

Trust in Queer Human-Robot Interaction

Raj Korpan¹

Abstract—Human-robot interaction (HRI) systems need to build trust with people of diverse identities. This position paper argues that queer (LGBTQIA+) people must be included in the design and evaluation of HRI systems to ensure their trust in and acceptance of robots. Queer people have faced discrimination and harm from artificial intelligence and robotic systems. Despite calls for increased diversity and inclusion, HRI has not systemically addressed queer issues. This paper suggests three approaches to address trust in queer HRI: diversifying human-subject pools, centering queer people in HRI studies, and contextualizing measures of trust.

I. INTRODUCTION

Acceptance of robots in the wild requires systems that build trust with people with diverse cultural and racial backgrounds, ages, sexes, and abilities. This position paper argues that queer people are an important and distinct population that must be explicitly included in the design of human-robot interaction (HRI) systems to address their particular barriers to trusting and accepting robots. Furthermore, measures of trust may not accurately capture the factors affecting trust for such marginalized minority populations even if they have been validated for the general population. Novel methods developed to assess trust should address the general lack of trust in institutions and technology among queer communities.

II. BACKGROUND

Recent calls for increased diversity, equity, and inclusion (DEI) in artificial intelligence (AI) and robotics have described how discrimination and bias exist at multiple levels: in the implementations of systems, the research community, and the targeted users [1]. Sex and gender diversity are at most risk in these systems, particularly in robotics [2]. To address these issues, workshops have promoted DEI in HRI [3]. Others called for the inclusion of feminist principles [4] and intersectional gender inclusivity [5] in HRI practices.

There have also been specific calls to address queer issues in AI [6] and robotics [7]. Queer in AI is an organization that has demonstrated the use of participatory design to both bring visibility to the harms that AI systems have on queer people and also to empower queer inclusion in the AI research community [8]. Recent work showed that advances in algorithmic fairness have not addressed queer concerns [9] and found that popular AI systems do not center the experiences of transgender and non-binary individuals [10].

Not only are queer issues ignored in the creation of HRI systems, their impacts on queer people are also not examined. One study found that no HRI studies centered sex and

gender issues across 415 papers from 2002 to 2023, and only two addressed sexuality but in the context of sex robots [11]. Similarly, queer people are not included as subjects in HRI studies: of 55,032 participants from 787 studies, only two trans people were reported and sexuality was only reported in one study [12]. Furthermore, (presumably cis-gender) men are over-represented as research participants [13]. Consequently, the impact of HRI systems on queer people are often completely unmeasured or, even worse, they may cause further discrimination and harm.

One recent work addressed the use of robots for elder care with queer people [14], [15]. In that work, queer participants expressed a desire for robotic systems that prioritize privacy and their physical and digital safety, and also understand, appreciate and respect the unique differences of the queer community. That research also found that in the care-giving context, trust is especially important between the caregiver and care recipient and that similar factors affect trust with a human caregiver as with a robot caregiver [16].

A robot’s gender presentation can also impact trust in HRI. One study showed that participants attributed more cognitive and affective trust to a robot with a female body shape [17]. Notably, however, that study did not ask participants about their gender identity or sexuality and so it may not generalize to the queer community. Moreover, the body shapes presented further stereotypes and exclude the diversity of queer bodies.

Measurement of trust is a major challenge in human-robot interaction because it cannot be directly observed [18]. Most commonly, self-report scales are used as a measurement tool for trust, such as the Muir trust questionnaire [19], the Trust in Automation Scale [20], the Human-Computer Trust Instrument [21], the Merritt scale [22], the Human-Robot Interaction Trust Scale [23], the Human-Robot Trust Scale [24], and the Multi-Dimensional Measure of Trust (MDMT) [25]. The MDMT, for example, measures trust along five dimensions: reliability, competency, ethicality, transparency and benevolence [26]. These tools have not all been assessed for reliability or validated [27], [28], especially on the same set of participants to determine the level of agreement across measures [29]. Although others have attempted to use physiological measures to infer trust (e.g., [30], [31]), it is unclear how reliable these approaches are as a measure of trust. Furthermore, none of these have been evaluated for their effectiveness and applicability to the queer community.

III. TRUST IN QUEER HRI

Despite calls for increased DEI in HRI, the literature shows that there has not been a sustained effort to address

¹Raj Korpan was with Iona University. He is currently with the Department of Computer Science, Hunter College, New York, NY 10065, USA. rk1665@hunter.cuny.edu

how systems are designed and evaluated with respect to their impact on queer people. The result is a limited understanding of how queer people interact with robots, what factors guide their acceptance and trust in these interactions, and how measures of trust could adapt to a queer audience. I offer three suggestions for how to address these issues: diversifying human-subject pools, centering queer people in HRI studies, and contextualizing measures of trust.

A. DIVERSIFY HUMAN SUBJECT POOLS

A major gap in the human-robot trust literature is the lack of human subjects with queer identities [12]. Visibility is the first step towards inclusion. As a starting point, I believe researchers should ask participants about their gender identity and sexuality. Instead of a lack of clarity about human subjects in HRI studies, inclusive practices like participatory design allow for better evaluation of trust with people [8]. The challenge with the collection of participants' queer characteristics is the potential for privacy violations or the misuse of this information for discriminatory or harmful purposes [32]. Despite these concerns, it is possible to build trust with vulnerable populations through intentionally designed studies that prioritize their issues [33].

Understanding the demographics of subjects is not enough to achieve improved DEI, a concerted effort must also be made to recruit queer individuals into HRI studies [13]. Instead of confronting the challenge of recruiting a representative sample, many HRI studies use convenience samples of university students [34], but the difficulty of recruitment should not be an excuse to exclude vulnerable and marginalized peoples. Without a diverse subject pool, we cannot begin to examine the impact on trust in queer HRI.

B. CENTER QUEER PEOPLE

Queer people trust technology designed for them more than systems created for the majority non-queer population [35]–[37]. Rarely have HRI systems focused on queer needs: only Poulsen et al. pursued such work [14]–[16]. Queer individuals face systematic discrimination and challenges in their lives distinct from non-queer people (e.g., in healthcare [38] and the workplace [39]). Robots could have a unique opportunity to support this marginalized community to address their issues, such as mental health, bullying, loneliness, lack of access to traditional systems of support [40]. To achieve these goals, however, HRI researchers must build systems designed for and evaluated with queer people, as proposed by Queer in AI [8]. Furthermore, the factors that affect trust will likely be different because of the discrimination queer people face (e.g., organizations face distinct challenges when engaging with queer communities [41]). Because AI and robotics systems have already been shown to be harmful and discriminatory toward queer people [7], I believe that it is important for HRI to center queer people and their needs while not furthering harmful stereotypes.

C. CONTEXTUALIZE MEASURES OF TRUST

Although measures of trust are treated as a standardized way to evaluate HRI systems across many settings and tasks,

regardless of the target population [18], the way queer people experience trust is different from non-queer people (e.g., [41]). While the standard approach provides consistency and reliability, it may ignore the contextualized, nuanced, and intersectional experience of queer individuals [42]. Instead, inclusive measures of trust need to address the aspects of trust specifically relevant to the queer community, such as differences in gender identities and orientations and use of preferred pronouns and inclusive language [43]. Simultaneously, these tools need to respect queer individuals' emotional and psychological well-being and privacy concerns.

The challenge with this approach is how to do this in a rigorous, verifiable, and reusable way because most HRI studies use a single trust measurement tool instead of separate demographic-specific measures. One approach could combine quantitative measures with qualitative data to provide clarity to the trust measure. Another approach is to use an iterative user-centered design process that involves queer people to produce a tool that is culturally competent [44]. A third approach could be a customizable questionnaire that includes context-based questions that address different factors that affect trust, such as the robot's appearance, performance, and behavior, the task, and the participant's identity. Although the MDMT is modular in that it provides five dimensions that can be selected based on the scenario [26], it does not adapt to other contextual factors, such as a participant's queer identity. A customizable approach could combine some common trust questions with context-specific questions. Alternatively, it could weigh the importance of trust factors differently. For example, queer people face unique concerns around privacy [32] and so a trust measure could weigh the importance of that higher than other factors.

IV. CONCLUSIONS

As robots are deployed in the real world, they need to interact with all kinds of humans. The challenge for these robots is to solve difficult tasks in unstructured dynamic environments while they are accepted and trusted by their human partners. This position paper focuses on trust in *queer human-robot* interaction. Despite calls for more inclusive approaches, queer people have not been broadly considered by HRI researchers in the design or evaluation of systems. Measures of human-robot trust also fail to consider the unique factors that could affect trust with queer individuals.

I offer three suggestions toward a more queer HRI. First, a person's queer identity is likely a confounding factor in the measure of trust, so studies should both ask about queer identity and recruit more queer participants. Second, HRI systems should be developed for the needs of the queer community. Third, a person's queer identity likely affects their perception of trust, so trust measures should consider that context. These three suggestions would expand who these systems are for, understand what affect they have on queer communities, and examine how trust could be reliably measured for them. Queer people also deserve the potential benefits of robots but to achieve that we must first understand how to build trust in queer human-robot interaction.

ACKNOWLEDGMENT

The author thanks the reviewers whose valuable comments and suggestions helped improve this paper.

REFERENCES

- [1] E. Fosch-Villaronga and A. Poulsen, "Diversity and inclusion in artificial intelligence," *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, pp. 109–134, 2022.
- [2] E. Fosch-Villaronga and H. Drukarch, "Accounting for diversity in robot design, testbeds, and safety standardization," *International Journal of Social Robotics*, pp. 1–19, 2023.
- [3] A. Tanevska, S. Chandra, G. Barbareschi, A. Eguchi, Z. Han, R. Korpan, A. K. Ostrowski, G. Perugia, S. Ravindranath, K. Seaborn, *et al.*, "Inclusive hri ii: Equity and diversity in design, application, methods, and community," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 956–958.
- [4] K. Winkle, D. McMillan, M. Arnelid, K. Harrison, M. Balaam, E. Johnson, and I. Leite, "Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical hri," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 72–82.
- [5] I. Hipólito, K. Winkle, and M. Lie, "Enactive artificial intelligence: Subverting gender norms in robot-human interaction," *Frontiers in Neurobotics*, vol. 17, p. 77, 2023.
- [6] E. Liliequist, A. A. Tubella, K. Danielsson, and C. Cocq, "Beyond the binary—queering ai for an inclusive future," *interactions*, vol. 30, no. 3, pp. 31–33, 2023.
- [7] A. Poulsen, E. Fosch-Villaronga, and R. A. Søråa, "Queering machines," *Nature Machine Intelligence*, vol. 2, no. 3, pp. 152–152, 2020.
- [8] A. OrganizersOfQueerin, A. Ovalle, A. Subramonian, A. Singh, C. Voelcker, D. J. Sutherland, D. Locatelli, E. Breznik, F. Klubicka, H. Yuan, J. Hetvi, H. Zhang, J. Shriram, K. Lehman, L. Soldaini, M. Sap, M. P. Deisenroth, M. L. Pacheco, M. Ryskina, M. Mundt, M. Agarwal, N. McLean, P. Xu, A. Pranav, R. Korpan, R. Ray, S. Mathew, S. Arora, S. John, T. Anand, V. Agrawal, W. Agnew, Y. Long, Z. J. Wang, Z. Talat, A. Ghosh, N. Dennler, M. Noseworthy, S. Jha, E. Baylor, A. Joshi, N. Y. Bilenko, A. McNamara, R. Gontijo-Lopes, A. Markham, E. Dong, J. Kay, M. Saraswat, N. Vytla, and L. Stark, "Queer in ai: A case study in community-led participatory ai," *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023.
- [9] N. Tomasev, K. R. McKee, J. Kay, and S. Mohamed, "Fairness for unobserved characteristics: Insights from technological impacts on queer communities," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021, pp. 254–265.
- [10] A. Ovalle, P. Goyal, J. Dhamala, Z. Jagers, K.-W. Chang, A. Galstyan, R. Zemel, and R. Gupta, "'i'm fully who i am': Towards centering transgender and non-binary voices to measure biases in open language generation," *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023.
- [11] K. Seaborn, "Diversity not discussed: Centring overlooked factors of inclusion within human-robot interaction," *Inclusive HRI II: Workshop on Equity and Diversity in Design, Application, Methods, and Community (DEI) at HRI '23*, 2023.
- [12] K. Seaborn, G. Barbareschi, and S. Chandra, "Not only weird but 'uncanny'? a systematic review of diversity in human-robot interaction research," *International Journal of Social Robotics*, pp. 1–30, 2023.
- [13] K. Winkle, E. Lagerstedt, I. Torre, and A. Offenwanger, "15 years of (who) man robot interaction: Reviewing the h in human-robot interaction," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 3, pp. 1–28, 2023.
- [14] A. Poulsen, E. Fosch-Villaronga, O. K. Burmeister, *et al.*, "Cybersecurity, value sensing robots for lgbtiq+ elderly, and the need for revised codes of conduct," *Australasian Journal of Information Systems*, vol. 24, 2020.
- [15] A. Poulsen, I. Skaines, S. McLaren, and O. K. Burmeister, "Value sensing robots: The older lgbtiq+ community," in *Societal Challenges in the Smart Society*. Universidad de La Rioja, 2020, pp. 275–288.
- [16] A. Poulsen, "The investigation of a new care robot design approach for alleviating lgbt+ elderly loneliness," Ph.D. dissertation, Charles Sturt University, 2022.
- [17] J. Bernotat, F. Eyssel, and J. Sachse, "The (fe) male robot: how robot body shape impacts first impressions and trust towards robots," *International Journal of Social Robotics*, vol. 13, pp. 477–489, 2021.
- [18] B. C. Kok and H. Soh, "Trust in robots: Challenges and opportunities," *Current Robotics Reports*, vol. 1, pp. 297–309, 2020.
- [19] B. M. Muir, "Operators' trust in and use of automatic controllers in a supervisory process control task," Ph.D. dissertation, University of Toronto, 1989.
- [20] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International journal of cognitive ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.
- [21] M. Madsen and S. Gregor, "Measuring human-computer trust," in *11th australasian conference on information systems*, vol. 53. Citeseer, 2000, pp. 6–8.
- [22] S. M. Merritt, "Affective processes in human-automation interactions," *Human Factors*, vol. 53, no. 4, pp. 356–370, 2011.
- [23] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human-robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, pp. 235–248, 2012.
- [24] K. Schaefer, "The perception and measurement of human-robot trust," Ph.D. dissertation, University of Central Florida, 2013.
- [25] B. F. Malle and D. Ullman, "A multidimensional conception and measure of human-robot trust," in *Trust in human-robot interaction*. Elsevier, 2021, pp. 3–25.
- [26] D. Ullman and B. F. Malle, "Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 618–619.
- [27] Z. R. Khavas, "A review on trust in human-robot interaction," *arXiv preprint arXiv:2105.10045*, 2021.
- [28] T. Ueno, Y. Sawa, Y. Kim, J. Urakami, H. Oura, and K. Seaborn, "Trust in human-ai interaction: Scoping out models, measures, and methods," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–7.
- [29] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, "Can you trust your trust measure?" in *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, 2021, pp. 92–100.
- [30] Q. Jenkins and X. Jiang, "Measuring trust and application of eye tracking in human robotic interaction," in *IIE Annual Conference. Proceedings*. Institute of Industrial and Systems Engineers (IISE), 2010, p. 1.
- [31] K. Akash, W.-L. Hu, N. Jain, and T. Reid, "A classification model for sensing human trust in machines using eeg and gsr," *ACM Transactions on Interactive Intelligent Systems (TiIS)*, vol. 8, no. 4, pp. 1–20, 2018.
- [32] T. Williams, C. Matuszek, K. Jokinen, R. Korpan, J. Pustejovsky, and B. Scassellati, "Voice in the machine: Ethical considerations for language-capable robots," *Commun. ACM*, vol. 66, no. 8, p. 20–23, jul 2023. [Online]. Available: <https://doi.org/10.1145/3604632>
- [33] H. R. Lee, E. Cheon, C. Lim, and K. Fischer, "Configuring humans: What roles humans play in hri research," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 478–492.
- [34] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme, "From characterising three years of hri to methodology and reporting recommendations," in *2016 11th acm/ieee international conference on human-robot interaction (hri)*. IEEE, 2016, pp. 391–398.
- [35] Z. Beare and M. Stone, "By queer people, for queer people: Folx, plume, and the promise of queer ux," in *Proceedings of the 39th ACM International Conference on Design of Communication*, 2021, pp. 20–25.
- [36] O. L. Haimson, D. Gorrell, D. L. Starks, and Z. Weinger, "Designing trans technology: Defining challenges and envisioning community-centered solutions," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.
- [37] O. L. Haimson, K. Nham, H. Thach, and A. DeGuia, "How transgender people and communities were involved in trans technology design processes," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.
- [38] C. H. B. Ayhan, H. Bilgin, O. T. Uluman, O. Sukut, S. Yilmaz, and S. Buzlu, "A systematic review of the discrimination against sexual and gender minority in health care settings," *International Journal of Health Services*, vol. 50, no. 1, pp. 44–61, 2020.
- [39] L.-C. Mara, M. Ginielis, and I. Brunet-Icart, "Strategies for coping with lgbt discrimination at work: A systematic literature review," *Sexuality Research and Social Policy*, vol. 18, pp. 339–354, 2021.
- [40] F. Mongelli, D. Perrone, J. Balducci, A. Sacchetti, S. Ferrari, G. Mattei, and G. M. Galeazzi, "Minority stress and mental health among

- lgbt populations: An update on the evidence." *Minerva Psichiatrica*, vol. 60, no. 1, pp. 27–50, 2019.
- [41] E. Ciszek, "'we are people, not transactions': Trust as a precursor to dialogue with lgbtq publics," *Public Relations Review*, vol. 46, no. 1, p. 101759, 2020.
- [42] B. Ruberg and S. Ruelos, "Data for queer lives: How lgbtq gender and sexuality identities challenge norms of demographics," *Big Data & Society*, vol. 7, no. 1, p. 2053951720933286, 2020.
- [43] S. J. Meyer and N. M. Elias, "Rainbow research: Challenges and recommendations for sexual orientation and gender identity and expression (sogie) survey design," *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, pp. 1–7, 2022.
- [44] M. Fine and M. E. Torre, "Critical participatory action research: A feminist project for validity and solidarity," *Psychology of Women Quarterly*, vol. 43, no. 4, pp. 433–444, 2019.