Pop-Up Encounters with Spot: Shaping Public Perceptions of Robots Through Hands-On Experience

Hae Won Park
haewon@media.mit.edu
MIT Media Lab
Cambridge, MA, USA

Georgia Van de Zande
gvandezande@theaiinstitute.com
RAI Institute
Cambridge, MA, USA

Xiajie Zhang
xiajie@media.mit.edu
MIT Media Lab
Cambridge, MA, USA

Dawn Wendell
dawn@rai-inst.com
RAI Institute
Cambridge, MA, USA

Jessica Hodgins
jkh@rai-inst.com
RAI Institute
Cambridge, MA, USA

Abstract

Public attitudes toward robots are often shaped by indirect exposure (e.g., media, staged demos), leaving open how direct, *hands-on* experience influences acceptance. In this study, we investigate how interacting with Boston Dynamics' Spot, an agile, state-of-the-art quadruped robot, in a public pop-up booth affects perceptions of comfort and suitability across everyday and high-stakes environments. In a walk-up, 10-week pop-up booth, participants ($N = 753$) completed pre–post surveys before and after driving Spot within curated Drive Scenes (Factory, Home, Hospital, Outdoor/Disaster). Measures captured *comfort encountering robots* and *perceived suitability* across Rated Contexts (RCs), affective reactions, and open-ended reflections. Hands-on control significantly increased comfort across all RCs, with the largest gains in Outdoor/Disaster, and increased perceived suitability—most in Home/Office/Hospital where baselines were lower. Improvements generalized beyond the experienced Drive Scene to other contexts. Age, gender, and prior familiarity moderated baseline levels and some changes, but hands-on exposure raised scores for all groups and attenuated several gaps. Thematic analysis showed memorable moments tied to locomotion, terrain adaptation, and expressive tilt; imagined roles consistently emphasized domestic assistance (e.g., cleaning, mobility), with *entertainment/play* and *companionship* emerging post-interaction. Together, these results demonstrate that brief, agency-granting encounters with a high-capability quadruped can broaden where people see robots as appropriate and diversify envisioned roles, offering a scalable model for public-facing HRI that fosters comfort, enthusiasm, and acceptance.

Keywords

Human–Robot Interaction (HRI), Public Attitude Toward Robots, Quadruped Robots, Hands-on Experience, Boston Dynamics Spot

1 Introduction

Robots are increasingly entering public spaces, workplaces, and homes, yet public attitudes remain ambivalent. On the one hand, robots are envisioned as valuable assistants in hazardous industries, healthcare, and household tasks. On the other, media portrayals and limited personal encounters can foster uncertainty about safety, usefulness, and social role. Understanding how such attitudes form—and how they can be shifted through *direct interaction*—is a central challenge in human–robot interaction (HRI).

Prior work has probed public perceptions via large-scale surveys, controlled laboratory encounters, and public exhibitions where passers-by observe, and sometimes interact with, robots in action. Each yields important insights, but few studies examine how *hands-on control by the general public* shapes *comfort* and *perceived suitability*, especially for *quadruped* robots like Boston Dynamics' Spot, whose physical agility can elicit both fascination and apprehension.

To address this gap, we developed *Drive-a-Spot*, a walk-up public booth where participants tele-operated Spot within curated *Drive Scenes* (DSs) reflecting real-world contexts—Factory, Home, Hospital, and Outdoor/Disaster. Structured *pre–post* surveys then captured perceptions across *Rated Contexts* (RCs) (e.g., "How comfortable would you feel if you encountered a robot in … ?" and "How well would this robot work in … ?"). Specifically, we asked: **How does hands-on interaction with a quadruped robot influence public perceptions of (a) comfort encountering robots, (b) perceived suitability across domains, (c) affective reactions, and (d) imagined roles for robots?** We hypothesize that:

- **Main effect of interaction:** Hands-on experience will increase *comfort* and *perceived suitability* ratings across Rated Contexts (RCs).
- **Demographic and exposure moderators:** Baseline levels and pre–post changes will vary by *age*, *gender*, and *prior exposure to Spot*; hands-on experience will raise scores for all groups and *attenuate* (but not necessarily eliminate) some gaps, with effects *depending on context.*
- **Role diversification:** After interaction, imagined robot roles will broaden beyond hazardous/industrial work to include a more diverse set of applications.

By situating *Drive-a-Spot* within the broader landscape of public-facing HRI studies, this work contributes new evidence of how direct, embodied control of an advanced quadruped robot can reshape not only attitudes but also the ways people imagine integrating robots into their everyday lives.

2 Related Work

2.1 Public Attitudes Toward Robots

Public attitudes toward robots have long been characterized by ambivalence: enthusiasm for robots' potential in hazardous and industrial work coexists with skepticism regarding their presence in care and domestic domains. The 2012 Eurobarometer reported strong support for manufacturing, demining, and space exploration but substantially lower support for healthcare, childcare, and education, alongside worries about job loss and diminished "human touch" [15]. A longitudinal analysis (2012–2017) further showed declining general attitudes toward autonomous systems, even as acceptance rose for specific applications (e.g., driverless cars, medical operations) and dropped for service robots in work and elder care [16]—underscoring the importance of domain framing.

Attitudes vary by who is surveyed and where. Demographic and cultural moderators are well-documented: men and higher-SES groups often report more favorable views than women and lower-income groups [2, 12, 16]; age effects are mixed (younger respondents sometimes more enthusiastic, older adults more receptive to assistance yet more cautious about autonomy/safety) [3, 16, 28]. Cross-regional differences across Europe, Asia, and the United States indicate culturally inflected perceptions rather than a single global pattern [25].

Beyond surveys, media and discourse shape and reflect public attitudes. Analyses of news and social platforms indicate a shift in public conversation from industrial imagery toward domestic/assistive narratives over the past decade [20]. At the same time, social-media reactions to high-profile robots (e.g., quadrupeds, drones) remain polarized, pairing admiration for agility with anxieties about militarization and surveillance [24]. The persistence of such ambivalence suggests that exposure alone is insufficient; rather, perceptions depend on how robots are framed (humanitarian vs. policing), who is imagined to deploy them, and what signals of intent and control are made legible. A systematic review (2005–2019) integrates these threads: coarse attitudes are often positive, but granular judgments depend on **application domain, morphology, and exposure modality**, with the literature still humanoid-centric and non-humanoid platforms underexamined [25]. Notably, evidence on exposure effects is inconsistent across studies (direct in-person vs. indirect video/simulation), reinforcing that **how the public "meets" robots** critically conditions attitudes.

2.2 Study Designs for Investigating Public Perception

Researchers probe attitudes using (a) representative surveys and discourse analyses that offer breadth, (b) lab-based experiments that enable controlled manipulation and validated scales (e.g., Godspeed [4, 17]), and (c) walk-up public studies in museums, malls, and other civic venues where participants are not pre-recruited. These walk-up studies capture spontaneous, on-site participation from heterogeneous publics and can involve surveys or observation only [29], spontaneous interaction [13, 21], or participant operation of the robot [29]. However, they rarely include within-subject pre–post measurement due to throughput and timing constraints, which limits causal inference on attitude change. In contrast, lab studies permit pre–post designs and construct-valid measurement (e.g., trust, safety, anthropomorphism) but often rely on smaller, less diverse samples or indirect exposure (video/simulation) whose effects can diverge from direct encounters [19, 30]. A cumulative lesson from this methodological spectrum is that exposure design—its modality, duration, and the degree of agency afforded—is a first-order driver of the attitudes researchers observe.

2.3 Quadruped Robots and Public Perception

Quadrupeds have gained prominence because their four-legged morphology affords stability, terrain adaptability, and recovery capabilities hard to achieve with wheeled or bipedal platforms, making them appealing for inspection, industrial, and search-and-rescue tasks. Quadruped robots have been widely disseminated through videos that highlight agility across varied contexts such as manufacturing and rough terrains, but such visibility also binds quadrupeds to policing/military narratives that can depress acceptance [24]. Empirical HRI studies show that public perception turns on appearance and framing: markers of human control (e.g., leashing, service vests) and benign roles increase perceived safety and approachability [19]; behavioral stylings perceived as submissive are judged safer than dominant ones during incidental encounters [18]. Yet much of this work uses staged or lab setups, small samples, and observation-based exposure. Research on hands-on operation of quadrupeds is emerging, however, focusing on the command modality comparisons [32]. Comparatively little is known about how direct control by general-public participants shifts perceptions across contexts. Boston Dynamics' Spot is the most visible exemplar among quadrupeds, with widely circulated demonstrations highlighting agility across diverse settings. Beyond company videos, public venues frequently feature a "drive Spot" experience that invites visitors to teleoperate the robot; while this has become a popular mechanism for showcasing the capabilities to broad audiences, there has been no formal evaluation of how such hands-on public demos affect perceptions of robots.

2.4 Contextual Variation in Robot Acceptance

Acceptance is strongly context-dependent. Consistent with survey and discourse work, robots are viewed more favorably in hazardous/physically demanding settings (manufacturing, disaster response) and more cautiously in care-intensive domains (home, hospital), where concerns about safety, autonomy, appropriateness, and loss of human contact are salient [15]. Recent large-scale evidence shows that acceptance of assistive robots can be inversely related to the vulnerability of the population being helped and moderated by personal factors such as religiosity [2]. For quadrupeds specifically, militarization/surveillance framings amplify unease [24], whereas assistive deployments are only beginning to be explored (e.g., eldercare awareness, navigation support) [11, 23]. These contextual effects motivate comparative designs that juxtapose favorably viewed domains (industrial, disaster) with socially sensitive ones (home, hospital) to test whether hands-on, agency-granting exposure can broaden acceptance beyond conventional roles.

In summary, prior research demonstrates that public perceptions of robots are shaped by demographic, the robot exposure design, and the contexts in which robots are deployed, yet most large-sample evidence comes from surveys or media analyses, and empirical studies rarely examine hands-on operation with publics. We address this gap with a walk-up, high-throughput study that pairs direct teleoperation of a quadruped with curated drive scenes spanning favorable (industrial, disaster) and socially sensitive (home, hospital) domains, and pre/post measures of comfort, perceived suitability, affect, and imagined roles. We next detail the study context, protocol, and measurement approach.

3 Methods

3.1 Study Context and Setting

Robot Lab: The study took place at the Robot Lab, a free public pop-up in a U.S. shopping mall, open for 10 weeks in summer 2025. The exhibit aimed to broaden public engagement with robotics and featured six stationary robots displayed with looping videos—Boston Dynamics' Atlas [7, 9], ANYbotics' ANYmal [1], and four industry prototypes—plus an interactive Boston Dynamics Spot (Figure 1) [10] that performed three short dance routines daily. Staff were on hand to answer questions and guide visits.

Figure 1: Boston Dynamics' Spot, the quadruped robot used in the "Drive-a-Spot" Experience (left & right), after a driver tilted its body (right).

Drive-a-Spot Experience: One side of the Robot Lab hosted a 14'×16' driving area with modular obstacles for guests to teleoperate Spot through core navigation tasks—walking and turning, maneuvering in tight vs. open spaces, crouching to pass under a barrier, and recovering height—using the standard controller [8]. Spot's built-in agility enabled it to step over low objects and avoid taller obstacles without user prompting (Figure 2).

Spot Control: Guests operated Spot from a desk outside the driving area using a custom, accessible controller built on the Xbox Adaptive Controller [31]. The interface condensed Spot's full control set to core, easy-to-learn commands: planar translation and rotation for navigation, height adjust (taller/shorter), posture commands (stand/sit), a tilt mode for body roll/pitch/yaw, and a reset to neutral pose (Figure 3). This simplification supported quick onboarding while retaining key capabilities for the driving tasks.

Participants: Over the 10-week exhibit, the Robot Lab logged 10,943 visitors (mean 171/day, weekends 249/day). Among guests who chose Drive-a-Spot, 840 consented to data collection. Eligibility required age 8+ and English fluency; participants (and guardians for minors) provided consent/assent under an IRB-approved protocol. Situated in a shopping mall in a major tourist city, the exhibit drew a demographically diverse audience. No compensation was provided.

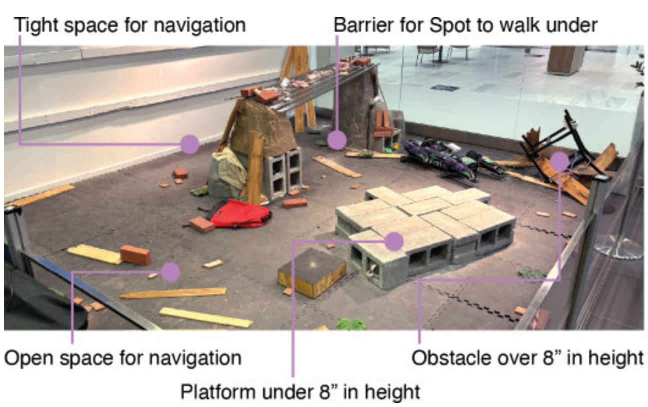

Figure 2: Drive-a-Spot area (Disaster scene) with shared navigation elements across scenes: open & tight passages, an under-barrier (requires crouch), low obstacles suitable for stepping over, and taller obstacles for autonomous avoidance.

3.2 Study Protocol

We administered pre- and post-surveys around a brief, hands-on driving interaction. Visitors either joined the Drive-a-Spot queue voluntarily or were invited by staff; upon queueing, they received a study overview and provided consent/assent.

Surveys ran on touchscreen tablets. Because the exhibit was open-plan, some guests observed others driving before their own pre-survey; thus, baseline responses reflect both prior expectations and in-situ exposure. Between surveys, participants engaged in free-play teleoperation of Spot with staff support (controller onboarding, Q&A) and safety oversight (emergency stop). Session durations varied with demand (0.23–26.0 min; M = 4.1, SD = 2.4, Median = 3.5), consistent with the high-throughput goals of the venue (more in Sec. 3.4).

A **Drive Scene (DS)** manipulation varied the driving area across Factory, Home, Hospital, and Outdoor/Disaster themes over 10 weeks (2–3 weeks each). Scenes were chosen to contrast favorably viewed (industrial, disaster) and socially sensitive (home, hospital)

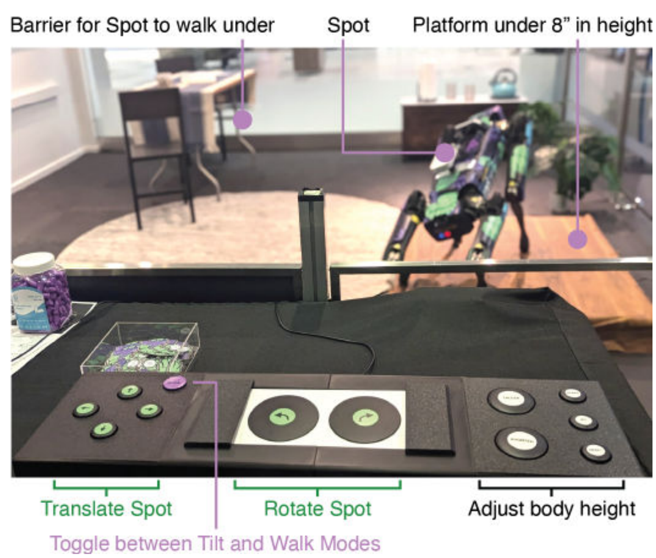

Figure 3: View from a participant's perspective. In the foreground, a controller that enabled participants to learn driving control quickly. In the background, the Home scene.

contexts, informed by prior public-attitude survey [15] and media portrayals of quadrupeds [24]. This rotation tested whether hands-on experience shifts perceptions across domains with differing baseline acceptance.

3.3 Data collection

Driving logs: Time-stamped controller events were recorded to compute engagement duration and drive/control patterns.

Pre- and post-surveys: Due to the Robot Lab's need for a high-throughput public exhibit, each survey had less than 8 items and targeted under 5 minutes, precluding the use of longer validated instruments (e.g., NARS [27], Godspeed [4]).

The survey items are summarized in Table 1. The pre-survey collected demographics, prior exposure (*"How often have you interacted with..."*), Spot familiarity, baseline comfort (*"How comfortable would you feel if you encountered a robot in ..."*) and suitability(*"How well do you think this robot would work in ..."*) across five **Rated Contexts (RC)**, and desired robot tasks (*"What task would you want it to do?"*). The post-survey repeated the comfort, suitability, and desired task items, and added ratings of the overall experience, affective responses (*"How do you feel after interacting with this robot?"*), recommendation intention, and open-ended reflections (*"Describe a memorable or surprising moment."*).

Together, the surveys provided concise quantitative measures of comfort, suitability, and desirability, alongside qualitative reflections on robot roles, experiences, and impressions.

Survey Item	Scale	Timing
Age (8-13,14-17,18-24,25-34,35-44, 45-54, 55-64, 65-74, 75+, decline)	Categorical	Pre
Gender (female, male, nonbinary, decline, other)	Categorical	Pre
Prior experience (robots, remote control toys, game controllers)	5-pt Likert (1:never—5:daily)	Pre
Spot familiarity (never, image, video, in-person, drove)	Multiple choice	Pre
Comfort encountering robot (Rated Contexts)	5-pt Likert (1:very uncomfortable – 5:very comfortable)	Pre/Post
Suitability (Rated Contexts)	5-pt Likert (1:very poor–5:very good)	Pre/Post
Desired robot tasks	Open-ended	Pre/Post
Rate experience	5-pt Likert (1:awful–5:brilliant)	Post
Affective response (excited, nervous, happy, scared, supportive)	Multiple choice	Post
Recommend to friend	Yes/No	Post
Memorable/surprising moment	Open-ended	Post

Table 1: Survey items. Rated Context (RC) = {Factory, Home, Hospital, Office, Outdoor/Disaster}.

3.4 Design Constraints in a Public Venue

The study ran under the Robot Lab's exempt IRB (later amended to include minors), which—together with the walk-up, high-throughput setting—limited data collection. To handle up to 370 visitors/day, participation (consent, surveys, driving) was capped at 15 minutes, necessitating brief surveys and short interactions, and precluding video/audio recording, and interviews. Session length was variable and adjusted to queue size (longer with no line, shorter under demand). These constraints reflect the trade-off of public-facing HRI: broader, more diverse reach but less intensive instrumentation. We return to these methodological constraints in the Discussion when interpreting the findings.

3.5 Analysis Methods

Comfort and suitability analysis: We tested pre–post changes with Wilcoxon signed-rank tests for ordinal Likert data. To address multiple comparisons across rated contexts (RC)/drive scenes (DS), we applied False Discovery Rate (FDR) correction to balance Type I error control with statistical power. Compared to conservative family-wise error corrections (e.g., Bonferroni), FDR provides a more appropriate safeguard when testing several related hypotheses, reducing the chance of false positives while still detecting meaningful effects [5, 14].

Subgroup analysis: We examined differences by age, gender, and Spot familiarity. Age groups followed U.S. Census-style bands adapted to our inclusion criteria: Children (8–17), Young Adults (18–44), Middle Aged (45–64), Older Adults (65+) [6]. Gender analyses focused on male/female due to small sample size in other categories. Familiarity levels: Driven, In-person, Indirect (images/videos), None. For each factor and outcome (comfort, suitability), we used Kruskal–Wallis tests with FDR correction; significant results were followed by Mann–Whitney U pairwise tests (FDR-corrected).

Qualitative thematic analysis: Two researchers built an inductive codebook for open-ended responses, iteratively refining definitions and inclusion/exclusion criteria. Two independent coders labeled the full dataset. Because responses could include multiple themes, we assessed inter-coder reliability using both Jaccard similarity (per-response overlap) [26] and Cohen's Kappa (overall label agreement) [22]. Discrepancies were resolved through discussion.

4 Results

For clarification, **Drive Scene (DS)** refers to the interaction condition in which participants drove Spot (levels: Factory, Home, Hospital, Outdoor/Disaster), and **Rated Context (RC)** refers to the survey prompt context in which participants judged comfort/suitability (levels: Factory, Home, Hospital, Office, Outdoor/Disaster).

4.1 Participants

Of 840 Drive-a-Spot guests who completed the pre-survey, 87 were excluded due to incomplete consent (n=14), missing survey (n=61), or other protocol issues (n=12). The remaining 753 (age 8+; median age band 18–24; gender: 56.7% male, 40.4% female, 1.6% nonbinary/other) completed both surveys and were analyzed (Figure 4). The sample skewed toward 18–34, with substantial child/adolescent participation. DS assignment: Factory 124, Home 253, Hospital 189, Outdoor/Disaster 187.

Prior technology experience (1=Never, 5=Daily) was lowest for robots ($\mu \pm std$ = 2.19±1.12), more balanced for remote control toys (2.79±0.99) and game controllers (2.79±0.99). Most had seen Spot via images (68%) or videos (66%) reflecting its popularity in media,

; 44.5% had seen it in person, 4.5% had driven it before (primarily robotics practitioners), and 8% reported no exposure.

Overall, the study reached a diverse and heterogeneous participant pool, spanning children to older adults, multiple genders, and varying levels of familiarity with robots and technology. This diversity highlights the advantages of a public-facing, high-throughput recruitment method compared to typical lab-based HRI studies.

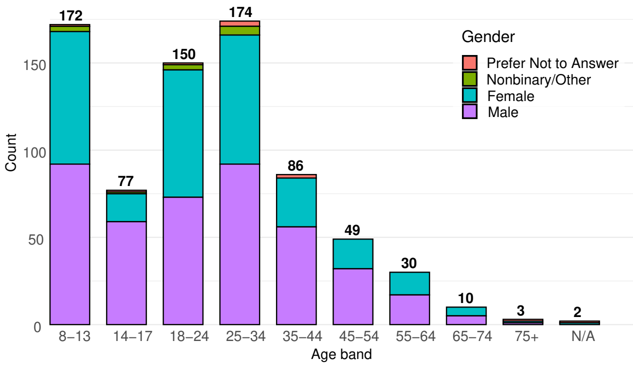

Figure 4: Participant age distribution by gender. Total appears above each bar.

4.2 Comfort Encountering Robots

Overall effect: Across all Rated Contexts (RCs), comfort increased from pre to post (Wilcoxon, all $p < .001$; Figure 5). Effect sizes ranged from small to moderate, with the largest improvement in the RC–Outdoor/Disaster ($\mu_\Delta \pm std_\Delta = 0.46 \pm 1.07$; $p < .001$; effect size $r = .434$, moderate), followed by smaller but reliable effects in the RC-Hospital (0.23±1.15; $p < .001$; $r = .218$), RC-Home (0.21±1.05; $p < .001$; $r = .209$), RC-Factory (0.21±1.01; $p < .001$; $r = .239$), and RC-Office (0.19±1.02; $p < .001$; $r = .194$).

Baseline comfort levels: Pre-survey comfort was highest for RC–Factory ($\mu \pm std = 4.03 \pm 1.02$) versus RC–Office (3.66±0.98), RC-Outdoor/Disaster (3.42±1.19), RC-Hospital (3.39±1.14), and RC-Home (3.31±1.13). This relative advantage for factory persisted in the post-survey. Also, observing others while queueing did not affect pre-survey comfort (Kruskal–Wallis, FDR-corrected).

Effect of interaction context: As shown in Figure 6, comfort generally significantly rose in the same context participants drove: DS–Home, DS–Hospital, and DS–Outdoor/Disaster each showed significant gains in their matching RCs (Fig. 6). DS–Factory was the exception: comfort with robots in RC-Factory did not significantly increase after interaction in the factory DS, likely reflecting already high baseline comfort in industrial contexts.

Age/Gender Effect: Pre-survey, children ($\mu \pm std = 3.82 \pm 1.04$) reported significantly lower comfort in RC–Factory than young (4.12±0.99; $p < .0001$; $U = 41738.0$) and middle-aged adults (4.21±1.05; $p < .001$; $U = 7341.0$). Children (3.58±1.02) also reported significantly lower comfort in RC–Office than middle-aged (3.94±0.85; $p < .01$, $U = 7501.0$). After the interaction, the factory age gap persisted, where children (4.01±0.98) rated lower comfort than young (4.33±0.88; $p < .0001$; $U = 40738.0$) and middle-aged adults (4.41±0.94; $p < .001$; $U = 7238.0$). The office gap did not persist. Children showed higher post comfort than young adults in RC–Home (C: 3.78±1.11; YA: 3.33±1.27; $p < .0001$; $U = 60790.5$). Although children's gains (post-pre) were numerically larger in several RCs, the differences in Δ were not significant after FDR correction.

By gender, men reported significantly greater pre comfort levels than women in all RC; RC-Factory (F: 3.88±1.09; M: 4.13±0.96; $p < .01$; $U = 71813.0$), RC-Home (F: 3.08±1.17; M: 3.48±1.08; $p < .0001$; $U = 75595.5$), RC-Hospital (F: 3.25±1.16; M: 3.49±1.11; $p < .01$; $U = 71309.5$), RC-Office (F: 3.56±0.97; M: 3.74±0.98; $p < .01$; $U = 70386.0$), and RC-Outdoor/Disaster (F: 3.20±1.21; M: 3.58±1.14; $p < .0001$; $U = 75092.5$). In post, gaps closed for RC–Factory and RC–Office but persisted (smaller) for RC–Home (F: 3.34±1.26; M: 3.68±1.16; $p < .001$; $U = 74315.0$), RC–Hospital (F: 3.51±1.22; M: 3.72±1.19; $p < .05$; $U = 70281.0$), and RC–Outdoor/Disaster (F: 3.71±1.25; M: 4.03±1.09; $p < .001$; $U = 73212.0$).

Spot familiarity Effect: Pre-survey comfort was higher among those familiar with Spot. For RC-Factory, the Never group's pre comfort (3.52±1.29) was significantly lower than Driven ($\mu \pm std = 4.33 \pm 0.85$; $p < .01$, $U = 1310.0$), In-person (4.03±1.03; $p < .01$, $U = 10914.0$), and Indirect (4.08±0.95; $p < .001$, $U = 12612.0$) groups. Also for RC-Office, Driven (4.24±0.75) had significantly higher pre comfort than In-person (3.67±1.03; $p < .01$, $U = 6575.0$), Indirect (3.62±0.94; $p < .001$, $U = 7803.5$), and Never (3.53±1.05; $p < .001$, $U = 1321.5$). Post, familiarity effects disappeared. Less-familiar groups showed larger numeric gains overall, but Δ differences were non-significant after correction with FDR.

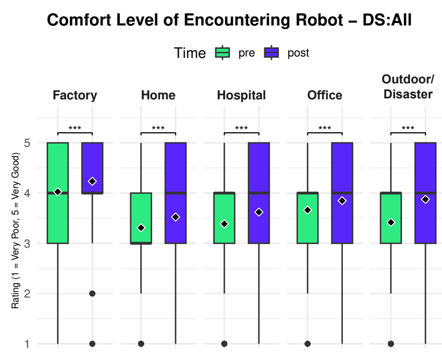

Figure 5: Comfort level of encountering a robot in each of the five rated contexts (RC) across all drive scenes (DS).

4.3 Perceived Suitability of Robots

Overall effect: Across all participants, perceived suitability increased from pre to post across all five Rated Contexts (RCs) (Wilcoxon, all $p < .05$; Figure 7). Effect sizes ranged from small to moderate, largest in RC-Home ($\mu_\Delta \pm std_\Delta = 0.33 \pm 0.95$; $p < .001$; effect size $r = .336$, moderate), followed by RC-Office (0.29±0.93; $p < .001$; $r = .297$, small), RC-Hospital (0.22±1.03; $p < .001$; $r = .251$, small), RC-Factory (0.15±0.95; $p < .001$; $r = .181$, small), and RC-Outdoor/Disaster (0.076±0.89; $p < .05$; $r = .093$, small).

Baseline patterns: Before interaction, participants rated robots as most suitable for RC–Outdoor/Disaster ($\mu \pm std = 4.20 \pm 0.99$) and and RC–Factory (4.02±1.03), consistent with prior survey research indicating public acceptance of robots in hazardous or industrial

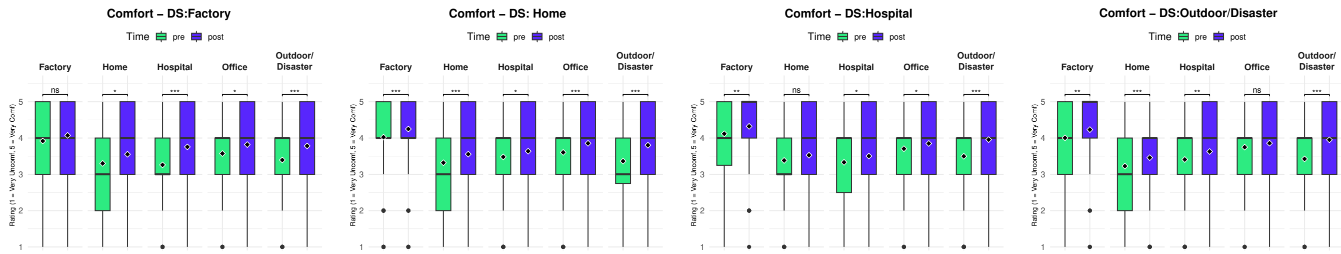

Figure 6: Comfort level of encountering a robot in each of the five rated contexts (RC), per drive scene (DS).

settings. By contrast, suitability ratings were lower for RC-Office (3.41±1.02), RC-Hospital (3.32±1.11), and RC-Home (3.23±1.03) reflecting greater ambivalence about robots in social or care-related environments. Pre-exposure (observing the DS while queueing) did not affect pre-survey suitability ratings by RC (Kruskal–Wallis, FDR-corrected).

Effect of interaction context: As shown in Figure 8, suitability generally increased post across RCs regardless of DS. The main exceptions were the RC-Factory and RC-Outdoor/Disaster domains, where baseline suitability was already high, resulting in smaller or non-significant changes within those same contexts.

Age/Gender Effect: Pre-survey, children rated RC–Factory significantly lower ($\mu \pm std$ = 3.79±1.07) than young (4.10±1.01; p < .0001; U = 42198.0) and middle-aged adults (4.31±0.89; p < .0001; U = 7006.0). Conversely, children rated RC–Home higher (3.57±0.89) than young (3.06±1.08; p < .0001; U = 65005.0) and middle-aged adults (3.12±0.93; p < .001; U = 12211.0). For RC–Outdoor/Disaster, the middle-aged (4.59±0.67) exceeded children (4.09±1.04; p < .0001; U = 6993.0) and young adults' (4.18±1.00; p < .001; U = 12348.0). These age effects persisted post. For RC-Factory, children (3.92±1.05) reported significantly lower than the young (4.27±0.90; p < .0001; U = 40881.5) and middle-aged adults (4.42±0.79; p < .0001; U =7104.0). For RC-Home, children's rating (3.89±1.02) was greater than young (3.35±1.18; p < .0001; U = 63682.5). For RC-Outdoor/Disaster, the middle-aged group (4.63±0.74) reported higher ratings than children (4.16±1.02; p < .0001; U = 7025.0) and young adults (4.26±1.02; p < .001; U = 12721.5).

By gender (male/female), men reported significantly higher pre suitability in RC–Home (F: 3.14±0.99; M: 3.32±1.03; p < .05; U = 69994.0) and RC-Outdoor/Disaster (F: 4.07±0.99; M: 4.30±0.99; p < .0001; U = 73028.0). Post, these differences persisted for RC–Home (F: 3.43±1.14; M: 3.69±1.12; p < .001; U = 72754.4) and RC-Outdoor/Disaster (F: 4.18±1.03; M: 4.35±0.97; p < .01; U = 69972.0), with additional gap emerged in RC-Office (F: 3.61±1.04; M: 3.77±1.11; p < .01; U = 70429.0).

Spot familiarity Effect: Mann–Whitney U tests by RC (FDR-corrected) showed no significant effects of Spot familiarity on suitability ratings in either the pre- or post-survey.

4.4 Sentiment and Recommendation

Emotional responses: Participants' immediate reactions after driving Spot were predominantly positive. The most common response was excitement (73.9%), followed by happiness (49.5%) and feeling supportive (29.4%). Negative emotions were much less frequent, with only 12.2% reporting nervousness and 4.8% reporting fear. These findings indicate that hands-on interaction with Spot elicited strong enthusiasm and enjoyment, with relatively little apprehension.

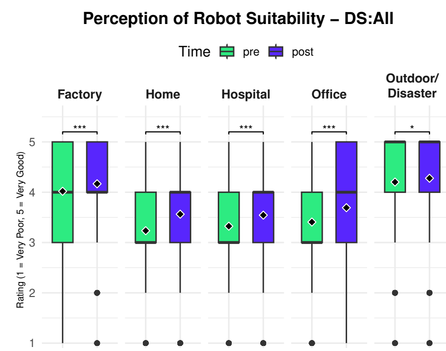

Figure 7: Perceived robot suitability in each of the five rated contexts (RC) across all drive scenes (DS).

Overall experience: The vast majority selected the top two categories (M = 4.46, SD = 0.68). Over 55% of participants rated the experience as "brilliant" (5), and 37% rated it as "good" (4), with very few lower ratings.

Likelihood to recommend: Responses were overwhelmingly positive (M = 4.49, SD = 0.76). Over 62% of participants rated "very likely" (5), and 27% rated it as "likely" (4) with the distribution heavily skewed toward the high end of the scale.

Taken together, these results show that the Drive-a-Spot experience generated highly positive affective responses, strong overall satisfaction, and robust word-of-mouth potential. While a small minority experienced nervousness or fear, the overall sentiment was enthusiastic, underscoring the effectiveness of brief, hands-on demonstrations in fostering public excitement around advanced robots.

4.5 Desired Robot Tasks

Coder reliability: Inter-coder agreement was substantial: pre Jaccard similarity M = 0.78 (SD = 0.33, Mdn = 1.00) and Cohen's Kappa κ = .76; post Jaccard M = 0.82 (SD = 0.32, Mdn = 1.00) and Cohen's Kappa κ = .82

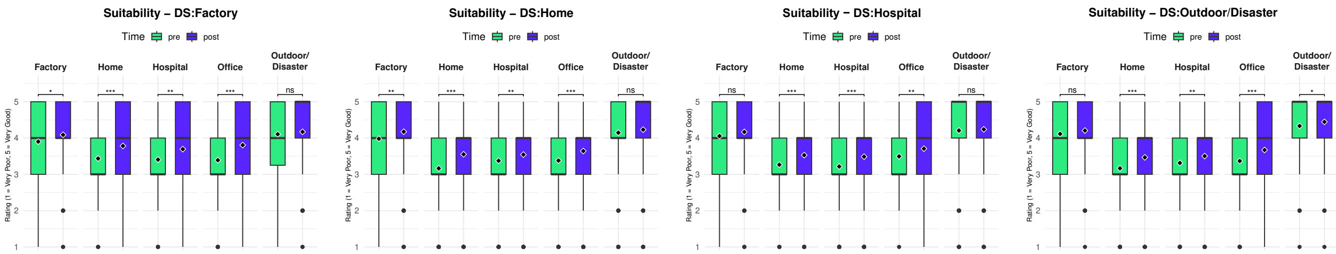

Figure 8: Perceived robot suitability in each of the five rated contexts (RC), per drive scene (DS).

Pre (N = 656): Dominant themes were domestic assistance—cleaning 36.7%, mobility support and shopping 19.2%, cooking 15.1%—alongside heavy labor 12.7%, hazardous tasks 8.7%, and outdoor/environmental 8.1%; entertainment and professional assistance were less frequent (7.5% each).

Pre (N = 665): Domestic assistance remained prominent (cleaning 38.2%; mobility/shopping 22.3%), with new emphases on entertainment/play 19.4% and companionship 5.0%. Mentions of cooking 9.0%, heavy labor 7.4%, and hazardous tasks 7.7% declined.

Summary: Thematic analysis indicates that while participants consistently emphasized domestic assistance across both surveys, hands-on interaction diversified their imagined robot roles. Post-experience responses shifted away from purely utilitarian or hazardous functions toward entertainment, play, and companionship, highlighting how encounters with Spot encouraged participants to imagine robots as both helpers and partners in everyday life.

4.6 Memorable and Surprising Moments

Participants (N = 158) described a wide range of memorable or surprising moments while interacting with the robot. Responses were thematically coded into eight categories, capturing both technical capabilities and social-emotional impressions.

The most frequently reported themes were Locomotion and Terrain Adaptation (22%) and Tilt and Expressive Movements (22%), with participants often remarking on the robot's ability to navigate steps, uneven terrain, and confined spaces, as well as its dance-like or doglike tilt behaviors. Obstacle Avoidance and Sensing (17%) was also commonly cited, reflecting admiration for the robot's use of sensors to avoid collisions, maintain balance, and adapt to environmental changes. A smaller proportion of responses highlighted the Ease of Control (9%), emphasizing how intuitive and responsive the driving interface felt.

Less frequently mentioned but notable themes included Sitting, Standing, and Postures (5%) and Playful/Entertaining Interactions (5%), such as making the robot dance or perform gestures. A minority of responses reflected Emotional Reactions and Anthropomorphism (3%), in which participants described the robot as "doglike," "silly," or expressed concern about "hurting" it. Finally, some participants noted its Size, Durability, and Physical Presence (2%), expressing surprise at its large, powerful form.

5 Discussion

5.1 Participants report increased comfort after hands-on experience.

Participants showed significantly higher comfort encountering robots after driving Spot across all Rated Contexts (RCs), consistent with work indicating that hands-on interaction reduces apprehension and normalizes robots in everyday settings [3]. Gains and effect size were largest in RC–Outdoor/Disaster, where baseline comfort was low. This finding suggests that hands-on experience may be particularly effective in domains where public comfort is initially low but the perceived utility of robots is high, such as humanitarian or emergency scenarios.

Baseline patterns clarify these shifts. Comfort was highest for RC–Factory, echoing survey evidence that robots are more accepted in industrial/hazardous roles [15], and lower for RC–Home/Hospital, aligning with caution in care-intensive contexts [16]. Notably, comfort was also low for RC–Outdoor/Disaster despite high perceived suitability—highlighting a gap between task appropriateness and personal ease. Media framings of quadrupeds as military/policing tools may contribute to this discomfort even when their search-and-rescue value is acknowledged [24].

The Drive Scene (DS) mattered: comfort rose most when participants drove in DS–Home, DS–Hospital, and DS–Outdoor/Disaster—the lowest-comfort domains—whereas DS–Factory showed no increase in RC–Factory, consistent with a ceiling effect. Overall, short, public-facing, hands-on encounters can meaningfully raise comfort in socially sensitive or risk-colored domains, helping extend acceptance beyond traditional industrial roles.

5.2 Participants report increased perceived suitability after hands-on experience.

Participants reported significantly higher perceived suitability after driving Spot across all Rated Contexts (RCs). This supports the idea that hands-on experience expands perceptions of what roles robots are appropriate for, extending beyond traditional industrial or hazardous contexts. The largest relative gain was observed in RC-Home, followed by RC-Office and RC-Hospital, while smaller but still significant increases appeared in RC-Factory and RC-Outdoor/Disaster, where baseline suitability was already high.

Baseline ratings confirm this interpretation: participants initially judged robots as most suitable for factory and outdoor/disaster roles, consistent with public narratives that emphasize robots' value in industrial production and emergency response [15, 24]. By contrast, home, office, and hospital domains were rated lower, reflecting ambivalence about robots in socially sensitive environments. The fact that these domains showed the largest post-interaction improvements suggests that hands-on experiences are especially effective in legitimizing robots in less conventional or more socially embedded roles.

Notably, suitability perceptions improved across multiple RCs regardless of which DS participants encountered, indicating that the impact of embodied interaction generalized beyond the immediate context of use. For example, participants driving Spot in the DS-Home not only rated robots as more suitable for home roles, but also for hospital and office roles. This pattern suggests that brief, hands-on encounters can broaden the imagined role space of robots, encouraging the public to see them as flexible technologies adaptable to diverse contexts.

5.3 Demographic and Exposure Moderates Robot Perception

Age had several impact on how people perceive robots in different rated contexts (RCs) in the pre-surveys. Children, particularly, reported less comfort and lower suitability for robots in RC-Factory than older age groups. This is possibly due to children are pre-workforce with little to no experience in contexts like factory or office. Further, **age effect is dependent on the contexts.** For instance, middle-aged group find robots more suitable for RC-Outdoor/Disaster than for RC-Home, compared to younger groups. This shows each group's distinct perception of robot's roles in the society, with younger people more open to robot's role in homes. Interestingly, most of the age effect on comfort and all on the suitability persisted in the post-experiment survey. The main change happened to the children group after exposure to unfamiliar scenarios to them. For instance, children had much larger growth in their comfort level to robots in the office compared to older groups. While it is promising that the exposure might benefit the inexperienced group more, those findings suggest that age-related perception differences are difficult to change with a simple hands-on exposure, as it might be rooted in an individual's life experiences.

Our gender analysis agrees with prior findings where men are more positive towards robots [2, 12, 16]. A strong gender effect was found on comfort levels across all scenarios, while the differences on suitability perception was only in home or outdoor/disastrous settings. Although the hands-on experience improved both gender's perception scores and possibly reduce the gender differences in post-survey, no significant differences was observed on the gender's effect on the changes in perception. This suggests that the perception change from the hands-on exposure is not biased towards a participant's gender.

Prior work reported positive relationship between people's perception of robot and their prior robot experiences [2]. We found similar phenomenon across two settings, factory and office. For the factory setting, we found any prior experience with the robot Spot – as long as the participant have seen or heard of it – makes a significant difference on their perception. This is possibly due to Spot's public image and media coverage on its industrial applications [24]. More interestingly, we found that for certain contexts (e.g., office), prior experience in driving the Spot leads to a significant difference with all other types of exposure. It is worth-noting that people who haven't driven Spot before acquired more growth from this exposure than people who has. These findings suggest that not only prior exposure to robots, but also the exposure method influences individual's perception of the robot.

5.4 Participants report diversified imagined robot roles after hands-on experience.

Brief, agency-granting tele-operation broadened imagined robot roles: while domestic assistance remained central, entertainment/play and companionship newly emerged, and references to heavy or hazardous work modestly receded. Likely drivers are the demo's legible agility (expressive tilt/posture, obstacle handling) and early control success, which make social, low-stakes uses feel plausible. Designwise, public demos that pair quick mastery loops with interpretable motion can catalyze richer role ideation without inflating risk perceptions.

5.5 Characteristics and Trade–offs of a Public, High–Throughput Design

Our study leverages the distinctive strengths of a *walk-up, high–throughput* setting: access to large, heterogeneous publics; immediate, agency–granting interaction (tele-operation); and exposure to a quadruped robot outside the lab. This design yielded a broad sample and robust pre–post inference at scale, while showcasing how brief, hands-on encounters can shift perceptions across diverse Rated Contexts (RCs) and Drive Scenes (DSs).

These benefits come with principled trade–offs. Participation was intentionally capped (≈15 minutes, including consent, surveys, and driving) to preserve throughput, which necessitated concise self-report instruments and short, variably timed interactions (Sec. 3.3). Richer instrumentation (e.g., audiovisual recording, interviews, participant media) was not pursued due to consent and bystander considerations in an open venue. Ambient features of the shopping mall (lighting, noise, passers-by) were present despite DS staging (Sec. 3.2) and may have shaped experience and self-reports—an inherent aspect of research situated in public spaces.

On balance, this approach complements controlled lab studies: it prioritizes *reach, diversity, ecological exposure, and immediacy of agency* over depth of instrumentation. We view the method and its results as evidence that scalable, public-facing HRI can produce meaningful shifts in comfort and perceived suitability, and as a template for future work that pairs high-throughput engagement with targeted enhancements (e.g., validated scene checks across all DSs, longitudinal follow-ups, or lightweight behavioral measures) to extend analytic depth while retaining public-scale benefits.

6 Conclusion

In a large, walk-up public study, brief hands-on teleoperation of a quadruped (Spot) produced reliable shifts in perception. Comfort increased across all Rated Contexts (RCs)—largest in RC–Outdoor/Disaster—and perceived suitability rose most in RC–Home/Office/Hospital, where baselines were lower. Effects generalized beyond the experienced Drive Scene (DS), indicating that short, agency-granting encounters broaden where people see robots as appropriate. Post-interaction affect was strongly positive, and open-ended responses showed role diversification: domestic help remained central while entertainment/play and companionship emerged. Memorable moments clustered around legible agility (locomotion, expressive tilt) and competent autonomy (obstacle sensing).

These shifts were moderated by age, gender, and familiarity—children started lower in industrial RCs but sometimes gained more; men reported higher pre-comfort than women with some gaps narrowing; prior exposure elevated some pre-ratings but hands-on control reduced familiarity differences. Our public, high-throughput format traded depth for reach, diversity, and ecological exposure, offering a practical template for rigorous, population-scale HRI.

References

[1] ANYbotics. Meet ANYmal, your new inspector, 2025. https://www.anybotics.com/robotics/anymal/.

[2] Laura Aymerich-Franch and Emilia Gómez. Public perception of socially assistive robots for healthcare in the eu: A large-scale survey. *Computers in Human Behavior Reports*, 15:100465, 2024.

[3] Uba Backonja, Amanda K Hall, Ian Painter, Laura Kneale, Amanda Lazar, Maya Cakmak, Hilaire J Thompson, and George Demiris. Comfort and attitudes towards robots among young, middle-aged, and older adults: a cross-sectional study. *Journal of Nursing Scholarship*, 50(6):623–633, 2018.

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[5] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[6] Laura Blakeslee, Zoe Caplan, Julie A Meyer, Megan A Rabe, and Andrew W Roberts. Age and sex composition: 2020. *Washington, DC*, pages C2020BR–06, 2023.

[7] Boston Dynamics. An Electric New Era for Atlas, 2024. https://bostondynamics.com/blog/electric-new-era-for-atlas/.

[8] Boston Dynamics. Spot® Product Specifications, May 2024. https://bostondynamics.com/wp-content/uploads/2020/10/spot-specifications.pdf.

[9] Boston Dynamics. Atlas® and beyond: the world's most dynamic robots, 2025.

[10] Boston Dynamics. Spot® - The Agile Mobile Robot, 2025. https://bostondynamics.com/products/spot/.

[11] Shaojun Cai, Ashwin Ram, Zhengtai Gou, Mohd Alqama Wasim Shaikh, Yu-An Chen, Yingjia Wan, Kotaro Hara, Shengdong Zhao, and David Hsu. Navigating real-world challenges: A quadruped robot guiding system for visually impaired people in diverse environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024.

[12] Marco Carradore. People's attitudes towards the use of robots in the social services: A multilevel analysis using eurobarometer data. *International Journal of Social Robotics*, 14(3):845–858, 2022.

[13] Yao-Cheng Chan. A field observation of incidental human-robot encounters in public. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1265–1268. IEEE, 2025.

[14] Shi-Yi Chen, Zhe Feng, and Xiaolian Yi. A general introduction to adjustment for multiple comparisons. *Journal of thoracic disease*, 9(6):1725, 2017.

[15] Special Eurobarometer. Public attitudes towards robots. *European Commission*, 1:1–1, 2012.

[16] Timo Gnambs and Markus Appel. Are robots becoming unpopular? changes in attitudes towards autonomous robotic systems in europe. *Computers in human behavior*, 93:53–61, 2019.

[17] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527, 2011.

[18] Nanami Hashimoto, Emma Hagens, Arkady Zgonnikov, and Maria Luce Lupetti. Safe spot: perceived safety of dominant and submissive appearances of quadruped robots in human-robot interactions. *arXiv preprint arXiv:2403.05400*, 2024.

[19] Elliott Hauser, Yao-Cheng Chan, Parth Chonkar, Geethika Hemkumar, Huihai Wang, Daksh Dua, Shikhar Gupta, Efren Mendoza Enriquez, Tiffany Kao, Justin Hart, et al. " what's that robot doing here?": Perceptions of incidental encounters with autonomous quadruped robots. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, pages 1–15, 2023.

[20] Alireza Javaheri, Navid Moghadamnejad, Hamidreza Keshavarz, Ehsan Javaheri, Chelsea Dobbins, Elaheh Momeni-Ortner, and Reza Rawassizadeh. Public vs media opinion on robots and their evolution over recent years. *CCF Transactions on Pervasive Computing and Interaction*, 2(3):189–205, 2020.

[21] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. A communication robot in a shopping mall. *IEEE Transactions on Robotics*, 26(5):897–913, 2010.

[22] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[23] Tyler Morris, Mengjun Wang, Yan Li, Songyan Liu, Shuai Li, and Xiaopeng Zhao. Awareness and acceptance of emerging technology and quadruped robots in dementia care: a survey study. In *Proceedings of the AAAI Symposium Series*, volume 2, pages 187–191, 2023.

[24] Jeremy Moses and Geoffrey Ford. See spot save lives: fear, humanitarianism, and war in the development of robot quadrupeds. *Digital War*, 2(1-3):64, 2021.

[25] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, 12(6):1179–1201, 2020.

[26] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. Using of jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 380–384, 2013.

[27] Tatsuya Nomura, Takayuki Kanda, and Tomohiro Suzuki. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society*, 20(2):138–150, 2006.

[28] Stefania Operto. Evaluating public opinion towards robots: a mixed-method approach. *Paladyn, Journal of Behavioral Robotics*, 10(1):286–297, 2019.

[29] David A Robb, Muneeb I Ahmad, Carlo Tiseo, Simona Aracri, Alistair C McConnell, Vincent Page, Christian Dondrup, Francisco J Chiyah Garcia, Hai-Nguyen Nguyen, Èric Pairet, et al. Robots in the danger zone: Exploring public perception through engagement. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 93–102, 2020.

[30] Nathan Tsoi, Rachel Sterneck, Xuan Zhao, and Marynel Vázquez. Influence of simulation and interactivity on human perceptions of a robot during navigation tasks. *ACM Transactions on Human-Robot Interaction*, 13(4):1–19, 2024.

[31] Xbox. Xbox Adaptive Controller: Game your way, 2025. https://www.xbox.com/en-US/accessories/controllers/xbox-adaptive-controller.

[32] Renchi Zhang, Jesse van der Linden, Dimitra Dodou, Harleigh Seyffert, Yke Bauke Eisma, and Joost de Winter. Walk along: An experiment on controlling the mobile robot 'spot'with voice and gestures. *ACM Transactions on Human-Robot Interaction*, 2025.