# Public Evaluation on Potential Social Impacts of Fully Autonomous Cybernetic Avatars for Physical Support in Daily-Life Environments: Large-Scale Demonstration and Survey at Avatar Land

Lotfi El Hafi[1,2,*], Kazuma Onishi[1], Shoichi Hasegawa[1], Akira Oyama[1], Tomochika Ishikawa[1],
Masashi Osada[1], Carl Tornberg[1], Ryoma Kado[1], Kento Murata[1], Saki Hashimoto[1],
Sebastian Carrera Villalobos[2], Akira Taniguchi[1], Gustavo Alfonso Garcia Ricardez[1,2],
Yoshinobu Hagiwara[1,3], Tatsuya Aoki[4], Kensuke Iwata[4], Takato Horii[4], Yukiko Horikawa[5],
Takahiro Miyashita[5], Tadahiro Taniguchi[1,6], and Hiroshi Ishiguro[4,5]

*Abstract*— Cybernetic avatars (CAs) are key components of an avatar-symbiotic society, enabling individuals to overcome physical limitations through virtual agents and robotic assistants. While semi-autonomous CAs intermittently require human teleoperation and supervision, the deployment of fully autonomous CAs remains a challenge. This study evaluates public perception and potential social impacts of fully autonomous CAs for physical support in daily life. To this end, we conducted a large-scale demonstration and survey during Avatar Land, a 19-day public event in Osaka, Japan, where fully autonomous robotic CAs, alongside semi-autonomous CAs, performed daily object retrieval tasks. Specifically, we analyzed responses from 2,285 visitors who engaged with various CAs, including a subset of 333 participants who interacted with fully autonomous CAs and shared their perceptions and concerns through a survey questionnaire. The survey results indicate interest in CAs for physical support in daily life and at work. However, concerns were raised regarding task execution reliability. In contrast, cost and human-like interaction were not dominant concerns. Project page: https://lotfielhafi.github.io/FACA-Survey/.

[1]Lotfi El Hafi, Kazuma Onishi, Shoichi Hasegawa, Akira Oyama, Tomochika Ishikawa, Masashi Osada, Carl Tornberg, Ryoma Kado, Kento Murata, Saki Hashimoto, Akira Taniguchi, Gustavo Alfonso Garcia Ricardez, Yoshinobu Hagiwara, and Tadahiro Taniguchi are with Ritsumeikan University; 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan. {lotfi.elhafi, onishi.kazuma, hasegawa.shoichi, oyama.akira, ishikawa.tomochika, osada.masashi, tornberg.carl, kado.ryoma, murata.kento, hashimoto.saki, a.taniguchi, garcia-g, yhagiwara, taniguchi}@em.ci.ritsumei.ac.jp

[2]Lotfi El Hafi, Sebastian Carrera Villalobos, and Gustavo Alfonso Garcia Ricardez are with Coarobo GK; 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0237, Japan. {lotfi, sebastian, tavo}@coarobo.com

[3]Yoshinobu Hagiwara is with Soka University; 1-236 Tangi, Hachioji, Tokyo 192-8577, Japan. hagiwara@soka.ac.jp

[4]Tatsuya Aoki, Kensuke Iwata, Takato Horii, and Hiroshi Ishiguro are with The University of Osaka; 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. {t.aoki@rlg., k.iwata@rlg., takato@, ishiguro@}sys.es.osaka-u.ac.jp

[5]Yukiko Horikawa, Takahiro Miyashita, and Hiroshi Ishiguro are with Advanced Telecommunications Research Institute International (ATR); 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan. {horikawa, miyasita, ishiguro}@atr.jp

[6]Tadahiro Taniguchi is with Kyoto University; Yoshida-Honmachi, Sakyo, Kyoto 606-8501, Japan. taniguchi@i.kyoto-u.ac.jp
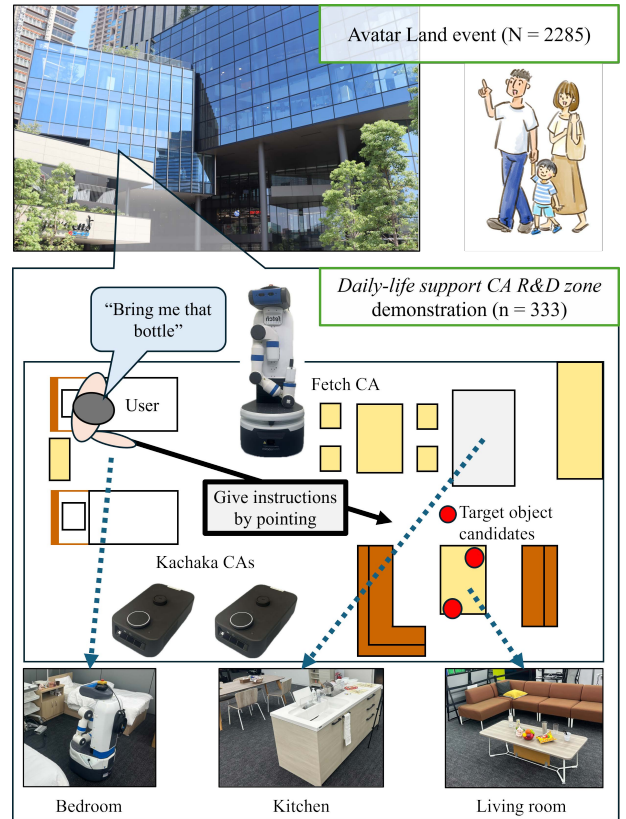
*Corresponding author.

Fig. 1. Overview of the *daily-life support CA R&D zone* demonstration at Avatar Land, where visitors interacted with fully autonomous cybernetic avatars (CAs) in a replicated home environment.

## I. INTRODUCTION

As birth rates decline and the population ages in developed societies, an increasing burden is expected to be placed on the working population to care for younger and elderly individuals while maintaining economic productivity and social welfare. In this context, cybernetic avatars (CAs) have been envisioned as a key component of an avatar-symbiotic society [1], aiming to free individuals from physical constraints by expanding their capabilities through teleoperated, semi-autonomous virtual agents and robotic assistants. While the application of CAs as conversational agents has been

actively researched and demonstrated [2]–[4], broader adoption is yet to be achieved, and further public assessment is required to demonstrate their necessity. In particular, most CA-related research has focused on semi-autonomous robots reliant on human teleoperation. However, the feasibility and social implications of fully autonomous robotic CAs and their physical interaction capabilities remain underexplored.

In this regard, various studies have investigated using service robots to provide physical support in daily life and work settings. Recent advancements in foundation models for robotics [5]–[7] aim to enable robots to acquire general-purpose physical skills by learning from large-scale datasets collected in home and office environments. However, achieving full substitution of human tasks remains a significant challenge. This is further highlighted in robotics competitions that turn service tasks, including object manipulation and instructions in natural language, into competitive benchmarks [8], [9].

Many robots that assist humans require high degrees of freedom to operate effectively across diverse environments and objects. While recent mobile manipulators and humanoid robots are capable of dexterous full-body movements, they must also dynamically perceive their surroundings at all times to ensure contextual awareness, shared understanding, and safe operations, which remain challenging in daily environments alongside humans. Therefore, gathering public data for validating future research and development directions toward fully autonomous CAs capable of physical interaction is essential, which requires publicly demonstrating prototypes to a representative sample of non-expert users.

To this end, we conducted a large-scale demonstration and survey during Avatar Land[1], a month-long event in a densely populated area of Osaka, Japan, where fully autonomous robotic CAs, alongside semi-autonomous robotic CAs, performed daily object retrieval tasks, as illustrated by Fig. 1. This demonstration was open to the public, and all visitors were invited to complete a survey questionnaire assessing their experiences with various CAs, including fully autonomous robotic systems. Following Avatar Land, we conducted a comprehensive analysis of the responses from 2,285 visitors, evaluating public perceptions, potential demand, and societal concerns regarding fully autonomous CAs. Hence, instead of focusing on quantitative metrics, our approach centered on assessing subjective public perception on a large scale.

Our contributions are threefold:

- We developed and deployed a fully autonomous CA demonstration, distinguishing it from the semi-autonomous approaches commonly explored with CAs.
- We conducted a large-scale public survey to assess the potential social impacts of fully autonomous CAs providing physical support in daily life.
- We analyzed the survey data to derive insights into public perception toward the future deployment of fully autonomous CAs.

## II. SURVEY METHODOLOGY

Avatar Land was held at Grand Green Osaka, a newly opened commercial and public facility located a 7-minute walk from Osaka Station, one of Japan's largest transportation hubs. The event ran from September 10 to 29, 2024, with daily opening hours from 10:00 to 16:00. The location and open nature of the event allowed for the gathering of a general audience, including families with young children and elderly individuals. Visitors could participate in 11 distinct interactive demonstrations illustrating various aspects of CAs' potential applications, at their own pace, with staff members providing guidance and explanations, and inviting them to fill a survey questionnaire at the end of their visit. The demonstrations were named as follows: *computer graphics (CG) CA zone*, *CA receptionist zone*, *communication training*, *caregiving support CA zone*, *CA teleoperation zone*, *daily-life support CA R&D zone*, *single operator controlling 15 CAs*, *facility guidance from teleoperated mobile CA*, *guide service from various CA cooperation*, *multi-language service for international cooperation*, and *chat service from paired CAs*.

Filling out the survey questionnaire was voluntary, and responses were collected from 2,285 participants who chose to answer ($N = 2285$). The participants who volunteered were asked the following multiple-choice questions:

- Q1: "Which CA demonstration did you participate in?"
- Q2: "Would you use the demonstrated CAs in your daily life?"
- Q3: "In what situations would you use the demonstrated CAs?"
- Q4: "Why do you not want to use the demonstrated CAs?"

Q1 was asked once to all survey volunteers, while Q2–Q4 were asked for each demonstration they participated in. The detailed possible answers are provided in Table I and Fig. 3. All data were anonymized, and visitors were informed that their data could be used for research purposes by the ethical guidelines of the Japan Science and Technology Agency (JST). Demographic data were also collected, but their analysis is outside the scope of this study. The total number of visitors is unknown as the demonstration site was deliberately open, without controlled entry points to encourage spontaneous interaction with the CAs. However, the number of visitors who did not respond to the survey questionnaire was likely an order of magnitude higher than those who did.

The demonstration described in this study, *daily-life support CA R&D zone*, was proposed by Group 4[2] within the JST Moonshot Research & Development Program project Avatar-Symbiotic Society [1]. Importantly, the *daily-life support CA R&D zone* was the only demonstration showing fully autonomous robotic CAs that were not teleoperated in any form, making it unique in its emphasis on autonomous physical support in daily life. To this aim, it took place in
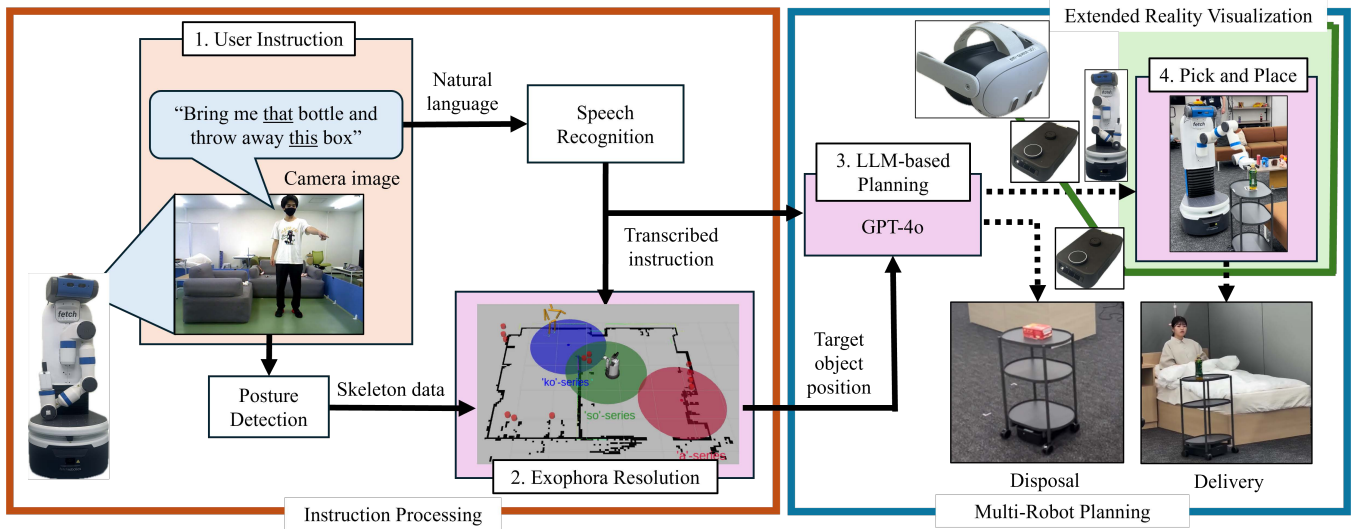
---

Fig. 2. Overview of the implementation of the demonstrated system. The user provides instructions to the robot by pointing at the target object. The system performs exophora resolution to identify the intended object and autonomously plans task execution among multiple robotic CAs.

a reproduced home environment that included a bedroom, kitchen, living room, and medical home care equipment, as described in Fig. 1. The demonstrated system consisted of three fully autonomous robotic CAs: a Fetch Robotics Fetch[3] mobile manipulator and two Preferred Robotics Kachaka[4] shelf-carrying robots.

Participants were invited to observe and interact with the system in a scenario where a user in bed would point to a target object and instruct the robot to retrieve it. The system performed exophora resolution to distinguish the target from similar objects using multimodal contextual information gathered from active exploration, ensuring accurate task execution. The robotic then CAs collaborated to execute the task autonomously, while visitors could monitor the process through a Meta Quest 3[5] extended reality (XR) headset, which provided real-time visualization of the robotic CAs' perception and execution processes. Additional TV displays were installed to share real-time visualization outside the XR headset and encourage visitor engagement.

The *daily-life support CA R&D zone* demonstration was designed as a real-world use case aligned with the envisioned future of an avatar-symbiotic society, showcasing how fully autonomous CAs could provide physical support in daily-life environments with:

- Multi-robot collaboration that allows a single user to interact with multiple CAs.
- Full autonomy that eliminates human supervision or teleoperation of the CAs.
- Intuitive interaction that relies on pointing gestures and natural language instructions to the CAs.
- Real-time visualization that facilitates shared understanding in XR between the user and CAs.

From the 2,285 survey questionnaires collected during Avatar Land, 333 visitors reported participating in *daily-life support CA R&D zone* demonstration ($n = 333$).

## III. SYSTEM OVERVIEW

This section details the flow and technical implementation of the *daily-life support CA R&D zone* demonstration, as described by Fig. 2. The system enables fully autonomous CAs to process user instructions, plan multi-robot actions, and execute object retrieval tasks. The entire demonstration was implemented using a containerized software development environment (SDE) [10], with all communication between CAs handled via ROS [11].

User speech was transcribed into text using Whisper [12], while the user posture was extracted from RGB-D images captured by Fetch CA's head camera using MediaPipe [13]. These inputs were processed by an exophora resolution model [14], enabling the system to determine the target object based on pointing gestures and natural-language instructions, and contextual information. The identified object location and user command were input into a large language model (LLM), GPT-4o [15], for multi-robot task planning.

The task execution sequence involved Fetch and one Kachaka robot CAs retrieving the specified object, while the second Kachaka CA handled object disposal. The entire process could be monitored in real-time using a Meta Quest 3 XR headset, providing users with an interactive visualization of robot CAs' perception and execution.

### A. Instruction Processing

User instructions were processed using an exophora resolution model [14], which resolves ambiguous linguistic commands, such as those including pronouns or demonstratives, through contextual information. Exophora information, in contrast to endophora information, refers here to information not included in the user's instruction itself but inferred

based on contextual information. The probability $p$ of the estimated target object location was computed as the product of three probabilities: $p_1$ related to the user's pointing gesture direction, $p_2$ related to the demonstrative word used in the instructions, and $p_3$ related to the object's category distribution in the environment.

In this demonstration, $p_1$ was determined using pose skeleton data, specifically the user's eyes and wrists, extracted from images captured by Fetch CA's head camera and processed with MediaPipe [13]. $p_2$ was assigned based on the detected demonstrative word in the user's instruction (*e.g.*, "this" or "that"). Ideally, $p_3$ should be derived from active exploration methods [16], to automatically acquire contextual information and facilitate the introduction of robotic CAs in new environments without user intervention.

However, to ensure the robustness of exophora resolution during the *daily-life support CA R&D zone* demonstration, we preloaded a user's pose skeleton and transcribed instruction data prepared offline, and manually recorded object locations as the model inputs. In addition, instead of demonstrating active exploration with visitors, which is time-consuming by nature, a pre-recorded video of the exploration performed on the site of Avatar Land was shown on TV displays.

### B. Object Manipulation

After moving near the target object inferred by the exophora resolution model, the Fetch CA picks up the object and places it on a Kachaka CA for delivery to the user, as the *daily-life support CA R&D zone* demonstration was designed to emphasize multi-CA collaboration. Although the position of the object is assumed to be known, in reality, there may be errors in self-position estimation, or the object position may have changed since the exophora resolution. Therefore, additional object detection was conducted with Detic [17] before grasping the target object for confirmation. The detected results were then combined with depth information to estimate the position of the target object. The position for placing the object on the Kachaka CA also needed to be determined similarly, so AR markers were attached to the shelf to compute the relative positions between the two CAs. MoveIt's RRT-Connect[6] was used for motion planning to pick and place the target object.

This conventional manipulation approach was also augmented with an experimental grasping method using infrared light-based proximity sensors [18]. We attached these sensors to the Fetch CA's fingertip and used them to control hand positioning and the grasping approach motion by inferring the target object's reflectance using a multimodal LLM, which is a unique property of each material [19]. However, this experimental proximity sensing-based grasping system was not demonstrated throughout the entire Avatar Land event, as we reverted to the conventional approach due to wiring limitations between the proximity sensors and the Fetch CA, which could have caused accidents.

[6]https://ompl.kavrakilab.org/

### C. Multi-Robot Planning

As the *daily-life support CA R&D zone* demonstration was designed to emphasize multi-CA collaboration, task allocation between Fetch and the two Kachaka robotic CAs was performed with an LLM-based approach [20] using GPT-4o. In this approach, appropriate action sequences were generated for each CA by providing GPT-4o with the capabilities of each robot and an example of task planning. However, to ensure reliable task planning during the *daily-life support CA R&D zone* demonstration, we precomputed task allocations using a set of predefined user instructions. The execution itself of each task was managed using a FlexBE behavior engine [21] state machine.

The demonstration also required ROS communication between all three robotic CAs. As Kachaka operates on ROS 2, we wrapped Kachaka's Python API in a ROS-compatible interface, enabling unified control of all robots inside our containerized SDE [10] running ROS.

### D. Extended Reality Visualization

For users to better understand the system's operation and monitor task execution of fully autonomous CAs in real time, a visualization interface was developed based on recent findings in human-robot interaction in XR [22]–[24]. Specifically, a Meta Quest 3 XR headset was used to display real-time sensor data and robot motion plans, allowing users to observe the CAs' self-localization, navigation paths, and planned arm trajectories. Importantly, the user could leverage the passthrough cameras of the XR headset to adjust the level of virtuality between augmented reality (AR), augmented virtuality (AV), and virtual reality (VR).

## IV. RESULT ANALYSIS

The survey results are shown in Table I and Fig. 3. Fig. 3(Q1) describes the participation of the 2,285 survey volunteers ($N = 2285$) across all 11 demonstrations, given the question "Which CA demonstration did you participate in?" (Q1). A total of 333 respondents participated in the *daily-life support CA R&D zone* demonstration described in this study ($n = 333$), ranking 8th out of the 11 demonstrations.

The usage likelihood regarding the demonstrated fully autonomous CAs is shown in Fig. 3(Q2). Among those who experienced the *daily-life support CA R&D zone*, 39.3% answered "Very likely to use" ($n_1^+ = 131$), and 35.4% responded "Use if conditions are right" ($n_2^+ = 118$) to the question "Would you use the demonstrated CAs in your daily life?" (Q2). This implies that 74.7% of the respondents ($n_1^+ + n_2^+ = 249$) had a positive attitude toward CAs providing daily-life support, indicating that the demonstration generated public interest. It is important to acknowledge that these positive responses may reflect initial impressions rather than views on long-term usability. Additionally, the results suggest that aversion to daily-life support CAs is minimal, with only 2.4% of "Do not want to use" responses ($n^- = 8$).

The scenarios in which people would like to use CAs are shown in Fig. 3(Q3). To the question "In what situations

| Demonstration Participation (N = 2285) | | |
|---|---|---|
| "Q1: Which CA demonstration did you participate in?" | | |
| Responses | Rate [%] | Sample |
| CG CA zone | 56.4 | 1289 |
| CA receptionist zone | 47.9 | 1095 |
| Communication training | 31.1 | 711 |
| Caregiving support CA zone | 39.7 | 907 |
| CA teleoperation zone | 16.9 | 386 |
| **Daily-life support CA R&D zone** | **14.7** | **n = 333** |
| Single operator controlling 15 CAs | 25.9 | 591 |
| Facility guidance from teleop. mobile CA | 31.1 | 710 |
| Guide service from various CA coop. | 12.4 | 283 |
| Multi-language service for int'l coop. | 9.1 | 207 |
| Chat service from paired CAs | 7.2 | 164 |
| No answer | 2.1 | 47 |
| **Usage Likelihood (n = 333)** | | |
| "Q2: Would you use the demonstrated CAs in your daily life?" | | |
| Responses | Rate [%] | Sample |
| **Very likely to use** | **39.3** | $n_1^+ = 131$ |
| **Use if conditions are right** | **35.4** | $n_2^+ = 118$ |
| Cannot tell if want to use or not | 20.8 | 69 |
| **Do not want to use** | **2.4** | $n^- = 8$ |
| No answer | 2.1 | 7 |
| **Usage Scenarios ($n_1^+ + n_2^+ = 249$)** | | |
| "Q3: In what situations would you use the demonstrated CAs?" | | |
| Responses | Rate [%] | Sample |
| Work | 32.5 | 81 |
| Education and studying | 17.3 | 43 |
| Hobby and leisure | 15.7 | 39 |
| Conversation and communication | 15.3 | 38 |
| Daily life | 47.8 | 119 |
| Others | 4.4 | 11 |
| No answer | 3.2 | 8 |
| **Usage Aversions ($n^- = 8$)** | | |
| "Q4: Why do you not want to use the demonstrated CAs?" | | |
| Responses | Rate [%] | Sample |
| Could not use well | 37.5 | 3 |
| Seemed fragile | 12.5 | 1 |
| Seemed expensive | 12.5 | 1 |
| No space for using or storing | 0.0 | 0 |
| Prefer human face-to-face interaction | 25.0 | 2 |
| Others | 25.0 | 2 |
| No answer | 25.0 | 2 |
| **Participant Comments** | | |

**Participant Comments**

C1: "I felt that life-support avatars are easy to imagine being used in various situations and can be utilized in many ways. However, I could not clearly envision their applications in communication, so I am looking forward to seeing how they will be used in the future."
C2: "I would love for them to be useful in households, especially for tasks like fetching items and tidying up."
C3: "It was fascinating to see the perspective of an autonomous avatar (robot). I was surprised at how much calculation goes into even picking up a single object."
C4: "I believe that if remote-support robots become more practical, they will greatly expand the possibilities in our daily lives."
C5: "I have high expectations for life-support and security robots (avatars). I think they could contribute to community safety."
C6: "Life-support assistance when caregiving is needed, as well as customer service in public facilities."
C7: "When I used a caregiving avatar to carry a plastic bottle, I wished it could also open the bottle cap."

exceed 50%. This suggests a gap in the perceived applicability of CAs for physical support in home environments. Further investigation is required to understand the factors contributing to this gap.

Fig. 3(Q4) highlights the reasons why some respondents were hesitant to use CAs. To the question "Why do you not want to use the demonstrated CAs?" (Q4), 37.5% of respondent mentioned "Could not use well" as their primary concern. From the results of Fig. 3(Q2) and Fig. 3(Q4), ensuring a high success rate and stability in demonstrations appears critical to meeting users' conditions for adopting fully autonomous CAs at home.

We also gathered feedback from the free-text section of the survey, as reported at the bottom of Table I. Some respondents, like in comment C7, suggested that adaptable manipulation for daily tasks should be prioritized to enhance perceived utility. Others expressed hope for the practical application of robotic CAs and noted that their interest in the technology increased after experiencing the demonstration, suggesting that such public engagement events are valuable in helping non-experts envision the future of CAs. However, free-text comments were too few to draw strong conclusions.

## V. CONCLUSION

This study presented a large-scale public demonstration and survey at Avatar Land, evaluating public perception and the potential social impact of fully autonomous CAs in daily-life environments. Unlike prior research on teleoperated CAs, this study assessed how fully autonomous robotic CAs were perceived when performing physical support tasks without human supervision. Specifically, we conducted a comprehensive analysis of responses from 2,285 visitors engaging with various CAs, from which a subset of 333 participated in the demonstration of fully autonomous CAs described in this study. Instead of focusing on quantitative metrics, our approach centered on assessing subjective public perception on a large scale via a survey questionnaire.

The survey results indicated public interest in integrating fully autonomous CAs into daily life and at work. However, concerns were raised about task execution reliability, with hesitation primarily centered on whether the robots could consistently complete tasks successfully. In contrast, factors such as cost and human-like interaction were not dominant concerns, although perceptions of cost might evolve with more prolonged user experience. These findings highlight the need to improve robot CA task performance to increase public adoption toward realizing an avatar-symbiotic society.

## REFERENCES

[1] H. Ishiguro, "The Realisation of an Avatar-Symbiotic Society Where Everyone can Perform Active Roles without Constraint," *RSJ Advanced Robotics (AR)*, vol. 35, no. 11, pp. 650–656, June 2021.
[2] K. Sakai, *et al.*, "Simultaneous Dialogue Services using Multiple Semiautonomous Robots in Multiple Locations by a Single Operator: A Field Trial on Souvenir Recommendation," *IEEE Robotics and Automation Letters (RA-L)*, vol. 9, no. 7, pp. 6280–6287, July 2024.
[3] Y. Horikawa, *et al.*, "Cybernetic Avatar Platform for Supporting Social Activities of All People," in *Proc. of 2023 IEEE/SICE International Symposium on System Integration (SII 2023)*, Atlanta, United States, Jan. 2023, pp. 1–4.

would you use the demonstrated CAs?" (Q3), the most frequent responses were for "Daily life" and at "Work", with 47.8% and 32.5% of responses, respectively. Since the demonstration focused on physical daily-life assistance, it is notable that the percentage of "Daily life" answers did not
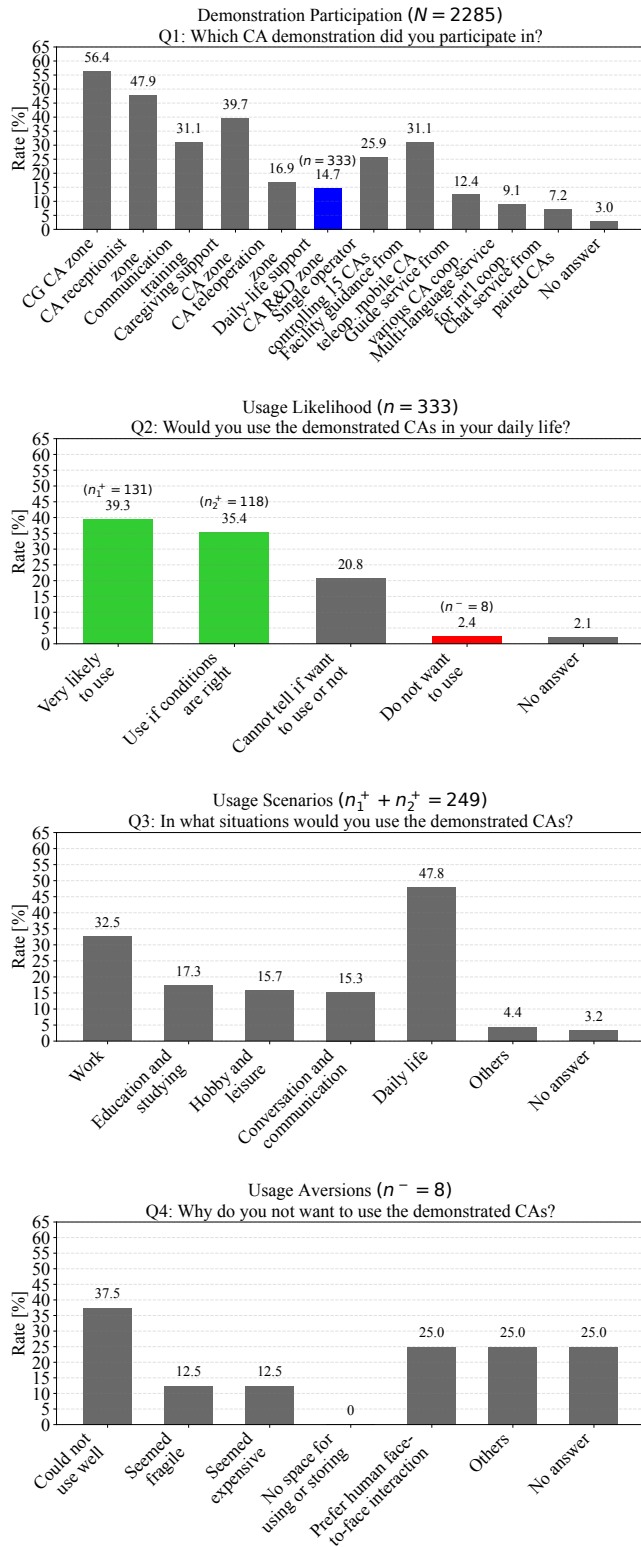
Fig. 3. Consolidated survey questionnaire results.

[4] Y. Fu, *et al.*, "Dual Variational Generative Model and Auxiliary Retrieval for Rmpathetic Response Generation by Conversational Robot," *RSJ Advanced Robotics (AR)*, vol. 37, no. 21, pp. 1406–1418, Nov. 2023.

[5] D. Ghosh, *et al.*, "Octo: An Open-Source Generalist Robot Policy," in *Proc. of 2024 Robotics: Science and Systems (RSS XX)*, Delft, Netherlands, July 2024, pp. 1–13.

[6] K. Black, *et al.*, "Pi0: A Vision-Language-Action Flow Model for General Robot Control," Oct. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2410.24164

[7] A. O'Neill, *et al.*, "Open X-Embodiment: Robotic Learning Datasets and RT-X Models," in *Proc. of 2024 IEEE International Conference on Robotics and Automation (ICRA 2024)*, Yokohama, Japan, May 2024, pp. 6892–6903.

[8] K. Wada, "New Robot Technology Challenge for Convenience Store," in *Proc. of 2017 IEEE/SICE International Symposium on System Integration (SII 2017)*, Taipei, Taiwan, Dec. 2017, pp. 1086–1091.

[9] L. Contreras, *et al.*, "Towards General Purpose Service Robots: World Robot Summit - Partner Robot Challenge," *RSJ Advanced Robotics (AR)*, vol. 36, no. 17-18, pp. 812–824, Sept. 2022.

[10] L. El Hafi, *et al.*, "Software Development Environment for Collaborative Research Workflow in Robotic System Integration," *RSJ Advanced Robotics (AR)*, vol. 36, no. 11, pp. 533–547, June 2022.

[11] M. Quigley, *et al.*, "ROS: An Open-Source Robot Operating System," in *Proc. of 2009 IEEE Workshop on Open Source Software*, Kobe, Japan, May 2009.

[12] A. Radford, *et al.*, "Robust Speech Recognition via Large-Scale Weak Supervision," in *Proc. of 40th International Conference on Machine Learning (ICML 2023)*, vol. 202, Honolulu, United States, July 2023, pp. 28 492–28 518.

[13] C. Lugaresi, *et al.*, "MediaPipe: A Framework for Perceiving and Processing Reality," in *Workshops of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, Long Beach, United States, June 2019, pp. 1–4.

[14] A. Oyama, *et al.*, "Exophora Resolution of Linguistic Instructions with a Demonstrative based on Real-World Multimodal Information," in *Proc. of 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023)*, Busan, South Korea, Aug. 2023, pp. 2617–2623.

[15] T. Brown, *et al.*, "Language Models are Few-Shot Learners," in *Proc. of 2020 Conference on Neural Information Processing Systems (NeurIPS 2020)*, vol. 33, (Virtual), Dec. 2020, pp. 1877–1901.

[16] T. Ishikawa, *et al.*, "Active Semantic Mapping for Household Robots: Rapid Indoor Adaptation and Reduced User Burden," in *Proceedings 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2023)*, Honolulu, United States, Oct. 2023, pp. 3116–3123.

[17] X. Zhou, *et al.*, "Detecting Twenty-Thousand Classes Using Image-Level Supervision," in *Proc. of 17th European Conference on Computer Vision (ECCV 2022)*, S. Avidan, *et al.*, Eds., Tel Aviv, Israel, Oct. 2022, pp. 350–368.

[18] Y. Suzuki, *et al.*, "Grasping Strategy for Unknown Objects based on Real-Time Grasp-Stability Evaluation using Proximity Sensing," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 8643–8650, Oct. 2022.

[19] G. A. Garcia Ricardez, *et al.*, "Reflectance Estimation for Pre-grasping Distance Measurement using RGB and Proximity Sensing," in *Proc. of 2023 IEEE/SICE International Symposium on System Integration (SII 2023)*, Atlanta, United States, Jan. 2023, pp. 1–6.

[20] S. Hasegawa, *et al.*, "Reducing cost of on-site learning by multi-robot knowledge integration and task allocation via large language models," *Journal of the Robotics Society of Japan (JRSJ)*, Jan. 2025.

[21] P. Schillinger, *et al.*, "Human-Robot Collaborative High-Level Control with Application to Rescue Robotics," in *Proc. of 2016 IEEE International Conference on Robotics and Automation (ICRA 2016)*, Stockholm, Sweden, May 2016, pp. 2796–2802.

[22] L. El Hafi, *et al.*, "Teaching System for Multimodal Object Categorization by Human-Robot Interaction in Mixed Reality," in *Proc. of 2021 IEEE/SICE International Symposium on System Integration (SII 2021)*, Iwaki, Japan (Virtual), Jan. 2021, pp. 320–324.

[23] H. Nakamura, *et al.*, "Multimodal Object Categorization with Reduced User Load through Human-Robot Interaction in Mixed Reality," in *Proc. of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022)*, Kyoto, Japan, Oct. 2022, pp. 2143–2150.

[24] C. Tornberg, *et al.*, "Mixed Reality-based 6D-Pose Annotation System for Robot Manipulation in Retail Environments," in *Proc. of 2024 IEEE/SICE International Symposium on System Integration (SII 2024)*, Ha Long, Vietnam, Jan. 2024, pp. 1425–1432.