EXTENDED-ABSTRACT

Exploring the Safety, Trust, Psychosocial, Legal, and Economic Impact of Cyber Attacks on Social Robots: What Should People Know and How Can We Mitigate Risks?

PHILLIP MORGAN, Cardiff University, Cardiff, South Glamorgan, Wales, U.K.

DAMITH HERATH, University of Canberra, Canberra, ACT, Australia

PRAMINDA CALEB-SOLLY, University of Nottingham, Nottingham, Nottinghamshire, U.K.

MATTHEW STUDLEY, University of the West of England, Bristol, U.K.

ELIZABETH WILLIAMS, The Australian National University, Canberra, ACT, Australia

AURORA AN-LIN HU, University of Canberra, Canberra, ACT, Australia

View all

Published: 16 March 2026

HRI '26: 21st ACM/IEEE International Conference on Human-Robot Interaction
March 16 - 19, 2026
Scotland, Edinburgh, UK

Conference Sponsors:
SIGAI
SIGCHI

https://doi.org/10.1145/3776734.3788837
ISBN: 9798400723216

Exploring the Safety, Trust, Psychosocial, Legal, and Economic Impact of Cyber Attacks on Social Robots: What Should People Know and How Can We Mitigate Risks?

Phillip Morgan
School of Psychology, AI, Robotics & Human-Machine Systems Centre, Cardiff University
Cardiff UK

Damith Herath
Collaborative Robotics Lab, University of Canberra
Bruce AU

Praminda Caleb-Solly
School of Computer Science, University of Nottingham
Nottingham UK

Matthew Studley
Bristol Robotics Laboratory, University of the West of England
Bristol UK

Elizabeth Williams
School of Engineering, Australian National University
AU

Aurora An-Lin Hu
Collaborative Robotics Lab, University of Canberra
Bruce AU

Eduardo B. Sandoval
School of Art and Design, University of New South Wales
Sydney AU

Maleen Jayasuriya
Collaborative Robotics Lab, University of Canberra
Bruce AU

Min Wang
Collaborative Robotics Lab, University of Canberra
Bruce AU

Janie Busby Grant
Collaborative Robotics Lab, University of Canberra
Bruce AU

Abstract

The increasing prevalence of interactive, mobile robots in domestic and social spaces requires not only a comprehensive examination of the security challenges inherent in their large-scale deployment, but understanding how we as a community can support safe and successful adoption of robots. This workshop provides a forum for researchers, practitioners, and stakeholders from a range of disciplines to build expertise and networks in safety, trust, psychosocial, legal and economic aspects for the secure deployment of social robots in domestic environments. The end goal of this workshop is to enable the development of methodologies and recommendations to empower people to understand and decide on how their robots are secured, and approaches for raising awareness of the impact of an attack. The workshop will incorporate keynote presentations from leading experts, a series of lightning talks, and a scenario planning activity in which teams of participants use custom-designed prospective scenarios to explore the precursors, processes, contexts and impacts of a breach of security in a robotic system. This will allow a deep-dive into the complexities of establishing and maintaining security and trust in interactive robotic applications, enabling long-term acceptance and adoption of trustworthy, secure, and safe interactive robots.

CCS Concepts

• **Security and privacy**; • **Social and professional topics** → *Privacy policies*;

HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3788837

Keywords

trust, safety, security, cybersecurity, secure-robotics, HRI, social robots, health care

ACM Reference Format:
Phillip Morgan, Damith Herath, Praminda Caleb-Solly, Matthew Studley, Elizabeth Williams, Aurora An-Lin Hu, Eduardo B. Sandoval, Maleen Jayasuriya, Min Wang, and Janie Busby Grant. 2026. Exploring the Safety, Trust, Psychosocial, Legal, and Economic Impact of Cyber Attacks on Social Robots: What Should People Know and How Can We Mitigate Risks?. 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/10.1145/3776734.3788837

1 Introduction

As robots become increasingly autonomous and interconnected, they are more susceptible to threats, including cyberattacks, physical intrusions, and AI manipulation [4]. These vulnerabilities not only jeopardize the operational integrity of robots but undermine human trust, which is necessary for widespread, long-term acceptance, adoption and continued use. Understanding the security of robots is critical to their development, deployment, and assurance (including via standards and regulation). Within the lens of human-robot interaction (HRI), secure robotics includes questions of perceived and actual trust, safety and risk, approaches to identify and address vulnerabilities, and crucially, the design of mechanisms and interfaces to support users to have oversight over this key component of the technology they develop and/or use. The necessity for a dedicated forum to explore the challenges associated with robotic security has been underscored by recent workshops (up to 50 participants each) including Robot Trust for Symbiotic Societies (RTSS) at IEEE ICRA 2024 and Trust, Acceptance and Social Cues in HRI (SCRITA) at IEEE RO-MAN 2022-2024, which focused on establishing and maintaining trust in human-robot interactions. The rapidly diversifying rollout of RAS internationally, and the

expanding scope of secure robotics, encompassing cybersecurity, safety, and trust, requires a comprehensive examination of these interconnected aspects to deploy interactive robots at scale [3]. **This half-day HRI 2026 workshop** will provide a platform and network for focused discussion of the human-facing impacts of security-based violations of interactive robots.

2 Secure Robotics: Interdependence of Trust, Safety, and Security

The escalating development and integration of interactive, assistive, and social RAS into our lives has brought a pressing need to address the issue of **human trust** in them and those developing them. Trust fundamentally impacts acceptance, adoption, and continued usage. *0ptimally calibrated* trust is crucial for effective human-robot collaboration and successful long-term acceptance and adoption, as well as continued use following situations where something goes wrong. Degradation of trust in the event of an error or accident, can substantially damage willingness to utilize robots and automated systems (RAS), especially in safety-critical contexts [6]. Such ironies of RAS have been warned about for over 40 years (e.g., [1, 5]).

Security is the non-negotiable foundation upon which both Trust and Safety are built. A security breach is not merely a data incident; it is a direct threat to the physical integrity and safety of the system, which can immediately (and possibly permanently) erode human trust in the system. Robust cybersecurity measures (including those based on assurance, ethical and responsible by design principles) are essential to optimize the confidentiality and integrity of the robot's control systems and data, thereby ensuring its predictability and preventing adversarial manipulation that could lead to physical harm.

Safety in this context must extend beyond mitigating engineering faults to include defenses against malicious attacks. The capacity for a robot to operate safely, even when under a cyber attack, is the final element that validates human trust. The synthesis of **Trust**, **Safety**, and **Security** into a unified and interdependent set of design requirements is the core definition of **Secure Robotics** and the focus of this workshop.

3 Workshop Activities

This half-day workshop is designed to foster awareness and understanding of user-centered secure robotics within HRI and related disciplines, and equip participants with appropriate foci, frameworks and collaborations to inform their ongoing robotics activities with this perspective. Table 1 outlines the workshop program.

Table 1: Tentative Workshop Program

Time	Speaker/s	Theme/Title
00:10	Organisers	Welcome and Introduction
00:10	Prof Tatsuhiko Inatani	Keynote 1 (Legal)
00:10	Prof Praminda Caleb-Solly	Keynote 2 (Safety)
00:10	Prof Matthew Studley	Keynote 3 (Ethics)
00:10	Prof David Cotterrell	Keynote 4 (Society)
00:40	Participants	Scenario Planning Activity
00:30	Break	Break
00:30	Early career/PhD (from public call)	Lightning Talks (x 12)
00:40	Participants	Small Groups: Applications
00:20	Organisers	Collation of Key Themes, Future Planning and Wrap-Up

Welcome and Introduction to Key Concepts (10 minutes) The organizers will present a succinct overview of the current state of security and safety in robotic systems, implications and importance of violations of these systems, and introduce the HEARS (human-centered, ethical, assured, responsible, secure: Morgan & Williams, 2025 - in press) framework as a means of conceptualizing the factors within such systems.

Keynote Presentations (40 minutes) Three brief keynote presentations by distinguished researchers will offer insights into cutting-edge advancements and real-world experiences and perspectives across domains spanning, but not limited to, assisted living, healthcare, emerging AI and automation solutions, and RAS legislation. The organizers who are international experts in these fields will present keynotes, alongside **Prof Tatsuhiko Inatani**, a legal expert on social robotics, HRI, robot ethics, and robots and autonomous systems at Kyoto University School of Law. They will summarize the fundamental tenets of secure robotics, encompassing trust, security, and safety, and explore the imperative of integrating these considerations into real-world HRI systems.

Scenario Planning Activity (40 minutes) First developed in defense settings, and now widely used across many fields, scenario planning is designed to scaffold strategic thinking, consideration of diverse options, and development of tactical approaches to problem-solving [2]. This activity will help develop participants' understanding of different contextual factors relevant to secure robotics, with a focus on conceptualising the user's experience and possible impacts from different examples of system violations. Workshop participants will form multidisciplinary teams provided with different initial problem state scenarios, tools and agendas. By situating the problem and context in real-world scenarios, this approach can engender consideration of a myriad of factors and outcomes that a more theoretical or discussion-based approach may not. The activity will include time for teams to discuss, plan, and provide notes and strategies, and report back to the broader workshop group for integration by organizers.

Lightning Talks (30 minutes) To further enrich the discourse and encourage active engagement, the workshop will invite submissions of position papers, industry case studies and early research findings pertinent to the workshop theme. This session specifically target the participation of Early Career Researchers (ECRs) and encourage disciplinary diversity. Selected contributions will be showcased in a series of lightning talks, providing a platform for concise, impactful presentations that stimulate further discussion and collaboration.

Small Group Discussion Activity: Applications Focus (40 minutes) The final discussion session will be in small groups and will focus on the application of the issues, frameworks and knowledge introduced across the workshop to real-world applications. Each group will be given a specific contextual framings within which to consider potential security harms and vulnerabilities, system strengths, mechanisms to support awareness and encourage security-conscious behavior, and methodologies to capture real-world intervention assessment.

Collation of Key Themes, Future Planning and Wrap-Up (20 minutes) The organisers will then draw the group back together to pull together key themes emerging from the discussions and initial mapping of key factors identified to a HEARS model for the scenario(s) focused on. We will wrap up with a discussion plans for further activity in HRI secure robotics.

4 Target Audience

Our audience will include researchers and practitioners interested in HRI, trust, secure robotics, creative robotics, cyber security, human factors, technology law, and related areas. No prior expertise is required, though foundational knowledge will enrich discussions and foster deep interactions.

5 Room Equipment

We require a room with audiovisual presentation equipment and the ability to host small group discussions. Based on previous recent workshops on this topic we anticipate between 30-40 participants, but we can accommodate smaller or larger numbers.

6 Recruiting Participants

To ensure a diverse and engaged group of participants, our recruitment strategy will leverage multiple approaches. We will: 1) tap into our extensive professional networks, including academic staff and PhD students and industry practitioners; 2) issue a call for presentations to attract those with relevant expertise and foster an environment of collaboration; and 3) utilise social media platforms.

7 Post-workshop Activities

The workshop will be documented, including emerging themes from the scenario planning activity and the applications small-group discussion. These will be integrated and presented on the workshop webpage at the conclusion of the workshop, alongside recommended readings, frameworks and methodologies from the organizers and keynote speakers an lightning speakers, to serve as an updating resource and central touchpoint for future discussions among participants, including a Part II of the workshop in 2027.

8 Workshop Organizers

Phillip Morgan is Professor of Human Factors and Cognitive Psychology (School of Psychology, Cardiff University, UK). He is DoR for the Centre for AI, Robotics, and Human-Machine Systems, and Director of an Airbus–Cardiff University Centre of Excellence in Human-Centric Cyber Security (seconded to industry part-time), with expertise in human aspects of AI, RAS, trust in disruptive technologies, Cyberpsychology, transportation, HMI/HCI, and adaptive cognition, with 50+ grants to date (£40mUK). morganphil@cardiff.ac.uk

Damith Herath is Professor in Robotics and Art at the UoC, where he leads the Collaborative Robotics Lab: an interdisciplinary, industry-led research group investigating HRI across diverse domains. He is an award-winning entrepreneur with 20+ years of experience in multidisciplinary and translational research in robotics. damith.herath@canberra.edu.au

Praminda Caleb-Solly is Professor of Embodied Intelligence at the University of Nottingham (UK) with expertise in assistive robotics and intelligent sensing, and research in HRI for inclusion. She is on the ISO TC299/WG2 Service Robot Safety and Ethics Working Groups, and co-founder of Robotics for Good. praminda.caleb-solly@nottingham.ac.uk

Matthew Studley is UWE Bristol's Engineering Director of Research, driving strategic initiatives and ensuring research aligns with ethical standards. His work spans robotics, machine learning and AI, autonomous weapons, standards and sustainability, embedding ethics into every stage. He has served as an advisor for the World Economic Forum and UK Government. matthew2.studley@uwe.ac.uk

Elizabeth Williams is an Associate Professor at the ANU School of Engineering. A physicist and complex systems scientist with expertise in qualitative and mixed-methods, technology design and development, she researches safety-critical systems - designing for safety, security and trust at scale. elizabeth.williams@anu.edu.au

Eduardo B. Sandoval is a Scientia Lecturer at UNSW. 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. e.sandoval@unsw.edu.au

Aurora An-Lin Hu is a PhD candidate at the UoC, pursuing her research at the Collaborative Robotics Lab. She has a background in psychology and criminology, and her PhD focuses on HRI, particularly the expectation gap and its impact on user adoption. aurora.hu@canberra.edu.au

Maleen Jayasuriya is a Lecturer in Robotics at the UoC's Collaborative Robotics Lab. His research focuses on human-robot collaboration, robot navigation and explainable AI in robotics. maleen.jayasuriya@canberra.edu.au

Min Wang is a Lecturer in AI at UoC and a Research Fellow at UNSW, working on human-machine interaction with a focus on privacy, security and trust. min.wang@canberra.edu.au

Janie Busby Grant is an Associate Professor in Cognitive Psychology and HRI at the Collaborative Robotics Lab, UoC. Her research examines the development and deployment of autonomous systems in applications such as health and aged care, with a focus on robot adoption, expectation and presence. janie.busbygrant@canberra.edu.au

References

[1] L. Bainbridge. 1983. Ironies of Automation. In *Analysis, Design and Evaluation of Man–Machine Systems*, G. Johannsen and J. E. Rijnsdorp (Eds.). Pergamon, 129–135. doi:10.1016/B978-0-08-029348-6.50026-9

[2] Kathya Cordova-Pozo and Etienne A.J.A. Rouwette. 2023. Types of Scenario Planning and Their Effectiveness: A Review of Reviews. *Futures* 149 (2023). doi:10.1016/j.futures.2023.103153

[3] Adam Haskard and Damith Herath. 2025. Secure Robotics: Navigating Challenges at the Nexus of Safety, Trust, and Cybersecurity in Cyber-Physical Systems. *ACM Comput. Surv.* 57, 9, Article 222 (April 2025), 48 pages. doi:10.1145/3723050

[4] Daniel M. Lofaro. 2016. Secure robotics. In *13th International Conference on Ubiquitous Robots & Ambient Intelligence (URAI)*. 311–313. doi:10.1109/URAI.2016.7734049

[5] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253. doi:10.1518/001872097778543886

[6] Qiyuan Zhang, Christopher D. Wallbridge, Dylan M. Jones, and Phillip L. Morgan. 2024. Public perception of autonomous vehicle capability determines judgment of blame and trust in road traffic accidents. *Transportation Research Part A: Policy and Practice* 179 (2024), 103887. doi:10.1016/j.tra.2023.103887

Received 2025-10-10; accepted 2025-11-21