The Road to Reliable Robots: Interpretable, Accessible, and Reproducible HRI Research

Megan Zimmerman[1], Ann Virts[1], Shelly Bagchi[1], Snehesh Shrestha[1,2], Patrick Holthaus[3], Emmanuel Senft[4], Daniel Hernandez Garcia[5], and Jeremy A. Marvel[1]
[1]*U.S. National Institute of Standards and Technology, Gaithersburg, MD, USA*
[2]*University of Maryland, College Park, MD, USA*
[3]*University of Hertfordshire, Hatfield, United Kingdom*
[4]*Idiap Research Institute, Martigny, Switzerland*
[5]*Heriot-Watt University, Edinburgh, Scotland*

***Abstract*—There are a multitude of robotic application domains that touch on the field of human-robot interaction (HRI). From modern manufacturing involving human-robot teams, to personal care robots assisting the elderly, the roles robots are being tasked with and the nature of interactions with humans are constantly shifting. Even the nature of interaction has changed to incorporate wearable technologies such as exosuits to enhance human capabilities, and advanced prosthetics to restore those abilities that have been lost. With this ever-evolving spectrum of HRI, the capacity of measurement science to evaluate, assess, and assure performance and safety struggles to keep up. Building on our previous five-workshop series on Test Methods and Metrics for Effective HRI, NIST presents a new series on evaluative methodologies for accelerating the pipeline from cutting-edge HRI research to state-of-practice. This workshop will address issues regarding 1) data collection and reporting for replicability and system validation, 2) test design and execution for performance verification, and 3) cross-modality artifact design for real-world application-adjacent technology transfer. The goal of this workshop is to accelerate and accommodate accessibility to HRI research results, and address the specific key performance indicators that would establish end-user trust and acceptance of emerging HRI technologies.**

I. Organizing Committee

Organizers and committee members will recruit and coordinate reviewers for author-submitted contributions. A website will be published upon acceptance to act as an archive of former workshops, schedules, and latest updates. For more up-to-date information, including confirmed speakers, please visit the webpage. A list of the organizers for 2025 follows.

Megan Zimmerman has been a computer science researcher at the National Institute of Standards and Technology (NIST) since 2016, and has been involved in the HRI community since 2015. Megan's primary area of expertise is in human-robot interaction (HRI) and alternative robot control interfaces. Currently, she is leading efforts at NIST to generate public datasets for human robot interaction research.

Ann Virts is the Acting Group Leader for the Manipulation and Mobility Systems Group in NIST's Intelligent Systems Division, and is the project leader of the Mobility Performance of Robotic Systems at NIST. She leads the project by developing test methods, artifacts, and datasets that measure the performance and safety characteristics of mobile and wearable robots within manufacturing environments. Ann is also the associate project leader for the Emergency Response Robots project, where she develops and implements robot test methods, interfaces with emergency responders and robot technology developers, and meets complex technical and logistical demands of demanding response robot field exercises.

Shelly Bagchi has been a robotics researcher at the Collaborative Robotics Lab at NIST since 2016. Shelly is the Project Lead for the Emerging Technology for Small- and Medium-sized Enterprises (SMEs) project at NIST. Her research interests are in human-robot interaction, replicability & reproducibility, and augmented reality. Her focus is on HRI, particularly evaluating new interface technologies (e.g., Augmented Reality and wearable devices) for HRI in manufacturing. She currently chairs the IEEE P3108 Standards Group, "Recommended Practice for Human-Robot Interaction Design of Human Subject Studies" and is the secretary for IEEE P3107, "Standard Terminology for Human-Robot Interaction."

Dr. Snehesh Shrestha (https://snehesh.com) is a postdoctoral researcher at NIST and Adjust Faculty in the Immersive Media Design department at the University of Maryland (UMD) College Park. His research is at the interaction of Human-Robot Interaction and artificial intelligence (AI) empowered education. He develops technology, tools, measurements, standards, and applications in these domains. He is interested in empowering people by creating supertools that augment their capabilities. He is an active contributor of IEEE P3017 and P3108 standards working group, "Standard Terminology for Human-Robot Interaction" and "Recommended Practice for Human-Robot Interaction Design of Human Subject Studies."

Dr. Patrick Holthaus is a Senior Research Fellow in the Robotics Research Group at the University of Hertfordshire (UK). His research revolves around social robotics and focuses on nonverbal interactive signals, social credibility and trust in assistive and companion robots. He is further interested in interaction architectures and behaviour coordination as well as systems integration in heterogeneous environments. Patrick has extensive expertise in social human-robot interaction and experimentation and is highly skilled with a large array of robotic and sensing technologies. As manager of the Robot House, a unique facility for human-robot interaction, he brings

together real-world applications and fundamental robotics research.

Dr. Emmanuel Senft is a research scientist and group leader at the Idiap Research Institute and member of the Swiss Young Academy. He leads the Human-centered Robotics and AI group. His group develops assistive robotic technologies and new paradigms to empower end users of AI technologies by exploring the interaction between robotics, participatory design, end-user programming, and interactive machine learning.

Dr. Daniel Hernandez Garcia is a research fellow at the Interaction Lab in Heriot-Watt University. His research focuses on the application of data-driven and deep learning approaches for developing intelligent and socially-aware assistive agents in real scenarios with human users.

Dr. Jeremy A. Marvel is a research scientist at NIST. He leads scientists and engineers in metrology efforts at NIST toward measuring HRI performance, and developing tools to enable enterprises to effectively deploy robot solutions. He currently chairs the IEEE P3107 standards working group, "Standard Terminology for Human-Robot Interaction."

II. Workshop Overview

The field of Human-Robot Interaction (HRI) is populated by a vast and varied collection of distinct research topics. From human-centric user interface design to applications in medical robotics, from robot safety in human-occupied spaces to psychological factors of robot presence, the nature of research within the HRI domain is diverse. The multidisciplinary nature of HRI therefore has diverse set of data collection needs. In an age where the quality of a given dataset is measured by its broad applicability as much as its thoroughness, being able to capture and describe collected data for a broad spectrum of domains is an even bigger challenge than ensuring said data meets one's own research needs.

In a previous series (https://hri-methods-metrics.github.io/) of five workshops, we brought together researchers and industry leaders to discuss the development and accessibility of HRI test methods and metrics for assessing and assuring the performance of HRI technologies. In the first of a new series of workshops, we are focused on establishing recommended practices for the design, collection, and dissemination of datasets for HRI research. In particular we will address factors such as:

- selection of sensors, artifacts, and survey tools for broad-applicability across application domains and use cases;
- verification and validation of novel survey tools, and the utility of existing survey tools that have already been validated;
- documentation of testing conditions, including environmental aspects, recruitment practices, and diversity in human subject studies;
- community engagement and collaboration, particularly with regards to conducting replication and repeatability studies to validate and grow community-driven data repositories; and
- practices and considerations for future-proofing new data sets, and mechanisms by which datasets can evolve to address future community needs.

A. Target Audience

This workshop continues to serve as a springboard for establishing a formalized and standardized HRI research community. This workshop expects to host approximately 30-40 participants from targeted interest groups including stakeholders in social, medical, response, and service robotics:

- researchers in the process of planning or building quality datasets for HRI community consumption;
- new members of the HRI community such as new students and visiting members from other research fields;
- researchers who are otherwise unable to develop their own datasets due to resource limitations or lack of specific sensing equipment;
- researchers developing frameworks and models of real-world, human-robot applications;
- businesses (both existing and prospective) who would leverage public datasets to accelerate their pipeline to product launch; and
- data scientists interested in assessing data quality and the impacts it has on community acceptance and adoption.

At this 6^{th} workshop, the first in our new series on HRI metrology, we are particularly interested in engaging new members of the HRI community. New students and visiting members from other research fields are poised to contribute the most to these discussions of dataset utility. We aim this workshop to help them to understand the community and provide a solid basis to produce quality research outputs to encourage the continued growth and development of the HRI corpus of research.

B. Schedule and Format

The half-day workshop schedule is shown in Table I, with a more detailed schedule to be posted on the workshop website. The structure is formatted such that it focuses on "data collection and documentation" (Session 1), and "application domains and end-users" (Session 2). Between the two sessions we will have an extended break to accommodate contributed posters and community networking. Each session features presentations from invited speakers followed by a panel discussion to solicit community discussions and feedback.

The themes of this workshop tie in with standardization efforts for HRI metrology. The IEEE Robotics and Automation Society has ongoing standards efforts in HRI measurement science. The first, IEEE P3108 ("Recommended Practice for Human-Robot Interaction Design of Human Subject Studies")[1], is focused on developing metrology best practices for human subject studies, designs of experiments, and HRI data set construction and dissemination. The second, IEEE P3107 ("Standard Terminology for

[1]https://standards.ieee.org/ieee/3108/10710/

TABLE I
HIGH-LEVEL WORKSHOP SCHEDULE

Session	Topic
Session 1	**Design and Collection of HRI Datasets**
	Welcome, intros, and ice breaker
	Keynote presentation
	Invited talks
	Panel discussion
Mid-way	**Extended Break**
	Posters
	Networking
Session 2	**HRI Dataset Users and Use Cases**
	Keynote presentation
	Invited talks
	Panel discussion

Human-Robot Interaction")[2] is developing standard definitions and conceptual ontologies for HRI applications. Information regarding these working groups can be found at https://www.nist.gov/el/intelligent-systems-division-73500/ieee-sg-metrology-human-robot-interaction.

This year's workshop will be held exclusively in-person, but keynotes and invited talks will be streamed live and recorded for greater accessibility. Recordings will be posted online following the conference.

C. Discussion Topics

Presentations by contributing speakers will focus on this year's theme of sustainability in their respective research fields. Keynote and invited speakers will be selected from a targeted list of HRI researchers across a broad spectrum of application domains. Poster session participants will be selected from contributors reporting late-breaking evaluations and their preliminary results.

Discussions are intended to highlight the various approaches, requirements, and opportunities of the research community toward supporting the HRI community through the development of quality datasets, with particular focus on formatting, artifact and sensor selection, and documentation of testing conditions and related metadata. Specific topics of discussion will include:

- human-centric design of data sets and subject studies to accommodate different demographics and a diverse research community;
- finding, accessing, understanding, and adopting standards, specifications, and recommended practices for HRI-relevant topics; and
- recommended practices in HRI research data collection, including how to recruit for and conduct human subject studies, creating statistical designs of experiments, and assembling knowledge bases to capture information that is not generally reported in published papers.

D. Outputs and Documentation

A workshop report documenting the presentations, discussions, and ensuing take-away and action items will be produced, and made publicly available as a NIST Report following the workshop. Additionally, recordings of talks will be kept publicly available via the workshop website.

[2]https://standards.ieee.org/ieee/3107/10709/

Additionally, insights and discussions from the workshop will contribute towards the IEEE Robotics and Automation Society Working Groups on Metrology for HRI, which are working to develop roadmaps and recommendations for the standardization of HRI test methods and metrics. We hope to encourage attendees to participate in the working groups, which conduct regular meetings throughout the year.

III. PARTICIPANT SOLICITATION PLAN

The organizing committee will be in attendance at the conference and will participate in the workshop. Targeted invitations to keynote and panel speakers will be sent directly via email. Solicitation of contributor-submitted posters will be achieved by means of broadcast advertisements, society and community channels, and directed invitations to submit. Specifically:

- solicitation via mailing lists such as Automation Worldwide, CHI Announcements, EU Robotics, HRI Announcements, and Robotics Worldwide;
- advertising in channels, including ACM Special Interest Group on Artificial Intelligence, IEEE RAS, IEEE Women in Engineering; and
- solicitation through standards bodies, including IEEE RAS, ASME, ASTM International, and ISO.