Sabine T. Koeszegi / Markus Vincze (Eds.)

TRUST IN ROBOTS

TU WIEN Academic Press

Sabine T. Koeszegi / Markus Vincze (Eds.)

TRUST IN ROBOTS

This publication emerged from the cooperation between TU Wien Bibliothek
and the Doctoral College “Trust in Robots – Trusting Robots”.

Sabine T. Koeszegi / Markus Vincze (Eds.)

TRUST IN ROBOTS

TU WIEN Academic Press

Cite as:

Koeszegi, S. T., & Vincze, M. (Eds.). (2022). *Trust in Robots.* TU Wien Academic Press.
https://doi.org/10.34727/2022/isbn.978-3-85448-052-5

TU Wien Academic Press 2022

c/o TU Wien Bibliothek
TU Wien
Resselgasse 4, 1040 Wien
academicpress@tuwien.ac.at
www.tuwien.at/academicpress

ISBN (print): 978-3-85448-051-8
ISBN (online): 978-3-85448-052-5

Available online: https://doi.org/10.34727/2022/isbn.978-3-85448-052-5

Media proprietor: TU Wien, Karlsplatz 13, 1040 Wien
Publisher: TU Wien Academic Press
Editors (responsible for the content): Sabine T. Koeszegi and Markus Vincze
Production: Facultas Verlags- und Buchhandels AG

Preface

Honesty, responsibility and accountability in all facets of research and university education are the foundations of society's faith in science and technology. These principles serve as the foundation for academic independence and the highest ethical and integrity standards. Therefore, "Technology for People" is the university's mission statement. We want to transform what is technologically feasible into what is desirable from a human-centered perspective. Innovative goods, services and procedures ought to make the world a better place to live in terms of compassion and social responsibility. TU Wien has made significant financial investments in multidisciplinary research carried out in doctoral colleges to address urgent societal concerns to further this purpose and uphold the highest standards of science and ethics. The doctoral college "Trust in Robots", led by Sabine Koeszegi & Markus Vincze, was established in 2018 to understand how we can build disruptive robotic and artificial intelligence (AI) technologies that people trust. Robotics and AI have the potential to help us overcome several problems, including the aging population crisis and climate catastrophe. The doctoral college "Trust in Robots" addresses this area of friction and bargains over the compatibility of technology and moral principles.

"Trust in Robots" has been set up as a transdisciplinary doctoral college in which postgraduate students and professors of various academic disciplines collaborate to understand the same phenomenon from different perspectives. From an institutional standpoint, the College's setup has been difficult because the systems and policies currently in place are not appropriate for admitting students with different academic backgrounds into the same study program for transdisciplinary research. However, the success of this doctoral college proves that this is how we must perform research in the future to overcome the existing silos of disciplines. The college has inspired certain changes that have been implemented in the Doctoral School of TU Wien and can serve as a model for future research projects. The introduction of the lecture "Responsible Research" for all doctoral students at TU Wien is one of the Trust Robots Doctoral College's most important accomplishments from our perspective. In this lecture, we consider ethical standards and the societal effects of innovation and science while preparing our students with morally sound design and trustworthy research techniques.

The doctoral college "Trust in Robots" is an unqualified success for TU Wien. The results of a four-year project at TU Wien are summarized in the twelve chapters of this book, which we are happy to release to the public.

Kurt Matyas (Vice Rector for Academic Affairs)
Johannes Fröhlich (Vice Rector for Research and Innovation)

Editorial

Robots are gradually becoming a part of our daily lives, populating our living and working spaces. We hope that robots will come to relieve us from chores and dangerous, dull, or dirty work. We believe that they can make our lives more comfortable, easier, and even more enjoyable by providing companionship and care. Hence, robots will change how we collaborate and assign tasks to human and machine agents and even—more fundamentally—how we live and perceive ourselves and our roles in society. Although we believe we have control over the machines we have built, this belief may fade as devices become more significant, autonomous, and influential. The independent actions of robots can be frightening. Thus, developing technology for people requires that we are—at all times—in control of the technology or that we can rely on the good intentions and safety of autonomous systems over which we have no control. Therefore, building trust in (autonomous) robot systems is necessary.

Trust has been an essential issue in automation and technology research since the 1980s. According to studies on interpersonal trust, trust as an attitude develops into reliance and so plays a crucial role in technology acceptance and appropriate use of automation. Furthermore, research indicates that the same social heuristics used in human–human interactions may apply to human–robot interactions (HRIs) because robots trigger similar social attributes as humans. Although previous research revealed disparities between trust in and reliance on technology and trust among people, this difference may become more blurred as robots increasingly mimic human interaction patterns and exhibit anthropomorphic appearance and behavior.

This problem is reflected in the title of this book, “Trust in robots—Trusting robots,” which carries different notions and unifies various research areas. While “Trust in robots” addresses the subject of how to develop technology that users are willing to rely on, “Trusting robots” focuses on the process of establishing a trusting relationship with robots, thereby extending previous research. This latter interpretation of trusting robots—although still to a great extent futuristic—poses the question of how to develop artificial intelligence and robotic technology that allows a robot to exhibit trusting skills when interacting with humans. It considers that humans may develop relationships with robots that go beyond technology acceptance and reliance. Thus, trust in this context does not only refer to the one-sided confidence of users toward robots but also to users’ need to be assured that robots incorporate notions of the meaning of objects and social norms, including biases, and have an understanding of scenes and situations to be capable of interacting with users socially. However, the mere possibility that we may develop bonding and emotional attachment to machines raises several ethical questions and concerns. Is it ethical to design devices that trigger trust and relationship building? Should robots simulate trustworthy behavior to start reciprocation by their users? Does trust in robots increase the vulnerability of users? How can we increase transparency regarding the capabilities of robots to ensure that users understand what robots can and should do? Should robots mimic other human qualities—such as empathy or emotions—to enhance trust?

These questions and topics have been the core of the “Trust Robots” doctoral college at TU Wien. The main aim was to comprehensively analyze trust in the context of robotic technology from various perspectives. The book presents the results of the 4 year endeavor of doctoral students—from fall 2018 to fall 2022. Before summarizing their contributions, let us briefly discuss the critical scientific challenges in transdisciplinary research.

Scientific Challenges and Transdisciplinary Research

On the one hand, building trust in robot systems entails endowing robots with capabilities and skills to perceive and understand human communication and behavior (for example, through natural language processing, by recognizing facial expressions, voice, gestures, and emotions); to recognize and ideally predict human intentions; and to adequately respond to all of these stimuli. Furthermore, any robot reaction must guarantee users that they are safe at all times and that human rights are respected and ensured. On the other hand, humans must perceive robots as safe and reliable. Since it is impossible to foresee or enumerate all possible situations, autonomous (social) robots must respond securely to unexpected and unforeseen encounters. They must be able to learn and adapt, as they will be tasked with making independent decisions that go far beyond the preprogrammed security rules and algorithms. In such a context, robots are ascribed and will have (social) agency.

To address these research issues, researchers from different disciplines must collaborate to pool their expertise, methodologies, and knowledge. The envisioned assistance from robots to improve the quality of life and work can only be realized responsibly when the issues associated with this technology are considered appropriately. Consequently, there is a need to discuss and understand possible future scenarios from different perspectives: technological (i.e., implementing aspects of trust on robots), human (i.e.,., deployment of trustworthy robots in work and social contexts), and societal (i.e., legal, ethical, political, and sociocultural aspects).

Our research is based fundamentally on the sociomateriality paradigm, which holds that sociocultural processes and technology and its applications are inherently entangled and cannot be analyzed separately.

Furthermore, since industrial and social robots are intelligent, autonomous machines that lack moral capacity, scientists and developers must assume responsibility for ethically aligned design from the outset. Hence, ethical robotics begins with R&D rather than mitigating the adverse effects and harm caused by new technologies after they are introduced. Therefore, our research is guided by the principles of the responsible robotics paradigm and focuses on the ethical concerns associated with the incorporation of robots into society.

The faculty and students of the doctoral college are truly interdisciplinary: they have backgrounds in the philosophy of science, design science, labor science, economics, social science, psychology, computer science, mechanical engineering, and electrical engineering, and they have worked on 12 different topics arching from a principle design-perspective on sociotechnical systems over joint attention and motion planning to adaptive task sharing in human–robot collaboration and a general reflection of trustworthy robots in society. Figure 1 shows an overview of the transdisciplinary research at the doctoral college.

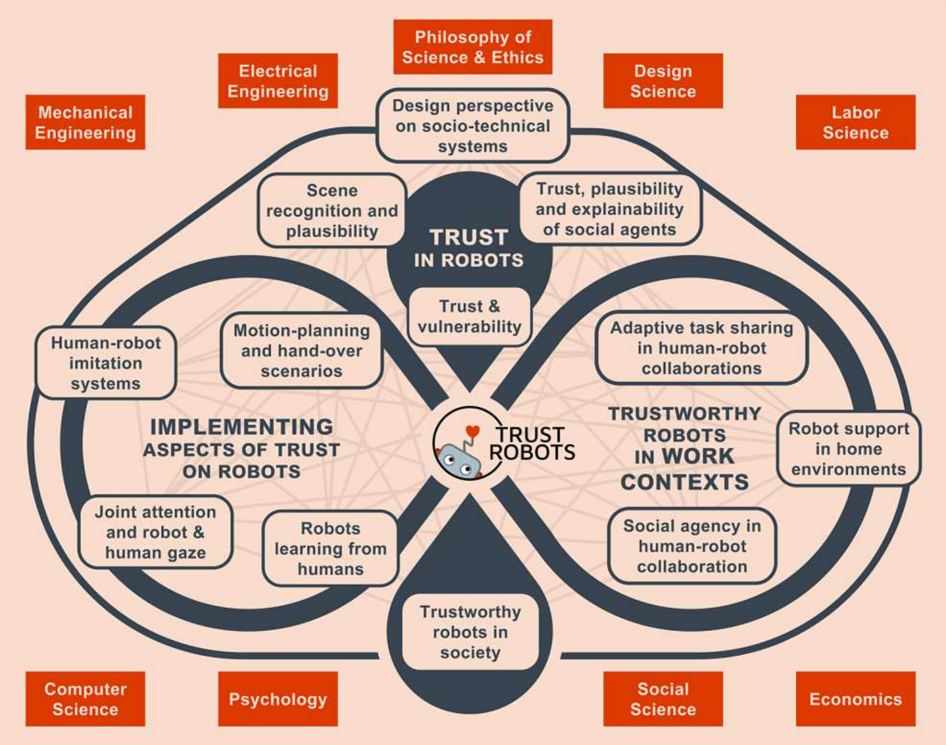
Philosophy of Science & Ethics
Electrical Engineering
Design Science
Mechanical Engineering
Labor Science
Design perspective on socio-technical systems
Scene recognition and plausibility
Trust, plausibility and explainability of social agents
TRUST IN ROBOTS
Trust & vulnerability
Human-robot imitation systems
Motion-planning and hand-over scenarios
Adaptive task sharing in human-robot collaborations
IMPLEMENTING ASPECTS OF TRUST ON ROBOTS
TRUST ROBOTS
TRUSTWORTHY ROBOTS IN WORK CONTEXTS
Robot support in home environments
Joint attention and robot & human gaze
Robots learning from humans
Social agency in human-robot collaboration
Trustworthy robots in society
Computer Science
Psychology
Social Science
Economics

Figure 1 Transdisciplinary fundamental technical and applied research on implementing aspects of trust in robots

This work completed in the doctoral college is genuinely transdisciplinary. Students from different disciplines collaborated to develop implementations on robots, design experiments and demonstrations, analyze data, and draw conclusions from the findings for the field of HRI and their core disciplines. This has led to the profound understanding that studying robots requires considering the entire sociotechnical system and context. This comprehensive perspective allows for designing meaningful and ethical robotic technology that will meet our expectations of making our lives easier and more enjoyable.

Summary of Results

The collection of articles in this book presents the highlights of the work on trustworthy robotics. We divide the summary into five sections: designing trustworthy robots, discussing trust and plausibility, implementing aspects of faith in robots, proposing that trustworthy robots must be viewed in the work context, and suggesting that trustworthy robots should be regarded in society.

Designing of Trustworthy Robots

In the first chapter of this book, Frijns & Schürer analyzed the contributions and importance of design work in the field of HRI research. They proposed that how interaction is conceptualized fundamentally impacts the design space and hence has to be considered in robotics research. Frijns et al. convincingly argued that the design space(s) for HRI must be extended beyond the individual aspects of humans and robots and encompass the sociotechnical system for which the robot is built. They make significant contributions to HRI through these design practice lenses.

Trust and Plausibility

The practical value of trust is founded on previous research findings that trust facilitates technology acceptance. Hannibal, Weiss & Purgathofer expanded on the perspective of "Trusting Robots" by providing a systematic identification of situational, robot-specific vulnerabilities in HRI. Hence, Hannibal, Weiss & Purgathofer shifted our focus to the contextual setting in which HRI occurs, challenging the prevalent negative association between interpersonal trust and vulnerability from both a theoretical—philosophical—and empirical perspective.

Based on the same fundamental idea of the relevance of context for HRI, Papagni & Koeszegi argued that for robots to be accepted within society, nonexpert users must find them valuable and trustworthy. They proposed to design robots that explain their decisions and actions to nonexpert users within the context of everyday interactions. Furthermore, they propose a model in which the plausibility of explanations resulting from contextual negotiations between the parties involved determines the understanding and supporting trust.

Bauer & Vincze applied this plausibility of explanations to the concrete scenario of scene interpretation, a core element of robots interacting in the world and with people. It first presents the technical approach to creating an object hypothesis using learned methods and then employs a verification process to obtain relationships between objects in the scene. The work shows that such scene-level information should be used to estimate object poses. Their primary assumption is that all object hypotheses concerning their visual observation and the physical scene in which they reside must be plausible. These scene interpretations are then employed in reasoning strategies to explain to the user what the robot perceives during HRI.

Implementing Aspects of Trust in Robots

The following studies focused on how to implement these various aspects of trust in robots and trusting robots into technology.

Stoeva & Gelautz presented a framework for a human–robot imitation system and examined the system requirements imposed by different interactions for communicative, functional, artistic, or abstract movements. The analysis identifies open challenges for designing and developing human–robot imitation systems, such as the difficulty of observing and accurately replicating human motions and how to transfer human to robot motions given different embodiments (correspondence) and even measuring the deviations. The study also addresses ethical issues, such as keeping privacy, not deceiving interactants, and correctly employing the robot system as intended and agreed upon.

Following the interpretation of human gestures, the robot might contact the human, as in a hand-over scenario. Beck & Kugi investigated motion planning specifically for such trustworthy human–robot collaboration, emphasizing the significance of ensuring human safety and comfort during the interaction. Concerning comfort, the study emphasizes fluency (a high level of coordination between humans and robots, resulting in accurately timed, and efficient sequences of action), legibility (a measure of how well the robot conveys its intent), and human-like motion. The study introduces a receding horizon trajectory optimization approach to achieve such behavior, where the requirements for safety and comfort during the interaction are formulated in objective functions.

Another critical aspect of fluent interaction is for the robot to understand the intention of the human user and to build a mutual understanding of the subsequent actions. Koller, Weiss & Vincze studied this joint attention perspective using a robot and human gaze behavior during collaborative actions. The study reviews research on joint attention and the theory of mind as foundational elements for the success of collaborative tasks in human–human interaction. The authors employ the research approaches of roboticists to provide robots with a joint attention capability or at least the technically feasible equivalent. The idea is that mechanical gaze behavior, which humans can easily comprehend, will improve the interaction capability of a social robot. This is evaluated in an already established HRI joint action benchmark scenario of collaboratively building a tower out of different blocks.

Finally, it would be great if robots could continue to learn from humans in everyday life scenarios. To achieve this, Hirschmanner & Vincze proposed using a grounded language learning approach to connect words and references in social spaces, such as objects. The authors presented a Pepper robot-based incremental word learning system. Then, they introduce how to learn specific low-level activities through demonstrations. Furthermore, they present systems with an industrial robotic arm and a dexterous robotic hand as concrete examples. Additionally, they address the role of the teacher in the learning process, determining which human factors are essential to facilitate the learning process.

Trustworthy Robots in Work Contexts

As previously stated, developing trustworthy robots requires considering the system's context. Thus, we must also study the context in which robots are deployed. The imagination of the role of robots is often driven by technology and top–down ideological agendas, without regard for the practical realities of everyday life and work contexts. Schwaninger, Weiss & Fitzgerald explored bottom–up HRI research in the context of home environments and robot support for older adults. Furthermore, the study presents an overview of assistive technology for home environments, the building blocks for HRI research in these contexts, and the issues of elderly support and care.

Zafari & Koeszegi addressed questions regarding the extent to which robots are accepted in work settings, as well as the impact human–robot collaboration has on workers and their perceptions of their own and the robot's role, agency, and efficacy. They show how agency is ascribed to nonhuman entities and present two experiments that analyze this impact. Zafari et al. provided valuable recommendations for both the design of artificial agents and organizational strategies in terms of which social practices and changes in the working context must provide opportunities for a successful collaboration.

Schmiedbauer & Schlund addressed another essential aspect of successful human–robot collaboration: how to allocate tasks between humans and robots. Instead, of automating all that can be automated and leaving the rest to humans, they employed a human factors approach and focused on the needs and capabilities of workers and economic targets at the center of analysis. They designed, developed, demonstrated, and evaluated a model for adaptive task sharing between humans and cobots (collaborative robots) and showed avenues for further development based on their insights.

Trustworthy Robots in Society

Finally, DePagter provided a macrolevel analysis, i.e., an analysis of the process of building trust in robots on a societal level. They proposed a narrative approach and argued that robots are a prominent example of a technology that has caught many people’s imagination of the future. The analysis of these future imaginaries of robots provides a deep understanding of how technology is perceived by the general public, what fears and hopes are associated with this technology, what roles are given to robots, and what challenges are associated with them. This narrative approach provides avenues for policymakers and developers to shape future imaginaries of robots.

Sabine T. Koeszegi, Markus Vincze

Table of Contents

Preface v

Editorial vii

Sabine T. Koeszegi, Markus Vincze

List of Abbreviations xvii

Designing of Trustworthy Robots 1

Design as a Practice in Human-Robot Interaction Research 3
Helena Anna Frijns, Oliver Schürer

Trust and Plausibility 31

Exploring the Situated Vulnerabilities of Robots for Interpersonal Trust
in Human-Robot Interaction 33
Glenda Hannibal, Astrid Weiss

Challenges and solutions for trustworthy explainable robots 57
Guglielmo Papagni, Sabine T. Koeszegi

Visual and Physical Plausibility of Object Poses for Robotic
Scene Understanding 81
Dominik Bauer, Timothy Patten, Markus Vincze

Implementing Aspects of Trust in Robots 105

Design, Requirements, and Challenges of a Human-Robot
Imitation System 107
Darja Stoeva, Margrit Gelautz

Motion Planning for Human-Robot Collaboration 129
Florian Beck, Andreas Kugi

I See What You Did There: Towards a Gaze Mechanism for
Joint Actions in Human-Robot Interaction 149
Michael Koller, Astrid Weiss, Markus Vincze

Robot Learning from Humans in Everyday Life Scenarios 179
Matthias Hirschmanner, Markus Vincze

Trustworthy Robots in Work Context **201**

Bottom-Up Research on Assistive Robots for the Aging Population 203
Isabel Schwaninger, Astrid Weiss, Geraldine Fitzpatrick

Agency in Sociotechnical Systems: How to Enact Human–Robot Collaboration 229
Setareh Zafari, Sabine T. Koeszegi

Adaptive Task Sharing between Humans and Collaborative Robots in a Manufacturing Environment 245
Christina Schmidbauer, Sebastian Schlund

Trustworthy Robots in Society **263**

Building trust in robots: A narrative approach 265
Jesse de Pagter

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AAL	Active and Assisted Living
ADD	Average Distance of Model Points
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANT	Actor-Network Theory
AR	Average Recall
ATS	Adaptive Task Sharing
AUC	Area Under the Curve
BPMN	Business Process Model and Notation
CCVSD	Care Centered Value Sensitive Design
CHOMP	Covariant Hamiltonian Optimization for Motion Planning
CNN	Convolutional Neural Network
CPPS	Cyberphysical Production System
CSCW	Computer Supported Cooperative Work
CV	Computer Vision
DDP	Differential Dynamic Programming
DDPG	Deep Deterministic Policy Gradient
DOF	Degrees of Freedom
DTMC	Discrete-Time Markov Chains
EC	European Commission
EED	Eye-Direction Detector
EJA	Ensuring Joint Attention
EU	European Union
GDPR	General Data Protection Regulation
GJK	Gilbert-Johnson-Keerthi

GMM	Gaussian Mixture Model
GUI	Graphical User Interface
GuSTO	Guaranteed Sequential Trajectory Optimization
HCI	Human-Computer Interaction
HER	Hindsight Experience Replay
HF/E	Human Factors / Ergonomics
HHI	Human-Human Interaction
HRC	Human-robot Collaboration
HRI	Human-Robot Interaction
HSR	Human Support Robot
ICP	Iterative Closest Point
ICT	Information and Communication Technology
ID	Intentionality Detector
IJA	Initiating Joint Attention
IL	Imitation Learning
IRL	Inverse Reinforcement Learning
ISO/TS	International Standards Organization / Technische Spezifikation
LfD	Learning from Demonstrations
LIDAR	Light Detection and Ranging Sensor
MCTS	Monte Carlo Tree Search
ML	Machine Learning
MPC	Model Predictive Control
MDP	Markov Decision Process
MTM	Methods Time Measurement
MTM-AUS	Methods Time Measurement - Universelles Analysier-System
npmi	normalized pointwise-mutual information
OMPL	Open Motion Planning Library
PD	Participatory Design
PDDL	Planning Domain Definition Language
PPO	Proximal Policy Optimization

STOMP	Stochastic Trajectory Optimization for Motion Planning
STRIPS	Stanford Research Institute Problem Solver
STS	Sociotechnical Systems
SUS	System Usability Scale
SVM	Support Vector Machines
ToM	Theory of Mind
ToMM	Theory of Mind Mechanism
UCB	Upper Confidence Bound
UI	User Interface
UX	UsereXperience
VSD	Visual Surface Discrepancy
XAI	Explainable Artificial Intelligence