Designing for Transparency in Human–Robot Interaction: A Dashboard and Custom Hardware for Mechanistic Interpretability

Phoebe Pinney*
Cambridge Consultants
Cambridge, UK
phoebe.pinney@cambridgeconsultants.com

Tom Shirley*
Cambridge Consultants
Cambridge, UK
tom.shirley@cambridgeconsultants.com

Emma Hughson
Cambridge Consultants
Cambridge, UK
emma.hughson@cambridgeconsultants.com

Ali Shafti
Cambridge Consultants
Cambridge, UK
ali.shafti@cambridgeconsultants.com

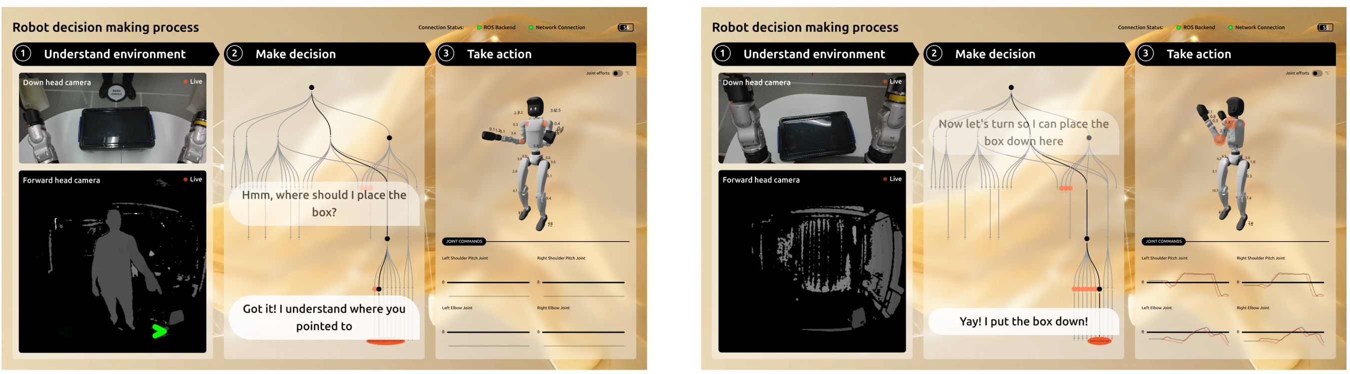

Figure 1: Screenshots of the dashboard whilst demo is running. First is during pointing detection, second is during AI pickup.

Abstract

As embodied AI systems transition from deterministic automation to fluid decision-making driven by systems such as Vision-Language-Action (VLA) models, the inherent opacity of these "black box" systems creates significant barriers to safe and effective human-robot interaction. This paper argues that transparency should not be a post-hoc addition, but a foundational design constraint integrated into the robot's physical and digital architecture. We present a humanoid demonstrator built on the Unitree G1 platform that embodies this holistic design philosophy. Our approach combines custom industrial hardware, including a specialised head unit with a Face User Interface (Face UI) for social signaling, with a real-time digital dashboard that translates complex AI reasoning into natural language and visualises the perception-to-actuation pipeline. To evaluate the efficacy of this multi-layered transparency design, we conducted a user study (N=10) comparing the fully transparent system against a baseline lacking the social interface and dashboard. Results indicate that our architecture improved transparency and understandability of the robot during interaction. These findings suggest that tailoring transparency mechanisms to specific user contexts is critical for the successful deployment of autonomous systems in shared human spaces.

Keywords

Human-robot interaction, XAI, mechanistic interpretability, humanoid robotics, transparent design

1 Introduction

The deployment of embodied AI into shared spaces necessitates a paradigm shift in how we architect autonomous systems. It is no longer sufficient for a robot to simply execute a task successfully [4]; for it to be safely operating next to and with humans, its actions must be legible and its internal reasoning accessible [10]. As systems evolve from deterministic automation to more fluid decision-making driven by Vision-Language-Action (VLA) models [17], the opacity of the "black box" becomes a primary barrier to interaction [3]. We posit that the solution lies not in post-hoc explanations, but in a holistic design philosophy where the mechanisms of transparency, from physical form to social signaling, and digital interpretability, are architected alongside the core autonomy stack [19].

This paper presents a proof of concept for this philosophy: a humanoid demonstrator built upon the Unitree G1 platform, engineered to maximize understandable interaction. Our methodology prioritises human-centric design as the foundational constraint.

*These authors contributed equally to this work

This began with the industrial design of a custom head and backpack unit, shaping the robot's physical affordances to be approachable rather than uncanny. Head hardware houses a dual-camera setup, downward-facing for manipulation and forward-facing for human engagement, integrated with a Face User Interface that uses eye animations for social cues, and back of the head LED lighting for robot state indication.

To bridge the gap between high-dimensional AI processing and human intuition, we developed a dashboard that runs in parallel to the robot (as seen in Figure 1). During a collaborative task, where the robot interprets human pointing gestures to move objects, this dashboard visualises the system's pipeline from perception to reasoning to actuation. Crucially, it translates the complex decision-making process into high-level natural language, offering a real-time narrative of why a decision was made and how it will be carried out. This multi-layered approach ensures that the robot communicates intent through both implicit behavioural cues and explicit explanatory feedback.

We emphasise that "understandability" is not a static property but is deeply context dependent [7]. The system presented here is optimised for a demonstration environment, where the user intent is driven by curiosity and engagement rather than industrial throughput[6, 18]. Consequently, our transparency features are designed to invite exploration and demystify the technology for a lay audience. This underscores a central theme of our work: the recognition that transparency mechanisms must be tailored to the specific persona of the user, whether a curious visitor, a logistics operator, or a domestic consumer.

To evaluate the efficacy of this integrated design, we conducted a user study (N=10) comparing the fully transparent system against a baseline condition lacking the Face UI and dashboard. By analysing participant experience across these conditions, we provide empirical evidence on how multi-modal explainability influences trust and the perceived intelligence of the system.

This paper is structured as follows: we first outline our design-led methodology and the hardware/software architecture of the transparent G1 platform. We then detail the implementation of the reasoning and visualisation pipeline, followed by the results of our user study. Finally, we discuss the implications of context-aware transparency for the future of human-robot interaction.

2 Methodologies

2.1 The HRI demonstrator

The transparency mechanisms described in this paper were developed for a box pick-and-place demonstrator built on the Unitree G1 platform. Users interact with the robot by pointing to a location on a U-shaped table; the robot interprets the gesture, picks up a box placed in front of it, and places it at the indicated position (as shown in Figure 2). The core manipulation capability is driven by a Vision-Language-Action (VLA) model based on NVIDIA's GR00T [14], fine-tuned on teleoperation data collected with the G1 for this task. A behaviour tree orchestrator manages higher-level execution, coordinating the VLA with waist rotation and placement once the pick is complete. Human pose estimation, performed using Microsoft's Azure Kinect SDK [12] with depth data from the robot's head-mounted Orbbec Femto Bolt camera, derives pointing vectors from the user's arm and maps them into the 3D coordinate space of the environment. It is the internals of this pipeline that the transparency mechanisms described below are designed to surface.

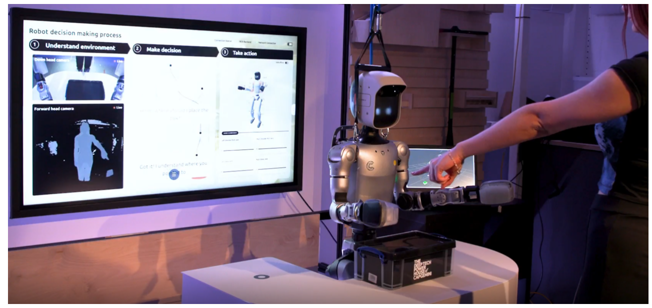

Figure 2: Our HRI demonstrator, a user points to where they would like the box moved to. The robot picks the box up and places it at the pointed destination.

2.2 Robot Hardware Design

Physical appearance is the first barrier to entry. It is one of the first and most noticeable qualities that users perceive when approaching the robot. In these initial moments it is vital that the robot appears friendly and somewhat familiar, setting the right expectation for follow-on interactions. Ultimately, if the first impression is good and future interactions are coherent and well aligned with user's expectation, there's a good chance the robot and user can form a productive and trusting relationship.

Figure 3: Example product usability features: access panels on the head and backpack including additional compute and networking hardware.

2.2.1 Robot Body Design

Robot body design considerations:

(1) **User Centric Perspective.** We start by understanding the user's role in the interaction, this forms the basis of our requirements. In this case we wanted the robot to take direction from a user to perform simple, manual tasks in a warehouse-like environment. To achieve intuitive human-robot interaction we applied the following principles:

- **Friendly appearance.** Humanoid robots are still perceived as intimidating by many [11], to support effective human-robot interaction we first need to overcome user's fear and distrust. To counter this image, we have created a soft, friendly form with warm colours and soft textured finishes. The addition of a tethered kill switch adds to the feeling of control and safety when interacting with the robot [2].
- **Intuitive human interaction with a coherent personality.** Humans intuitively create expectations about an individual's personality and capability based on appearance [5]. We have leveraged this innate skill, ensuring the appearance of the robot accurately reflects it's capability, allowing users to accurately judge it's capability at a glance. For example, the simplicity of the casework design, coupled with the over-sized head gives the robot a gentle, child-like quality. This persona signals that humanoid robotics is in it's infancy, therefore robots may have limited intelligence, mobility and awareness of their surroundings.
- **Intuitive product usability.** By applying design heuristics from traditional consumer products [13], we created a secondary level of product communication that uses lighting to indicate robot status, and graphics that highlight access panel touch-points (as seen in Figure 3).

(2) **Safety.** In normal operation, we prioritise fostering a safe-feeling environment through clear, intuitive signals of intent (such as eye movement, head lighting and body movement), this is supported by the live dashboard screen adding further transparency to the robot's decisions and actions. In the event of an error, head lighting changes colour, providing a very visible indication something is wrong. Finally as a last resort, we have integrated a tethered or wireless kill switch that powers down the system in a controlled manner. The hardware is covered in smooth, rounded panels, minimising the risk of finger traps and sharp edging catching users when in an error state.

(3) **Efficient operation.** The hardware design is relevant to real world logistics and manufacturing use-cases, with useful features that reduce downtime and enable rapid maintenance and repair. Example features include: a grab handle for ease of lifting, access panels throughout for testing and maintenance, fast disassembly allowing a complete head to be replaced within minutes by removing a single bolt (as seen in Figure 3).

2.2.2 Robot Head Design

The standard Unitree G1 head was replaced with a fully custom design. This was done to allow us to integrate additional cameras, increasing the robot's sensing capabilities to support perception required by the AI policies we wanted to train and deploy. At the same time, the new head enables facial animations, allowing the robot to convey internal states and emotions through an expressive, human centered Face UI, improving legibility and social presence in human–robot interaction.

The robot's head (as seen in Figure 4) integrates two RGB-D cameras, a display, and a peripheral LED strip, all connected to an onboard NVIDIA Jetson Orin Nano, which serves as the primary compute unit for perception and interaction. An Orbbec Femto Bolt camera provides a forward facing view used for environmental perception and human pose tracking, while an Orbbec Gemini 2 camera is oriented downward toward the robot's hands to support object pick and place tasks. The face display is a 4.3-inch Waveshare HDMI LCD used to display animated eyes. Eye animation and LED control are implemented as decoupled components within a ROS 2 lifecycle node running on the Jetson, enabling coordinated yet independent control of facial and peripheral visual feedback.

The new head design gives the robot a child-like silhouette, with a larger head relative to its body size and soft rounded form. The brow section extends past the face like a cap, shielding the cameras from environmental noise that could confuse sensors, while hiding the technology inside to create an overall simpler interface for users. The single wrap around white shell is reminiscent of a construction hard hat, a reference to ruggedness and possible warehouse or manufacturing use cases. Rectangular ear covers allow quick access to adjust camera location and tilt inside the head, the covers have orange bands indicating a touch point directing the operator to lift at that location. To the rear of the head is a transparent frosted window showing a glimpse of the technology hidden within. To the front of the head, the Face UI is covered with a tinted and frosted panel, this acts as a diffuser, hiding technical components and softening the UI lighting effects, adding to the generally soft and gentle feel of the robot. All of these elements combine to create a friendly appearance that intuitively expresses robot status and intent in a clear and simple way to operators.

Figure 4: Front and Rear view of the new head and Face UI

Peripheral LED indicators at the rear of the head are controlled via a compact serial protocol to an Arduino-based controller. In the default state, the LEDs display a subtle animated white-light pattern. The colour and dynamism of the light pattern signals high-level robot state, giving operators immediate feedback on whether the robot is 'thinking', 'confused' or in an error state requiring operator intervention. The LEDs are also used to communicate system status, such as battery level, by subscribing to a battery state topic and displaying yellow or red light under low-power conditions. This cue is intended primarily for operators running demonstrations, providing actionable feedback while remaining unobtrusive to participants.

2.3 Face User Interface

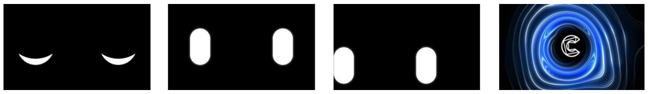

Figure 5: Face UI animation states: sleepy, idle, interactive, busy

The eye animation engine supports multiple animation modes, selected based on the current state of the robot's behaviour tree (as seen in Figure 5). Behaviour state is published by the robot control system and received by the Face UI node over the robot's network. Smooth transitions between animation modes are used to preserve visual continuity. In its default idle state, the Face UI renders eyes as simple rounded rectangular shapes with soft glow effects. Blinking is implemented at random intervals, and the eyes occasionally shift gaze laterally, this subtle low-level motion subconsciously humanises the robot and makes people feel more at ease interacting with it [1]. The large rounded shape of the eyes give the robot a friendly, cartoon-like feel. This minimalist approach avoids undesirable uncanny valley effects and accurately communicates the child-like intelligence of early humanoid robot systems. Communicating robot capability accurately is critical to avoiding user frustration and building trust [9].

Additional animation modes communicate higher level system state. A sleep state, featuring closed eyes with gentle swaying motion, is activated when the robot is halted or resting. A busy state replaces the eyes with an abstract wave animation incorporating our company logo, explicitly signalling that the robot is executing autonomous behaviour driven by our AI policy. The Face UI also supports interactive modes tailored to specific demonstrations. In one mode, the 3D camera coordinates are converted into normalised gaze targets using camera intrinsic parameters and field-of-view calculations to allow the eyes respond to human tracking data and follow the closest detected person. In a box pick and place demo, the eyes respond to the direction of a participant's pointing gesture to indicate recognition, supporting transparency of perception and intent. These responsive gaze behaviours help humanise the robot and make its perception and actions more interpretable to observers [16].

2.4 AI Dashboard

The AI dashboard (as seen in Figure 1) makes the robot decision making process more transparent by providing a real-time, web based interface that exposes the internal states of ROS2 robotic systems. By visualising sensor inputs, behaviour tree flow, and robot actions, the dashboard serves as a tool for mechanistic interpretability, enabling observers to understand how the robot arrives at its decisions.

Communication between the web interface and the robot is achieved via a WebSocket bridge. The React/TypeScript client uses `roslibjs` [15] to establish and manage WebSocket connections to ROS2, subscribing to relevant topics in real time. Component state is maintained using React hooks, with shared references for 3D transforms to support efficient, real-time visualisation.

Visualisations are organised into three sections, with visual cues clarifying how information from each stage is passed to the next, making the robot's perception–decision–action pipeline explicit:

- **Understand Environment:** Displays live camera feeds for both cameras - an RGB view for the Orbecc Gemini 2 and a depth view the for Orbecc Femto Bolt. It also visualises the direction a tracked person is pointing to, allowing an observer to explicitly see how their actions are being recognised and processed by the robot.
- **Make Decision:** Provides an intuitive visual representation of the behaviour tree, with nodes colour coded by execution state (active, running, dormant). Time limited speech bubble messages of the robot's 'internal monologue' provide a human readable interpretation of node states and selected actions, making the robot's reasoning mechanistically interpretable to observers.
- **Take Action:** Visualises the URDF robot model with overlays for joint efforts and temperatures. The model mirrors the physical robot, making it clear that the data shown on the dashboard is directly mapped from the real world. There are also graph displays for real time joint commands versus states for key joints, providing an explicit link between internal commands and observable behaviour.

Through this design, the dashboard explicitly makes the robot's inputs, internal decision processes, and outputs transparent to observers. It also functions to assist operators with system monitoring, such as battery state, network connections and motor temperatures, without obscuring the dashboard's primary role in promoting interpretability.

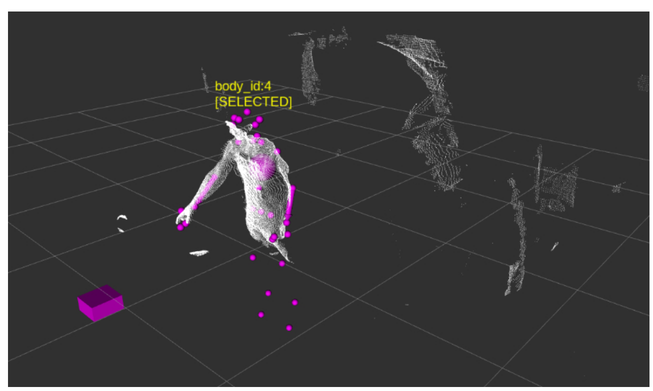

Figure 6: RVIZ view of pointing recognition

On a separate screen alongside the dashboard, the RVIZ view of the robot's perceived environment is also visualised (as seen in Figure 6). This highlights the human pose detection markers and pointing recognition.

2.5 User Study

We conducted a within-participant user study with 10 participants recruited from our company, Cambridge Consultants, to evaluate the effect of the enhancements designed to improve our HRI system. The study consisted of two conditions, A and B. In Condition A, participants interacted with the robot without any of the interface enhancements; only the robot body and its head (with the face UI turned off) were visible. In Condition B, all enhancements were enabled. The first five participants received Condition A followed by Condition B and the final five participants received Condition B before Condition A. This swapping was to control for learning and spillover effects. Participants completed a set of questions immediately after experiencing each condition.

In both conditions, the robot was positioned at the center of a semicircular table with a box placed directly in front of the robot. Participants were instructed to point either to the left or the right side of the table. The robot then picked up the box and placed it at the location indicated by the participant. This was repeated 3 times within each condition. Before the study commenced, we had participants stand and point left and right in front of the robot. This was just to check that the pointing detection worked for them and they were confident with where they needed to aim their pointing to.

The questions asked were on a 5-point Likert scale (e.g. 5=strongly agree; 1 = strongly disagree) and designed to measure the following five areas:

(1) Robot perception
- The robot was responsive to its surroundings
- The robot knew exactly where I was pointing
- I clearly understood what the robot was looking at
- The robot's movement was confusing (inverse clarity)

(2) Predictability
- I had clear context for how the robot made decisions
- I understood the robot's point-of-view
- I could predict the robot's next move
- I understood what progress the robot had made
- The robot seems intelligent

(3) Social Perception
- The robot is good-natured
- The robot is honest
- The robot seems intelligent
- The robot could be a friend of mine
- The robot behaved like a living creature
- I have confidence the robot is able to get the job done

(4) Cognitive Load
- Interacting with the robot required a lot of mental effort
- I struggled to focus on the robot

(5) Safety and Privacy
- The robot made me feel uncomfortable
- I feel safe when the robot moves around
- I am afraid the robot will hurt me
- My privacy was not considered

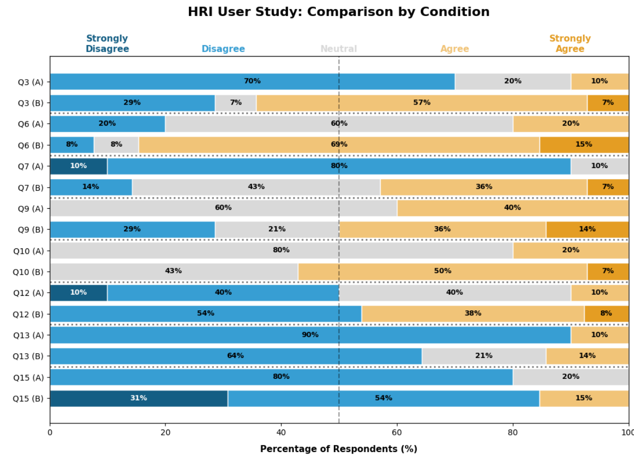

Figure 7: Distribution of 5-point Likert scale responses for selected questions. Condition A involves interacting the robot without our enhancement. Condition B includes all our enhancements.

3 Results

To assess how our interface enhancements impacted users, we performed a comparative analysis to measure the differences between the two conditions. Given our small sample size (N=10) and ordinal nature of the Likert scale data, we could not assume normality. As participants experienced both conditions, we used the Wilcoxon signed-rank test, the non-parametric equivalent of the paired t-test, to compare responses between conditions (Table 1).

Questions regarding robot perception illustrated a significant increase for gaze/attention clarity (Q3, $p < 0.05$). Figure 7 shows significantly more disagree ratings in Condition A compared to B where more participants rated it as agree. Clearly indicating that our enhancements increased users understanding of where the robot was looking at, however did not have a significant impact on helping users perception of the robot movement (Q4, $p = 0.16$) nor feel they are being responded to (Q2, $p = 0.77$; Q1, $p = 1.00$).

Predictability-focused questions demonstrated the strongest overall improvement in performance. Participants reported significant gains in perspective-taking (Q6, $p < 0.05$, $r = 0.86$) and predictability (Q7, $p < 0.05$, $r = 0.75$). Figure 7 shows that Condition B had significantly more ratings of strongly agree and agree for Q6 when compared to Condition A. In addition, more participants strongly disagreed in Condition A for Q7. These findings suggest that increasing transparency elements for participants fosters a deeper and more accurate understanding of the robot's current actions. All four items (Q9, Q10, Q12, $r = 0.84$; Q13, $r = 0.82$) related to social perception point in a consistent, socially positive direction, but none reached significant ($p < 0.05$) in this sample. Similarly, questions related to cognitive load indicated no reliable differences between conditions, but one did have a high r value (Q15, $p = 0.06$). Finally questions regarding safety and privacy were stable across conditions, indicating no impact with or without the enhancements. With a larger sample size, several of these could plausibly become significant.

Group Comparison	z	p	r
Q1. the robot was responsive to its surroundings	0.18	1.00	0.09
Q2. The robot knew exactly where I was pointing to	-0.42	0.77	0.16
Q3. I clearly understood what the robot was looking at	-2.45	0.01	0.87
Q4. The robot movement was confusing	-1.57	0.16	0.64
Q5. I had clear context for how the robot made decisions and executed tasks	-1.44	0.19	0.54
Q6. I understood the robot's point-of-view	-2.10	0.03	0.86
Q7. I could predict the robot's next move	-2.25	0.02	0.75
Q8. I understood what progress the robot had made during the task	-1.89	0.06	0.84
Q9. The robot is good-natured	-1.89	0.06	0.84
Q10. The robot is honest	-1.89	0.06	0.84
Q11. The robot seems intelligent	-0.91	0.50	0.46
Q12. The robot could be a friend of mine	-1.89	0.06	0.84
Q13. The robot behaved in a way I would expect from a living creature	-1.64	0.13	0.82
Q14. I have confidence the robot is able to get the job done	-1.08	0.38	0.48
Q15. Interacting with the robot required a lot of mental effort	-1.89	0.06	0.84
Q16. I struggled to focus on the robot	-0.27	1.00	0.15
Q17. The robot made me feel uncomfortable	0.00	1.00	0.00
Q18. I feel safe when the robot moves around	-1.08	0.38	0.48
Q19. I am afraid the robot will hurt me	0.45	1.00	0.32
Q20. My privacy was not considered	0.00	1.00	0.00

Table 1: Wilcoxon signed-rank test results, comparing questionnaire responses when participants experienced a box-picking human-robot interaction demo with, and without our transparency enhancements.

Since we used a Wilcoxon signed-rank test, we compared the positive (w^+) and negative (w^-) rank sums to determine the direction of differences between conditions. Four questions (Q15, Q17, Q20, and Q4) showed lower scores in Condition A than in Condition B. One question (Q19) showed no difference between the two conditions. The remaining 15 questions demonstrated higher scores in Condition B compared with Condition A.

Taken together, the results provide preliminary indications that the interface enhancements may have supported users' understanding of certain aspects of the robot's behavior, particularly its gaze direction, perspective-taking, and predictability. Other areas such as social perception, cognitive load, safety, and privacy showed more mixed patterns, with several items trending positively but not reaching significance in this sample. The directionality of the Wilcoxon comparisons largely favored Condition B, i.e. including our enhancements, although a small number of items showed no difference or favored Condition A. Overall, these findings offer an initial view of where the enhancements may be having an effect and highlight the need for a larger study to more confidently assess their impact across the full range of user perceptions.

4 Conclusion

This initial comparison highlights the potential effects of our enhancements in an HRI setting. Participants appeared better able to take the robot's perspective and understand how the robot perceived the environment through its sensors. This perspective-taking may have helped participants adjust their pointing behavior, allowing the robot to more accurately infer the intended target location during the box-moving task. Participants also showed improved ability to anticipate the robot's next actions, which could support more seamless interaction in collaborative contexts, such as in industrial settings where heavy lifting would be required between a robot and human. Although not statistically significant, the r-value associated with the face UI suggests that it may have contributed to a more socially positive impression of the robot, with several participants rating it as friendly, honest, and relatable. The face UI is applicable to any context that requires a social interaction between a human and robot. A friendlier, honest, and relatable robot can improve the bond and trust between a robot and human [8]. Measures related to privacy and safety did not show significant differences with and without our enhancements, but the generally neutral responses indicate that participants did not view the enhancements as negatively affecting these dimensions.

Although our results suggest that the enhancements had several beneficial effects, it is important to acknowledge some limitations. The depth camera used in the study was unable to reliably detect some items of clothing that absorbed infrared light. As a result, the pose detection system did not correctly identify the torso of one participant, which affected the accuracy of the pointing estimation for that trial. In addition, one participant experienced inconsistencies in the demo performance may have introduced biases. We also did not isolate the individual components of the enhancement package, making it difficult to determine which specific element contributed to each observed effect.

Altogether, we have developed an HRI dashboard, Face UI, and robot body designed to support greater transparency and understanding during human–robot interactions. The broader goal is to contribute to environments where users feel safe, understood, and able to collaborate with robots through shared mental models and effective perspective taking.

References

[1] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63. doi:10.5898/JHRI.6.1.Admoni

[2] Neziha Akalin, Annica Kristoffersson, and Amy Loutfi. 2022. Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures. *International Journal of Human-Computer Studies* 158 (2022), 102744. doi:10.1016/j.ijhcs.2021.102744

[3] Victoria Alonso and Paloma de la Puente. 2018. System Transparency in Shared Autonomy: A Mini Review. *Frontiers in neurorobotics* 12, 83 (Nov. 2018). doi:10.3389/fnbot.2018.00083

[4] Georgios Angelopoulos, Dimitri Lacroix, Ricarda Wullenkord, Alessandra Rossi, Silvia Rossi, and Friederike Eyssel. 2025. Measuring transparency in intelligent robots. *Scientific Reports* 15, 43809 (2025). doi:10.1038/s41598-025-29685-w

[5] Benedict Tay Tiong Chee, Park Taezoon, Qianli Xu, Jamie Ng, and Odelia Tan. 2012. Personality of social robots perceived through the appearance. *Work* 41, Suppl 1 (2012), 272–276. doi:10.3233/WOR-2012-0168-272

[6] Yuhao Chen, Yue Luo, Chizhao Yang, Mustafa Ozkan Yerebakan, Shuai Hao, Nicolas Grimaldi, Song Li, Read Hayes, and Boyi Hu. 2022. Human mobile robot interaction in the retail environment. *Scientific Data* 9, 673 (2022). doi:10.1038/s41597-022-01802-8

[7] Pauline Chevalier, Bob R. Schadenberg, Amir Aly, Angelo Cangelosi, and Adriana Tapus. 2022. Context-Awareness in Human-Robot Interaction: Approaches and Challenges. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 1241–1243. doi:10.1109/HRI53351.2022.9889584

[8] L. Christoforakos, A. Gallucci, T. Surmava-Große, D. Ullrich, and S. Diefenbach. 2021. Can Robots Earn Our Trust the Same Way Humans Do? A Systematic Exploration of Competence, Warmth, and Anthropomorphism as Determinants of Trust Development in HRI. *Frontiers in robotics and AI* 8, 640444 (2021). doi:10.3389/frobt.2021.640444

[9] Nicholas Conlon, Daniel Szafir, and Nisar R. Ahmed. 2022. "I'm confident this will end poorly": Robot proficiency self-assessment in human–robot teaming. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction (HRI)*. ACM.

[10] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 301–308. doi:10.1109/HRI.2013.6483603

[11] Martina Mara, Markus Appel, and Timo Gnambs. 2022. Human-Like Robots and the Uncanny Valley: A Meta-Analysis of User Responses Based on the Godspeed Scales. *Zeitschrift für Psychologie* 230, 1 (2022), 33–46. doi:10.1027/2151-2604/a000486

[12] Microsoft Corporation. [n. d.]. Azure Kinect Body Tracking SDK Documentation. https://microsoft.github.io/Azure-Kinect-Body-Tracking/release/1.1.x/index.html. Accessed: 20 February 2026.

[13] Donald A. Norman. 2002. *The Design of Everyday Things*. Basic Books.

[14] NVIDIA Corporation. [n. d.]. Isaac GR00T - Generalist Robot 00 Technology. https://developer.nvidia.com/isaac/gr00t. Accessed: 18 February 2026.

[15] RobotWebTools. 2026. roslibjs: The Standard ROS JavaScript Library. GitHub repository. https://github.com/RobotWebTools/roslibjs

[16] Mohamed Salem et al. 2021. Bidirectional Multi-modal Signs of Checking Human-Robot Engagement and Interaction. *International Journal of Social Robotics* 13 (2021), 1234–1250. doi:10.1007/s12369-021-00855-w

[17] Ranjan Sapkota, Yang Cao, Konstantinos I. Roumeliotis, and Manoj Karkee. 2026. Vision-Language-Action (VLA) Models: Concepts, Progress, Applications and Challenges. arXiv:2505.04769 [cs.CV] https://arxiv.org/abs/2505.04769

[18] Xiang Zhi Tan, Elizabeth J. Carter, and Aaron Steinfeld. 2024. Person Transfer in the Field: Examining Real World Sequential Human-Robot Interaction Between Two Robots. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. 909–915. doi:10.1109/RO-MAN60168.2024.10731281

[19] Robert Wortham, Andreas Theodorou, and Joanna Bryson. 2016. What Does the Robot Think? Transparency as a Fundamental Design Requirement for Intelligent Systems. In *25th International Joint Conference on Artificial Intelligence*.