Pax: A Queer-Affirming Robot Companion for the Home

Alexandria Thylane
City University of New York
New York USA
contact@alexandriathylane.com

Daniel J. Foulen
City University of New York
New York USA
daniel.foulen@cuny.edu

Keys K. Rigual
City University of New York
New York USA
keys.rigual82@myhunter.cuny.edu

Raitah A. Jinnat
City University of New York
New York USA
raitah.jinnat@macaulay.cuny.edu

Jackie Yee
City University of New York
New York USA
jackie.yee@cuny.edu

Kaylee Nam
Rice University
Houston USA
kn36@rice.edu

Raj Korpan
City University of New York
New York USA
raj.korpan@hunter.cuny.edu

Abstract

Pax is a queer-affirming robot companion prototype co-designed with LGBTQIA+ users. Implemented as a Unity-based embodied agent with a FastAPI backend, it translates community-identified requirements into a working interactive system. *Pax* combines queer-affirming natural language interaction, safety guardrails, and user-controlled adaptability to support identity affirmation and emotional well-being in the home. This demo highlights the core technical pipeline and ethical design choices behind queer-inclusive robot companions.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; • **Social and professional topics** → **Sexual orientation**; **Gender**; *Socio-technical systems*; • **Computer systems organization** → *Robotics*.

Keywords

Human–Robot Interaction, Queer-Affirming Design, Participatory Design, Ethical HRI, Robot Companions, LGBTQIA+

ACM Reference Format:
Alexandria Thylane, Daniel J. Foulen, Keys K. Rigual, Raitah A. Jinnat, Jackie Yee, Kaylee Nam, and Raj Korpan. 2026. Pax: A Queer-Affirming Robot Companion for the Home. 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/10.1145/3776734.3794588

1 Introduction

Interactive robot companions are increasingly encountered as part of everyday life, where they are used for emotional support, identity affirmation, and casual social engagement [1]. Rather than functioning solely as assistive tools, these systems are often experienced as sources of comfort and continuity across domestic, healthcare, and educational contexts [1, 14]. For queer (LGBTQIA+) individuals, whose lives are frequently shaped by discrimination, social isolation, and barriers to affirming care, such technologies may hold particular relevance for supporting mental health and everyday resilience [8, 17]. Prior research demonstrates that people readily form social bonds with virtual and robotic agents, engaging them as relational partners capable of fostering rapport, emotional regulation, and therapeutic alliance [5, 7, 11]. Building on these insights, we introduce *Pax*, a queer-affirming robot companion co-designed with queer adults to support everyday well-being in ways that reflect their identities, values, and lived experiences [10].

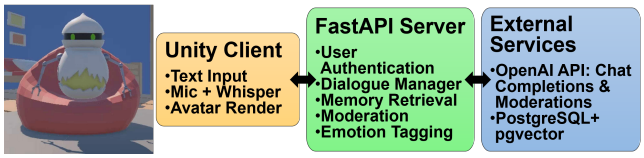

Figure 1: *Pax*, the queer-affirming robot companion, shown alongside a high-level overview of its system architecture.

Companionship with robots is never solely a dyadic human-machine relationship [16]. Companion robots are embedded within sociotechnical infrastructures that may mediate human–human ties, redistribute care labor, or raise concerns around surveillance, privacy, and liability [2, 16]. For queer communities, these risks intersect with long-standing harms in AI and robotics, including misgendering, erasure, and the encoding of cisnormative defaults [4]. Although participatory design (PD) is foundational in HRI for fostering identity-centered and culturally competent technology [3, 9], queer lived experience remains underrepresented in human-robot interaction research [6, 15]. This gap reflects broader critiques about whose values shape data, models, and embodiment in interactive systems [13, 18].

Participants emphasized the home as a setting for identity expression, emotional decompression, and safety, motivating *Pax*'s non-anthropomorphic form and supportive conversational style

HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3794588

[10]. The prototype operationalizes queer-centered values using large-language-model–driven dialogue with emotional expressivity, persistent memory-based personalization, scripted onboarding for names, pronouns, and boundaries, and safety guardrails for sensitive topics. This demo showcases *Pax*'s expressive behaviors, conversational architecture, and integration of community-derived design insights into a working domestic companion robot.

2 System Overview

Pax is implemented as a three-dimensional embodied virtual robot developed in Unity and supported by a FastAPI backend that manages dialogue, behavior generation, and system-level services, as shown in Figure 1. The Unity front end renders a stylized bedroom environment that situates the robot within its intended domestic use context. *Pax* appears as a small, abstract, non-humanoid agent with a rounded white shell, a soft conical head, and a black visor displaying simple green eyes. This form deliberately avoids gendered or animal-like cues while supporting subtle expressive behaviors through eye changes, arm motions, and posture adjustments.

Users interact with *Pax* through either text or speech. The system includes login and authentication to initialize personalized sessions. Once authenticated, users are guided through an introductory script that prompts them to share their name, pronouns, and preferred interaction style. This onboarding sequence helps establish safe and affirming conversational norms while enabling the robot to reference the user respectfully throughout the interaction.

For voice interactions, user utterances are transcribed locally within the Unity client using an embedded instance of Whisper [12]. The resulting text, whether originating from speech or typed input, is then routed from Unity to the FastAPI backend for processing. This backend integrates with a large language model (LLM; a GPT-family model) via the OpenAI API to interpret user text and generate responses. In addition to producing conversational content, the LLM assigns an emotion label to each turn, selected from a predefined set (*neutral, happy, sad, confused, surprise, dead, disgust, anger, cry, love*). These labels are sent back to the Unity client and mapped to corresponding animations and visual cues, allowing the robot's embodied behavior to reflect the emotional tone of the exchange in a controlled and stylized manner.

To enable continuity across sessions, *Pax* includes a memory-augmented personalization pipeline that extracts and stores user-disclosed information (e.g., names, pronouns, preferences, or life events). Salient facts are embedded and saved in a user-scoped PostgreSQL database. During each turn, the system retrieves semantically relevant memories via vector similarity search and injects them into the prompt, allowing *Pax* to reference prior conversations and maintain context over time. All memories derive from explicit user disclosures rather than inferred attributes, supporting identity affirmation while preserving user control.

To support safe deployment in sensitive and identity-relevant contexts, the system incorporates content moderation and safety-aware response handling. All user utterances are evaluated using the OpenAI Moderation API, which returns category-level assessments for risks such as self-harm, abuse, or hate-related content. Thresholds for several categories, particularly sexual content and hate speech, are deliberately tuned to better accommodate queer users' discussions of identity, embodiment, and lived experience, which may otherwise be misclassified as harmful. When elevated risk signals are detected (e.g., indications of acute distress or self-harm), *Pax* responds with predefined supportive and de-escalatory language, while avoiding clinical advice or claims beyond its scope. These guardrails ensure consistent, ethically bounded behavior across interactions while demonstrating how queer-affirming conversational design can be paired with robust safety infrastructure.

This architecture enables *Pax* to integrate queer-affirming dialogue, expressive embodiment, and safety-aware behavior into a cohesive prototype suitable for real-time demonstration. The modular design further supports future extensions, including additional emotional states, alternative moderation strategies, new interaction modalities, or deployment on physical robot platforms.

3 Demo Description

The demo presents a live interactive session with *Pax* running in Unity. Attendees are invited to log in to the system, either as themselves or using a guest profile, and are then guided through the introductory script in which *Pax* asks for their name, pronouns, and preferred interaction style. This onboarding sequence showcases how the robot establishes queer-affirming norms from the outset and orients the conversation around user-defined identity terms.

After onboarding, participants can converse with *Pax* via text input or spoken dialogue, with speech transcribed locally through Whisper before being processed by the backend. Our system processes this input, and the resulting response and associated emotion label are rendered in real time: *Pax* speaks its reply, displays the text on screen, and updates its eyes, arms, and posture to reflect the selected emotional state. Because the system maintains a persistent memory store, returning users can observe how *Pax* recalls previously shared preferences or identity information, demonstrating continuity across sessions. The demo highlights this full interaction loop, allowing attendees to see how queer-affirming language, emotional expression, and non-normative embodiment are coordinated.

To illustrate safety-aware behavior, we also provide scripted example prompts that trigger the system's guardrails, such as references to self-harm or acute distress. In these cases, *Pax* shifts to predefined supportive responses that acknowledge the user's feelings, avoid overstepping its role, and model safer handoffs (e.g., encouraging reaching out to trusted people or professional resources) without offering clinical advice. Through these interactions, the demo concretely demonstrates how *Pax* integrates identity affirmation, expressive embodiment, and safety mechanisms within a cohesive home-oriented companion prototype.

4 Discussion

Pax contributes to ongoing work in inclusive companion robotics and identity-affirming HRI by demonstrating how queer-centered design values can be concretely instantiated in an LLM-driven robot companion. Rather than treating personalization as a surface-level aesthetic layer, the prototype embeds deeper commitments to user-defined names, pronouns, boundaries, and interaction styles that avoid normative assumptions and center explicit self-disclosure. This approach reduces risks of misgendering, stereotyping, or invasive inference, harms disproportionately experienced by queer

individuals, while allowing users to determine how they are represented and addressed. Community participants also expressed discomfort with humanoid or animal-like forms that evoke gendered cues or anthropomorphic expectations [10]. Accordingly, *Pax*'s abstract, non-human form mitigates parasocial dependence, supports safer emotional boundaries, and creates a more open interpretive space for users to define the relationship.

The prototype operationalizes these commitments by integrating scripted onboarding, persistent personalization, emotional tagging, expressive behaviors, and safety-aware LLM responses, which together support identity affirmation and emotional decompression in the home, an environment participants described as central to well-being [10]. By embedding moderation checks and guardrails that constrain generative outputs during distress or sensitive conversations, *Pax* illustrates how ethical constraints can operate alongside expressive capabilities to protect users as conversational agents enter intimate spaces. These layered design choices demonstrate that warmth and responsiveness can be delivered without implying human equivalence or encouraging unhealthy dependence, advancing ethically grounded approaches to companion robot design that prioritize safety, autonomy, and identity affirmation.

As a live demonstration and design instantiation, this work emphasizes architectural integration and interaction design rather than formal empirical evaluation. We therefore do not yet report quantitative performance metrics, comparative benchmarks, or longitudinal measures of user experience, and the current implementation as a virtual embodied agent may limit direct generalization to fully physical robotic platforms. Our goal at this stage is to establish technical feasibility and concretely illustrate how queer-affirming principles, such as explicit self-disclosure, persistent memory-based personalization, expressive embodiment, and safety-aware moderation, can be operationalized in a working companion system.

Future work will include controlled performance evaluation and longer-term, in-home studies to assess usability, trust, and well-being outcomes, in addition to expanding emotional responses and sensing modalities, exploring adaptive behaviors over extended interaction, and evaluating how queer-affirming design principles translate across different embodiments, including physical robots. Additional community-based participatory design can refine guardrails, clarify expectations around autonomy and agency, and deepen understanding of how identity-affirming companions contribute to daily well-being.

5 Conclusion

Pax offers a foundation for developing companion robots that empower marginalized users by making space for their identities, preferences, and lived experiences. Developed through participatory design with queer adults and implemented as a functional, safety-aware, LLM-driven system for domestic contexts, the prototype demonstrates that identity-affirming principles can be translated into concrete architectural and embodiment choices in companion robotics, such as explicit self-disclosure, personalization, and respectful boundary management. By foregrounding queer community values, *Pax* challenges dominant assumptions embedded in social robot design and shows that queer-affirming interaction is not merely an aesthetic layer but a technical and methodological orientation with implications for trust, safety, and emotional sustainability. The system illustrates how companion technologies can support autonomy and well-being without reinforcing normative identity expectations or encouraging dependence, modeling more socially responsive and justice-oriented directions for HRI in domestic and intimate settings.

Acknowledgments

This research was supported by the CUNY Chancellor's Strategic Investment Initiative Fund and the CUNY BRES Collaboration Hub.

References

[1] Eshtiak Ahmed, Oğuz 'Oz' Buruk, and Juho Hamari. 2024. Human–robot companionship: Current trends and future agenda. *International Journal of Social Robotics* 16, 8 (2024), 1809–1860.

[2] Andrea Bertolini and Giuseppe Aiello. 2018. Robot companions: A legal and ethical analysis. *The Information Society* 34, 3 (2018), 130–140.

[3] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in AI design: Theoretical foundations and the current state of practice. In *ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–23.

[4] Tamanna Hossain, Sunipa Dev, and Sameer Singh. 2023. MISGENDERED: Limits of large language models in understanding pronouns. *arXiv preprint arXiv:2306.03950* (2023).

[5] Sooyeon Jeong, Laura Aymerich-Franch, Sharifa Alghowinem, Rosalind W Picard, Cynthia L Breazeal, and Hae Won Park. 2023. A robotic companion for psychological well-being: A long-term investigation of companionship and therapeutic alliance. In *Proceedings of the 2023 ACM/IEEE international conference on human-robot interaction*. 485–494.

[6] Raj Korpan. 2023. Trust in Queer Human-Robot Interaction. In *Trust, Acceptance and Social Cues in Human-Robot Interaction (SCRITA) Workshop Proceedings at RO-MAN 2023*.

[7] Nicole C Krämer, Astrid M Rosenthal-von der Pütten, and Laura Hoffmann. 2015. Social effects of virtual and robot companions. *The handbook of the psychology of communication technology* (2015), 137–159.

[8] Liviu-Catalin Mara, Matías Ginieis, and Ignasi Brunet-Icart. 2021. Strategies for coping with LGBT discrimination at work: A systematic literature review. *Sexuality Research and Social Policy* 18 (2021), 339–354.

[9] Michael Muller and Allison Druin. 2002. Participatory Design: The Third Space in HCI. *Handbook of HCI* (01 2002), 32.

[10] Kaylee Nam, Jackie Yee, Raitah A. Jinnat, Keys K. Rigual, Alexandria Thylane, and Raj Korpan. 2026. Not the Intended User: Queer Perspectives on Identity, Risk, and Trust in Robot Companions. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI '26)*. ACM.

[11] Adam Poulsen, Eduard Fosch-Villaronga, and Roger Andre Søraa. 2020. Queering machines. *Nature Machine Intelligence* 2, 3 (2020), 152–152.

[12] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.

[13] Katie Seaborn, Giulia Barbareschi, and Shruti Chandra. 2023. Not Only WEIRD but "Uncanny"? A Systematic Review of Diversity in Human–Robot Interaction Research. *International Journal of Social Robotics* (2023), 1–30.

[14] Andrew Specian, Ross Mead, Simon Kim, Maja Matarić, and Mark Yim. 2021. Quori: A Community-Informed Design of a Socially Interactive Humanoid Robot. arXiv:2109.00662 [cs.RO] https://arxiv.org/abs/2109.00662

[15] Michael Stolp-Smith and Tom Williams. 2024. More Than Binary: Transgender and Nonbinary Perspectives on Human Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[16] Ellen Van Oost and Darren Reed. 2010. Towards a sociological understanding of robots as companions. In *International conference on human-robot personal relationship*. Springer, 11–18.

[17] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring care: How trans and non-binary people meet health and well-being needs through technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[18] Katie Winkle, Erik Lagerstedt, Ilaria Torre, and Anna Offenwanger. 2023. 15 Years of (Who) man Robot Interaction: Reviewing the H in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–28.

Received 2025-12-08; accepted 2026-01-12