ISSN 2186-7437

NII Shonan Meeting Report

No. 177

Natural Interaction with Humanoid Robots

Kristiina Jokinen
Wolfgang Minker
Catherine Pelachaud

January 23–26, 2023

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Natural Interaction with Humanoid Robots

Organizers:
Kristiina Jokinen (National Institute of Advanced Industrial Science and Technology, Japan)
Wolfgang Minker (Ulm University, Germany)
Catherine Pelachaud (Centre National de la Recherche Scientifique, Sorbonne University, France)

January 23–26, 2023

1 Description of the Meeting

Robotics has been a hot research topic for the previous decade and a daydream of many fantasy writers and readers. Significant advancements have been achieved in this field, introducing robots to wide range of applications in everyday life. The scope of applications ranges from restaurant kitchens to prepare the food, over elderly care to help the elderly get out of bed and provide a sense of companionship, up to vacuum cleaners. However, most robotics applications are intended to substitute well-defined and structured human labor or activities. Nevertheless, robotics is an interdisciplinary field of research, that attracts developments from various areas, such as computer vision, mechanical and control engineering, artificial intelligence, etc. Mostly, the applications of robotics focus on supporting a human or replacing him in some specific tasks. In order to integrate robots seamlessly into everyday life and to render it more empathic, humanoid robots have been introduced. They have increased both the research interest and the customer's demand over the last few years. Still, one important part that is missing is the natural interaction between robots and humans. Robots are controlled by a set of commands that usually have to be pronounced precisely in order to have an effect, which introduces a gap between a human user and a robot.

On the other hand humanoid robots if integrated into an everyday life, have a great potential to collect the data and interact with humans in the widest range of situations. This enables to conduct research beyond laboratory conditions, to which current research is predominantly limited, towards naturalistic ones. With an ease of data collection, the data-driven approaches will get prerequisites to be developed and solve the task successfully. Having information from other types of devices and sensors available, robots are more likely to produce a relevant, context-based, meaningful response, compared to an isolated dialogue system. It integrates the rapidly developing area of pervasive computing.

Interdisciplinary work, combining research on humanoid robotics and spoken dialogue-based interactions, may have a significant impact on the development of natural interaction between humans and human-like robots. Spoken dialogue systems interact with users employing speech in order to e.g. provide them with specific automated services, give intuitive access to information/data and enable to adapt to their preferences and expectations. Integrating dialogue systems into robots will help to broaden the group of its users, which makes it possible to greatly expand the amount of applications e.g. sensitive listening, emotion monitoring, stress detection and prevention, etc. Using the insights into interaction between users and humanoid robots and improving the naturalness of the latter, robots will be perceived as more trustworthy and reliable, barriers and fears of contact will be dismantled. To some extent robots will become more human-like or humanoid. However, this will entail the need for the exchange with additional research fields, such as psychology, sociology and medicine.

While the aforementioned research fields have recently been of interest for the respective research groups, questions of Natural Interaction with Humanoid Robots (NIHR) have not been discussed sufficiently in each community. An active cooperation between these communities will create a platform to exchange ideas and benefit from complementary work. We aim to create the first and unique venue for discussion and collaboration between experts from these disciplines. Therefore we plan to divide the participants according to their interests and research fields into Working Groups which are concerned with the topics mentioned below. Each group will be asked to document the outcome of their discussions and prepare a presentation for the other groups, which will be held at the end of each meeting day. The planned Shonan Meeting will help to explore possible challenges and jointly develop a research agenda for main directions. The meeting will establish a platform for an international collaboration for the next three to five years.

Research challenges in development:

- Challenges of NIHR.
- Role of multi-modal in- and output (speech, gestures and emotions).
- Dialogue modeling and appropriate response generation for expressive and adaptive human-robot systems.
- Single and multiple user identification, modelling and tracking in NIHR.
- Context awareness (automatic detection of stress, user status and environmental context).
- Personalization and user-centered development.
- Adaptation mechanisms.
- Evolution of the technology and ideas for future research and application scenarios.
- Role of grounding and embodiment in language and dialogue.
- Psychological issues and learning effects in NIHR (robots vs. users).
- Role of personality in adaptive human-robot systems.
- Trust and reliability in NIHR.
- Theory of mind for NIHR.

Use cases, user groups and industrial applications:

- Appropriate user groups for NIHR (e.g. elderly, youngsters, and school kids).
- Specific application domains for NIHR:
  - robot-assisted stress prevention.
  - robots as sensitive listeners.
  - public space (e.g. interactive digital signage, guidance systems).
  - assistive environments (e.g. elderly care, hospitals).
  - education
- Success stories, functional systems and industrial challenges.

Development, testing and evaluation:

- Experimental design, user studies and evaluation of NIHR.
- Investigation of long-term vs. short-term relation in NIHR.

Ethics and societal impact:

- Legal issues.
- Social responsibility.
- Data protection and privacy by design and default.
- Social design and development of naturally interacting human-robot systems.

2 Meeting Schedule

Natural Interaction with Humanoid Robots					
Time Table	Arrival Day	1st Day	2nd Day	3rd Day	Final Day
	January 22nd (Sunday)	January 23rd	January 24th	January 25th	January 26th
7:00 - 7:30					
7:30 - 8:00		Breakfast	Breakfast	Breakfast	Breakfast
8:00 - 8:30					
8:30 - 9:00			**Keynotes:** Kaoru Sumi (Topic 1)	Working Groups - Session 4 (Topic	Working Groups - Elaboration of final
9:00 - 9:30		Introduction	Keichii Yasumoto (Topic 2)	-related)	final results
9:30 - 10:00		Pecha Kucha 1	Silvia Rossi (Topic 3)		
10:00 - 10:30		Break	Break		
10:30 - 11:00		Pecha Kucha 2	WG - Session 2	Break	Break
11:00 - 11:30		WG Allocation & Photo Shoot		Discussion of final results	Presentation of final results
11:30 - 12:00					Wrap up and Farewell
12:00 - 12:30	Early check-in (negotiable)	Lunch	Lunch	Lunch	Lunch
12:30 - 13:00					
13:00 - 13:30				Excursion:	
13:30 - 14:00		**Keynotes:** Michael Spranger (Topic 2)	**Keynotes:** Satoshi Nakamura (Topic 1)	Visiting Jomyoji and Hokokuji temple	
14:00 - 14:30		Maria Ines Torres (Topic 1)	Graham Wilcock (Topic 2)	with Japanese Tea ceremony	
14:30 - 15:00		Nick Campbell (Topic 3)	Tony Belpaeme (Topic 3)		
15:00 - 15:30	Regular check-in	Break	Break		
15:30 - 16:00		WG - Session 1	WG - Session 3		
16:00 - 16:30					
16:30 - 17:00					
17:00 - 17:30			Discussion of		
17:30 - 18:00			Preliminary Results		
18:00 -18:30		Dinner	Dinner		
18:30 - 19:00					
19:00 - 19:30	Welcome Banquet			Banquett	
19:30 - 20:00					
20:00 - 20:30					
20:30 - 21:00					

3 Working Groups

The participants of this meeting were divided into three Working Groups (WG) and discussed the following three topics.

- Research challenges in the development
- Use cases, user groups, and industrial applications
- Development, testing, evaluation (short-/long-term in the wild), ethics and societal impact

Overview of Talks

Speech aware Dialogue management

Prof. Dr. M. Ines Torres, Universidad del País Vasco UPV/EHU, Spain

Spoken Dialogue systems need to focus on close domains to be useful to solve specific tasks in collaboration with humans. In this framework, context information is essential for the DM to make appropriate decisions. Thus, defining which context is necessary and how to get and manage it, is an open research question. The talk focused on the information that speech, and audio, analysis can provide. Therefore, the role of emotion and other related features that can be extracted from the speech was stressed. Afterward, the aspect of how speech representations can benefit the policy optimization when they directly fed the DM, was discussed.

Outracing Champion Gran Turismo Drivers with Deep Reinforcement Learning

Dr. Michael Spranger, Sony AI Inc., Japan

Many potential applications of artificial intelligence involve making real-time decisions in physical systems while interacting with humans. Automobile racing represents an extreme example of these conditions; drivers must execute complex tactical manoeuvres to pass or block opponents while operating their vehicles at their traction limits. Racing simulations, such as the PlayStation game Gran Turismo, faithfully reproduce the non-linear control challenges of real race cars while also encapsulating the complex multi-agent interactions. It was described how Spranger et al. trained agents for Gran Turismo that can compete with the world's best e-sports drivers using Reinforcement Learning. The capabilities of their agent, Gran Turismo Sophy, was demonstrated by winning a head-to-head competition against four of the world's best Gran Turismo drivers. By describing how they trained championship-level racers, the possibilities and challenges were demonstrated to control complex dynamical systems in domains where agents must respect imprecisely defined human norms.

“Natural” social interaction with non-human entities

Prof. Dr. Nick Campbell, Trinity College Dublin, Ireland

The talk included slides from various stages of Nick Campbell's career which illustrate the corpora he has collected and drawn conclusions from these studies. Especially it was shown that eyes are as important as ears when talking (either with humans or with robots) and that certain e-words might be more relevant than others when it comes to spoken interaction. Thus, this keynote was focused on “social interaction” but also encompassed many aspects of human-human and human-machine discourse.

Design of Human-Interacting Robots and Virtual Agents that learn from Affective Computing

Prof. Dr. Kaoru Sumi, Future University of Hakodate, Japan

Because humans are emotional creatures, we treat artifacts as having feelings. Today, humans and robots aretoot on equal footing. When humans interact with robots, we treat them simply as machines. We can thus examine how to make humans and artifacts such as robots communicate as equals, like friends. To be treated as equal, such artifacts should be sufficiently intelligent, useful, and trustworthy. For an artifact to be considered intelligent, useful, and trustworthy, it should be able to sense a human's current situation. That is, we can have a good relationship with artifacts that sometimes do things that make us want to thank them, and that understand us and our feelings. For humans and artifacts to interact with each other on an equal footing, the most desirable artifact might be one that can sense a human's current situation, empathize with it, and take some action in response. For example, an intelligent salesperson artifact could detect and empathize with a human being's current situation, thereby increasing sales. According to our research results, the impression of empathy through facial expressions is very important in establishing such relationships with humans. We have found that facial expressions and verbal responses affect persuasiveness. Using these results, we developed a human-agent interaction-based facial expression training system and applied it in customer service. Similar ideas have been considered in recent years, and there seems to be a need for such systems. Recently, our laboratory has also conducted research in the field of education. Specifically, we detect emotions from learners' facial expressions, gestures, and actions, and we give them appropriate hints. In the past two to three years, distance education has rapidly become popular, and communication using virtual space and avatars (virtual agents) seems to be gaining popularity. People enjoy communication by controlling their avatars remotely with 6DoF, thus becoming a different person. In the future, distance education and experiential learning will use virtual spaces. In such cases, we should consider what the avatars' characteristics should be. There have been various studies on how avatars should be used, and performance differences based on gender, personality, and shape have also been studied. Currently, models of virtual agents are becoming more realistic and of higher quality. The conditions are now in place for a human being to be able to replace another human being. Because virtual space is indeed virtuto, it can be changed and displayed in any way. I would like to pursue the linkage between virtual space and real space, including what kind of virtual space or virtual agent can improve a person's performance and his/her perception of his/her current situation in real space.

Title: Past efforts and future challenges in AIoT research toward realizing smarter homes, life, and cities

Prof. Dr. Keiichi Yasumoto, NAIST, Japan

The development of IoT and AI technologies has made it possible to understand people's activities and environmental conditions in various locations in real time. In the home, by recognizing people's daily activities, it is possible to understand their health status and recommend activities to improve their quality of life (QoL). In urban areas, the system can help people decide where to go and when to travel by recognizing the conditions of each location. However, to make these context-aware technologies pervasive in our living space, various issues must be resolved, especially how efficiently/effectively data collection and feedback are performed from/to humans/environments. In this talk, we will introduce our past efforts and future challenges on how these issues can be solved through IoT and AI technologies.

Development, testing, evaluation (short-/longterm in the wild), ethics and societal impact

Prof. Dr. Silvia Rossi, Universita' degli Studi di Napoli Federico II, Italy

In this talk, some examples of healthcare applications with children and elderly people were presented, discussing the challenges in the development and testing of autonomous applications to be deployed into the wild in comparison to lab evaluations. Additional challenges are required when aiming at long-term interaction and relationships, whereas a model of the user must be learned in terms of privacy and legal issues. Finally, the talk focused on the evaluation of NIHR in terms of legibility and predictability. This kind of non-functional evaluation will require new metrics and may highlight relevant differences in terms of the naturalness of the interaction between people or between people and robots.

Research challenges in embodied virtual agent system for social skills training

Prof. Dr. Satoshi Nakamura, NAIST, Japan

Communicating with others, and giving a presentation in front of a (even small) public is not always an easy task. Some participants may have strong difficulties standing in front of others, and some may suffer from social anxiety disorders. Social skills training (SST) is a type of behavioral therapy for people with mental disorders or developmental disabilities. SST is normally given by psychiatrists, therapists, or other professionals. This talk introduced their project on developing a conversational virtual agent for Social Skill Training (SST) in various situations. Target populations are healthy control, social anxiety disorder, and an Autism spectrum disorder. It was discussed how to design the character, behavior, and feedback of the agent. The talk also included a demo video of the system.

Natural Interaction with Humanoid Robots (NIHR) - Topic 2 use cases, user groups and industrial applications

Prof. Dr. Graham Wilcock, University Helsinki, Finland

This keynote talk was given with respect to Topic 2:“Use cases, user groups and industrial applications”. To start with the talk reflected on which user groups are appropriate for NIHR (e.g. elderly, youngsters, school kids) and more explicitely if it is in the interest of those users or rather the researchers to conduct respective experiments implying to focus on the benefit of the users when taking part in experiments. Afterward potential specific application domains for NIHR are reviewed. Firstly, robot-assisted stress prevention was discussed as well as the question how to prevent robot-aggravated stress underpinned with demo videos. Secondly, the application of robots as sensitive listeners was shown by giving an overview from ELIZA to ERICA, and from WikiTalk to WikiListen. Thirdly, guidance systems in public space were discussed illustrated by the example of a coinlocker guidance system (From Irrashaimase! Irrashaimase! to Coin lockers are in 3 locations). Finally, different success stories, functional systems and industrial challenges. And with regard to the latter the specific aspect wether one can use ROS for HRI.

The bumpy road towards autonomous Human-Robot Interaction

Prof. Dr.Tony Belpaeme, Universiteit Gent, Belgium

While interactive systems and specifically Human-Robot Interaction have been studied for over 20 years, we do not yet see autonomous systems that can support open-ended social interaction. Instead, people have been taking shortcuts in restricted domains which for constrained applications seem to work well. However, the dream -of course- is to build true autonomous HRI rivalling human-to-human interaction. This talk will look into what would be needed for that and speculates if recent advances in data-driven AI can bring us there. Specifically, the very recent demonstrations of Large Language Models, such chatGPT, GPT3.5 and Claude, have shaken the field of Natural Language Interaction to its core, and with it the field of AI. What will this mean for the interactions one has with humanoid robots?

4 Summary of discussions

This section shall give an overview of the very fruitful and wide-ranging discussions during the Meeting.

Natural interaction with humanoid robots has the potential to provide a wide range of services to users. For robots to be economically viable and likely to be adopted, they should be able to provide multiple services to users. Natural interaction depends on the use cases, activities, and application, but in general we define natural interaction to refer to the robot's interaction capabilities that allow the users to have interaction in a smooth and intuitive manner without a long learning phase, conducted with multimodal natural interactions taking into account social requirements. We explored some of the potential use cases for natural interaction with humanoid robots, as well as the key requirements that must be met for these interactions to be effective. These are reported in the following sections: use cases, requirements, key beneficiaries and main challenges, possible short-term commercial applications, open research questions.

A survey [13] shows that "robots were more persuasive and perceived more positively when physically present in a user's environment than when digitally-displayed on a screen either as a video feed of the same robot or as a virtual character analog; robots also led to better user performance when they were collocated as opposed to shown via video on a screen."

This raises the question: How do we design robots that improve the world we live in? Any attempt to answer this question needs to start with a discussion about what "improve" means or who decides what is desirable or not. In other words, it requires a discussion about morality. In human-robot interaction (HRI), the discourse on ethics and values is in its infancy compared to human-computer interaction (HCI), though many facets on how technology should and can ethically develop translate over. For instance, humanoid robots as embodied AI systems enforce specific norms, such as what is an appropriate "body" size, type, or color, as well as choices on how a robot should talk, e.g., accent, gendered voice, that require an evolving line of research that is specific to HRI.

Established approaches like value sensitive design [58] can be overly broad to address specific ethical quandaries in HRI, like the legal establishment of intelligent robots as potential electronic persons by the Civil Law Rules on Robotics (European Parliament, 2017). Should robots be granted the same moral status as people, and in what ways, especially if human-human moral norms do not necessarily apply to human-robot interactions (e.g. [67])? It is important to build understanding about the values we apply when designing machines, what societal impact those machines have, and therefore, how we can evaluate these machines.

What deserves attention is what we mean by pursuing natural interaction with humanoid robots. "Naturalness" and "human-like" can be equated in how humanoid robots are designed, used, and presented. Natural interaction refers to the way in which humans interact with technology in a way that feels intuitive and comfortable, similar to how they interact with other humans or with their physical environment. In the context of human-robot interaction, natural interaction could mean that the robot is able to respond to human behavior and cues in a way that is similar to how a human would respond. However, the way a robot is behaving may not necessarily be fully similar to the one of a human as far as its behavior is legible and transparent to the human interacting partner.

Interaction capabilities may be asymmetrical but still natural. The discussion on naturalness and human-likeness of robots influences the path towards a specified research agenda on ethical considerations by the HRI community. This includes perspectives on broadening research methods, e.g., long-term field studies, and evaluation methods, e.g., privacy concerns during multi-modal, sensor-based studies, and what it means to be a human participant.

5 Summary of new findings

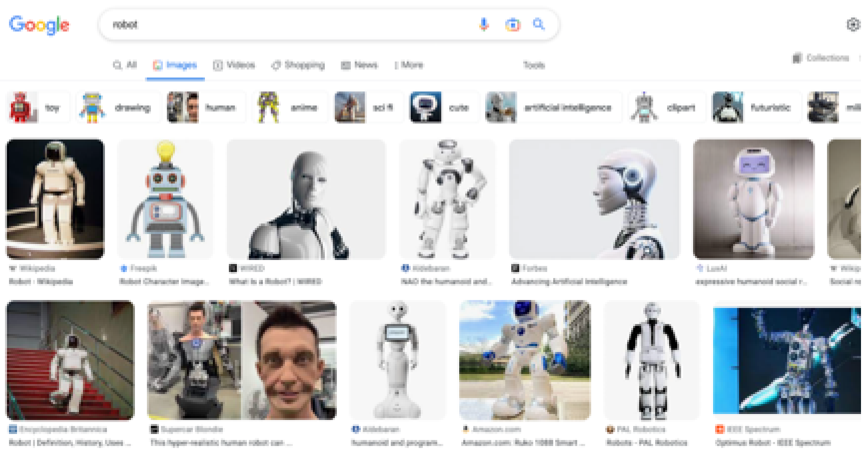
There were multiple interesting findings derived from exchange within each working group, which are presented in the following.

5.1 Ethics and grey areas

Normative ideas on what counts as “natural” and “human-like” can limit our approach to ethics with robots and ethics for robots. Ethics for robots means that humans identify and promote how robots should be designed. Ethics with robots stands for how humans should design ourselves through robots. The two are related concepts. Ethics for robots: Ethics and aesthetics are intertwined [59]. When we consider the form factor of robots, i.e., what robots look like, the most common humanoid robots have been criticized for containing ableist, sexist, and racists norms [54, 82]. For instance, commercial humanoid robots, so far, tend to look white with working appendages, such as arms and fingers that humans have (5.1 - Pepper robot). They contain ”ideals” on what bodies are worth having, which are laden with and promote racist notions [54]. A Google image search (January 25th, 2023 - Fig. 5.1) is filled with white robots, able-bodied robots as well, showing the limitation of our view of what counts as “being human” in terms of form factor.

5.2 Use Cases for NIHR

One potential use case for natural interaction with humanoid robots is providing support and assistance for technologically naive and untrained users. Robots can be designed to provide social companionship [18], for instance they can be attentive listeners [22] and provide psychological support to people who may



be isolated or lonely, such as older adults or people with disabilities, while providing support for daily activities [20]. Studies have shown that older adults who have established regular social relations have a lower risk of developing depression [44], dementia [43], or cognitive decline [45]. Robots that can understand and respond to human emotions can provide emotional support and act as conversational partners. Human-robot interaction during activities of daily living can be monitored and used to assess cognitive [33] and physical skills [37] of users in order to adapt to their level and to provide crucial information to carers and doctors about the mental and physical skills to tailor health interventions. Another potential use case is using humanoid robots with natural interaction to help people learn to move and interact with others in a more human-like way. This can include physical rehabilitation and cognitive rehabilitation, e.g. humanoid robots can potentially support the physical rehabilitation of post-stroke patients [25]. Natural interaction can also support social skills training for people with autism spectrum disorder, e.g by imitating human behaviour, robots can help people learn how to interact with others and learn new skills like physical imitation [23]. Intuitive natural interaction with humanoid robots can also enable their use in a variety of public services and facilities [28], such as receptionists in hotels, hospitals, and convenience stores [35], in which users are not familiar with the robot but must be able to interact and be served immediately. In addition, robots can be used to provide assistance in education by providing practical explanations, educational content, and psychological support for students and teachers [31]. Entertainment is another potential use case for humanoid robots with social abilities, which are being deployed in amusement parks [34], more details about this use case are provided in the section about short-term commercial applications. Robots can be used for entertainment, including sex robots, which may also be a feature of companionship [30]. Additionally, robots can be used as avatars, deadbots, and realistic telepresence, such as in the form of Hibari Misora [14] and Yumi Matsutoya [15] where AI avatar of the singers with realistic appearance and singing is developed. Deadbots are chatbots that simulate having a conversation with a deceased person by imitating their responses. Some are based on an approach patented by Microsoft [17] for copying the personality traits of an individual into a conversational bot. Others are based on OpenAI's GPT-3 or ChatGPT.

Finally, we propose that robots can be used in rescue operations. However, it can be debated whether humanoid robots are the most suitable for this purpose or not as this work is often performed by mobile robotic platforms. Dogs have long been important in rescue situations due to their powerful smell capabilities and ability to crawl through narrow spaces, while robot dogs have been developed mainly for domestic companionship and entertainment. Perhaps dog-like robots can be developed for rescue operations, and this raises questions about how to provide natural interaction with people to be rescued, and whether robot dogs should be able to talk. Probably dog-like search and rescue robots will be tele-operated by a human rescue team viewing images from the robot cameras, who will then talk to the located person to reassure them and plan the next actions. In [36], the current state of rescue robots is surveyed. It is important to note that these use cases are based on current research and development and may change as technology and societal needs evolve. Additionally, further research is needed to fully understand the potential benefits and limitations of natural interaction with humanoid robots in these different scenarios.

5.3 Evaluation of Natural Human-Robot Interaction

McGrath (1981) proposed to organize research strategies along two dimensions: obtrusiveness –the quality of being too noticeable– and universality. However, evaluations in HRI predominantly rely on laboratory experiments and survey studies, making the objects of study obtrusive and non-universal. Calls for a wider range of research approaches, especially towards studies in the field, have been made frequently [75, 73, 63]. In the wild studies are less common, but potentially provide richer data. Indeed, the development and testing of humanoid robots aimed at naturally interacting with people typically requires different stages and phases spanning functional testing, lab studies, and into-the-wild user studies, so requiring different approaches. For example:

- Early stage of the development: pilot testing, performance evaluation, lab studies, WOZ testing, ethnographic field studies, qualitative evaluations; lab user studies: behavioral analysis, user observations and interviews, functional development testing;

- Into the wild experimentation: interaction evaluation through questionnaires, objective measurements of performance. In this case, how to design objective, quantitative measures that do not remove ecological validity is a significant challenge.

There are many practical considerations for running diverse types of studies, which involve different research methods, quantitative and qualitative. Every study is going to be unique, but there is value in trying to follow methods that have historically proven to be effective [62]. The notion of a "user study" relates to addressing utility and usability of systems in HCI [61], from which studies in human-robot interaction (HRI) have evolved from.

Experimental methods used for research in psychology and behavioral science largely influenced how HRI researchers view studies, and many recent efforts can serve as points of departure. There are concerns about the validity of lab-based studies when it comes to developing useful systems for real people, even though lab-based studies serve well to test well-described questions.

Concerns include the sample size of one's study, given that a justification is required based on the desired or the smallest effect size of importance, a-priori power analysis, or resource constraints, among others [65, 55]. The intentions behind a study should be reported in advance, e.g., registered reports [70] and resources of many labs can be combined for collaboratively addressing issues regarding replicability [64]. Qualitative studies in HRI can also have well-defined research questions at the start, but also include generative designs in an iterative manner [79]. Many qualitative studies have focused on user observations, interviews and focus groups, for instance [79]. Designerly approaches to HRI in which designs are embodied knowledge utilize qualitative data, e.g., workshop results, as research contributions [68, 83]. There is a distinction between qualitative research that begins with a clear beginning and end points (qualitative research questions) and those that do not have a linear process. However, these types of distinctions should not be seen as having clear boundaries that drive mutually exclusive research paradigms. Rather than motivating the overall process, the type of evidence used can also be either quantitative or qualitative; Yin [81] argues that this distinction better applies to the evidence that is used to support an argument. For example, an ethnographic field study on robots that uses video recordings and surveys as part of the data collection method can rely on quantitative evidence to support its arguments. Or, a set of graphs generated from sensor data as part of a controlled laboratory study are discussed and serve as qualitative evidence to support an argument.

Current research does not cover the full range of available research strategies. Both objective and subjective evaluations are critical, although it is common for much research in HRI to focus predominantly on one or the other. Models of classification and interaction which show good "objective" performance in the lab or pilot studies may be completely unusable in real world conditions. Quantitative performance metrics often do not provide any indication to whether this is even noticeable by the end user.

Similarly, subjective evaluations are often done using Likert scale questionnaires, which reduces the complexity of interactions to a single number. Measuring human behavior is not enough to classify interactions as "good" and "bad" given how much variance there is between individuals. In several scenarios, a richer form of subjective analysis (e.g. interviews with end-users) would yield more valuable information than traditional questionnaires, even if the sample sizes of the former are relatively low.

5.3.1 What evaluation methods should be used?

Evaluation methodologies of robots for industrial and academia are crucial to distinguish. Lab studies are often evaluated in terms of statistical tests and behavioral performance improvements. The "evaluation" of commercial interests is arguably whether the robot sells more than the previous version. Typically, evaluation is considered in regards to the former, but if adoption of humanoid robots in society is the end goal, the latter framework should also be kept in mind. Expert involvement, for example psychologists and educators, are crucial in the development and implementation of robots and research setups in order to ensure that they are targeted towards the needs of the final user. In recent years, co-design or participatory development have become proven to be particularly effective in not only designing robots and their interactions, but also in formulating relevant evaluation questions. The power of iterative design, in which successive rounds of development and evaluation power system improvement, is too often ignored in academic research. Often due to the short-term nature of research, dictated by the short duration of research funding or personal research funds, and the desire to produce novel contributions, rather than incremental contributions. But iterative design, in which systems are developed over many years, have an important role to play in our field. In HRI there is a need for specific metrics properly validated and not merely extended from HCI. Moreover, experimentation in ecologically valid settings requires metrics should be suitable (or adaptable) to provide an evaluation at different stages of the interaction in terms of evaluation of expectations before the interaction, evaluation of the "progress" during the experimental phase (extremely important in case of long-term interaction), and evaluation after the end of the experimentation. Finally, some studies, such as the study on the link between watching television, physical activity and clinical depression [66], can take 10 years. We can expect studies to extend in time in HRI. Long-term user studies involve recording a history of the interactions while the data continuously increases during the experiment. A major issue is how to store, organize, and analyze this huge amount of collected data (sensory information, real-time ASR/TTS/dialogue models/language models. etc.). Summarization and preview methods should be developed in advance to aid in evaluation.

5.4 Methodologies for NIHR studies

Designing experimental natural interactive human-robot setups can be a complex task, as it involves a combination of elements from multiple disciplines, including robotics, human-computer interaction, and psychology. Some key factors to consider when designing such a setup include:

- Determining interaction scenarios and/or task: The specific scenarios in which the human and robot will interact should be defined in advance. With human-based studies, It is clearly not possible to fully "control" all aspects of the interaction and this should be recognized when designing the study.

- Identifying the ecological setting where the final setup will be tested, e.g., public vs. private settings, single vs. multi-users, silent vs. noisy, etc. All too often this is convenience based, a lab study is easier to set up than a real-world study, but careful consideration should be given to what environment will be most likely to add value to the evaluation, within the constraints of what is practically possible. Testing the intelligibility of a shopping mall robot is best done in a shopping mall, but practicalities might make it easier (or force us, in case of a pandemic) to evaluate our robot in the lab. Moreover, in lab settings the naturalness and quality of the interaction, as well as the performance are hard to assess since the users are typically involved in simulated scenarios and not committed towards the end goal of the application. Therefore, testing the robot in the ecological setting for which it is being developed should be preferred.

- Identify the users and stakeholder of the defined interaction scenarios. Different "categories" of users would have an impact on the evaluation methods to be considered.

- Articulating research questions: Having a clearly articulated research question is crucial when designing experimental setups. Research questions can help focus and guide research efforts. Research questions often evolve during research projects. Thus it is less important to have a clearly defined research question at the beginning of the project than having "a" research question that can be articulated at any point during a project.

- Selecting, designing, and configuring robots. Coming to an understanding about what robot to use for a research study is crucial as it can greatly impact the outcome of the research. The choice or design of a robot will depend on the specific research question and the type of interaction being studied. For instance, a full humanoid robot may be more suitable for studies on social interactions, while a more simplified robot may be better suited for studies on task-oriented interactions. However, it's important to consider whether a humanoid robot is even necessary for a particular interaction, as a text-based chatbot may serve a similar purpose. Additionally, practical constraints such as the availability and capabilities of the robot can also play a significant role in shaping the research question that can be explored. For example, a robot with poor speech recognition capabilities may not be suitable for studying autonomous spoken interactions with a specific population. The appearance of a robot will also have to be considered.

- Measuring the human, robot and system behavior: To evaluate the effectiveness of the human-robot interaction, it is important to have a way to measure the behavior of both the human and the robot. This can be done through a variety of methods such as questionnaires, physiological measures, and video or audio recordings, and evaluation of performance. Although questionnaires can be used for subjective measurements, this tends to be limited due to the complexity of interactions (e.g. measuring conversational engagement). Additionally, there are few, if any, standardized measures for HRI research and it is unlikely that these can be developed to reliably cover the wide range of experiments that can be conducted. Qualitative methodologies such as interviews may produce a richer form of analysis.

- User testing: Before the main into the wild experiment, it is important to conduct pilot studies with a small group of participants to test the usability of the setup and to make any necessary adjustments. Pilot studies often reveal bugs and unexpected behavior of the system, but also serve as a general repetition of the entire study pipeline. From bringing the participants in all the way to debriefing, a pilot study is likely to throw up major problems in the evaluation process.

- Ethical considerations: There are many ethical considerations when conducting human-robot interaction research, such as participant safety, privacy, and informed consent. With the uptake of large scale data collection from corporate entities, many people are unaware about how their data will be used. Considerations should be made for elderly, children and others for whom simply signing an agreement may not signify full understanding and consent.

- Data Management and Analysis: After the experiment, the data collected need to be stored and analyzed properly, with appropriate tools and methodologies. Data on human behavior requires that privacy and transparency is critical at this stage. Increasingly anonymised data is made public through open data repositories, such as Zenodo or the Open Science Framework.

- Reporting Results: Results should be reported in a clear, concise and unbiased manner, highlighting the main findings and the contribution made to the field. “Negative” results should also be reported: research time can be used efficiently by reporting what does not work.

It is important to note that these steps do not necessarily have to be followed in the order listed above. For example, some research projects might not require all the steps articulated here, or projects might start with insights gained from user testing rather than with the initial steps listed here.

There is an increasing demand for more ecologically valid evaluations, the so-called “in the wild” studies, in which an interactive device is tested in the real world, with its associated dynamics and noise. While lab-based studies do still have a role to play, they have limitations which do not sit easy with the goals of our field. In lab-based studies, the participants are unlikely to respond naturally, feedback is likely biased, the complexity of the interaction is constrained and the duration of the interaction is severely limited. In the wild studies come with their own challenges. The noise and variability in real-world environments makes it difficult to collect clean data, and in the wild studies seldom results in clear answers. Additionally, conducting these types of studies has greater costs both financially and in terms of time. However, the richness of the results and the relevance to the end user more than compensate for this.

5.5 Summary of research challenges

An important result of this meeting’s discussions were the following identified issues and research challenges in the context of “natural interaction with humanoid robots”. Even if the level of urgency is always difficult to determine we would still propose that it is necessary to deal with research challenges at two different levels at the same time, working in parallel and synchronising achievements.

5.5.1 Cognitive abilities

There is a need to study and subsequently implement theory of mind-like abilities as it appears to underlie most of the necessary cognitive abilities of a social agent, and are needed for personalisation, adaptation, and producing suitable explanations that the end user needs, these mechanisms need attention. Without learning this is not possible. As adaptation to individual users and their preferences and demands is important, robots should learn in interaction with and from lay users. Interactive capabilities that allow for a fluent, contingent and incremental process are still missing. On the technological side, they include the need for hardware advances, better and/or novel sensors and sensory processing and interpretation algorithms.

5.5.2 Technical challenges

Natural interactions, online learning and adaptation are concepts that still need a great effort on technology development. Signal gathering and processing: some of the current technologies are somewhat considered mature, but they are still in their infancy outside the laboratory, especially in real and adverse environments where robots have to work: audio and visual noisy environments, speaker accents, multilinguality, etc. Moreover, we are still far away to understand and decode all the information that these signals about the current circumstances of the users and their environment, which still are difficult to be individually understood and well identified. Decision making: even though we already have technical methods to optimise goals we still need that these goals match the mentioned cognitive abilities. Natural behaviour of the robot: natural language, emotional voices and gestures have still to be improved to get robots implementing a natural and well accepted behaviour during interaction with humans.

5.5.3 Multimodal and "realistic" data

There is a mismatch between desired cognitive abilities and technology capacities. And one of the main reasons is that nowadays technologies need data, and more data and always data. Since spontaneous interactions are very difficult, or impossible, to render, scientists tends to simulate such interactions either through professional actors or through the Wizard of Oz methodology. As a community we need to stop obtaining useless data and reflect on how to acquire data on interactions between people as well as on interactions between people and machines that really provide us with scenarios and useful data for the good work of machine learning systems. And this is urgent.

5.6 Addressing research challenges

It is important to think about what kind of society is desirable and how this technology can contribute to society. For example, to address the issues of children, the physically challenged, and other minorities, it will be necessary to consider use cases.

Other researchers are needed besides computer scientists and robotics researchers, it will be necessary to work with experts in cognitive science, developmental psychology, medicine, psychiatry and other fields. More work at all levels and from multiple disciplines is required. The actual application of the humanoid robot throughout the entire creation process involves user participation, user surveys, and long-term research.

Further research in artificial intelligence, machine learning, natural language processing, human robot interaction, human agent interaction, affective computing, biological information processing, or other areas of expertise listed above is needed as fundamental research. Possible applications include education, care for the elderly, care for the sick, etc. All places of use will be targeted, including schools, hospitals, homes, public facilities, and stores.

5.7 Possible solutions in practice

To achieve successful technology and knowledge transfer, communication is key. It is crucial to build up collaborations with all necessary stakeholders to carry out the research activities and obtain the necessary resources. According to Klemme et al. [5] for the example of healthcare technology including humanoid robots, economic, academic and social partners should be involved in an equal partnership already early on in a knowledge transfer process (including technology transfer), even before a joint project starts [5]. Open, early, and regular communication between all relevant stakeholders as equals is crucial for the success of transfer. The joint acquisition of appropriate funding depends on the availability of respective funding opportunities. Often, especially in research, technical solutions are developed based on the current state-of-the-art and advances in the technical field. With the embedding of multiple stakeholders and user groups in the development process, the successful implementation of the developed technology into practice is realistic. Technology partners including academic and industry partners have an understanding of what can actually be achieved whereas practice partners have an understanding of everyday life in the context of employment but also of financial requirements. To be able to formulate needs, practice partners have to learn about technical possibilities in a process of mutual learning (cf. [12]). After prototype development, industry partners have to ensure product development and the development of working, potentially new, business models for commercialisation. For this, other stakeholders, as for instance regulatory bodies for innovation in medicine, should be tied in as well. At the same time diversity is important also in the group of involved researchers as research questions have to be answered from different perspectives and disciplines. This interdisciplinarity does not only refer to a small interdisciplinarity as between informatics and mathematics but a big interdisciplinarity involving for example the natural sciences, humanities and social sciences.

6 Summary of identified issues

There are a number of challenges associated with this topic. The lack of clear notions of both "natural interaction" and "humanoid robot" highlights that it is not at all clear what the aims of the field actually are, so these need to be identified. For example, although HRI has been around for a few decades already and we do have plenty of "narrow" studies that focus on specific applications, it is not evident what a credible use case for natural interaction with a humanoid robot would actually be.

An important criterion here is that both "natural interaction" and "humanoid robot" need to be aspects that cannot be abstracted away without loss of something essential in the use case. It is still to be identified what this would actually be (though there are clearly suggestions in the other topic); for example, is a use case that could also be carried out with a virtual avatar on a screen a compelling use case for a humanoid robot, and if so, why?

Both the "natural interaction" and "humanoid robot" requirements set relatively high standards on all technical aspects. This is clear when we consider that even relatively constrained scenarios likely require significant cognitive abilities. We have previously highlighted theory of mind-like abilities as a key underpinning of much of human cognition in scenarios in which one might want to deploy a humanoid robot. For example, ToM is necessary for personalisation and adaptation to a user as well as for ensuring transparency in the sense that the user understands the abilities and the limitations of the robot through tailored explanations [9]. However, this ability is poorly understood even among psychologists, so we are in a situation where we would like advanced cognitive abilities that we do not yet understand. It may well be that building these robots will, in fact, lead to progress in understanding human cognition too - this is, after all, one of the aims of cognitive robotics [2] and cognitive systems research (Vernon, 2014) - but the fact remains that it is currently not clear how to deliver these abilities.

The field traditionally circumvents this by resorting to the Wizard of Oz paradigm. This has the side effect that all "hard" problems appear to be not as urgent as they might otherwise be since they are effectively outsourced to the human wizard. It also means that potential fundamental differences between humans and robots in terms of cognitive abilities are ignored while they might have critical consequences for what kind of tasks can and cannot be done with a robot. For example, theories of embodied cognition, if taken seriously, imply that cognitive abilities are shaped by sensorimotor abilities [10]; it does not automatically follow that a robot can do anything the human wizard could.

It is therefore an issue that the WoZ paradigm allows us to sidestep all of this as we might spend significant amounts of resources (including the time of the participants) collecting data that does, at the end of the day, not pertain to a realistic use case.

It does appear clear that the envisioned robots will also need advanced sensory abilities if they are to, one day, replace the human wizard. This requires advances in sensory and social signal interpretation algorithms, which relate again to theory of mind that is currently not well understood. There may also be a need for new sensory hardware, as well as considerations of which modalities to include, and how they relate to each other.

The need for multimodal data is also recognised in other fields, such as machine learning; however the notion is often simplified to mean e.g. both text and image data. In reality, the notion is more complex, especially in the context of HRI, and needs further exploration [7]. In any case, it remains tempting to make use of standard machine learning algorithms in HRI; however these often require large amounts of training data and we currently lack reasonable datasets, both in terms of the necessary multimodal data and simply in terms of covering realistic scenarios of natural interaction with a humanoid robot. It remains therefore unclear to what degree the field can rely on standard machine learning approaches; however the alternative requires significant algorithmic advances, potentially finding approaches that do not rely on big data in the first place.

In terms of implicit requirements the field assumes, the hardware design of present-day robots may also not be the most suitable to achieve natural interactions. For example, one use case one could consider would be physical rehabilitation training (e.g. [1]). This, however, requires the robot demonstrator to demonstrate human movements at a high degree of fidelity, which no current robot is capable of, either because the motors used do not allow demonstrating biological kinematics, or because the morphology does not map onto the human body, or both.

As usual, commercial value is likely to have precedence when translating research into real world products. This means that the first users of natural interaction with humanoid robots are likely to be those who can afford the technology, rather than those who need it the most. This raises ethical questions about the commercialization of this technology, as it may lead to a situation where the most vulnerable and marginalised groups are left behind [53].

When it comes to human embodiment, it can be a limitation for the robot in some situations, particularly when interacting inappropriately. These robots are designed to have some human characteristics and skills, but not all, as it depends on the application. For example, a humanoid robot designed for customer service in a store may not be suitable for providing companionship, as it may not be able to perform the necessary tasks or may be perceived as out of place. The question of whether humanoid robots can replace humans completely in some tasks is a complex and controversial one. Some experts argue that robots can eventually be programmed to perform many tasks that are currently done by humans, while others believe that robots will only serve as additional assistants, augmenting human capabilities rather than replacing them.

One important aspect to consider when discussing the use of humanoid robots in commercial applications is the field of Computer-Supported Cooperative Work (CSCW) and ethics. CSCW focuses on the design and use of technology to support social interaction, and it is crucial to consider the impact of humanoid robots on human interactions and collaboration. Additionally, there are ethical considerations surrounding the commercial value of humanoid robots versus their ethical value, as well as the potential impact on employment and the economy.

Another important factor to consider is the issue of expectation setting, from both commercial and developers' perspectives. Humanlike robots with social behaviour invite people to treat them as if they were humans, therefore they expect them to have the same capability of humans [27]. For example, the robot Pepper was marketed as being able to understand and respond to human emotions, but it was later found that it could not meet these expectations, leading to disappointment among users. Therefore, it is important to better set expectations at the first place even before the development of the robotic project, based on the form-factor of the robot, and ensure that the capabilities and limitations of the robot are clearly communicated to users.

A final challenge that sits alongside all these needs for significant fundamental research activities before the envisioned humanoid robots become realistic concerns the current research funding landscape. The emphasis is currently on innovations and applications, or, more generally, aspects that are clearly measurable by KPIs. This is a challenging environment in which to propose projects that focus on fundamental research. Since funding programs shape research fields, attempts are made to deliver applications before it is really feasible to achieve them. This may eventually have a detrimental effect on the perception of robotics since the resulting products will be limited by necessity in a manner that is not congruent with how they are advertised nor the general hype around them, including hype that is pushed by exaggerated claims in AI and machine learning and hype pushed by certain, often US based, technological companies. Obvious examples include marketing stunts such as awarding citizenship to the robot Sophia or Musk's promises around the Optimus platform.

List of Participants

- Kristiina Jokinen, AIST Tokyo
- Wolfgang Minker, Ulm University
- Catherine Pelachaud, CNRS-ISIR, Sorbonne University
- Matthias Kraus, University of Augsburg
- Annalena Aicher, Ulm University
- Nick Campbell, Trinity College Dublin
- Tony Belpaeme, Ghent University
- Malte Jung, Cornell University
- Junpei Zhong, The Hong Kong Polytechnic University
- Alessandro Di Nuovo, Sheefield Hallam University
- Michael Spranger, Sony AI Inc.,
- Keiichi Yasumoto, Nara Institute of Science and Technology (NAIST)
- Satoshi Nakamura, Nara Institute of Science and Technology (NAIST)
- Silvia Rossi, Universitat degli Studi di Napoli Federico II
- Serge Thill, Donders Institute for Brain, Cognition and Behaviour
- Divesh Lala, Kyoto University
- Kaoru Sumi, Future University of Hakodate
- Dimosthenis Kontogiorgos, Potsdam University
- Maria Ines Torres, Universidad del Pais Vasco UPV/EHU
- Graham Wilcock, University Helsinky
- Anna-Lisa Vollmer, Bielefeld University
- Minha Lee, Technische Universiteit Eindhoven (Eindhoven University of Technology)
- Hung-Hsuan Huang, University of Fukuchiyama

References

[1] Bayon, C., Ramírez, O., Serrano, J. I., Del Castillo, M. D., Pérez-Somarriba, A., Belda-Lois, J. M.,... & Rocon, E. (2017). Development and evaluation of a novel robotic platform for gait rehabilitation in patients with Cerebral Palsy: CPWalker. Robotics and autonomous systems, 91, 101-114. DOI:10.1016/j.robot.2016.12.015

[2] Cangelosi, A. & Asada, M. (2022). Cognitive Robotics. MIT Press. DOI:10.7551/mitpress/13780.001.0001

[3] Clark, H. H., & Brennan, S. E. (1991). Grounding in communication.

[4] Kendon, A. (1990). Conducting interaction: Patterns of behavior in focused encounters (Vol. 7). CUP Archive.

[5] Klemme, I., Richter, B., De Sabbata, K., Wrede, B., & Vollmer, A. L. (2021). A Multi-Directional and Agile Academic Knowledge Transfer Strategy for Healthcare Technology. Frontiers in Robotics and AI, 8, 418.

[6] Kontogiorgos, D., Van Waveren, S., Wallberg, O., Pereira, A., Leite, I., & Gustafson, J. (2020, April). Embodiment effects in interactions with failing robots. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-14).

[7] Lagerstedt, E., & Thill, S. (2022). Multiple roles of multimodality among interacting agents. ACM Transactions on Human-Robot Interaction. DOI:10.1145/3549955

[8] Pelikan, H., Robinson, F. A., Keevallik, L., Velonaki, M., Broth, M., & Bown, O. (2021, March). Sound in Human-Robot Interaction. In Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (pp. 706-708).

[9] Riveiro, M., & Thill, S. (2021). "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. Artificial Intelligence, 298, 103507. DOI:10.1016/j.artint.2021.103507

[10] Windridge, D.,& Thill, S. (2018). Representational fluidity in embodied (artificial) cognition. BioSystems, 172, 9-17. DOI:10.1016/j.biosystems.2018.07.007

[11] Vernon, D. (2014). Artificial cognitive systems: A primer. MIT Press. ISBN:9780262028387

[12] Weiss, A., & Spiel, K. (2022). Robots beyond Science Fiction: mutual learning in human–robot interaction on the way to participatory approaches. AI & society, 37(2), 501-515.

[13] Li, Jamy. "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents." International Journal of Human-Computer Studies 77 (2015): 23-37.

[14] Ysamaha: VOCALOID: AI - Misora Hibari -, 2019 (in Japanese) https://www.yamaha.com/ja/about/ai/vocaloid_ai/

[15] University of Tokyo press release: "Singing voice of 50 years ago revived beyond time -Recreation of singing by task mixture deep learning using small data", 2022 (in Japanese), https://www.i.u-tokyo.ac.jp/news/press/2022/202210032121.shtml

[16] Marchang, Jims, and Alessandro Di Nuovo. 2022. "Assistive Multimodal Robotic System (AMRSys): Security and Privacy Issues, Challenges, and Possible Solutions" Applied Sciences 12, no. 4: 2174, https://doi.org/10.3390/app12042174

[17] Microsoft Technology Licensing. Creating a conversational chatbot of a specific person., https://patents.google.com/patent/US10853717B2/en?oq=us10853717b2

[18] Dario, P., Verschure, P.F., Prescott, T., Cheng, G., Sandini, G., Cingolani, R., Dillmann, R., Floreano, D., Leroux, C., MacNeil, S. and Roelfsema, P., 2011. Robot companions for citizens. Procedia Computer Science, 7, pp.47-51.

[19] Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. Philosophical Transactions of the Royal Society B, 374(1771), 20180032.

[20] Cavallo, Filippo, Raffaele Esposito, Raffaele Limosani, Alessandro Manzi, Roberta Bevilacqua, Elisa Felici, Alessandro Di Nuovo, Angelo Cangelosi, Fabrizia Lattanzio, and Paolo Dario. "Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study." Journal of medical Internet research 20, no. 9 (2018): e9460.

[21] Charalampous, K., Kostavelis, I., & Gasteratos, A. (2017). Recent trends in social aware robot navigation: A survey. Robotics and Autonomous Systems, 93, 85-104.

[22] Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara, 2020. An Attentive Listening System with Android ERICA: Comparison of Autonomous and WOZ Interactions. Proceedings of the SIGdial 2020 Conference, 118–127.

[23] Conti, Daniela, Grazia Trubia, Serafino Buono, Santo Di Nuovo, and Alessandro Di Nuovo. "An empirical study on integrating a small humanoid robot to support the therapy of children with Autism Spectrum Disorder and Intellectual Disability." Interaction Studies 22, no. 2 (2021): 177-211.

[24] Erol, B. A., Majumdar, A., Benavidez, P., Rad, P., Choo, K. K. R., & Jamshidi, M. (2019). Toward artificial emotional intelligence for cooperative social human–machine interaction. IEEE Transactions on Computational Social Systems, 7(1), 234-246.

[25] Lee, M.H., Siewiorek, D.P., Smailagic, A. et al. Enabling AI and Robotic Coaches for Physical Rehabilitation Therapy: Iterative Design and Evaluation with Therapists and Post-stroke Survivors. Int J of Soc Robotics (2022), https://doi.org/10.1007/s12369-022-00883-0.

[26] Kok, B. C., & Soh, H. (2020). Trust in robots: Challenges and opportunities. Current Robotics Reports, 1, 297-309.

[27] You, robot: on the linguistic construction of artificial others. AI & SOCIETY, 26(1), 61-69, doi:10.1007/s00146-010-0289-z.

[28] Mahdi, Hamza, Sami Alperen Akgun, Shahed Saleh, and Kerstin Dautenhahn. "A survey on the design and evolution of social robots—Past, present and future." Robotics and Autonomous Systems (2022): 104193.

[29] McCarthy, J. (2022). Artificial Intelligence, Logic, and Formalising Common Sense. Machine Learning and the City: Applications in Architecture and Urban Design, 69-90.

[30] Fosch-Villaronga, Eduard and Poulsen, Adam. "Sex care robots: Exploring the potential use of sexual robot technologies for disabled and elder care" Paladyn, Journal of Behavioral Robotics 11, no. 1 (2020): 1-18. https://doi.org/10.1515/pjbr-2020-0001

[31] Belpaeme, Tony, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. "Social robots for education: A review." Science robotics 3, no. 21 (2018): eaat5954.

[32] The role of robots in entertainment. Industrial Robot, Vol. 49 No. 4, pp. 667-671, 2022. https://doi.org/10.1108/IR-02-2022-0054

[33] Di Nuovo, A., Varrasi, S., Lucas, A., Conti, D., McNamara, J., & Soranzo, A. (2019). Assessment of Cognitive skills via Human-robot Interaction and Cloud Computing. Journal of Bionic Engineering, 16(3), 526–539.

[34] C. Lytridis et al., "Social Robots as Cyber-Physical Actors in Entertainment and Education," 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2019, pp. 1-6, doi:10.23919/SOFTCOM.2019.8903630.

[35] Johanson, Deborah L., Ahn, Ho Seok, Sutherland, Craig J., Brown, Bianca, MacDonald, Bruce A., Lim, Jong Yoon, Ahn, Byeong Kyu and Broadbent, Elizabeth. "Smiling and use of first-name by a healthcare receptionist robot: Effects on user perceptions, attitudes, and behaviours" Paladyn, Journal of Behavioral Robotics 11, no. 1 (2020): 40-51, https://doi.org/10.1515/pjbr-2020-0008

[36] Delmerico, J., Mintchev, S., Giusti, A., Gromov, B., Melo, K., Horvat, T., ... & Scaramuzza, D. (2019). The current state and future outlook of rescue robotics. Journal of Field Robotics, 36(7), 1171-1191.

[37] Fiorini, Laura, Federica G. Cornacchia Loizzo, Alessandra Sorrentino, Jaeseok Kim, Erika Rovini, Alessandro Di Nuovo, and Filippo Cavallo. "Daily gesture recognition during human-robot interaction combining vision and wearable systems." IEEE Sensors Journal 21, no. 20 (2021): 23568-23577.

[38] Wilcock, G. and Jokinen, K. Conversational AI and knowledge graphs for social robot interaction. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022). Sapporo, Japan (2022).

[39] Wilcock, G. and Jokinen, K. Cooperative and uncooperative behaviour in task-oriented dialogues with social robots. In: 31st IEEE International Conference on Robot & Human Interactive Communication (RO-MAN 2022). Naples, Italy (2022).

[40] Wilcock, G. and Jokinen, K. Should robots indicate the trustworthiness of information from knowledge graphs? In: 10th International Conference on Affective Computing and Intelligent Interaction (ACII 2022), Nara, Japan (2022).

[41] Wilcock, G. Generating more intelligent responses and explanations with conversational AI and knowledge graphs. In: Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2022). Kyoto, Japan (2022).

[42] Buyukgoz S, Grosinger J, Chetouani M and Saffiotti A (2022), Two ways to make your robot proactive: Reasoning about human intentions or reasoning about possible futures. Front. Robot. AI 9:929267. doi: 10.3389/frobt.2022.929267

[43] Kuiper, J. S., Zuidersma, M., Voshaar, R. C. O., Zuidema, S. U., van den Heuvel, E. R., Stolk, R. P., & Smidt, N. (2015). Social relationships and risk of dementia: A systematic review and meta-analysis of longitudinal cohort studies. Ageing research reviews, 22, 39-57.

[44] Santini, Z. I., Koyanagi, A., Tyrovolas, S., Mason, C., & Haro, J. M. (2015). The association between social relationships and depression: A systematic review. Journal of affective disorders, 175, 53-65.

[45] Piolatto, M., Bianchi, F., Rota, M., Marengoni, A., Akbaritabar, A., & Squazzoni, F. (2022). The effect of social relationships on cognitive decline in older adults: an updated systematic review and meta-analysis of longitudinal cohort studies. BMC Public Health, 22(1), 1-14.

[46] Selvaggio, M., Cognetti, M., Nikolaidis, S., Ivaldi, S., & Siciliano, B. (2021). Autonomy in physical human-robot interaction: A brief survey. IEEE Robotics and Automation Letters, 6(4), 7989-7996.

[47] Yaacoub, J. P. A., Noura, H. N., Salman, O., & Chehab, A. (2022). Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. International Journal of Information Security, 1-44.

[48] McTear, M., Jokinen, K., Dubey, M., Chollet, G., Boudy, J., Lohr, C., Roelen, S. D., Mössing, W. & Wieching, R. (2022). Empowering Well-Being Through Conversational Coaching for Active and Healthy Ageing. Participative Urban Health and Healthy Aging in the Age of AI. Cham, Springer International Publishing, pp. 257–265r, `doi:10.1007/978-3-031-09593-1\_21`

[49] Zhong, J., Ling, C., Cangelosi, A., Lotfi, A., & Liu, X. (2021). On the gap between domestic robotic applications and computational intelligence. Electronics, 10(7), 793.

[50] Savage, Neil. "Robots rise to meet the challenge of caring for old people." Nature 601, no. 7893 (2022): 8-10.

[51] Williams, T., Matuszek, C., Jokinen, K., Korpan, R., Pustejovsky, J., Scassellati, B. (2023). Voice in the Machine: Ethical Considerations for Language-Capable Robots. Communications of ACM (Association for Computing Machinery).

[52] Jokinen, K., Wilcock, G. (2021). Do you remember me? Ethics in Social Robot Interactions. The 30th IEEE International Conference on Robot and Human Interactive Communication (ROMAN-2021), 10.1109/RO-MAN50785.2021.9515399

[53] Hernandez, Kevin, and Tony Roberts. "Leaving no one behind in a digital world." (2018).

[54] Benjamin, R. (2019). Race after technology: Abolitionist tools for the new jim code. Social forces.

[55] Bartlett, M. E., Edmunds, C. E. R., Belpaeme, T., & Thill, S. (2022). Have I got the power? analysing and reporting statistical power in HRI. ACM Transactions on Human-Robot Interaction (THRI), 11(2), 1-16.

[56] Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: a review. Gerontechnology, 8(2), 94-103.

[57] Danaher, J. 2019. Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. Science and Engineering Ethics (2019), 1–27

[58] Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. Early engagement and new technologies: Opening up the laboratory, 55-95.

[59] Germak, C., Lupetti, M. L., & Giuliano, L. (2015). Ethics of robotic aesthetics. Design and semantics of form and movement, 165.

[60] Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass. Eamon Dolan Books.

[61] Grudin, J. (1992). Utility and usability: research issues and development contexts. Interacting with computers, 4(2), 209-217.

[62] Hoffman, G., & Zhao, X. (2020). A primer for conducting experiments in human–robot interaction. ACM Transactions on Human-Robot Interaction (THRI), 10(1), 1-31.

[63] Jung, M., & Hinds, P. (2018). Robots in the wild: A time for more robust theories of human-robot interaction. ACM Transactions on Human-Robot Interaction (THRI), 7(1), 1-5.

[64] Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Sowden, W. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. Advances in Methods and Practices in Psychological Science, 1(4), 443-490.

[65] Lakens, D. (2022). Sample size justification. Collabra: Psychology, 8(1), 33267.

[66] Lucas, M., Mekary, R., Pan, A., Mirzaei, F., O'Reilly, É. J., Willett, W. C., ... & Ascherio, A. (2011). Relation between clinical depression risk and physical activity and time spent watching television in older women: a 10-year prospective follow-up study. American journal of epidemiology, 174(9), 1017-1027.

[67] Lee, M., Ruijten, P., Frank, L., de Kort, Y., & IJsselsteijn, W. (2021, May). People may punish, but not blame robots. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-11).

[68] Lupetti, M. L., Zaga, C., & Cila, N. (2021, March). Designerly ways of knowing in HRI: Broadening the scope of design-oriented HRI through the concept of intermediate-level knowledge. In Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (pp. 389-398).

[69] McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. American Behavioral Scientist, 25(2), 179-210.

[70] Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results.

[71] Nyholm, S. 2018. Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. Science and Engineering Ethics 24, 4 (2018), 1201–1219.

[72] Rossi S, Santini SJ, Di Genova D, Maggi G, Verrotti A, Farello G, Romualdi R, Alisi A, Tozzi AE, Balsano C . Using the Social Robot NAO for Emotional Support to Children at a Pediatric Emergency Department: Randomized Clinical Trial. J Med Internet Res 2022;24(1):e29656, doi: 10.2196/29656

[73] Sabanovic, S., Michalowski, M. P., & Simmons, R. (2006, March). Robots in the wild: Observing human-robot social interaction outside the lab. In 9th IEEE International Workshop on Advanced Motion Control, 2006. (pp. 596-601).

[74] Scassellati, B., & Vázquez, M. (2020). The potential of socially assistive robots during infectious disease outbreaks. Science Robotics, 5(44), eabc9014.

[75] Sung, J., Christensen, H. I., & Grinter, R. E. (2009, March). Robots in the wild: understanding long-term use. In Proceedings of the 4th ACM/IEEE international conference on Human robot interaction (pp. 45-52).

[76] West, M., Kraut, R., & Ei Chew, H. (2019). I'd blush if I could: closing gender divides in digital skills through education.

[77] Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.

[78] Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., & Dogramadzi, S. (2020). Designing ethical social robots—a longitudinal field study with older adults. Frontiers in Robotics and AI, 7, 1.

[79] Veling, L., & McGinn, C. (2021). Qualitative research in HRI: A review and taxonomy. International Journal of Social Robotics, 13(7), 1689-1709.

[80] Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., ... & Pandey, A. K. (2019, March). Second language tutoring using social robots: a large-scale study. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 497-505). IEEE.

[81] Yin, R. K. (1981). The case study crisis: Some answers. Administrative science quarterly, 26(1), 58-65.

[82] Zaga, C., & Lupetti, M. L. (Eds.) (2022). Diversity Equity and Inclusion in Embodied AI: Reflecting on and Re-imagining our Future with Embodied AI . 4TU.Federation. https://doi.org/10.3990/1.9789036553599

[83] Zimmerman, J., Forlizzi, J., & Evenson, S. (2007, April). Research through design as a method for interaction design research in HCI. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 493-502).