Errors, Mistakes, and Failures in Humans and Robots

Christopher D. Wallbridge
wallbridgec@cardiff.ac.uk
Cardiff University
United Kingdom

Kavyaa Somasundaram
kavyaa.somasundaram@oru.se
Örebro University
Sweden

Frank Förster
f.foerster@herts.ac.uk
University of Hertfordshire, UK
United Kingdom

Maia Stiber
maiastiber@microsoft.com
Microsoft Research
United States of America

Eduardo B. Sandoval
e.sandoval@unsw.edu.au
University of New South Wales
Australia

Chinmaya Mishra
chinmaya.mishra@mpi.nl
MPI for Psycholinguistics
Netherlands

Christina E. Stimson
cestimson1@sheffield.ac.uk
University of Sheffield
United Kingdom

Patrick Holthaus
p.holthaus@herts.ac.uk
University of Hertfordshire
United Kindgdom

Hatice Gunes
hatice.gunes@cl.cam.ac.uk
University of Cambridge
United Kingdom

Abstract

This workshop proposes looking at errors in humans and robots. The workshop will focus along three axes: Human, Robot and Interaction induced errors. Errors in Human-Robot Interaction (HRI) pose a crucial challenge for the deployment of robots. Errors can affect the safety and trust of users, therefore it is important for empowering society that these robots can handle errors robustly. The workshop intends to foster interdisciplinary discussion among researchers in robotics, HCI, cognitive science, and social sciences, to encourage the development of robust, user-centered, and socially responsible HRI systems. The invited speakers, paper presentations and discussions will focus around topics of detecting errors, handling errors, adapting to errors, restoring trust, using intentional errors, communication strategies, issues of expectation as well as look at design, interdisciplinary and ethical approaches to research. In this way we will help inform research into errors, and develop robotic systems capable of robust interaction and collaboration.

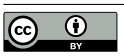
CCS Concepts

• **Computing methodologies** → **Cognitive robotics**; • **Human-centered computing** → **HCI theory, concepts and models**.

Keywords

Human errors, Robot errors, Failures in HRI

ACM Reference Format:
Christopher D. Wallbridge, Kavyaa Somasundaram, Frank Förster, Maia Stiber, Eduardo B. Sandoval, Chinmaya Mishra, Christina E. Stimson, Patrick Holthaus, and Hatice Gunes. 2026. Errors, Mistakes, and Failures in Humans and Robots. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI Companion '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3776734.3789489



HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3789489

1 Motivation

Errors in human–robot interaction (HRI) pose a crucial challenge for the deployment of robots in everyday, safety-critical and socially sensitive settings. Communicative errors, chat breakdowns, and the social credibility of robots impact how humans perceive risk, trust, and safety, highlighting that not only technical error, but social framing and recovery strategies are crucial [5, 15]. There is also the issue of human-induced interaction errors, highlighting that some errors originate not in a robot's hardware, but in human commands, assumptions or intent, and that systems must be designed to detect and respond to these errors [9]. Natural communication and dynamic language repair can improve how robots and humans collaborate, particularly when miscommunication or misunderstanding can lead to errors [13, 14].

Collectively, these perspectives motivate a systematic study of error typologies, repair mechanisms, and social framing in HRI; not simply to minimise errors, but to build resilient systems that empower society by fostering trust and safety in robotic technologies. This workshop proposes to explore errors along three axes:

(1) *Human-induced errors* - An error is caused by a person; based on a mistake or knowledge, execution or planning failure [7]
(2) *Robot-induced errors* - Errors occur in the system due to hardware or software, causing the system to misunderstand or act inappropriately [6].
(3) *Interaction errors* - Error occurs in the interaction between human and robot; it is not uniquely attributable to one agent.

Human-induced errors in human–robot interaction (HRI) arise when the human operator provides incorrect, ambiguous, or incomplete input, misinterprets robot capabilities, or operates the system outside its intended context. Such errors often stem from gaps in the user's mental model, cognitive overload, or lack of training, and they can significantly impact the safety and effectiveness of human-robot teams [7]. In teleoperation and collaborative robotics human-induced interaction failures can inadvertently trigger unsafe or unintended robot behaviors, underscoring the need for intelligent error detection and mitigation strategies [10]. Addressing these errors requires designing interfaces that support clear communication, provide feedback, and guide users in decision-making,

thereby reducing execution and knowledge-based failures while promoting trust and reliable collaboration.

Robot-induced errors occur when failures originate within the robotic system itself, such as misperception, faulty reasoning, hardware malfunctions, or incorrect task execution. These can result from limitations in sensors, inaccuracies in algorithms, unexpected environmental conditions, or insufficient adaptability to novel scenarios [4, 8]. Addressing robot-induced errors requires robust design, fault detection and recovery mechanisms, and adaptive systems able to learn from previous mistakes, ensuring reliability and safety while maintaining effective human-robot collaboration.

Interaction and communication errors arise from breakdowns in the shared space between humans and robots, even when both parties are operating correctly. These errors often involve misinterpretations of verbal, gestural, or multimodal signals, mismatched timing in collaborative tasks, or failures to establish common ground, which can lead to coordination problems and reduced trust [12]. Research by Helal et al [5] emphasizes that social credibility and the robot's ability to repair conversational or task-related misunderstandings are critical to maintaining effective collaboration. Addressing these errors requires designing robots that are socially aware, capable of dialogue repair, and able to provide transparent feedback, thereby supporting more resilient and trustworthy human-robot interactions. Interaction errors and mistakes are common especially in social HRI. The ERR@HRI challenge [2, 11] series has been pioneering machine learning models using benchmark multimodal datasets. Axelsson et al. [1] highlight the need for design-oriented approaches, especially for long-term interactions.

The workshop aims to advance understanding of errors in human–robot interaction by examining human-induced, robot-induced, and interaction/communication errors, their causes, and their impact on collaboration, trust, and safety. It seeks to share methods for detecting, mitigating, and recovering from these errors, including intelligent disobedience, adaptive learning, and dialogue repair strategies. By fostering interdisciplinary discussion among researchers in robotics, HCI, cognitive science, and social sciences, the workshop encourages the development of robust, user-centered, and socially responsible HRI systems. By looking at ways to handle errors and failures, identifying future research in this area and emerging challenges, we can empower users and society by promoting reliable, transparent and accessible HRI.

2 Organisers

Christopher D. Wallbridge is a Lecturer in HRI at Cardiff University; focusing on communication and trust - especially in the context of errors. He was recently awarded a RAICo fellowship investigating errors human-robot teams in a nuclear glovebox. He was previously an organiser of the Future Leaders Academy workshop for IROHMS (2022) and the Student Volunteer Chair for HRI'20.

Kavyaa Somasundaram is a final-year PhD candidate at Örebro University, Sweden. Her research focuses on intelligent error mitigation strategies in the context of human-induced errors during human-robot interaction. She co-led the workshop "The imperfectly relatable robot: An interdisciplinary workshop on the role of failures in HRI", organised at the ACM HRI'2023 conference.

Frank Förster is a Senior Lecturer in Computer Science performing research in developmental robotics and HRI and focussing on speech-involving interactions. He is a Co-Investigator of the EPSRC *FLUIDITY* project and has led two workshops on '*Working with Troubles and Failures in Conversation with Humans and Robots*' in '22 and '23, the latter associated with CUI 2023.

Maia Stiber is a Senior Researcher at Microsoft Research. Her research focuses on leveraging implicit behavioral responses to robot actions to create human-aware, error-aware robots. She received her Ph.D. in Computer Science from Johns Hopkins University. She was previously a co-lead for the 'SS4HRI: Social Signal Modeling in Human-Robot Interaction' workshop at HRI'24 and an organizer of the ERR@HRI Challenges in 2024 and 2025.

Eduardo B. Sandoval is a social robotics researcher, his work spans different aspects of social robotics, including Reciprocity in HRI, robots and education, robots and games, and creative interactions with robots. His work incorporates insights from behavioural economics and social psychology to explore different approaches in social robotics.

Chinmaya Mishra is a Postdoctoral Researcher at the Multimodal Language Department, Max Planck Institute for Psycholingusitcs. His research uses an interdisciplinary approach by combining insights from psychology, psycholinguistics, artificial intelligence, and cognitive science to model/ automate robot behaviors that are needed to facilitate a seamless HRI with a special focus on nonverbal behaviors. He co-organized the "Do Social Robots Need a Face?" workshop at ICSR 2024.

Christina E. Stimson is a Postdoctoral Research Fellow at Kyushu Institute of Technology. She recently completed her PhD at The University of Sheffield. Her thesis and current project focus on co-designing and evaluating an arts-based methodology for involving non-technical stakeholders in robot and assistive technology design with disabled people including children with *osteogenesis imperfecta* and older adults with early-stage Alzheimer's disease.

Patrick Holthaus is a Reader in Interactive Assistive Technology. He currently leads the UK-RAS topic group on *Human-Robot Interaction: Best practices and methods* and is Co-I of the Horizon Europe project *SWAG* and *Hospital@home* funded by Dinwoodie. Patrick co-led the three-workshop series "Troubles and Failures in HRI", organised as standalone events and at CUI 2023.

Hatice Gunes is a Professor at the University of Cambridge, directing the Affective Intelligence and Robotics Lab (AFAR) and serving as Associate Director of CHIA. Her research advances social and affective AI with focus on ethics, fairness, and explainability. She has been named among the World's Top 2% Scientists by Elsevier. Hatice is also a co-chair of ERR@HRI and guest edits a special issue in ACM Transactions on errors and failures in HRI.

3 Overview

The workshop will take place over half a day, taking place on the morning of the 16^{th} of March 2026. The following is an outline of the schedule for the workshop:

Introduction - 5 mins
Invited Speaker - 25 mins
Papers selected for presentation 1 - 45 mins
Coffee Break - 10 mins

Papers selected for presentation 2 - 45 mins
Papers select for presentation 3 - 45 mins
Invited Speaker - 25 mins
Breakout Groups - 25 mins
Form report group and Outro - 15 mins

Format: We will be aiming for presenters to have talks between 7-10 minutes, corresponding to 4-5 speakers per session grouped by theme. The final part of each session would then invite all the speakers back for a panel to promote open discussion. In the case of a large number of submissions, we will change the format to 1-2 minute 'elevator pitches' with half the session then dedicated to poster presentations.

Invited Speakers:

- Tom Williams, Colorado School of Mines, USA - Tom directs the Mines Interactive Robotics Research Lab. Tom will talk about errors in ethics and recovery from them.
- Ilaria Torre, Chalmers Tekniska Högskola, SWE - Ilaria is an assistant professor for Chalmers Research. The talk will be on her work regarding resolving ambiguities and repair.

Topics: This is the list of potential submission topics that we list on the website and advertising materials. This list is non-exhaustive, and we may choose to include other topics relevant to the workshop:

- Human-induced errors
  - Detecting human errors.
  - Handling human errors.
  - Communication strategies for human errors.
  - Errors relating to human factors.
- Robot-induced errors
  - Error analysis and detection.
  - Error Repair.
  - Restoring trust.
  - Wilful errors to prevent worse ones or to help learning.
- Interaction-induced errors
  - Errors during communication.
  - Production errors, e.g. incorrect or missing social signals.
  - Errors in expectation, e.g. incorrect assumption of what the other agent is aware of.
  - Design-oriented approaches, e.g.participatory and interdisciplinary design, evaluation strategies, and ethics.

4 Audience

The topic of this workshop should be of interest to all researchers in HRI and attendees from industry who work in robotics. Ensuring that robots are abled to handle errors and mistakes is key to enabling wider adoption to empower society.

There have been a number of workshops on similar topics. A previous workshop in 2022 "Working with troubles and failures in conversation between humans and robots" attracted 30 attendees [4]. The workshop "The Imperfectly Relatable Robot: An interdisciplinary workshop on the role of failure in HRI" had 40 in-person participants as part of HRI 2023. The workshop "Working with Trouble and Failures in Conversation between Humans and Robots & Is CUI Design Ready Yet?" was held at the ACM CUI conference with 30 participants [3]. With interest in the field growing further, such as a special issue in ACM Transactions on Human-Robot Interaction - "Understanding, Detecting and Repairing Errors and Failures in Human-Robot Interactions" and the growth of the HRI conference, we are hoping to recruit 40 to 50 in-person participants.

We expect to recruit participants through a number of avenues. We have a website with details of the workshop. Multiple calls for papers through HRI and robotics-themed mailing lists (e.g. HRI-Announcments) will be made to attract participants and paper submissions, advertising keynotes and other details of the workshop. We have also been endorsed and advertised by the IEEE RAS Technical Committee for Cognitive Robotics.

5 Documenting

We will ask permission from authors to record talks and share them on the workshop website. We will archive submissions using ArXiv. We will seek to publish a report in a journal publication.

Acknowledgments

H. Gunes is supported by CHANSE & NORFACE through the MICRO project, funded by UKRI/ESRC (grant no: UKRI572).

References

[1] Minja Axelsson, Micol Spitale, and Hatice Gunes. 2024. "Oh, Sorry, I Think I Interrupted You": Designing Repair Strategies for Robotic Longitudinal Well-being Coaching. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, HRI 2024, Boulder, CO, USA, March 11-15, 2024*, Dan Grollman, Elizabeth Broadbent, Wendy Ju, Harold Soh, and Tom Williams (Eds.). ACM, 13–22. doi:10.1145/3610977.3634948

[2] Shiye Cao, Maia Stiber, Amama Mahmood, and et al. 2025. ERR@HRI 2.0 Challenge: Multimodal Detection of Errors and Failures in Human-Robot Conversations. *CoRR* abs/2507.13468 (2025). arXiv:2507.13468 doi:10.48550/ARXIV.2507.13468

[3] Frank Förster, Marta Romeo, Patrick Holthaus, and et al. 2023. Working with Trouble and Failures in Conversation between Humans and Robots: Is CUI Design Ready Yet? doi:10.48550/arXiv.2401.04108

[4] Frank Förster, Marta Romeo, Patrick Holthaus, and et al. 2023. Working with Troubles and Failures in Conversation between Humans and Robots: Workshop Report. *Frontiers in Robotics and AI* 10 (2023). doi:10.3389/frobt.2023.1202306

[5] Manal Helal, Patrick Holthaus, Gabriella Lakatos, and Farshid Amirabdollahian. 2023. Chat failures and troubles: Reasons and solutions. *arXiv preprint arXiv:2309.03708* (2023).

[6] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology* 9 (2018), 861.

[7] James Reason. 1990. *Human error*. Cambridge university press.

[8] Thomas B Sheridan and Raja Parasuraman. 2005. Human-automation interaction. *Reviews of human factors and ergonomics* 1, 1 (2005), 89–129.

[9] Kavyaa Somasundaram, Katherine Harrison, Andrey Kiselev, and Amy Loutfi. 2023. An interdisciplinary approach to intelligent disobedience: A Nuanced Exploration of User Experience in Human-Induced Interaction Failures during Teleoperation. In *Human Agent Interaction*.

[10] Kavyaa Somasundaram, Andrey Kiselev, and Amy Loutfi. 2023. Intelligent disobedience: A novel approach for preventing human induced interaction failures in robot teleoperation. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 142–145.

[11] Micol Spitale, Maria Teresa Parreira, Maia Stiber, and et al. 2024. ERR@HRI 2024 Challenge: Multimodal Detection of Errors and Failures in Human-Robot Interactions. In *Proceedings of the 26th International Conference on Multimodal Interaction, ICMI 2024, San Jose, Costa Rica, November 4-8, 2024*. ACM, 652–656. doi:10.1145/3678957.3689030

[12] Christopher D. Wallbridge. 2021. *Dynamic Generation of Spatial Referring Expressions for Social Robots*. Ph. D. Dissertation. University of Plymouth.

[13] Christopher D Wallbridge, Séverin Lemaignan, Emmanuel Senft, and Tony Belpaeme. 2019. Generating spatial referring expressions in a social robot: Dynamic vs. non-ambiguous. *Frontiers in Robotics and AI* 6 (2019), 67.

[14] Christopher D Wallbridge, Alex Smith, Manuel Giuliani, Chris Melhuish, Tony Belpaeme, and Séverin Lemaignan. 2021. The effectiveness of dynamically processed incremental descriptions in human robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 1 (2021), 1–24.

[15] Xinyi Zhang, Sun Kyong Lee, Whani Kim, and Sowon Hahn. 2023. "Sorry, it was my fault": Repairing trust in human-robot interactions. *International Journal of Human-Computer Studies* 175 (2023), 103031.