

ERR@HRI 3.0 Challenge: Multimodal Detection of Errors and Anticipation in Human-Robot Interactions

Maria Teresa Parreira
Cornell University
Ithaca, NY, USA

Shiye Cao
Johns Hopkins University
Baltimore, MD, USA

Hatice Gunes
University of Cambridge
Cambridge, UK

Micol Spitale
Politecnico di Milano
Milan, Italy

Amama Mahmood
Johns Hopkins University
Baltimore, MD, USA

Wendy Ju
Cornell Tech
New York, NY, USA

Maia Stiber
Microsoft Research
Redmond, WA, USA

Chien-Ming Huang
Johns Hopkins University
Baltimore, MD, USA

ABSTRACT

As robots become increasingly integrated into human environments, their ability to detect and respond to errors remains critical for maintaining user trust and interaction quality. While recent advances in machine learning have improved error detection capabilities, most approaches are limited to specific contexts, controlled settings, or pre-extracted features, limiting their generalizability and applicability to real-world conditions. To address this challenge, the third edition of the ERR@HRI Challenge (ERR@HRI 3.0) provides researchers with two complementary datasets that enable end-to-end innovation in methods for both detecting and preventing errors in human-robot interaction. The challenge offers raw, non-anonymized video data from naturalistic settings: (1) the Bystander Affect Detection (BAD) dataset, containing webcam recordings of 45 participants' spontaneous reactions to robot and human failure scenarios; and (2) the Bad Idea dataset, featuring 29 participants' anticipatory facial responses while predicting action outcomes *before* failures occur. Both datasets were collected via crowdsourcing, capturing the inherent variability of real-world conditions—diverse lighting, camera angles, participant positioning, and environmental contexts. This naturalistic variability, while challenging, provides an authentic testbed for developing robust error detection systems. Participants are invited to develop machine learning models that can generalize across these diverse contexts and temporal stages. Submissions will be evaluated on standard classification metrics (e.g. F1-score) with consideration for real-world deployment constraints. This challenge is a step toward developing robust error detection and prevention systems that can operate in the variable conditions of real-world human-robot collaboration.

KEYWORDS

Robot Failure, Error Detection, Human-Robot Interaction, Multimodal Interaction, Benchmarking.

ACM Reference Format:

Maria Teresa Parreira, Micol Spitale, Maia Stiber, Shiye Cao, Amama Mahmood, Chien-Ming Huang, Hatice Gunes, and Wendy Ju. 2026. ERR@HRI 3.0 Challenge: Multimodal Detection of Errors and Anticipation in Human-Robot Interactions. In *Proceedings of 28th ACM International Conference on Multimodal Interaction (ICMI '26)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Robot errors, deviations from expected or intended behavior [5], are not merely technical malfunctions but social events that can disrupt interaction flow, diminish user trust, and negatively impact the overall quality of human-robot collaboration [15]. While recent advances in machine learning and sensor technologies have enhanced robots' perceptual and decision-making capabilities, autonomous systems still struggle to reliably detect when they have made a mistake, particularly in dynamic, real-world settings.

The challenge of error detection in human-robot interaction is multifaceted. Errors manifest in diverse ways—from functional failures like navigation mistakes or object manipulation errors, to social missteps such as interrupting users or misunderstanding conversational intent [11, 20]. Traditional approaches often rely on task-specific metrics or domain knowledge that limit generalizability across different robots, tasks, and contexts [4]. Additionally, methods that depend on users to explicitly report errors introduce delays that hinder timely recovery.

An alternative approach leverages the multimodal behavioral signals that humans naturally exhibit in response to unexpected events. Just as people infer errors from bystanders' reactions—confusion, concern, or surprise signaling that something went wrong—robots might detect their own mistakes through observable social cues [1]. Recent work has demonstrated that user reactions, including facial expressions, speech patterns, and body language, contain information about interaction failures [8, 9, 11, 16, 18, 19]. However, most existing approaches have been constrained in two important ways: first, they rely on pre-extracted features rather than raw sensor data, limiting the space of applicable methods; and second, they

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICMI '26, 5–9 October 2026, Napoli, Italy

© 2026 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

117 focus on controlled laboratory settings that may not reflect the
 118 variability of real-world deployments.
 119

120 2 ERR@HRI 3.0 CHALLENGE

121 2.1 Previous ERR@HRI Editions

122 The ERR@HRI initiative was established as a platform for bench-
 123 marking both datasets and machine learning models designed to
 124 detect robot errors from multimodal behavioral signals. The inaugu-
 125 ral ERR@HRI challenge [17], held in conjunction with ICMI'24, fo-
 126 cused on detecting interaction ruptures during longitudinal robotic
 127 well-being coaching sessions, providing a dataset of 89 sessions
 128 across 23 participants. The challenge tracks included 1) detection of
 129 robot mistakes (e.g., interrupting or not responding), 2) detection
 130 of user awkwardness (e.g., when the coachee feels uncomfortable
 131 interacting with the robot) and 3) detection of interaction ruptures
 132 (IR) (disjunction of 1 and 2), and a total of 3 teams participated.
 133 Building on this foundation, ERR@HRI 2.0 [3], held at ACM MM'25,
 134 expanded the scope to include conversational errors from LLM-
 135 powered robots across two distinct embodiments (social robot [2]
 136 and voice assistant[10]) and five collaborative tasks, while intro-
 137 ducing separate detection challenges for system-level errors versus
 138 user-perceived errors, and a total of 2 teams submitted results. To-
 139 gether, these challenges have demonstrated both the feasibility and
 140 the challenges of developing generalizable error detection models
 141 from human behavioral responses, with important contributions to
 142 the field in the form of participants' reports [6, 7, 13, 14, 21].
 143

144 2.2 ERR@HRI 3.0

145 Building on prior editions, ERR@HRI 3.0 addresses two challenges
 146 in existing error detection research. First, we provide raw, non-
 147 anonymized video data rather than pre-extracted features, enabling
 148 participants to explore end-to-end learning approaches, modern
 149 computer vision techniques, and novel feature representations.
 150 Second, we focus on naturalistic data collected via crowdsourc-
 151 ing, which inherently contains the variability—in lighting, cam-
 152 era positioning, participant behavior, and environmental context—
 153 characteristic of real-world deployments.
 154

155 The specific goals of ERR@HRI 3.0 are to:

- 156 • Advance multimodal error detection by enabling end-to-
 157 end learning from raw visual data
- 158 • Promote development of robust models that handle natu-
 159 ralistic data variability
- 160 • Explore the relationship between anticipatory and reactive
 161 error responses
- 162 • Establish benchmarks for generalizable error detection across
 163 contexts and temporal stages
- 164 • Foster community development of open, reproducible error
 165 detection methods

167 This challenge contributes to the broader goal of creating robots
 168 that can operate more safely and effectively in human environments
 169 by better understanding when and how their actions deviate from
 170 expectations.

171 Specifically, this challenge offers:

- 172 (1) **The Bystander Affect Detection (BAD) Dataset [1]:**

173 Audiovisual webcam recordings of 45 participants' facial

175 reactions to 46 robot and human failure scenarios, capturing
 176 observer responses to errors. This non-anonymized dataset
 177 enables fine-grained analysis of how people react when
 178 witnessing mistakes, providing a foundation for detecting
 179 errors through bystander behavior.

180 (2) **The Bad Idea Dataset [12]:** Audiovisual webcam record-
 181 ings of 29 participants' anticipatory reactions while predict-
 182 ing whether 30 action scenarios would end well or poorly,
 183 captured *before* the outcomes were revealed. This dataset
 184 uniquely focuses on anticipatory responses, enabling re-
 185 search into how observable human behavior might signal
 186 impending errors before they fully occur.

187 The BAD and Bad Idea datasets introduce unique challenges
 188 beyond traditional generalization concerns. Both datasets were
 189 collected via crowdsourcing ¹, rather than controlled laboratory
 190 conditions, meaning they inherently contain high variability in data
 191 quality. While such variability presents modeling challenges, it also
 192 offers a more realistic testbed for developing robust error detection
 193 systems that must function across diverse deployment conditions.
 194 Furthermore, these datasets are provided in non-anonymized form,
 195 preserving the original visual information rather than extracting
 196 only pre-computed features. This design decision expands the space
 197 of applicable methods, enabling researchers to apply modern com-
 198 puter vision techniques such as convolutional neural networks
 199 (CNNs) directly on image data, explore end-to-end learning ap-
 200 proaches, or develop novel feature representations tailored to error
 201 detection tasks.

202 Participants are invited to develop machine learning models
 203 that can leverage one or more of these datasets to advance our
 204 understanding of multimodal error detection across diverse HRI
 205 contexts.

207 2.3 Materials

208 A challenge website and GitHub repository will be set up with a
 209 commitment to be maintained for at least the next 3 years. For the
 210 BAD and Bad Idea datasets, which contain non-anonymized visual
 211 data, participants must agree to strict usage terms, including: no
 212 right to redistribute; datasets can only be used for the challenge; no
 213 right to use datasets to defame human participants; and proper data
 214 security measures. Access to the datasets will be provided through
 215 signing EULA agreements.

217 2.4 Datasets Overview

219 The ERR@HRI 3.0 Challenge provides two complementary datasets,
 220 each capturing different aspects of error detection in human-robot
 221 interaction.

223 **2.4.1 BAD (Bystander Affect Detection) Dataset. Data format:**
 224 Raw video files (.mp4) containing participants' faces as they watched
 225 failure scenarios (average 17.84 s); Video dataset (46 videos) used
 226 as stimuli (.mp4).

227 **Labels:** Videos in the BAD dataset are labeled according to the
 228 stimulus video content: *Failure* (human or robot) or *Control* (no
 229 failure) (6 videos).

230 ¹<https://www.prolific.com/>

Table 1: Comparison of the datasets in ERR@HRI 3.0

Characteristic	BAD Dataset	Bad Idea
Participants	45	29
Total recordings	2,054 videos	951 videos
Total time (s)	36650.58	1850.54
Temporal focus	During failure	Before failure
Error types	Observed failures	Anticipated outcomes
Setting	Crowdsourced online	
Data format	Video and audio (.mp4)	
Anonymized	No	

2.4.2 *Bad Idea Dataset. Data format:* Raw video files (.mp4) of participants watching and reacting to scenarios, along with their outcome predictions (average 1.95 s). Video dataset (30 videos) used as stimuli (.mp4).

Labels: The Bad Idea dataset includes annotations for the participant's prediction of whether the scenario will end *Well* (scenario shown ends in good outcome, e.g., a human on a bike does not fall) or *Poorly* (scenario ends in a bad outcome, e.g. robot jumping fails the landing). Importantly, these labels are based on *what the participant predicted was going to happen* after watching the video with the outcome cut-off, not what really happens. Nonetheless, data is mostly balanced (1.15 good-to-bad outcome prediction ratio).

Table 1 provides an overview of the two datasets.

2.5 Challenge Tasks

The ERR@HRI 3.0 Challenge consists of multiple sub-challenges organized around the two datasets:

2.5.1 *Track 1: Bystander Reaction Detection (BAD Dataset).* Binary classification of whether the participant is observing a failure (vs. control scenario).

2.5.2 *Track 2: Anticipatory Response Prediction (Bad Idea Dataset).* Binary classification of participants' predicted outcome (well/poorly) from their anticipatory behavior.

2.5.3 *Cross-Dataset Generalization (Optional).* Participants are encouraged to explore transfer learning and generalization across datasets, e.g. train on BAD dataset and test on Bad Idea dataset.

2.6 Evaluation Metrics

Models will be evaluated using:

- **Offline metrics:** F1-score, Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC)
- **Windowed predictions with video-level labels:** Use fixed-length sliding windows, and consider a video correctly classified if any window in the video predicts the correct label.
- **Earliest Detection Time:** For correctly classified videos, report the percentage of video elapsed before the first correct prediction (lower is better).
- **False Negative Rate per video:** Count how many windows "miss" positive predictions in error/bad outcome videos.

3 CHALLENGE LOGISTICS

3.1 Challenge Format

The ERR@HRI 3.0 Challenge will be held as a half-day event. The challenge will include:

- Presentation of challenge results and winning submissions
- Invited talk on error detection in human-robot interaction
- Panel discussion on future directions for generalizable error detection systems
- Presentations from participating teams
- Winners announcement
- Networking session for participants and organizers

3.2 Soliciting Participation

We will recruit challenge participants through multiple channels:

- Email advertisements to relevant mailing lists (e.g., robotics-worldwide, HRI-announcements, SIGCHI announcements)
- Announcements on social media platforms (X, LinkedIn, research-focused communities)
- Direct outreach to research groups working in human-robot interaction, affective computing, and multimodal machine learning
- Promotion through affiliated workshops and conferences (ACM/IEEE HRI, IEEE ICRA, IEEE/RSJ IROS, IEEE ACII)
- Challenge website with registration form and comprehensive information

3.2.1 Evaluation Protocol

- (1) **Training data:** Participants receive training and validation splits for each dataset. Data sets will be subject-independent to evaluate generalizations to unseen participants
- (2) **Test data:** Test sets (without labels) will be released one month before submission deadline
- (3) **Submission limits:** Each team may submit up to 3 times on the test set
- (4) **Ranking:** Teams will be ranked separately for each track
- (5) **Paper submission:** Participating teams must submit a description paper detailing their approach
- (6) **Code availability:** Teams are strongly encouraged to release code repositories. Code will be requested from participants for reproducibility

Teams will be ranked separately for each track based on F1-score.

Winners will be selected for:

- Track 1: Bystander Reaction Detection
- Track 2: Anticipatory Response Prediction
- Best Overall Performance (across tracks)
- Best Cross-Dataset Generalization (optional track)

3.2.2 *Review Process.* All submitted papers will undergo peer review by members of the Technical Program Committee. Proposed reviewers include Rutherford Agbeshi Patamia (Deakin University), Xun Jiang (University of Electronic Science and Technology of China, Tongji University), Peter Tisnikar (King's College London), Lennart Wachowiak (King's College London), Pradip Pramanick (University of Naples Federico II), Ruben Janssens (Ghent University).

349 3.3 Important Dates

350 The proposed timeline for the challenge is as follows:

- 351 • **March 1, 2026:** Challenge announcement and call for par-
352 ticipation
- 353 • **March 15, 2026:** Registration opens; training and valida-
354 tion data released
- 355 • **May 1, 2026:** Baseline models and evaluation scripts re-
356 leased
- 357 • **June 1, 2026:** Test data released (without labels)
- 358 • **June 15, 2026:** Submission deadline for predictions
- 359 • **June 22, 2026:** Paper submission deadline
- 360 • **July 15, 2026:** Notification of acceptance
- 361 • **August 6, 2026:** Camera-ready papers due
- 362 • **October 5, 2026:** Challenge workshop at ICMI 2026

364 3.4 Funding

366 Currently, this challenge does not have dedicated funding. Organiz-
367 ers are affiliated with institutions that support their participation
368 in this event. We will provide certificates and a symbolic award to
369 the challenge winners.

370 3.5 Organizers

372 Maria Teresa Parreira

373 PhD Candidate at the Information Science Department at Cornell
374 University, USA

375 Short bio: Maria Teresa Parreira is a PhD candidate at the Informa-
376 tion Science Department at Cornell University, supervised by Pro-
377 fessors Wendy Ju and Malte Jung. Her research focuses on building
378 tools for socially competent agents through direct interaction with
379 the environment, leveraging multimodality and human-in-the-loop
380 sensing. She was an organizer of the 2024 and 2025 ERR@HRI chal-
381 lenges.

383 Micol Spitale

384 Assistant Professor, DEIB, Politecnico di Milano, Italy

385 Short bio: Micol Spitale is an Assistant Professor (tenure-track) at
386 the Department of Electronics, Information and Bioengineering at
387 the Politecnico di Milano. In recent years, her research has been
388 focused on the field of Social Signal Processing, Human-Robot Inter-
389 action, and Affective Computing, exploring ways to develop robots
390 that are socio-emotionally adaptive in different contexts with the
391 goal of having a positive impact on the society. She was an organ-
392 izer of the 2024 and 2025 ERR@HRI challenges.

394 Maia Stiber

395 Senior Researcher at Microsoft Research, USA

396 Short bio: Maia Stiber is a Senior Researcher at Microsoft Research.
397 Her research focuses on leveraging multimodal behavioral signals
398 to develop robots that are human-aware to assist human and sup-
399 port well-being. She received her Ph.D. in Computer Science from
400 Johns Hopkins University. She was an organizer of the 2024 and
401 2025 ERR@HRI challenges.

403 Shiye Cao

404 PhD Candidate at Johns Hopkins University, USA

405 Short bio: Shiye Cao is a PhD candidate in the Intuitive Computing

407 Lab at Johns Hopkins University. Her research focuses on enhanc-
408 ing the social capabilities of robots to more closely emulate natural
409 human-human interactions and empower stakeholders to co-design
410 personalized and meaningful interactions. She was an organizer of
411 the 2025 ERR@HRI challenge.

413 Amama Mahmood

414 Postdoctoral Fellow at Johns Hopkins University, USA

415 Short bio: Amama Mahmood is a Postdoctoral Fellow at Malone
416 Center for Engineering in Healthcare at Johns Hopkins University.
417 Her research focuses on the design, development, evaluation, and
418 integration of AI assistants that support health and well-being. She
419 has been awarded Malone Postdoctoral Fellowship, JHU Computer
420 Science Department Fellowship, and Creel Family Fellowship. She
421 is a Fulbright Scholar. She was an organizer of the 2025 ERR@HRI
422 challenge.

424 Chien-Ming Huang

425 Associate Professor at Johns Hopkins University, USA

426 Short bio: Chien-Ming Huang is the John C. Malone Associate
427 Professor in the Department of Computer Science at the Johns Hop-
428 kins University. His research focuses on designing interactive AI
429 aimed to assist and collaborate with people. His research has re-
430 ceived media coverage from MIT Technology Review, Tech Insider,
431 and Science Nation. Huang completed his postdoctoral training
432 at Yale University and received his Ph.D. in Computer Science at
433 the University of Wisconsin–Madison. He is a recipient of the NSF
434 CAREER award. He was an organizer of the 2024 ERR@HRI and
435 2025 ERR@HRI challenges.

437 Hatice Gunes

438 Full Professor of Affective Intelligence & Robotics at the University
439 of Cambridge, UK

440 Short bio: Hatice Gunes is a Full Professor at the University of
441 Cambridge's Department of Computer Science and Technology,
442 where she directs the Cambridge Affective Intelligence and Robotics
443 Lab (AFAR Lab) and leads award-winning research on multimodal,
444 social, and affective intelligence for AI systems. Her work spans
445 Machine Learning, Affective Computing, Social Signal Processing,
446 and Robotics. She has received numerous awards, including best
447 paper awards at prestigious conferences such as IEEE ACII and
448 IEEE FG. She has also played key and leading roles on such events.
449 Her research on creating robotic coaches for assessing/promoting
450 mental wellbeing received extensive media coverage with >1,700
451 global reports in The Guardian, BBC News, Medical News Today,
452 Science Daily, Telegraph, Sky News, ITV News, Bloomberg, New
453 York Post etc. Most recently, she was named a Finalist for Sony
454 Women in Tech Award with Nature 2025. She was an organizer of
455 the 2024 and 2025 ERR@HRI challenges.

456 Wendy Ju

457 Associate Professor of Information Science at the Jacobs Technion-
458 Cornell Institute at Cornell Tech, USA

459 Short bio: Wendy Ju is an Associate Professor of Information Sci-
460 ence at the Jacobs Technion-Cornell Institute at Cornell Tech and
461 inaugural faculty in Cornell's new multi-college Design Tech de-
462 partment. Prof. Ju has innovated numerous methods for early-stage

prototyping of automated systems to understand how people will respond to systems before the systems are built. In her research, Prof. Ju has worked closely with industrial partners such as Toyota, Spotify, Intel, Ford, Bosch, Renault, Fiat Chrysler, Panasonic, Volvo, Nissan and Mitsubishi. She has a PhD in Mechanical Engineering from Stanford, and a Master's in Media Arts and Sciences from MIT. Her monograph on The Design of Implicit Interactions was published in 2015. She was an organizer of the 2024 and 2025 ERR@HRI challenges.

3.6 Call for Participation

The ERR@HRI 3.0 challenge addresses the problem of error detection in human-robot interaction (HRI) by providing the community with complementary datasets that span the temporal spectrum of error management, from anticipatory responses before failures occur to reactive responses during observed errors.

Upon acceptance of the ERR@HRI 3.0 terms and conditions and signing the End User Licence Agreement (EULA), we will share two datasets containing raw video recordings: (1) the Bystander Affect Detection (BAD) dataset with webcam recordings of 45 participants' reactions to 46 robot and human failure scenarios, and (2) the Bad Idea dataset with webcam recordings of 29 participants' anticipatory reactions while predicting action outcomes before failures occur. Both datasets are provided as raw video files to enable end-to-end learning approaches.

The ERR@HRI 3.0 challenge tasks are:

- **Track 1:** Bystander Reaction Detection (BAD Dataset) - Binary classification of whether the participant is observing a failure vs. control scenario
- **Track 2:** Anticipatory Response Prediction (Bad Idea Dataset) - Binary classification of participants' predicted outcome (well/poorly) from their anticipatory behavior
- **Cross-Dataset Generalization (Optional):** Transfer learning across datasets

We invite participants to collaborate in teams to submit their multi-modal ML models for evaluation, which will be benchmarked based on various performance metrics, including F1-score, and early detection metrics. Teams can participate in all tracks and they will be evaluated separately.

For more information about the challenge, check our website [PLACEHOLDER: URL]. To register and access the challenge data, please fill the registration form you will find on the website. For each task, training and validation data will be made available to participants. All submissions will be evaluated on a held-out test dataset to ensure a fair comparison. Participants will also be encouraged to submit a conference-style paper describing their proposed approach for tackling the challenge task(s) as well as the results obtained following the template of the ICMI 2026 main conference papers. All accepted challenge papers will be published in the main ACM ICMI 2026 proceedings.

Challenge timeline:

- **March 1, 2026:** Challenge announcement and call for participation
- **March 15, 2026:** Registration opens; training and validation data released

- **May 1, 2026:** Baseline models and evaluation scripts released
- **June 1, 2026:** Test data released (without labels)
- **June 15, 2026:** Submission deadline for predictions
- **June 22, 2026:** Paper submission deadline
- **July 15, 2026:** Notification of acceptance
- **August 6, 2026:** Camera-ready papers due
- **October 5, 2026:** Challenge workshop at ICMI 2026

The Organisers:

- Maria Teresa Parreira, Cornell University, USA
- Prof. Micol Spitale, Politecnico di Milano, Italy
- Dr. Maia Stiber, Microsoft Research, USA
- Shiye Cao, Johns Hopkins University, USA
- Dr. Amama Mahmood, Johns Hopkins University, USA
- Prof. Chien-Ming Huang, Johns Hopkins University, USA
- Prof. Hatice Gunes, University of Cambridge, UK
- Prof. Wendy Ju, Cornell University, USA

Contact us via email: [PLACEHOLDER]

4 CONCLUSION

The ERR@HRI 3.0 Challenge addresses critical gaps in robot error detection research by providing two complementary datasets that span the temporal spectrum of error management—from anticipatory responses before failures occur to reactive responses during observed errors. By offering raw, non-anonymized video data collected in naturalistic crowdsourced settings, the challenge enables end-to-end innovation in error detection methods while exposing models to the variability characteristic of real-world deployments.

This challenge represents a step toward developing more adaptive, context-aware robots that can operate safely in human environments. The techniques developed through this challenge have potential to advance our understanding of multimodal behavioral analysis, improve robot error awareness, and ultimately enhance the quality and safety of human-robot collaboration in real-world applications.

We invite researchers from the robotics, computer vision, affective computing, and human-robot interaction communities to participate in this challenge. We look forward to the innovative approaches that participants will develop and the insights this challenge will generate about multimodal error detection in human-robot interaction.

REFERENCES

- [1] Alexandra Bremers, Maria Teresa Parreira, Xuanyu Fang, Natalie Friedman, Adolfo Ramirez-Aristizabal, Alexandria Pabst, Mirjana Spasojevic, Michael Kuniavsky, and Wendy Ju. 2023. The Bystander Affect Detection (BAD) Dataset for Failure Detection in HRI. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11443–11450.
- [2] Shiye Cao, Jiwon Moon, Amama Mahmood, Victor Nikhil Antony, Ziang Xiao, Anqi Liu, and Chien-Ming Huang. 2025. Interruption handling for conversational robots. *Robotics: Science and Systems* (2025).
- [3] Shiye Cao, Maia Stiber, Amama Mahmood, Maria Teresa Parreira, Wendy Ju, Micol Spitale, Hatice Gunes, and Chien-Ming Huang. 2025. ERR@HRI 2.0 Challenge: Multimodal Detection of Errors and Failures in Human-Robot Conversations. In *Proceedings of the 33rd ACM International Conference on Multimedia* (Dublin, Ireland) (MM '25). Association for Computing Machinery, New York, NY, USA, 14130–14135. doi:10.1145/3746027.3762073
- [4] Jennifer Carlson and Robin R Murphy. 2005. How UGVs physically fail in the field. *IEEE Transactions on robotics* 21, 3 (2005), 423–437.

- 581 [5] Shabee Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures
582 in human-robot interaction: Literature review and model development. *Frontiers
583 in psychology* 9 (2018), 861.
- 584 [6] Ruben Janssens, Eva Verhelst, and Mathieu De Coster. 2024. Predicting Errors
585 and Failures in Human-Robot Interaction from Multi-Modal Temporal Data. In
586 *Proceedings of the 26th International Conference on Multimodal Interaction* (San
587 Jose, Costa Rica) (ICMI '24). Association for Computing Machinery, New York,
588 NY, USA, 671–676. doi:10.1145/3678957.3688388
- 589 [7] Xun Jiang, Shuang Li, Chong Liu, and Xing Xu. 2025. Multimodal Time Series
590 Alignment for Error Detection in Human Robot Interactions. In *Proceedings of the 33rd ACM
591 International Conference on Multimedia* (Dublin, Ireland) (MM '25). Association for
592 Computing Machinery, New York, NY, USA, 14143–14149. doi:10.1145/3746027.3762075
- 593 [8] Dimosthenis Kontogiorgos, Andre Pereira, Boran Sahindal, Sanne van Waveren,
594 and Joakim Gustafson. 2020. Behavioural responses to robot conversational
595 failures. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-
596 Robot Interaction*, 53–62.
- 597 [9] Dimosthenis Kontogiorgos, Minh Tran, Joakim Gustafson, and Mohammad So-
598 leymani. 2021. A systematic cross-corpus analysis of human reactions to robot
599 conversational failures. In *Proceedings of the 2021 International Conference on
600 Multimodal Interaction*, 112–120.
- 601 [10] Amama Mahmood, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-
602 Ming Huang. 2025. User interaction patterns and breakdowns in conversing
603 with llm-powered voice assistants. *International Journal of Human-Computer
604 Studies* 195 (2025), 103406.
- 605 [11] Nicole Mirmig, Manuel Giuliani, Gerald Stollnberger, Susanne Stadler, Roland
606 Buchner, and Manfred Tscheligi. 2015. Impact of Robot Actions on Social Signals
607 and Reaction Times in HRI Error Situations. In *Social Robotics*, Adriana Tapus,
608 Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.).
609 Springer International Publishing, Cham, 461–471.
- 610 [12] Maria Teresa Parreira, Sukruth Gowdru Lingaraju, Adolfo Ramirez-Artizabal,
611 Alexandra Bremers, Manaswi Saha, Michael Kuniavsky, and Wendy Ju. 2024. “Bad
612 Idea, Right?” Exploring Anticipatory Human Reactions for Outcome Prediction
613 in HRI. In *2024 33rd IEEE International Conference on Robot and Human Interactive
614 Communication (ROMAN)*, 2072–2078. doi:10.1109/RO-MAN60168.2024.10731310
- 615 [13] Rutherford Agbeshi Patamia, Ha Pham Thien Dinh, Ming Liu, and Akansel
616 Cosgun. 2025. Beyond Technical Failures: Multimodal Time-Series Modelling
617 for Detecting Social Breakdowns and User Repair Attempts in Human-Robot
618 Interaction. In *Proceedings of the 33rd ACM International Conference on Multimedia*
619 (Dublin, Ireland) (MM '25). Association for Computing Machinery, New York,
620 NY, USA, 14136–14142. doi:10.1145/3746027.3762074
- 621 [14] Pradip Pramanick and Silvia Rossi. 2024. PRISCA at ERR@HRI 2024: Multimodal
622 Representation Learning for Detecting Interaction Ruptures in HRI. In *Proceed-
623 ings of the 26th International Conference on Multimodal Interaction* (San Jose,
624 Costa Rica) (ICMI '24). Association for Computing Machinery, New York, NY,
625 USA, 666–670. doi:10.1145/3678957.3688387
- 626 [15] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dauten-
627 hahn. 2015. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and
628 Personality on Human-Robot Cooperation and Trust. In *Proceedings of the Tenth
629 Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland,
630 Oregon, USA) (HRI '15). Association for Computing Machinery, New York, NY,
631 USA, 141–148. doi:10.1145/2696454.2696497
- 632 [16] Micol Spitali, Minja Axelsson, and Hatice Gunes. 2023. VITA: A Multi-modal
633 LLM-based System for Longitudinal, Autonomous, and Adaptive Robotic Mental
634 Well-being Coaching. *arXiv preprint arXiv:2312.09740* (2023).
- 635 [17] Micol Spitali, Maria Teresa Parreira, Maia Stiber, Minja Axelsson, Neval Kara,
636 Garima Kankariya, Chien-Ming Huang, Malte Jung, Wendy Ju, and Hatice Gunes.
637 2024. ERR@HRI 2024 Challenge: Multimodal Detection of Errors and Failures in
638 Human-Robot Interactions. *arXiv:2407.06094 [cs.RO]* <https://arxiv.org/abs/2407.06094>
- 639 [18] Maia Stiber. 2022. Effective Human-Robot Collaboration via Generalized Robot
640 Error Management Using Natural Human Responses. In *Proceedings of the 2022
641 International Conference on Multimodal Interaction* (Bengaluru, India) (ICMI '22).
642 Association for Computing Machinery, New York, NY, USA, 673–678. doi:10.
643 1145/3536221.3557028
- 644 [19] Maia Stiber, Russell H Taylor, and Chien-Ming Huang. 2023. On using social
645 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
646 International Conference on Human-Robot Interaction*. 222–230.
- 647 [20] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
648 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
649 doi:10.1145/3439720
- 650 [21] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
651 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
652 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
653 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
654 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
655 3688386
- 656 [22] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2024. A Taxonomy of
657 Social Errors in Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE
658 International Conference on Human-Robot Interaction*. 222–230.
- 659 [23] Maia Stiber, Russell H Taylor, and Chien-Ming Huang. 2023. On using social
660 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
661 International Conference on Human-Robot Interaction*. 222–230.
- 662 [24] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
663 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
664 doi:10.1145/3439720
- 665 [25] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
666 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
667 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
668 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
669 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
670 3688386
- 671 [26] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
672 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
673 International Conference on Human-Robot Interaction*. 222–230.
- 674 [27] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
675 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
676 doi:10.1145/3439720
- 677 [28] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
678 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
679 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
680 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
681 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
682 3688386
- 683 [29] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
684 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
685 International Conference on Human-Robot Interaction*. 222–230.
- 686 [30] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
687 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
688 doi:10.1145/3439720
- 689 [31] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
690 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
691 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
692 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
693 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
694 3688386
- 695 [32] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
696 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
697 International Conference on Human-Robot Interaction*. 222–230.
- 698 [33] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
699 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
700 doi:10.1145/3439720
- 701 [34] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
702 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
703 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
704 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
705 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
706 3688386
- 707 [35] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
708 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
709 International Conference on Human-Robot Interaction*. 222–230.
- 710 [36] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
711 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
712 doi:10.1145/3439720
- 713 [37] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
714 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
715 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
716 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
717 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
718 3688386
- 719 [38] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
720 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
721 International Conference on Human-Robot Interaction*. 222–230.
- 722 [39] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
723 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
724 doi:10.1145/3439720
- 725 [40] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
726 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
727 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
728 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
729 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
730 3688386
- 731 [41] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
732 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
733 International Conference on Human-Robot Interaction*. 222–230.
- 734 [42] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
735 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
736 doi:10.1145/3439720
- 737 [43] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
738 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
739 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
740 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
741 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
742 3688386
- 743 [44] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
744 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
745 International Conference on Human-Robot Interaction*. 222–230.
- 746 [45] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
747 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
748 doi:10.1145/3439720
- 749 [46] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
750 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
751 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
752 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
753 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
754 3688386
- 755 [47] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
756 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
757 International Conference on Human-Robot Interaction*. 222–230.
- 758 [48] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
759 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
760 doi:10.1145/3439720
- 761 [49] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
762 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
763 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
764 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
765 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
766 3688386
- 767 [50] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
768 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
769 International Conference on Human-Robot Interaction*. 222–230.
- 770 [51] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
771 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
772 doi:10.1145/3439720
- 773 [52] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
774 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
775 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
776 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
777 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
778 3688386
- 779 [53] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
780 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
781 International Conference on Human-Robot Interaction*. 222–230.
- 782 [54] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
783 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
784 doi:10.1145/3439720
- 785 [55] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
786 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
787 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
788 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
789 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
790 3688386
- 791 [56] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
792 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
793 International Conference on Human-Robot Interaction*. 222–230.
- 794 [57] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
795 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
796 doi:10.1145/3439720
- 797 [58] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
798 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
799 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
800 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
801 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
802 3688386
- 803 [59] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
804 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
805 International Conference on Human-Robot Interaction*. 222–230.
- 806 [60] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
807 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
808 doi:10.1145/3439720
- 809 [61] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
810 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
811 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
812 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
813 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
814 3688386
- 815 [62] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
816 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
817 International Conference on Human-Robot Interaction*. 222–230.
- 818 [63] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
819 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
820 doi:10.1145/3439720
- 821 [64] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
822 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
823 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
824 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
825 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
826 3688386
- 827 [65] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
828 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
829 International Conference on Human-Robot Interaction*. 222–230.
- 830 [66] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
831 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
832 doi:10.1145/3439720
- 833 [67] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
834 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
835 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
836 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
837 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
838 3688386
- 839 [68] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
840 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
841 International Conference on Human-Robot Interaction*. 222–230.
- 842 [69] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
843 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
844 doi:10.1145/3439720
- 845 [70] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
846 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
847 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
848 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
849 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
850 3688386
- 851 [71] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
852 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
853 International Conference on Human-Robot Interaction*. 222–230.
- 854 [72] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
855 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
856 doi:10.1145/3439720
- 857 [73] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
858 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
859 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
860 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
861 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
862 3688386
- 863 [74] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
864 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
865 International Conference on Human-Robot Interaction*. 222–230.
- 866 [75] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
867 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
868 doi:10.1145/3439720
- 869 [76] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
870 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
871 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
872 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
873 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
874 3688386
- 875 [77] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
876 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
877 International Conference on Human-Robot Interaction*. 222–230.
- 878 [78] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
879 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
880 doi:10.1145/3439720
- 881 [79] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
882 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
883 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
884 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
885 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
886 3688386
- 887 [80] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
888 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
889 International Conference on Human-Robot Interaction*. 222–230.
- 890 [81] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
891 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
892 doi:10.1145/3439720
- 893 [82] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
894 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
895 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
896 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
897 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
898 3688386
- 899 [83] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
900 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
901 International Conference on Human-Robot Interaction*. 222–230.
- 902 [84] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
903 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
904 doi:10.1145/3439720
- 905 [85] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
906 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
907 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
908 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
909 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
910 3688386
- 911 [86] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
912 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
913 International Conference on Human-Robot Interaction*. 222–230.
- 914 [87] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
915 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
916 doi:10.1145/3439720
- 917 [88] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
918 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
919 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
920 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
921 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
922 3688386
- 923 [89] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
924 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
925 International Conference on Human-Robot Interaction*. 222–230.
- 926 [90] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
927 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
928 doi:10.1145/3439720
- 929 [91] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
930 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
931 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
932 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
933 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
934 3688386
- 935 [92] Maria Teresa Parreira, Maia Stiber, and Chien-Ming Huang. 2023. On using social
936 signals to enable flexible error-aware hri. In *Proceedings of the 2023 ACM/IEEE
937 International Conference on Human-Robot Interaction*. 222–230.
- 938 [93] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-
939 Robot Interaction. *J. Hum.-Robot Interact.* 10, 2, Article 13 (feb 2021), 32 pages.
940 doi:10.1145/3439720
- 941 [94] Lennart Wachowiak, Peter Tisnikar, Andrew Coles, Gerard Canal, and Oya
942 Celikutan. 2024. A Time Series Classification Pipeline for Detecting Interaction
943 Ruptures in HRI Based on User Reactions. In *Proceedings of the 26th International
944 Conference on Multimodal Interaction* (San Jose, Costa Rica) (ICMI '24). Association
945 for Computing Machinery, New York, NY, USA, 657–665. doi:10.1145/3678957.
946 3688386
- 947 [9