Designing, Developing, & Deploying Social Robots for Therapy

Rebecca Ramnauth · rramnauth2220.github.io/

The reason we build robots is to help people. Where the field of robotics has traditionally focused on the reliability and precision of motion to achieve functional task assistance, my research explores how robots can provide *social* value and assistance to people through therapy. Unlike other forms of technology, robots possess a physical body: they can turn to face you, wait for your reply, and hold space in ways that organically elicit social engagement. This allows them to support on-demand, personalized interactions that are both socially situated and physically co-present. However, designing robots capable of delivering effective therapy requires progress across three key frontiers: **design** (How should the system look and interact with users?), **development** (How can we technically realize these design choices for robust, real-world use?), and **deployment** (How effectively does the system work in real-world contexts with real users with real social needs?).

My research engages each frontier. I explore the *design* of socially assistive robots—systems capable of operating autonomously in dynamic, unstructured spaces and engaging meaningfully with users of diverse cognitive profiles and social needs. I then apply those design choices to *develop* robot-based therapies targeting key social skills (e.g., joint attention, conversational reciprocity, resilience to interruptions, deep breathing, and emotional de-escalation). In doing so, I tackle several technical challenges such as enabling robots to disambiguate social presence [1], discern when to appropriately engage or offer assistance [6], and safeguard generated actions when relying on foundation models [5]. Finally, I *deploy* these systems for extended periods—from a week to several months—with diverse populations, including those with autism [4, 7, 9], with co-occurring neurodevelopmental conditions [7, 8], or in specialized programs [8]. These deployments occur under real-world conditions such as in homes during a global pandemic [4, 7, 10] or in a dual-setting school (general and special education) [8].

Figure 1: Studies in which robots operated fully autonomously with vulnerable users in personal, real-world spaces over extended periods. From left to right: a robot system to combat the effects of social isolation during the COVID-19 lockdown among young children [10]; a month-long in-home study to support gaze and social skill development in children with autism [9]; a week-long in-home intervention for adults with autism to practice small talk skills [7]; and a month-long deployment in a K–5 public school to support behavioral de-escalation [8].

Research Themes & Why This Work is Challenging

Throughout the evolution of the field (as reviewed in [2] and [3]), human-robot interaction research has been conducted largely in controlled laboratory or clinical settings. These studies tend to be short: a single session, often less than an hour. They typically involve one human (usually a neurotypical adult, or someone considered "high-functioning") interacting with a robot in a supervised, constrained, and isolated manner. Because there is significant engineering effort that goes into building robots for interaction, the goals of these studies have largely been feasibility or proof-of-concept—can the robot work, at least under these ideal experimental conditions? Although such controlled studies allow researchers to isolate and test specific interaction parameters, these approaches fail to capture the complexity and sustainability of long-term use in the real world. For instance, in extended interactions, users are habituated to novelty, expectations evolve, and the utility of a system is increasingly judged by its ability to provide meaningful, contextually appropriate support.

Therefore, we need to build robots that demonstrate **technical robustness**—operating reliably amid environmental variability and uncertainty—and **interactional competence**, interpreting and responding appropriately to human social signals. They must function *autonomously* in *dynamic, real-world contexts* and maintain *therapeutic relevance* *over time*. This is because truly effective therapy does not unfold within a short, single session; instead, it develops over days to months, through exposure to novel, real-world situations that test the ongoing relevance and adaptability of learned strategies. It leads to progress beyond designated "therapy time," sustains engagement without researcher involvement or oversight, generalizes beyond robot interaction to improve human-human relationships, and accommodates diverse cognitive profiles and needs.

However, each of these points highlights both a research gap and the central challenge of this work. Moving beyond the laboratory to the inherently messy, dynamic, and unpredictable real world means confronting questions that seem deceptively simple. For example, even with state-of-the-art perception systems, questions like "*Is my user looking at me?*" [9] or "*Is it my user or the TV speaking?*" [1] remain difficult to answer reliably, requiring advances in multimodal perception, representation learning, and robustness to noise. Beyond accurate detection, long-term autonomy demands that robots reason about the social context to act appropriately—deciding, for example, "*is it socially appropriate to engage my user right now?*" [1, 6]. This entails probabilistic reasoning, planning under uncertainty, and contextual inference, all classic challenges in artificial intelligence and autonomy. Ultimately, once embedded in real-world spaces, robots encounter people and situations that cannot be exhaustively anticipated or preprogrammed. Still, they must improvise appropriate responses in real time, fully independently and without researcher involvement [7, 8]. This requires generalization beyond training data, fault–tolerant design, sustained hardware and software reliability, and more.

My research addresses these computational and non-computational challenges to enable robots to operate fully autonomously and deliver relevant therapy to diverse, vulnerable users in their personal spaces over extended periods of time. As a result, **my work introduces several firsts for the field.** It features the first robotic interventions designed specifically for adults with autism [4, 7]; one of the only robotic studies to demonstrate continuous learning progression tied to clinical measures of therapeutic efficacy [9]; and the first use of foundation models to deliver unscripted, improvised therapy [7, 8]. It also presents the first robots capable of providing personalized feedback on users' social skills [7], and the first robot-led intervention to support behavioral de-escalation in public settings while remaining agnostic to users' age or diagnostic profile [8].

Among the selection of five robot deployments summarized in my Ph.D. thesis [4, 7–10], my work has directly engaged **284 families**, with participants ranging in age **from 4 to 104 years**. Collectively, these studies produced **12,984 hours** of continuous, fully autonomous robot operation outside of the lab, in real-world environments. My robots delivered more than **3,000 sessions** total with users, resulting in **950 hours** of active therapy time, and yielding approximately **84 terabytes** of multimodal vision and audio data. All user studies submitted to ACM, IEEE, or joint venues have been recognized as finalists or recipients of Best Paper Awards.

Next 5 Years: Research Agenda and Funding Fit

I recognize that the funding climate in the United States has shifted dramatically in the recent year. I am grateful that the interdisciplinary nature of my research (in bridging psychology, engineering, education, and cognitive science) continues to attract support from diverse and competitive funding sources.

This work has been supported by the National Science Foundation (NSF), including Expedition in Computing and CISE/IIS programs, with continued alignment to their Mind, Machine, and Motor Nexus (M3X) and Future CoRE Human-Centered Computing (HCC) initiatives. Additional support has come from the National Institutes of Health (NIH) and the U.S. Office of Naval Research (ONR). My Ph.D. was awarded both the NSF Graduate Research Fellowship and the Ford Foundation Fellowship from the National Academies of Sciences, Engineering, and Medicine (NASEM). Furthermore, my research has fostered collaborations with the Organisation for Economic Co-operation and Development (OECD), the U.S. Department of Education, Yale School of Medicine, Seattle Children's Research Institute, and several local partners such as therapy centers and non-profit organizations. These collaborations create pathways for new partnerships and private support from foundations such as Autism Science Foundation, Simons Foundation, and other mission-aligned organizations.

Building on this foundation, I outline three directions for future research and then conclude with my lab vision.

Robots for Novel Users, Spaces, and Skills. My research began with exploring robots for autism therapy. Because autism is defined by differences in social functioning, it can be a valuable lens through which the underlying mechanisms of human social cognition become more observable and testable. For example, when a robot designed to elicit joint attention succeeds with neurotypical children but fails with those with autism, it suggests a need to revisit our understanding of how joint attention is triggered or maintained. Yet, as a spectrum, effective care for autism is rarely one-size-fits-all; it demands highly individualized approaches that are difficult to implement at scale and require significant time, expertise, and involvement from families, clinicians, and other stakeholders. Given this dual complexity, autism represented a testbed for developing robots that are socially adaptive and broadly applicable across diverse user needs. Still, the vast majority of autism research and clinical programs continue to focus almost exclusively on children. Among the 304 studies presenting a robot for autism [3], only two describe robots deployed in the homes of *adults* with autism—both of which are my own work [4, 7].

I can identify additional user groups and contexts that fit this pattern: those that remain understudied and under-supported but hold strong potential to uncover new facets of human cognition and to drive broader rethinking of systemic and methodological paradigms. For instance, environments such as foster housing, youth detention centers, and homeless shelters remain absent from the current literature. Yet, these spaces represent critical contexts where robots could deliver meaningful, high-impact support. Similar research gaps exist in healthcare and mental health domains: How can we build robots that effectively ease anxiety during chemotherapy, aid veterans coping with post-traumatic stress, or assist teens with eating disorders? What technical and social considerations are required to build systems that meaningfully support users living with clinical depression? Each example inspires novel lines of inquiries across all three frontiers of design, development, and deployment.

Robust Guardrails and Introspective Mechanisms. With the advent of foundation models and rapid trend toward greater autonomy, new risks emerge as robots gain the ability to generate unscripted, context-sensitive behaviors in open-ended environments. My research introduced domains that resist overly rigid or pre-scripted control, such as robots that can seamlessly maintain small talk, determine when emotional de-escalation is needed, or decide when best to switch activities. While these systems demonstrate the value of generative and social flexibility, they also highlight the need for robust guardrails to ensure interactions remain safe.

Future work will focus on developing technical safeguards that enable the ethical deployment of autonomous robots in socially sensitive environments. Building on my Grounded Observer framework [5], which enabled behavioral guarantees and real-time oversight mechanisms in several of my deployment studies, I aim to design introspective robots that monitor, evaluate, and adapt their behaviors in-situ. These methods will integrate symbolic reasoning, probabilistic inference, and user feedback to effectively prevent harmful or socially inappropriate outputs. Beyond technical control, this research also engages with questions of moral responsibility, consent, and user agency to ensure that generative autonomy advances human well-being rather than undermining it.

Rethinking Robot Intelligence. The field is developed around a stable set of assumptions about what robots are and how they should behave. These tacit rules quietly govern research and design: robots should always engage, always help, always be productive, always remain polite, never lie, never err, and never ignore. While these prevailing norms have merit, they also constrain our imagination of the interactions robots can meaningfully support. In fact, I can imagine several contexts in which defying these norms can produce interactions that are more ethical, effective, and socially intelligent. For instance, should robots politely wait their turn to speak, or should they dare to interrupt, confront, or even withhold help? Consider the expectation that robots should always provide complete and accurate information. However, in socially or ethically complex situations, unconditional disclosure can be harmful. In elder care, for example, a robot may know where the car keys are but withhold that information if its user with dementia could attempt to drive unsafely. Here, withholding is not deception; it is a protective measure that prioritizes user safety over immediate compliance. Likewise, a therapeutic robot might soften or delay responses that could cause distress, or in educational contexts, intentionally pause before offering answers to encourage independent problem-solving. Hence, future work will critically and carefully evaluate such prevailing norms to develop computational models that expand our notion of "robotic intelligence," such as through reasoning about context, intent, and social consequence.

. . .

The vision I have for my own lab is to bridge technical innovation with human-centered design, training the next generation of researchers to approach robotics with both computational rigor and social responsibility. We will develop intelligent robots that operate safely and autonomously in real-world settings, fostering positive social outcomes that extend into everyday human–human relationships. Through open and community-based collaboration with clinicians, educators, and policymakers, my lab will ensure that advances in robotics and artificial intelligence meaningfully serve the people they are designed to help.

References

[1] Nicholas C Georgiou, **Rebecca Ramnauth**, Emmanuel Adeniran, Michael Lee, Lila Selin, and Brian Scassellati. Is someone there or is that the TV? Detecting social presence using sound. *ACM Transactions on Human-Robot Interaction (THRI)*, 12(4):1–33, 2023.

[2] Kayla Matheus, **Ramnauth, Rebecca**, Brian Scassellati, and Nicole Salomons. Long-term interactions with social robots: Trends, insights, and recommendations. *ACM Transactions on Human-Robot Interaction (THRI)*, 2025.

[3] **Rebecca Ramnauth**. *Chapter 3: Robots for Autism Therapy*. Ph.D. Dissertation, Yale University, 2025. URL `https://rramnauth2220.github.io/ramnauth-20250719-f2.pdf`. Systematic review appearing in *Building Intelligent Robots for Social Regulation Therapy*.

[4] **Rebecca Ramnauth**, Emmanuel Adéníran, Timothy Adamson, Michal A Lewkowicz, Rohit Giridharan, Caroline Reiner, and Brian Scassellati. A social robot for improving interruptions tolerance and employability in adults with ASD. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 4–13. IEEE, 2022.

[5] **Rebecca Ramnauth**, Dražen Brščić, and Brian Scassellati. A grounded observer framework for establishing guardrails for foundation models in socially sensitive domains. *arXiv preprint arXiv:2412.18639*, 2024.

[6] **Rebecca Ramnauth**, Dražen Brščić, and Brian Scassellati. Should I help?: A skill-based framework for deciding socially appropriate assistance in human-robot interactions. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 2051–2058. IEEE, 2024.

[7] **Rebecca Ramnauth**, Drazen Brščić, and Brian Scassellati. A robot-assisted approach to small talk training for adults with ASD. In *Proceedings of the Robotics: Science and Systems Conference (RSS)*, June 2025.

[8] **Rebecca Ramnauth**, Drazen Bršćić, and Brian Scassellati. From fidgeting to focused: Developing robot-enhanced social-emotional therapy (RESET) for school de-escalation rooms. In *Proceedings of the 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2025.

[9] **Rebecca Ramnauth**, Frederick Shic, and Brian Scassellati. Gaze behavior during a long-term, in-home, social robot intervention for children with ASD. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, page 949–957. IEEE Press, 2025.

[10] Nathan Tsoi, Joe Connolly, Emmanuel Adéníran, Amanda Hansen, Kaitlynn Taylor Pineda, Timothy Adamson, Sydney Thompson, **Rebecca Ramnauth**, Marynel Vázquez, and Brian Scassellati. Challenges deploying robots during a pandemic: An effort to fight social isolation among children. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 234–242, 2021.