

Algorithmic Transparency for Accessible Robotics

Mavis Murdock
Department of Computer Science
Tufts University
Medford MA, USA
mavis.murdock@tufts.edu

Abstract

Despite technological advances, the rejection rates of assistive technology remain quite high [1]. Our work focuses on assistive robots, aiming to bridge the gap between their technical potential and real-world adoption by end-users. We seek to improve how robots communicate their behavior, enabling users to better anticipate how a robot will act in various environments. Understanding a robot's decision-making process can help users determine when and how to use the robot effectively and confidently. Our research centers on two key threads: disability and explainability. By making robot behavior more transparent, we aim to empower users to engage with the technology in ways that most benefit them.

We define "explainability," as a robot's ability to communicate how it will behave in an environment according to its underlying policy. The human-understandability of these policies may vary; a task learned using a neural network, for instance, tends to be less human-understandable than one that uses symbolic methods. We aim to distill robot policies into a format which is more "transparent". That is, one which helps the user predict the robot's general behavior without extensive technical expertise or prior knowledge about how the policy was trained.

There has been a notable lack of research into robot explainability tailored specifically to people with disabilities. Many user studies rely on convenience populations, but these findings are not necessarily transferable. Disabled and non-disabled users may not use a piece of assistive technology with the same frequency or for the same tasks. Even the user's role in the interaction, whether active or passive, may differ. For instance, one user may use a robot as an assistant for helping them clean dishes whereas another orders the robot to complete the chore in their place while they engage in another task.

Robot explainability can take many forms, including natural language descriptions, video snippets, or images. The challenge of

choosing examples which are optimally representative of a robot's overall behavior remains a difficult and open problem [2]. Our current project develops a novel approach, using visual examples from the robot's task execution which are presented in the form of an analogy. This takes advantage of humans' innate use of analogical reasoning, avoiding the need for the user to have advanced technical knowledge and avoiding the assumption that robot behavior can map neatly into natural language explanations.

We hypothesize that when users find robot behavior more predictable, they will feel more empowered and willing to use the technology in a wider range of situations. Ultimately, our goal is to support individuals with disabilities in fully utilizing assistive robots by communicating information about their behavior and creating more informed interactions.

CSS Concepts

• Human-centered computing; • Human-centered computing → Accessibility; • Computer systems organization → Embedded and cyber-physical systems → Robotics

Keywords

AI Explainability, Human-Robot Interaction, Accessibility

ACM Reference Format

Mavis Murdock. 2025. Algorithmic Transparency for Accessible Robotics. In *Proceedings of the 2025 Conference for Research on Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT 2025)*, July 14–16, 2025, Newark, NJ, USA. ACM, NY, NY, USA. 2 pages. <https://doi.org/10.1145/3704637.3734751>

1 Questions and Connections

At this stage of my doctoral research, I am particularly interested in developing stronger research collaborations with end-users (in this case, individuals from disabled communities). I would like to connect with researchers who have experience in building trust and co-creating research with communities outside of the university context. I am curious about strategies others have used to initiate and sustain these relationships, especially in ways that ensure community members feel ownership, agency, and long-term value from their involvement in the research process. What approaches have been effective in making these collaborations more equitable and reciprocal?

Permission to make digital or hard copies of part or all 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).

RESPECT 2025, July 14–16, 2025, Newark, NJ, USA
© 2025 Copyright is held by the owner/author(s).
ACM ISBN 979-8-4007-1355-2/2025/07. <https://doi.org/10.1145/3704637.3734751>

In my current research, I am exploring questions related to algorithmic transparency, particularly in the context of human-robot interaction. I'm eager to learn how others conceptualize what it means for a system—such as a robot—to be “understandable.” What kinds of explanations do people want when a robot behaves in unexpected or undesirable ways? How might we design systems that allow users to request or interpret explanations in ways that align with their expectations, needs, or levels of technical expertise? I'm especially interested in hearing diverse perspectives on what transparency should look like in real-world interactions.

Ultimately, I hope to connect with DC participants and mentors whose work touches on human-centered design in AI and who can help me challenge and refine the questions I ask, the methods I use, and the communities I work with as I move forward.

2 Additional Projects

In addition to my project focusing on analogical robot explanations, I am working on two additional projects. The first examines the effects of a robot's physical embodiment and autonomy level during a collaborative household alongside a

human. I am conducting the study with both disabled and non-disabled participant pools to assess the differences between how users prefer to interact with the robot and the tasks they might choose to use it for. My other project is a literature review examining the role of AI in assistive technology from 2014-2024. This has included a roughly-equal split between both human-centered recommendations for AI tools and new AI tools/systems.

Acknowledgements

This work was supported in part by funding from the National Science Foundation (#2137312)

References

- [1] Stefan Salminger, Heiko Stino, Lukas H. Pichler, Clemens Gstoettner, Agnes Sturma, Johannes A. Mayer, Michael Szivak, and Oskar C. Aszmann. 2020. Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance? *Disability and Rehabilitation* 44, 14 (Dec 2020), 1–12. <https://doi.org/10.1080/09638288.2020.1866684>
- [2] Serena Booth, Sanjana Sharma, Sarah Chung, Julie Shah, and Elena L Glassman. 2022. Revisiting human-robot teaching and learning through the lens of human concept learning. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (Sapporo, Hokkaido, Japan) (HRI '22)*. IEEE Press, 147–156.