., service robots) which are increasingly human-like, studies suppose that customer satisfaction is higher within a pure human-human interaction as opposed to an HRI [6, 20]. The *uncanny valley* paradigm further predicts a positive relationship between a robot's human-likeness and the robot's likability by the human [3, 20, 27]. However, in case that a robot "has an almost, but not perfectly, realistic human appearance" [28, p. 337], humans report negative responses towards the robot, up to a feeling of eeriness [3, 20, 27].

Consequently, an anthropomorphic service robot's *physical design* has an important impact on a

Figure 1. Interdisciplinary Framework for Designing Responsible HRI with Anthropomorphic Service Robots

user's perception and should be determined, when programming and placing it (see Figure 1, *physical design*).

2.2. Responsible Programming

Transparency: With the increasing complexity of anthropomorphic service robots, their application areas and artificial reasoning increase as well. This enables the integration of systems in dynamic real-world settings with a variety of interaction contexts and partners. Here, *transparency* can be a means to approach a responsible robot design and, thus, establish trust and social acceptance for anthropomorphic service robots.

Transparency has often been considered concerning robots interacting with an operator or teammate ([29, 30]) to understand and discover the robotic decisions. However, in the case of anthropomorphic service robots, legal and social requirements need to be considered as well. This includes the investigation of incidents, caused by undesired robot behavior in dynamic interaction scenarios. Thus, transparency has different meanings depending on the HRI stakeholder and needs for separated requirement analyses and assessments [31, 32]. For example, the appropriate information disclosed and conveyed to understand the service robot's behavior is likely to differ between an end-user and the robotic service provider.

A black box recorder [32, 33] could facilitate retrospectively inspect the decisions made and incidents that occurred for different stakeholders. Here, efficient storage and semantic access [34] on the collected data from manifold sources, as well as, data security and fulfilling legal conditions of personal data [35] is required. Further, the authenticity and integrity of the recordings must be ensured to resolve the legal liability of accidents by lawyers or accident witnesses. A technical solution for the demanded black box recorder with appropriate tools to investigate malfunction or incident causes is needed for a responsible HRI. This aims for filling the technical gap to facilitate the realization of transparency requirements tailored to the stakeholder and at the same time responsibly handle the collected data.

To make the decision-making and other interaction information transparent to end-users, with no deep understanding of the robot's internal mechanisms, anthropomorphic service robots have beneficial capabilities for HRI [36, 37]. As they can express information using a combination of gestures, mimic and speech, the intuition of performed and future actions can be expressed in a usual way for the user. In addition to transfer gestures between different anthropomorphic or humanoid robots [38], a special challenge consists of modeling and programming these whole-body expressions depending on the current interaction context in a generic way, that also allows to adapt gestures online. This would enable to abstract motions from a specific robot and develop general interaction strategies for a transparent and context sensitive interaction.

Emotional Responsiveness: Emotions play an essential role in the interaction between people and the interaction between humans and anthropomorphic robots [9, 39]. Especially for the interaction with anthropomorphic robots, humans expect the robot to express emotions and to behave in an empathetic manner [40]. In this paper, emotions are defined as "an episode of interrelated synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [41, p. 697]. This definition indicates that emotions are time-limited (approximately between 0.5 and 4 seconds [42]) and are evoked via stimulus events (like a robot coming closer). The autonomic nervous system is an example of an organismic subsystem that is responsible for system regulation and therefore affects the physiology of the individual. These changes can be detected by sensors and inferred to the emotion of the person by trained models. This "ability to encode an ensemble of sensory stimuli providing information about the emotional state

of another individual" [43, p. 1] is referred to as *emotion recognition*.

Much has happened in the field of emotion recognition in terms of transparency and ethics in experimental setups. At the latest with the paper [44], in which they name five factors that influence data collection in emotion recognition, the community has to deal with these questions. One of these factors, the authors identify, raises the question of whether or not the subject is aware of the purpose of the experiment. They point out that because of privacy and ethics concerns, it is practically impossible to collect data in which the subject neither knows about the data collection nor the purpose of the experiment. [45] show that ethical issues can change if an in-the-wild study is conducted. As one advantage, the problem of eliciting negative emotions, is named which is actually problematic in a laboratory study. However, the collection of real-life data reduces the responsibility for eliciting negative emotions even if they appear.

2.3. Responsible Placement

Since the prevalence of service robots is steadily increasing, human responses to anthropomorphic service robots have become a critical question [3]. *Responsible placement* in this paper aims to improve humans' perception of likeness (e.g., a user's acceptance) of anthropomorphic service robots in HRI in a human-oriented and responsible manner. Anthropomorphic service robots co-exist with humans, they perform a variety of tasks in different areas, interact and communicate with people, and indicate some basic intelligent behaviors to autonomously perform assigned tasks [1, 3, 46, 47].

The anthropomorphic design as well as increasingly advanced robot embodiment critically influence users' comfort and acceptance of anthropomorphic service robots [2, 20] (see Section 2.1). For a responsible design of HRI, the acceptance of robots by users and the reduction of users' discomfort are fundamental. However, the extent to which humans are willing to interact with robots does not only depend on robotic features [3, 20] (see Section 2.1, 2.2), or technical, safety and legal issues [48] (see Section 2.2), but also to a large extent on the selected scenario of the HRI (i.e., the location and tasks assigned to the robot) [49].

From a user perspective, preferences regarding activities, services, and tasks which should or should not be performed by anthropomorphic service robots, exist. Several studies considering the attitude of users towards service robots have been conducted in the fields of travel, tourism, hospitality, and healthcare [2, 49–51].

These studies point out that users feel comfortable, when robots perform repetitive tasks, such as providing information or performing housekeeping activities (e.g., delivering items and cleaning), whereas users tend to not accept activities requiring a user to subordinate her/his body to a service robot (e.g., babysitting, massages, or hairdressing) [49, 50, 52].

To determine the responsible placement of anthropomorphic service robots, users' psychological needs and responses to HRI should be appropriately considered. Consequently, the scenario which should be investigated in the experimental setting should be critically scrutinized beforehand (see Section 3).

3. Responsible Experimental Design

The proposed multi-method approach in this Section (see Figure 2) shows the application of the interdisciplinary framework to achieve a responsible HRI with anthropomorphic service robots. Regarding Design Science Research, the outcome of the studies, conducted within the frame of project "RoboTrust", will be considered to adjust the requirements of the *Environment*, to create additions to the *Knowledge Base* (cf. *Relevance* and *Rigor Cycle*), and evaluate the design process.

3.1. Applying the Framework

Following the first two goals (see Figure 2) the experimental scenario as well as the anthropomorphic service robots (i.e., humanoid and android robots) had to be selected (see *study 1*). In one online study (*study 1*), data were collected from Amazon Mechanical Turk (MTurk) workers. With this study, we evaluated the needs concerning the *physical design* and *responsible placement* of our robots. The participants had to rate different service scenarios in terms of their importance for and likelihood to occur in the near future. The survey was visible only to workers with an historic acceptance rate greater than 95% and who were residents of the United States (US). As an incentive, all participants received a financial remuneration of $ 2.00 for completing the study. We compared four different scenarios: (1) retailing, (2) pharmacy, (3) traveling agency, and (4) restaurant in a between-subject design with 434 participants. To ensure the quality of the responses, we placed five screening questions in a random order in the questionnaire (e.g., "Please check 'totally disagree' here"). Cases of participants with incorrect answers to these screening questions were deleted from our database. After considering

these control questions, we received data from 392 US-citizens[1] (60% men, 39% women, 1% diverse, mean age = 39.69 years, SD = 12.79). The major professional status was employed (67%), followed by self-employed (10%), not-employed (8%), and others (14%). The participants were randomly assigned to the four different scenarios. First, every participant answered the same general (e.g., demographic) questions. Afterwards, the participants were briefly introduced into the topic of service robots and were asked to answer various questions regarding the use of service robots for their allocated scenario (i.e., retailing, pharmacy, traveling agency, or restaurant). Considering the COVID-19 pandemic, we asked the participants how realistic they assessed the implementation of service robots in each area to become reality in companies in general using a 7-point Likert-type scale. They predominantly evaluate the use of service robots as a likely future scenario in each area, namely in a furniture store (74%), a pharmacy (63%), a traveling agency (66%), and a restaurant (70%). The results revealed that the retail setting was the most likely and relevant by the US-participants. This finding was further underlined by open remarks, the participants made in the questionnaire.

Furthermore, we examined the uncanny valley paradigm [53] (see Section 2.1). Specifically, it was examined which service representative, i.e., a human frontline employee, humanoid service robot, or android service robot, was preferred. The participants were asked to rank the three possible service representatives according to their preference when imagining to interact within a retail scenario. The study confirmed the predictions of the uncanny valley paradigm in that humanoid and android robots were perceived as less likable than humans. Not surprisingly, the participants predominantly ranked humans as their preferred interaction partner in the first place (73%), followed by humanoid robots (23%) and android robots (3%). However, other studies show a more differentiated picture in the behaviour towards android and humanoid robots. For instance, in an experiment in a real-company setting employees disclosed more personal information towards android robots than towards humanoid robots in complex tasks [54]. Thus, humanoid robots as well as android robots are suitable research robots, as our findings indicate differences regarding the acceptance of these robots. As a consequence, two types of robots were selected for the HRI experiment, namely the humanoid robot TIAGo++ and the android robot Elenoide (see Figure 3).

In *study 2* we attempt to validate the emotion recognition (i.e., facial expression) of the service robot (see Figure 1, *responsible programming*). Human facial expression recognition is one possible method for emotion recognition during HRI as most robots come with an inbuilt camera. Among different machine learning and deep learning techniques for facial expression recognition in HRI (see [39] for an overview), Convolutional Neural Network (CNN) [55] is the most popular algorithm [39]. We implemented a CNN where we first detect the landmarks (eyes, nose and mouth) using dlib [56] and feed a simplified face image into the CNN for recognizing the emotion from five categories (i.e., angry, happy, neutral, sad, and surprise). These five categories are chosen among the seven universally known facial expressions in Ekman's work [57]. The CNN is trained on simplified frontal face images of the Karolinska Directed Emotional Faces database [58]. To test our algorithm, we conducted a preliminary study with six participants, who were shown a video displaying a certain emotion. We recorded their facial expressions while watching the videos. Before each video, the subject was told which emotion was to be triggered by it. The kind of video (e.g., a happy video) we showed to the participants and their expected reactions from watching those videos became the ground truth label. Later, the emotion labels given by the CNN were compared with these ground truth labels to determine the overall accuracy. The participants were recorded in varying lighting conditions to test if our method is robust. We evaluated our algorithm on 30 recorded videos which were used as a test set and as a result 67% of overall accuracy was achieved. Some of the participants were not very expressive with their faces, especially the emotions sad and angry could not even be appropriately recognized by some of the humans. As the facial expressions of the participants can easily be controlled [59], it would be beneficial to perform facial expression recognition combined with audio data and data from physiological sensors to improve the emotion recognition (see Section 3.2).

To gather important features of a responsible anthropomorphic service robot from a user's perspective, we will next perform another online user study *(study 3)* with MTurk workers. The survey will be visible only to workers with an historic acceptance rate greater than 95% and who are residents of the US. As an incentive, all participants will receive a financial remuneration for completing the study. In this study, we will evaluate the feature which are the most relevant for users when interacting with anthropomorphic service robots in a responsive, e.g., human-oriented, manner, such as transparency, privacy compliance, security

---

[1]392 participants, allocated to retailing (90 participants), pharmacy (106 participants), traveling agency (105 participants) and restaurant (91 participants)

#	Goal of Study	Setting	Framework Stage
1	* **Selection of experimental scenario**	Retailing scenario as most likely scenario for service robots in the future	Responsible Placement
1	* **Selection of anthropomorphic service robots**	Anthropomorphic service robots (i.e., humanoid robot TIAGo++, android robot Elenoide) with different human acceptances	Physical Design
2	* **Validation of Emotion Recognition – Facial Expression**	Evaluation of CNN model by elicitation of emotions via videos	Responsible Programming
3	** **User-centric validation of important features of a responsible service robot**	*Preliminary features:* i.e., transparency, privacy compliance, emotional reponsiveness	Physical Design / Responsible Programming
4	** **Validation of Emotion Recognition– Physiological Signals**	*Into-the-wild-study:* recording of physiological data and emotions in everyday life	Responsible Programming
5	** *HRI in laboratory:* **Assess responsible service robot features**	*Vignette-based experiment:* artificial retailing setting in which robot advises and serves customers	Physical Design / Responsible Programming / Responsible Placement
6	** *HRI in the wild:* **Asses responsible service robot features**	*Real-life experiment in retailing:* artificial retailing-setting in which robot advises and serves customers	Physical Design / Responsible Programming / Responsible Placement

* = study completed, ** = study in progress

Figure 2. Multi-method study overview for the investigation of responsible HRI with anthropomorphic service robots

and emotional responsiveness (see Figure 1, *physical design, responsible programming*).

Study 4 will aim to validate the emotion recognition through physiological signals (see Figure 1, *responsible programming*). For this we implemented a smartphone application recording physiological signals with experienced emotions in real-time. To record the physiological signals, we use the E4 wristband from Empatica [60], which is connected via Bluetooth with the application. As soon as the algorithm on the smartphone detects an emotion, the app shows a notification to the user. This notification asks the user to confirm and label the emotion. If the user confirms the emotion, the app navigates to a questionnaire. This questionnaire contains the Self Assessment Manikins (SAM) [61] to enter the felt emotion and a text field to name the emotional trigger or context. SAM requires the opportunity to validate the felt emotions in three scales regarding the circumplex model of affect from Russel [62]: Arousal ranging from calm to excited, valence ranging from pleasant to unpleasant, and dominance ranging from controlling to controlled. Finally, the entered emotion will be stored with the physiological signals fitting it. In our study, we leave the laboratory and record emotions via physiological signals using the developed application. The goal of the in-the-wild study is to evaluate the algorithms for recording the data by first analyzing the associated signals and the number of false recognized emotions. Second, the experiment will record an initial data

set linking physiological signals, the trigger, and the associated emotion via a questionnaire (as ground truth). The time-frame of this field experiment is limited to five days. The intended number of participants is about 120 participants with different demographic characteristics, such as ages and gender. This diversity of participants is important because emotion recognition models are subject-dependent [63], and thus bias can be reduced. First, we will give a brief introduction regarding the goal of the study, data collection, and processing, technical handling of the E4, and the application. The privacy protection measures and the rights of the participants will be introduced (e.g., discontinuation of study participation at any time) before the participant will be asked to sign an informed consent. Second, we introduce the questionnaire before we can send the participants out to do the measurement by themselves during their everyday life.

3.2. Evaluating the Design Process

By conducting *studies 5 & 6*, "RoboTrust" aims to assess the findings from the previously performed *studies 1 to 4* regarding our interdisciplinary framework (*physical design*, *responsible programming*, and *responsible placement*).

Therefore, we suggest a two-step test approach. This will include an experimental laboratory study *(study 5)* prior to an experimental field setting *(study 6)*. Here, the scenarios, techniques, and methods applied will

be critically evaluated under legal and ethical aspects. Findings of *study 5* will be used to adapt the final HRI experiment in-the-wild (*study 6*), according to the interdisciplinary framework's aspects. The experiments are planned along the guidelines, suggested by [5].

The experimental study set up for both experiments will represent a service encounter in a retail setting, more precisely in a construction market (see Figure 3 for a sample setting). Prior to the interaction, participants will be instructed based on a vignette and take the role of a retail customer. A vignette is defined as a "short, carefully constructed description of a person, object, or situation, representing a systematic combination of characteristics" [64, p. 128]. In the role of the customer, the participants will need help from a service representative in the retail scenario.

Generally, the experimental design follows a three step's procedure: (1) In the *preparation phase*, the customers will have to answer a pre-questionnaire, containing questions about their demographics, their prior experience with robots and their current affective state. Furthermore, the E4 wristband [60] will be introduced, and the use of the data explained. The participant will be asked to sign an informed consent after the privacy protection measures and the rights of the participants were introduced (e.g., erasing their collected data on request). (2) In the *experimental manipulation phase*, the participants will carry out specific tasks requiring an interaction with the robots. The specific interaction task will be introduced with an experimental vignette. To ensure participants understanding the scenarios, they will have to indicate their task; if their description is inaccurate, the experimenter will correct their understanding and will ask them to provide another written summary. The participants are given a smartphone on which an app for recording emotion via a questionnaire will be installed. This questionnaire will be explained to the participants, and questions about it will be answered. Since the robot is intended to trigger emotions, the participants will not be familiarized with it before the measurement. The anthropomorphic service robot will act as a consultant and assistant and will perform activities, such as answering customer questions or advising customers. Consequently, the robot will not have to complete any physical tasks (e.g., moving goods). The agent as the independent variable will experience a manipulation and might either be a robot (i.e., TIAGo++, Elenoide) or another human (as control group). Finally, (3) the *post-experimental phase* will focus on the completion of a post-questionnaire, asking about the affective states during the interaction with the robot, and will reveal the true purpose of the experiment to the participants.

In the laboratory setting, the experimental lab will be free of any external sounds, and the room lighting and temperature will remain constant. All visual displays and sounds will be recorded by external high-definition cameras, positioned throughout the room. The experimenter will not be visible to the participants but observe video streams from the cameras, through a separate screen. The participants will be told that the robot acts autonomously. For standardization and consistency, the interaction actually will be controlled by a programmer in a separate place (e.g., an adjoining room). This Wizard of Oz method is widely employed in HRI (see [5] for an overview). To improve the natural flow of conversation during the experiment, the experimenter will established natural language processing, so the robots will respond appropriately to things said or questions raised by participants.

To reduce any confounding effects due to personal communication styles and to standardize the interaction, we will adobe standardized service scripts for the robots, similar to those commonly used by retail stores in practice. The participants will be students, enrolled in psychology, management, or computer science courses, who volunteer to participate. To increase the realism of the experiments, the robots will refer to an existing service offer by a retail store. Furthermore, realistically relevant material will be available during the conversation, such as shelves, products, and advertising about the store.

The research design is a critical decision [64,65] (see Section 2). We will conduct a between-subject design to avoid learning effects. Therefore, we will test user responses (i.e., trust, acceptance) to different robotic features (such as transparency, privacy compliance, security, and emotional responsiveness) in various conditions; each participant will be randomly assigned to one condition.

4. Conclusion and Implications

Within the frame of project "RoboTrust", we have developed an interdisciplinary framework for designing responsible HRI with anthropomorphic service robots that fulfills the identified research objectives (see Section 1). Three core steps were framed by four specialized disciplines, ensuring that ethical, psychological, legal, and technical insights are considered for the responsible, human-oriented application and placement of anthropomorphic service robots (Objective 1). By applying the discussed framework, we realize an example socio-technical system that shows the usability of the interdisciplinary approach, also underlined by the conducted experiments

Figure 3. Android robot Elenoide (left) and humanoid robot TIAGo++ (right) in a retail setting

so far (Objective 2). Furthermore, we offered a proposal and guideline for the conduction of a responsible experimental design, consisting of a multi-method approach including six studies (Objective 3).

This paper presents implications from HRI for design science with respect to the procedure on how to design HRI with anthropomorphic service robots in a responsible manner. In particular, the presented interdisciplinary framework represents a new artifact and covers the three design science research cycles. Further, the relevant *Environment* and *Knowledge Base* are determined by the identified four disciplines.

Beyond this we offer implications for further design science research focusing on HRI with anthropomorphic service robots, in a responsible manner:

Implication 1: We have developed an interdisciplinary research framework for the responsible investigation of HRI with anthropomorphic service robots (see Section 2). This framework enables further investigation of aspects from ethics, psychology, law, and computer science to strengthen the *Knowledge Base* in the field of responsible HRI. With this, based on the gained findings more studies according to the framework are needed (cf. *Rigor, Relevance Cycle*). Since the interdisciplinary framework is built in accordance to the established design science cycles, we believe that the application to other HRI scenarios is desirable without any limitations.

Implication 2: We have demonstrated the usability and an evaluation of our interdisciplinary framework by proposing a multi-method approach. Using this approach is well suited, since it takes individual requirements of all identified disciplines into consideration. Thus, it should be used to investigate HRI research questions, addressing the *physical design* of robotic features, as well as the *responsible programming* and *placement* of anthropomorphic service robots.

Implication 3: The presented experimental design is restricted by cross-sectional data collected at a certain point in time. This kind of experimental design may suffer from the human's first impression of the anthropomorphic service robot by neglecting long-term human responses to anthropomorphic service robots. Therefore, it is necessary to additionally study HRI in longitudinal experiments and thus, adjusting the requirements of the *Environment*, amending the *Knowledge Base*, and evaluating the design activity.

Acknowledgement

References

[1] J. Holland, L. Kingston, C. McCarthy, E. Armstrong, P. O'Dwyer, F. Merz, and M. McConnell, "Service Robots in the Healthcare Sector," *Robotics*, vol. 10, no. 1, p. 47, 2021.

[2] F. Fusté-Forné and T. Jamal, "Co-Creating New Directions for Service Robots in Hospitality and Tourism," *Tourism and Hospitality*, vol. 2, no. 1, pp. 43–61, 2021.

[3] M. Mara and M. Appel, "Effects of Lateral Head Tilt on User Perceptions of Humanoid and Android Robots," *Computers in Human Behavior*, vol. 44, pp. 326–334, 2015.

[4] S. H. Ivanov, C. Webster, and K. Berezina, "Adoption of Robots and Service Automation by Tourism and Hospitality Companies," *Revista Turismo & Desenvolvimento*, vol. 27, no. 28, pp. 1501–1517, 2017.

[5] N. Homburg, "How to Include Humanoid Robots into Experimental Research: A Multi-Step Approach," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 4423–4432, January 2-6, 2018.

[6] M. Merkle, "Customer Responses to Service Robots–Comparing Human-Robot Interaction with Human-Human Interaction," in *Proceedings of the 52nd*

Hawaii International Conference on System Sciences, pp. 1396–1405, January 8-11, 2019.

[7] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch, and A. Martins, "Brave New World: Service Robots in the Frontline," *Journal of Service Management*, vol. 29, no. 5, pp. 907–931, 2018.

[8] T. Fong, C. Thorpe, and C. Baur, *Collaboration, Dialogue, Human-Robot Interaction*, pp. 255–266. Berlin, Heidelberg, Germany: Springer, 2003.

[9] R. Stock-Homburg, "Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research," *International Journal of Social Robotics*, pp. 1–23, 2021.

[10] R. Wullenkord and F. Eyssel, "Societal and Ethical Issues in HRI," *Current Robotics Reports*, pp. 1–12, 2020.

[11] C. Lutz, M. Schöttler, and C. P. Hoffmann, "The Privacy Implications of Social Robots: Scoping Review and Expert Interviews," *Mobile Media & Communication*, vol. 7, no. 3, pp. 412–434, 2019.

[12] H. Zech, "Liability for AI: Public Policy Considerations," in *ERA Forum*, vol. 22, pp. 147–158, Springer, 2021.

[13] R. Stock-Homburg, J. Kirchhoff, J. Heinisch, A. Ebert, P. Busch, N. Rawal, K. David, J. Wendt, I. Spiecker gen. Döhmann, O. von Stryk, M. Hannig, and M. Knof, "Responsible Human-Robot Interaction with Anthropomorphic Service Robots: State of the Art of an Interdisciplinary Research Challenge," in *Proceedings of the 55th Hawaii international Conference on System Sciences*, January 3-7, 2022.

[14] J. Simon, "Value-Sensitive Design and Responsible Research and Innovation," *The Ethics of Technology-Methods and Approaches*, pp. 219–236, 2016.

[15] X. Deng, K. D. Joshi, and R. D. Galliers, "The Duality of Empowerment and Marginalization in Microtask Crowdsourcing: Giving Voice to the Less Powerful through Value Sensitive Design," *MIS Quarterly*, vol. 40, no. 2, pp. 279–302, 2016.

[16] M. Liegl, A. Boden, M. Büscher, R. Oliphant, and X. Kerasidou, "Designing for Ethical Innovation: A Case Study on ELSI Co-Design in Emergency," *International Journal of Human-Computer Studies*, vol. 95, pp. 80–95, 2016.

[17] A. Hevner, S. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly: Management Information Systems*, vol. 28, no. 1, pp. 75–105, 2004.

[18] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–78, 2008.

[19] A. R. Hevner, "A Three Cycle Vview of Design Science Research," *Scandinavian Journal of Information Systems*, vol. 19, no. 2, p. 4, 2007.

[20] M. Mori, K. F. MacDorman, and N. Kageki, "The Uncanny Valley [from the Field]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[21] E. Phillips, X. Zhao, D. Ullman, and B. F. Malle, "What is Human-Like? Decomposing Robots' Human-Like Appearance Using the Anthropomorphic RoBOT (ABOT) Database," in *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pp. 105–113, 2018.

[22] C. Nass and Y. Moon, "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[23] Y. Kim and S. S. Sundar, "Anthropomorphism of Computers: Is it Mindful or Mindless?," *Computers in Human Behavior*, vol. 28, no. 1, pp. 241–250, 2012.

[24] A. Gambino, J. Fox, and R. A. Ratan, "Building a Stronger CASA: Extending the Computers are Social Actors Paradigm," *Human-Machine Communication*, vol. 1, no. 1, p. 5, 2020.

[25] L. Damiano and P. Dumouchel, "Anthropomorphism in Human-Robot Co-Evolution," *Frontiers in Psychology*, vol. 9, p. 468, 2018.

[26] J. Fink, "Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction," in *International Conference on Social Robotics*, pp. 199–208, Springer, 2012.

[27] C.-C. Ho and K. F. MacDorman, "Measuring the Uncanny Valley Effect," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 129–139, 2017.

[28] J. Seyama and R. S. Nagayama, "The uncanny valley: Effect of realism on the impression of artificial human faces," *Presence: Teleoperators and virtual environments*, vol. 16, no. 4, pp. 337–351, 2007.

[29] A. Bhaskara, M. Skinner, and S. Loft, "Agent Transparency: A Review of Current Theory and Evidence," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 3, pp. 215–224, 2020.

[30] F. Rajabiyazdi and G. A. Jamieson, "A Review of Transparency (seeing-into) Models," in *2020 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 302–308, October 11-14, 2020.

[31] H. Felzmann, E. Fosch-Villaronga, C. Lutz, and A. Tamo-Larrieux, "Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies," *IEEE Robotics & Automation Magazine*, vol. 26, no. 2, pp. 71–78, 2019.

[32] IEEE Standards Association, "IEEE Draft Standard for Transparency of Autonomous Systems," *IEEE P7001/D1*, pp. 1–70, June 2020.

[33] A. F. Winfield and M. Jirotka, "The Case for an Ethical Black Box," in *Annual Conference Towards Autonomous Robotic Systems*, pp. 262–273, July 19-21, 2017.

[34] T. Niemueller, G. Lakemeyer, and S. S. Srinivasa, "A Generic Robot Database and its Application in Fault Analysis and Performance Evaluation," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 364–369, October 7-12, 2012.

[35] V. Vasylkovskyi, S. Guerreiro, and J. S. Sequeira, "BlockRobot: Increasing Privacy in Human Robot Interaction by Using Blockchain," in *2020 IEEE International Conference on Blockchain*, pp. 106–115, November 2-6, 2020.

[36] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 313–323, 2013.

[37] E. Park, K. J. Kim, and A. P. del Pobil, "The Effects of Robot's Body Gesture and Gender in Human-Robot Interaction," *Human-Computer Interaction*, vol. 6, pp. 91–96, 2011.

[38] G. V. de Perre, M. V. Damme, D. Lefeber, and B. Vanderborght, "Development of a Generic Method to Generate Upper-Body Emotional Expressions for Different Social Robots," *Advanced Robotics*, vol. 29, no. 9, pp. 597–609, 2015.

[39] N. Rawal and R. M. Stock-Homburg, "Facial Emotion Expressions in Human-Robot Interaction: A Survey," *To appear in International Journal of Social Robotics*, 2021.

[40] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew, "How People Anthropomorphize Robots," in *Proceedings of the 3rd international conference on Human robot interaction - HRI '08*, pp. 145–152, March 2008.

[41] K. R. Scherer, "What are Emotions? And How Can They Be Beasured?," *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.

[42] R. W. Levenson, *Emotion and The Autonomic Nervous System: A Prospectus for Research on Autonomic Specificity.*, ch. xiii, pp. 17–42. Oxford, England: John Wiley & Sons, 1988.

[43] V. Ferretti and F. Papaleo, "Understanding Others: Emotion Recognition in Humans and Other Animals," *Genes, Brain and Behavior*, vol. 18, no. 1, p. e12544, 2019.

[44] R. W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[45] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward Emotion Recognition From Physiological Signals in the Wild: Approaching the Methodological Issues in Real-Life Data Collection," *Frontiers in Psychology*, vol. 11, p. 1111, July 2020.

[46] T. Haidegger, M. Barreto, P. Gonçalves, M. K. Habib, S. K. V. Ragavan, H. Li, A. Vaccarella, R. Perrone, and E. Prestes, "Applied Ontologies and Standards for Service Robots," *Robotics and Autonomous Systems*, vol. 61, no. 11, pp. 1215–1223, 2013.

[47] J. Forlizzi and C. DiSalvo, "Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 258–265, 2006.

[48] N. Savela, T. Turja, and A. Oksanen, "Social Acceptance of Robots in Different Occupational Fields: A Systematic Literature Review," *International Journal of Social Robotics*, vol. 10, no. 4, pp. 493–502, 2018.

[49] C. Webster and S. Ivanov, *Robots in Travel, Tourism and Hospitality: Key Findings from a Global Study*. Varna, Bulgaria: Zangador, 2020.

[50] S. Ivanov, C. Webster, and P. Seyyedi, "Consumers' Attitudes Towards the Introduction of Robots in Accommodation Establishments," *Tourism: An International Interdisciplinary Journal*, vol. 66, no. 3, pp. 302–317, 2018.

[51] C.-A. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "Older Adults' Preferences for and Acceptance of Robot Assistance for Everyday Living Tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, pp. 153–157, 2012.

[52] S. Ivanov and C. Webster, "What Should Robots Do? A Comparative Analysis of Industry Professionals, Educators and Tourists," in *Information and Communication Technologies in Tourism 2019: Proceedings of the International Conference 2019*, pp. 249–262, Springer, January 30 - February 1, 2019.

[53] K. F. MacDorman and H. Ishiguro, "The Uncanny Advantage of Using Androids in Cognitive and Social Science Research," *Interaction Studies*, vol. 7, no. 3, pp. 297–337, 2006.

[54] R. Stock-Homburg and M. Hannig, "Is There A Privacy Paradox in The Workplace?," in *Proceedings of the Forty-First International Conference on Information Systems*, pp. 1–17, December 12-17, 2020.

[55] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional Networks and Applications in Vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pp. 253–256, 2010.

[56] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[57] P. Ekman and W. V. Friesen, "Constants Across Cultures in the Face and Emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, p. 124, 1971.

[58] D. Lundqvist, A. Flykt, and A. Öhman, "The Karolinska Directed Emotional Faces (KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, vol. 91, no. 630, pp. 2–2, 1998.

[59] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," *International Journal of Synthetic Emotions*, vol. 1, no. 1, pp. 68–99, 2010.

[60] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3 - A Wearable Wireless Multi-Sensor Device for Real-Time Computerized Biofeedback and Data Acquisition," in *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies*, pp. 39–42, November 3-5, 2014.

[61] M. M. Bradley and P. J. Lang, "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, pp. 49–59, March 1994.

[62] J. A. Russell, "A Circumplex Model of Affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. Publisher: American Psychological Association.

[63] F. Alamudun, J. Choi, R. Gutierrez-Osuna, H. Khan, and B. Ahmed, "Removal of Subject-Dependent and Activity-Dependent Variation in Physiological Measures of Stress," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, (San Diego, CA, USA), pp. 115–122, IEEE, May 2012.

[64] C. Atzmüller and P. M. Steiner, "Experimental Vignette Studies in Survey Research," *Methodology*, vol. 6, no. 3, pp. 128–138, 2010.

[65] H. Aguinis and K. J. Bradley, "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies," *Organizational research methods*, vol. 17, no. 4, pp. 351–371, 2014.