What Are They Worried About? Initial Insights into Users' Mental Models of Robot Hazards

James M. Berzuk
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada
berzukj@myumanitoba.ca

Samuel R. M. Barrett
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada
barret15@myumanitoba.ca

James E. Young
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada
young@cs.umanitoba.ca

Abstract

We are conducting design workshops with potential users of robots to discover their mental models of safety risks in human-robot interaction. Robots may pose many dangers, such as physical harm, deception, privacy concerns, and more. Initial results suggest that people are already aware of, and concerned by, a wide range of specific risks associated with robots, which we enumerate in this paper, together with the organization and framing they use to understand them. Further, our ongoing study investigates their ideas for strategies to communicate these risks to users. Through this study, we ultimately aim to synthesize these results into a blueprint toward the transparent communication of risks in human-robot interaction.

CCS Concepts

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI theory, concepts and models**;

Keywords

human-robot interaction, social robots, human-robot hazards

ACM Reference Format:
James M. Berzuk, Samuel R. M. Barrett, and James E. Young. 2026. What Are They Worried About? Initial Insights into Users' Mental Models of Robot Hazards. In *Proceedings of the 14th International Conference on Human-Agent Interaction (HAI '26), November 16–19, 2026, Osaka, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3841580.3845790

1 Introduction

Interaction with robots presents many dangers which may not be apparent from their designs, such as physical impact and injury [12, 26, 40], invasive contact [8, 27, 34], social manipulation [11, 21, 31, 32, 42], and surveillance [13, 18]. Some capabilities, such as generative AI, may be ethically concerning or have behavioural implications (e.g., unpredictable, offensive utterances). We argue users must be informed of these risks to make informed decisions [6, 29, 33], and that robots must systematically communicate hazards of interaction (e.g., via warning labels).

Much work has investigated perceptions of robot safety, whether through scales of perceived risk [4, 20], or through factors contributing to sense of safety [1]. These support designing robots which make users *feel* safe, but actual capabilities may not match perceptions [30, 36]. While a robot's dangerous behaviour may be constrained (e.g., [16, 24]), this can limit the interaction and user experience [5, 23, 41]. Users must be informed to choose which risks are acceptable.

Prior work has noted the importance of robots communicating safety information to users [15], whether by verbal warning [39], gestures [28], or displays [35]. Berzuk & Young [7] presented an initial concept of a hazard labelling system for robots, highlighting key challenges including the need for a defined set of hazards which are relevant to users.

To effectively communicate interaction hazards to users, we must understand how they conceive of those risks, to align the presentation of the information with their preconceptions. Employing research-through-design [10, 38, 43], we engage potential users in design workshops [3] (Figure 1) to discover their mental models of hazards and hazard communication, and explore the design space. While this study is ongoing, our initial results provide insight into risks users are concerned by and how they categorize them. We aim to produce guidelines to support designers in the transparent and effective communication of safety information.

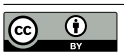
2 Study Design

We engage participants in group design workshop sessions [3, 9] (Figure 1) to elicit preferences and preconceptions about robot hazards and how to communicate them. They were required to be at least 18 years old and were grouped into sessions of ~8 participants, following standard focus group study guidelines [22]. They receive an honorarium of 35 CAD each.

2.1 Procedure

We conduct one 120-minute workshop per session group. Participants then proceed through design activities, both in small breakout groups (2-4 people) and with the whole session group. We collect paper artifacts (index cards, sketches, prototypes), audio recordings, and written notes by the moderators. We are in the process of analyzing these using open coding thematic analysis [14, 17, 19] to identify dominant themes.

Activity 1: The moderator gives a brief lecture on social robots, including live example robots (Pepper [2], NAO [37], aibo [44], and SnuggleBot [25]). In breakout groups, participants brainstorm what hazards social robots may pose on index cards (Figure 1). The moderator collects the ideas on a white board and gives an additional lecture on robot hazards, followed by a second brainstorming round.



This work is licensed under a Creative Commons Attribution 4.0 International License.
HAI '26, Osaka, Japan
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2575-3/2026/11
https://doi.org/10.1145/3841580.3845790



Figure 1: Our workshop space during a brainstorming activity (mock session, left), with participant-generated hazards (right) written on index cards (Activity 1) and sorted into thematic groupings (Activity 2).

Table 1: Initial thematic categories of risks proposed by participants in Activity 2, with examples of specific hazards raised in Activity 1. Closely overlapping sets with different names have been amalgamated (e.g., *emotional, social, psychological*, etc. into Social).

Thematic Category	Hazard Examples
Physical	direct impact, pinching in joints, hair caught in joints, harming pets
Security	recording user, robot may be hacked, used to hack other devices in home
Social	emotional dependence, manipulation of the user, grief on loss of robot
Environmental	wasting finite resources, unsustainable disposal, dangerous materials
Economic	replacing people's jobs, financial burden of ownership
Legal	liability, lack of regulation, robots in law enforcement

Activity 2: Participants sort their hazard ideas into possible categories. The moderator collects these ideas and closes with another group discussion.

Activity 3: The moderator proposes systematically communicating hazards, including a comparison to hazard labels, while emphasizing that labels are just one possibility. Using craft supplies, participants create sketches and paper prototypes of their own system for communicating hazards.

Activity 4: The moderator leads a semi-structured group discussion on trade-offs in hazard communication, focused on granularity and contextuality.

3 Preliminary Results

We have so far conducted four sessions with 32 participants, and present early results of Activities 1 and 2: participants' categories of hazards, with specific examples of each (Table 1). We look forward to presenting the full results in future.

Participants in all sessions have readily identified a wide range of hazards, highlighting broad pre-existing awareness of robot safety concerns. They frequently expressed risks in terms of other technologies, conflating robots with AI and comparing them to more familiar concerns such as "e-waste", monetization in modern entertainment (e.g., robot manufacturers offering decorations for robots for "microtransactions", as in video games), and the emotional attachment encouraged by Tamagotchi virtual pets [45].

Participants generally regarded robot safety as a real, imminent concern, not merely hypothetical, though in most sessions participants also raised concerns which they believed to be more rooted in science fiction, like robots "exploding" or "shooting lasers", which they typically raised in a humorous way with reference to fictional portrayals.

4 Discussion

The hazards participants identified in our study sessions (Table 1), prior to receiving examples or other prompting, suggest users will bring considerable awareness and prior expectations when interacting with a robot. This corpus spans the breadth of thematic safety concerns from prior HRI safety taxonomies [1, 3, 7].

Given that participants conflate robot hazards with those of other technologies, these results suggest communicating robot safety information may be more effective if designers adhere to users' expectations and frame safety risks in familiar terms, adopting vocabulary from other technologies. As we continue our analysis, we aim to construct a larger set of concrete design recommendations.

5 Conclusion

We are conducting design workshops with potential users of robots to discover their mental models of robot safety risks and how to communicate them. Initial results highlight the salience of the dangers of robots within public consciousness and underscore the need for transparent systems to inform users about risks. With the forthcoming results of this study, we aim to support designers in informing users about robot risks.

Acknowledgments

This project was funded by the Natural Sciences and Engineering Research Council of Canada through their Discovery Grants, as well as the University of Manitoba Graduate Fellowship.

References

[1] Neziha Akalin, Andrey Kiselev, Annica Kristoffersson, and Amy Loutfi. 2023. A Taxonomy of Factors Influencing Perceived Safety in Human–Robot Interaction. *Int J of Soc Robotics* 15, 12 (December 2023), 1993–2004. https://doi.org/10.1007/s12369-023-01027-8

[2] Aldebaran. Pepper the humanoid and programmable robot |Aldebaran. Retrieved September 30, 2023 from https://www.aldebaran.com/en/pepper

[3] Minja Axelsson, Raquel Oliveira, Mattia Racca, and Ville Kyrki. 2021. Social Robot Co-Design Canvases: A Participatory Design Framework. *J. Hum.-Robot Interact.* 11, 1 (October 2021), 3:1-3:39. https://doi.org/10.1145/3472225

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int J of Soc Robotics* 1, 1 (January 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[5] Steve Benford, Eike Schneiders, Juan Pablo Martinez Avila, Praminda Caleb-Solly, Patrick Robert Brundell, Simon Castle-Green, Feng Zhou, Rachael Garrett, Kristina Höök, Sarah Whatley, Kate Marsh, and Paul Tennent. 2025. Somatic Safety: An Embodied Approach Towards Safe Human-Robot Interaction. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2025. 429–438. https://doi.org/10.1109/HRI61500.2025.10973822

[6] James M. Berzuk, Lauren Corcoran, Brannen McKenzie-Lefurgey, Katie Szilagyi, and James E. Young. 2025. Knowledge Isn't Power: The Ethics of Social Robots and the Difficulty of Informed Consent. https://doi.org/10.48550/arXiv.2509.07942

[7] James M. Berzuk and James E. Young. 2025. Interact at Own Risk! Developing a Prototype of a Robot Hazardous Capability Labeling System. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction* (*HRI '25*), March 04, 2025. IEEE Press, Melbourne, Australia, 1240–1244.

[8] Skyla Y. Dudek and James E. Young. 2022. Fluid Sex Robots: Looking to the 2LGBTQIA+ Community to Shape the Future of Sex Robots. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2022. 746–749. https://doi.org/10.1109/HRI53351.2022.9889580

[9] Chris Elsden, Ella Tallyn, and Bettina Nissen. 2020. When Do Design Workshops Work (or Not)? In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*, July 06, 2020. ACM, Eindhoven Netherlands, 245–250. https://doi.org/10.1145/3393914.3395856

[10] William Gaver. 2012. What should we expect from research through design? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '12*), May 05, 2012. Association for Computing Machinery, New York, NY, USA, 937–946. https://doi.org/10.1145/2207676.2208538

[11] Denise Y. Geiskkovitch, Derek Cormier, Stela H. Seo, and James E. Young. 2016. Please continue, we need more data: an exploration of obedience to robots. *J. Hum.-Robot Interact.* 5, 1 (March 2016), 82–99.

[12] Sami Haddadin. 2015. Physical Safety in Robotics. In Formal Modeling and Verification of Cyber-Physical Systems: 1st International Summer School on Methods and Tools for the Design of Digital Systems, Bremen, Germany, September 2015, Rolf Drechsler and Ulrich Kühne (eds.). Springer Fachmedien, Wiesbaden, 249–271. https://doi.org/10.1007/978-3-658-09994-7_9

[13] Woodrow Hartzog. 2014. Unfair and Deceptive Robots. *Md. L. Rev.* 74, (2015 2014), 785.

[14] Jennifer Morey Hawkins. 2017. Thematic Analysis. In *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc, 1757–1760. Retrieved from https://doi.org/10.4135/9781483381411.n624

[15] Hooman Hedayati, Mark D. Gross, and Daniel Szafir. 2021. What Information Should a Robot Convey? In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 2021. 6232–6239. https://doi.org/10.1109/IROS51168.2021.9635999

[16] Jochen Heinzmann and Alexander Zelinsky. 2003. Quantitative Safety Guarantees for Physical Human-Robot Interaction. *The International Journal of Robotics Research* 22, 7–8 (July 2003), 479–504. https://doi.org/10.1177/02783649030227004

[17] Helene Joffe. 2011. Thematic Analysis. In *Qualitative Research Methods in Mental Health and Psychotherapy*. John Wiley & Sons, Ltd, 209–223. https://doi.org/10.1002/9781119973249.ch15

[18] Margot E. Kaminski, Matthew Rueben, William D. Smart, and Cindy M. Grimm. 2016. Averting Robot Eyes. *Md. L. Rev.* 76, (2017 2016), 983.

[19] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann.

[20] Jaclyn Molan, Laura Saad, Eileen Roesler, J. Malcolm McCurry, Nathaniel Gyory, and J. Gregory Trafton. 2025. The Perceived Danger (PD) Scale: Development and Validation. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2025. 420–428. https://doi.org/10.1109/HRI61500.2025.10973898

[21] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *J CONSUM RES* 26, 4 (March 2000), 323–339. https://doi.org/10.1086/209566

[22] David L. Morgan, Richard A. Krueger, and Jean A. King. 1998. *The Focus Group Kit*. Sage Publications, Inc, Thousand Oaks, CA, US.

[23] Ugo Pagallo. 2013. Robots in the cloud with privacy: A new threat to data protection? *Computer Law & Security Review* 29, 5 (October 2013), 501–508. https://doi.org/10.1016/j.clsr.2013.07.012

[24] Danika Passler Bates, Skyla Y. Dudek, James M. Berzuk, Adriana Lorena González, and James E. Young. 2024. SnuggleBot the Companion: Exploring In-Home Robot Interaction Strategies to Support Coping With Loneliness. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (*DIS '24*), July 01, 2024. Association for Computing Machinery, New York, NY, USA, 2972–2986. https://doi.org/10.1145/3643834.3660702

[25] Danika Passler Bates and James E Young. 2020. SnuggleBot: A Novel Cuddly Companion Robot Design. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, November 10, 2020. ACM, New York, NY, USA, 260–262. https://doi.org/10.1145/3406499.3418772

[26] Aslam Pervez and Jeha Ryu. 2008. Safe physical human robot interaction-past, present and future. *J Mech Sci Technol* 22, 3 (March 2008), 469–483. https://doi.org/10.1007/s12206-007-1109-3

[27] Brian S. Peters, Priscila R. Armijo, Crystal Krause, Songita A. Choudhury, and Dmitry Oleynikov. 2018. Review of emerging surgical robotic technology. *Surg Endosc* 32, 4 (April 2018), 1636–1655. https://doi.org/10.1007/s00464-018-6079-2

[28] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2016. 101–108. https://doi.org/10.1109/HRI.2016.7451740

[29] Julia Rosén and Denise Y. Geiskkovitch. 2025. Reimagining Informed Consent in Human-Robot Interaction: Introducing the RoboConsent Framework. In *2025 34th IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, August 2025. IEEE, Eindhoven, Netherlands.

[30] Julia Rosén, Jessica Lindblom, and Erik Billing. 2022. The Social Robot Expectation Gap Evaluation Framework. In *Human-Computer Interaction. Technological Innovation* (*Lecture Notes in Computer Science*), 2022. Springer International Publishing, Cham, 590–610. https://doi.org/10.1007/978-3-031-05409-9_43

[31] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *ACM/IEEE International Conference on Human-Robot Interaction* (*HRI '15*), 2015. Association for Computing Machinery, New York, NY, USA, 141–148. https://doi.org/10.1145/2696454.2696497

[32] Elaheh Sanoubari, Stela H. Seo, Diljot Garcha, James E. Young, and Veronica Loureiro-Rodriguez. 2019. Good Robot Design or Machiavellian? An In-the-Wild Robot Leveraging Minimal Knowledge of Passersby's Culture. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 22, 2019. IEEE, 382–391. https://doi.org/10.1109/HRI.2019.8673326

[33] Vasanth Sarathy, Thomas Arnold, and Matthias Scheutz. 2019. When Exceptions Are the Norm: Exploring the Role of Consent in HRI. *J. Hum.-Robot Interact.* 8, 3 (July 2019), 14:1-14:21. https://doi.org/10.1145/3341166

[34] Matthias Scheutz and Thomas Arnold. 2016. Are we ready for sex robots? In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2016. 351–358. https://doi.org/10.1109/HRI.2016.7451772

[35] Sarah Schömbs, Saumya Pareek, Jorge Goncalves, and Wafa Johal. 2024. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (*CHI '24*), May 11, 2024. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3613904.3642911

[36] Lena T. Schramm, Derek Dufault, and James E. Young. 2020. Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design. *ACM/IEEE International Conference on Human-Robot Interaction* (2020), 439–441. https://doi.org/10.1145/3371382.3378280

[37] SoftBank Robotics America Inc. NAO: Personal Robot Teaching Assistant | SoftBank Robotics America. Retrieved December 11, 2024 from https://us.softbankrobotics.com/nao

[38] Pieter Jan Stappers and Elisa Giaccardi. 2014. 41. Research through Design. In *The Encyclopedia of Human-Computer Interaction* (2nd edition). Interaction Design Foundation - IxDF. Retrieved September 26, 2025 from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design

[39] Eric J. Vanman and Arvid Kappas. 2019. "Danger, Will Robinson!" The challenges of social robots for intergroup relations. *Social and Personality Psychology Compass* 13, 8 (2019), e12489. https://doi.org/10.1111/spc3.12489

[40] Milos Vasic and Aude Billard. 2013. Safety issues in human-robot interactions. In *2013 IEEE International Conference on Robotics and Automation*, May 2013. 197–204. https://doi.org/10.1109/ICRA.2013.6630576

[41] Jochen Wirtz and Ruth Stock-Homburg. 2025. Generative AI Meets Service Robots. *Journal of Service Research* (May 2025), 10946705251340487. https://doi.org/10.1177/10946705251340487

[42] James Young. 2021. Danger! This robot may be trying to manipulate you. *Science Robotics* 6, 58 (September 2021). https://doi.org/10.1126/scirobotics.abk3479

[43] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '07*), April 29, 2007. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704

[44] aibo. Retrieved August 16, 2022 from https://us.aibo.com/

[45] Official Tamagotchi Site. *Official Tamagotchi Site*. Retrieved September 16, 2025 from https://tamagotchi-official.com/gb/