2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Enhancing the NAO: Extending Capabilities of Legacy Robots for Long-Term Research

Austin Wilson[1,2], Sahar Kapasi[1,3], Zane Greene[1,2], and Alexis E. Block[1,2]

***Abstract*— Legacy (unsupported) robotic platforms often lose research utility when manufacturer support ends, preventing integration of modern sensing, speech, and interaction capabilities. We present the *Enhanced NAO*, a revitalized version of Aldebaran's NAO robot featuring upgraded beamforming microphones, RGB-D and thermal cameras, and additional compute resources in a fully self-contained package. This system combines cloud-based and local models for perception and dialogue, while preserving the NAO's expressive body and behaviors. In a pilot user study validating conversational performance, the Enhanced NAO delivered significantly higher conversational quality and elicited stronger user preference compared to the *NAO AI Edition*, without increasing response latency. The added visual and thermal sensing modalities established a foundation for future perception-driven interaction. Beyond this implementation, our framework provides a platform-agnostic strategy for extending the lifespan and research utility of legacy robots, ensuring they remain valuable tools for human-robot interaction.**

I. Introduction

The Aldebaran NAO [1] has been a staple in human–robot interaction (HRI) research for its approachable design, ease of use, and multimodal sensors. With Aldebaran's recent bankruptcy and uncertainty around ongoing support, many labs risk losing a valuable research and education platform [2]. This challenge is not unique to NAO: Pepper was discontinued in 2021 [3], Rethink Robotics shut down in 2018, leaving Baxter obsolete [4], [5], and Willow Garage closed in 2014, ending support for the TurtleBot (first and second generations) and PR2 [6]–[8]. To address this broader problem of platforms becoming unsupported, we present a platform-agnostic framework for sustaining and enhancing obsolete robots, using the NAO as a central example.

Updating obsolete robots is challenging due to closed designs and outdated onboard compute resources. For NAO, prior work has added external computers [9], updated the operating system, [10], or integrated cloud services [11], but these approaches trade off portability, scope, or robustness.

We address these challenges with an integrated upgrade package (Fig. 1), featuring dual head-mounted cameras for multimodal perception, a Raspberry Pi 5 [12], a Seeed Studio ReSpeaker 4 Mic Array [13] for robust speech pipelines, and a dedicated battery for peripheral power. This modular design extends sensing, processing, and interaction capabilities without reliance on manufacturer support or tethering.

[1]SaPHaRI Lab, Human Fusions Institute,
[2]Department of Electrical, Computer, and Systems Engineering,
[3]Department of Psychological Sciences/Department of Cognitive Science
Case Western Reserve University, Cleveland, OH, USA
{amw223, snk83, zdg14, alexis.block}@case.edu

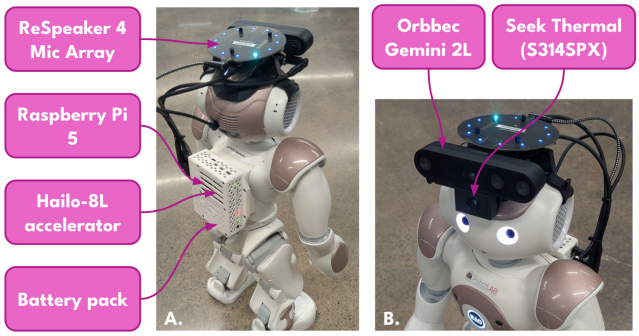

Fig. 1. Enhanced NAO robot with integrated sensing, computing, and power systems. A. Back isometric view showing the ReSpeaker 4 Mic Array, Raspberry Pi 5, Hailo-8L, and battery pack. B. Front view highlighting the Orbbec Gemini 2 L depth camera and Seek Thermal camera (S314SPX).

Here, we contribute: (1) a complete hardware design, (2) an integration framework, and (3) supporting resources available upon request to enable widespread adoption. In a pilot validation study, the Enhanced NAO (with the custom upgrade package) demonstrated *richer verbal interaction* (e.g., conversation quality, participant preference) *without added latency* compared to the NAO AI, while *expanded sensing and processing functionalities* establish a practical blueprint for sustaining and enhancing legacy robots.

II. Related Work

Researchers have extended the life of legacy research platforms (e.g., NAO [1], Baxter [4], PR2 [7], TurtleBots [6]) through added compute resources [9], [14], [15], updated sensors [14], [16]–[21], modified end-effectors [17], [22], [23], and alternative software stacks [10], [21], [24]–[29].

For the NAO, researchers pursued software and hardware strategies [9], [10], [14], [24]. Software efforts replaced the proprietary NAOqi head-unit with Ubuntu [24] and ROS2 [10], but these require flashing new firmware [24] and are limited by the NAO's small eMMC, requiring a permanently attached USB stick for operation [24]. Hardware approaches preserved the factory firmware but added external compute and peripherals via short, wired connections [9], [14] , enabling capabilities not supported by the stock head unit.

Similar modifications exist for other humanoid and non-humanoid robots. Prior software upgrades have replaced ROS1 with ROS2 for Baxter [25], abstracted low-level ROS2 interfaces for PR2 [27], and introduced new inverse kinematic libraries for smoother control of both robots [26], [28]. Hardware upgrades have included custom end-effectors [17], [22], [23], [30], clothing and tactile sensors to improve human-robot hugs [19], and RGB-D cameras for

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

teleoperation and object detection [16]–[18]. Non-humanoid robotic platforms, such as the early TurtleBots have received comparable software [20], [29] and hardware upgrades [15], [21], extending their utility beyond their official support.

III. Materials and Methods

Prior work demonstrates the potential and limitations of sustaining unsupported robots via piecemeal upgrades. These efforts highlight the need for integrated solutions. To address these limitations, we developed a platform-agnostic, custom hardware and software upgrade framework that introduces new data streams and enhances on-board compute resources for legacy, yet capable, robots. We demonstrate its usability on the NAO v6, a widely used platform in social and physical HRI research with uncertain long-term support.

A. NAO Robot

We use the NAO v6 humanoid robot [1], a 58 cm-tall, 5.6 kg bipedal platform with 25 degrees of freedom (DOF), powered by an Intel Atom E3845 CPU with 4 GB RAM and a 32 GB eMMC storage running a Gentoo-based Linux OS. The robot has two head-mounted HD cameras, a torso-mounted inertial measurement unit (IMU), and two ultrasonic sonars, four directional microphones, two speakers, nine capacitive touch sensors (three in the head, three per hand), and four bumper switches (two per foot). The platform supports Ethernet, Wi-Fi, and USB. Its lithium-ion battery provides approximately 60-90 minutes of operation.

The NAO is controlled over a socket interface exposed by the proprietary NAOqi framework. Aldebaran provided a C++ library, Python bindings, and the Choregraphe GUI, though these require a local network connection to communicate with the socket interface. While the NAO v6 supports local code execution, it is restricted to Python 2.7, limiting developer flexibility [10].

RobotLab further markets the NAO AI Edition (hereafter referred to as "NAO AI") with added cloud-based conversational features [31]. In contrast, our work builds on the *base* NAO v6 with custom hardware and software upgrades, collectively referred to as the "Enhanced NAO:"

B. Hardware upgrades

1) Computing and Power System: A Raspberry Pi 5 serves as the main processor, featuring a Broadcom BCM2712 quad-core 64-bit Arm Cortex-A76 (Armv8) running at 2.4 GHz [12]. The Raspberry Pi 5 provides USB, GPIO pins, and PCIe expansion, supporting a Hailo-8L accelerator for 13 TOPS of onboard vision-based inference [32].

Peripheral connections include USB ports and CSI/DSI interfaces for additional sensors (e.g., Seek Thermal Camera). The system is powered by a 5 V/5 A USB-C battery pack that sustains all peripherals. Mechanical enclosures, including backpack and sensor housings, were 3D printed, with mounts adapted from open-source designs [14] and custom components, as needed.

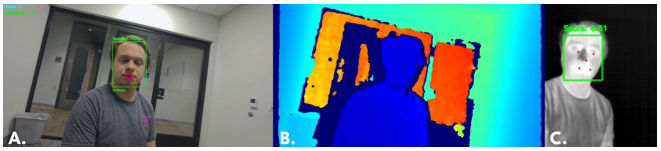

Fig. 2. Sample camera frames produced by the enhanced camera system. A) Processed RGB and B) depth frames from the Orbbec Gemini 2 L and (C) processed thermal frame from the Seek Thermal (S314SPX) camera.

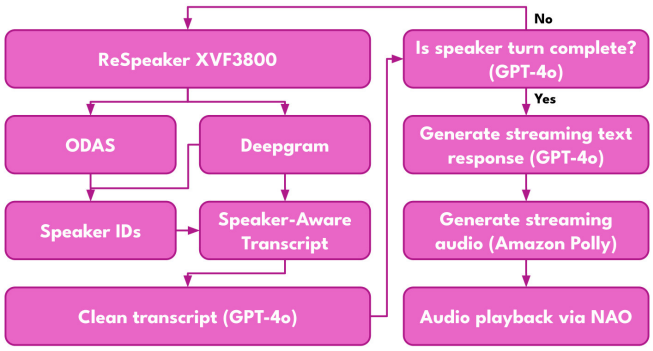

Fig. 3. Audio processing pipeline, from audio input via the ReSpeaker to a response audio played back through the NAO speakers.

2) Audio System: Audio capture is provided by the ReSpeaker with an embedded XMOS XVF-3000, offering built-in echo cancellation, noise suppression, and direction-of-arrival estimation [13].

3) Camera System: The camera sensing suite integrates an Orbbec Gemini 2 L depth camera [33] and a Seek Thermal camera (S314SPX) [34]. The Orbbec provides synchronized RGB-D streams with native OpenCV compatibility, while the Seek Thermal delivers thermography data and RGB thermal visualizations.

C. Software upgrades

The Raspberry Pi 5 runs Raspberry Pi OS with system nodes implemented in Node.JS, Python, and C++.

1) Camera Fusion: The Raspberry Pi 5 processes RGB, depth, and thermal streams locally. Frames are synchronized and passed through the Hailo-8L for human detection, face region extraction, and landmark identification in thermal and RGB data. This pipeline establishes integration of thermal and visual cues for future interaction logic. Because the Enhanced NAO's upgraded cameras augment rather than replace the existing NAO AI camera system, they were not used in the pilot study. Their integration expands sensing capabilities (adding depth and thermal perception) (Tab. II) without interfering with the existing system. However, we include a passive vision pipeline integrating the Orbbec Gemini 2 L camera, Seek Thermal module, and the Hailo-8L for demonstration (Fig. 2).

2) Verbal Interaction Pipeline: The audio pipeline (Fig. 3) begins with the ReSpeaker, which outputs raw, processed, and echo-canceled streams. These signals are processed by the Open Embedded Audition System (ODAS) [35], for real-time speaker separation and sound source localization.

Deepgram [36] performs speech recognition and completion, bypassing NAO's built-in speech-to-intent engine to enable unconstrained dialogue. Following Deepgram's best

Dimension	1 - Low	2 - Medium	3 - High
Relevance (R)	Off topic or mostly unrelated	Partially relevant but incomplete	Fully relevant to the prompt and context
Specificity (S)	Vague, generic	Includes some surface level detail	Clear and specific
Clarity (C_l)	Hard to follow, confusing	Understandable, some minor issues	Concise, clear, and well-structured
Coherence (C_o)	No logical flow from previous turn	Mostly logical	Flows naturally and connects to previous turns
Appropriateness (A)	Flat tone	Moderately engaging tone	Warm, polite, and engaging tone
Timing (t)	Long pause before response	Robot interrupts user	No issues

TABLE I

RUBRIC FOR EVALUATING CONVERSATION QUALITY ACROSS SIX DIMENSIONS, EACH RATED ON A 1-3 SCALE, WITH HIGHER SCORES INDICATING BETTER QUALITY.

Fig. 4. A user engaging in a multi-turn conversation with the Enhanced NAO during the pilot validation study.

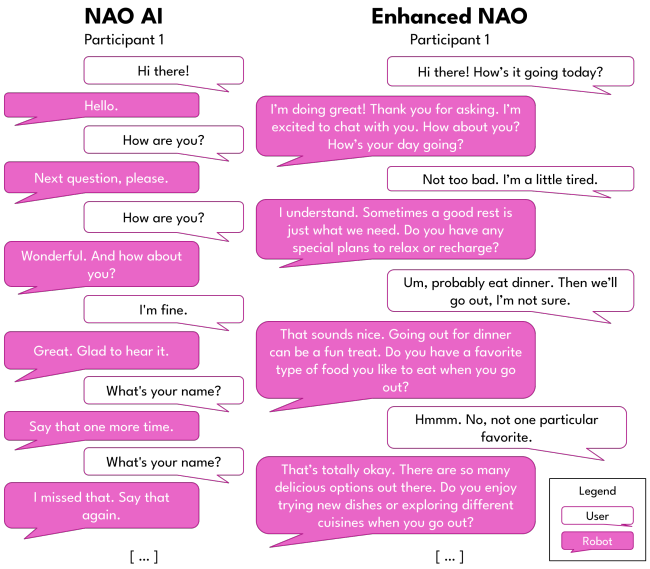

Fig. 5. Example of a participant conversation with the NAO AI and the Enhanced NAO.

practices, speech completion is detected when either the Endpointing feature identifies the end of natural speech in the raw audio stream or when the UtteranceEnd event indicates a sufficient silence gap in the transcript. Once a complete transcript is received, we align the word timings with ODAS outputs to produce speaker-aware transcripts.

Although Deepgram achieves a low transcription error rate, performance can degrade with low speaker volume, microphone quality, distance, or background noise. Additionally, a complete transcript does not always indicate that a participant has finished speaking. To address these challenges, we created three GPT-4o [37]-based LLM agents with distinct roles: a) correcting transcription errors in speaker-aware transcripts, b) detecting turn completion, and c) generating a contextually appropriate response. We used prompt engineering to structure the system and user prompts for each agent, ensuring accurate task execution and coherent, contextually relevant replies that sustained natural, open-ended dialogue.

To further reduce latency, responses are streamed immediately using Amazon Polly [38]. Because Polly does not natively support streaming, we developed a sentence-segmentation method that sends ordered requests, enabling a "think and speak" behavior and reducing the delay between user input and robot output.

D. Pilot validation study

We conducted a within-subjects, counter-balanced pilot study with eight participants (four male presenting, four female presenting) to compare interactions with the NAO AI and the Enhanced NAO. The study focused on evaluating conversational capabilities, the only functionality directly replaced and improved by the Enhanced NAO, using equivalent audio inputs and outputs. Each 15-minute session included two conditions, presented in counterbalanced order, with no compensation provided. In each condition, participants held a short conversation (≥ 4 *turns*; one turn defined as a user utterance followed by a robot response). As shown in Fig. 4, the NAO was placed on a table facing the seated participant, bringing it closer to eye level. To ensure consistency, the first turn was always a user greeting to the robot; subsequent turns were open-ended, with the option to prompt the robot for a joke, if needed. A comparison of example conversations from both conditions is shown in Fig. 5. After both conversations, each participant was verbally asked which conversation they preferred (e.g., in terms of ease and naturalness of conversation). Their responses were manually recoded by a research assistant. All sessions were video-recorded, manually transcribed, segmented into speaker turns, and timestamped for conversational quality assessment.

E. Quantifying quality of responses

We assessed conversational quality using a two-level framework adapted from LLM evaluation methods [39]. The original framework included coherence, accuracy, clarity, relevance, and efficiency; we adapted it for conversational contexts by replacing accuracy with specificity, removing efficiency, and adding appropriateness and timing. This framework allows us to capture objective response quality and subjective user experience.

At the *turn level*, each robot response was rated on six dimensions (Table I): *relevance* (addresses user input), *specificity* (detail vs. vagueness), *clarity* (ease of understanding), *coherence* (logical flow), *appropriateness* (politeness and

tone), and *timing* (delay before response).

At the *conversation level*, we computed mean scores for each dimension:

$$\overline{D} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{1}$$

where $\overline{D}$ is the dimension mean, d_i is the rating for the dimension at turn i, and n is the number of turns rated. Once we obtain mean values for all six dimensions, we calculate an overall conversational quality score as:

$$C_q = \frac{\overline{R} + \overline{S} + \overline{C_l} + \overline{C_o} + \overline{A} + \overline{t}}{6} \tag{2}$$

where C_q is the conversation quality score, and $\overline{R}$, $\overline{S}$, $\overline{C_l}$, $\overline{C_o}$, $\overline{A}$, and $\overline{t}$ represent the mean ratings for relevance, specificity, clarity, coherence, appropriateness, and timing, respectively.

We also examined *turn balance*, measuring whether conversations were dominated by the robot or the user. We define this ratio as:

$$R = \frac{W_r}{W_u} \tag{3}$$

where W_r and W_u denote the number of words spoken by the robot and the user, respectively. Their ratio quantifies relative talkativeness. Using Equation 3, an $R = 1$ indicates perfectly balanced turn lengths, $R > 1$ indicates robot-dominated speech, and $R < 1$ indicates user-dominated speech. This value is averaged across the entire interaction. Because natural conversation generally involves shared participation, we considered $0.5 < R < 2$ an acceptable range. Values outside this range signal issues in turn-taking, such as overly strict end-of-turn detection (user-dominated conversation) or overly verbose robot responses (robot-dominated conversation), both of which can lead to a less comfortable interaction.

Finally, we tracked *topic shifts* in each conversation, defined as major disruptions to conversational flow. A topic shift was recorded when the robot: 1) introduced an unrelated subject, 2) ignored or cut off the user's input, stopping any elaboration, 3) returned to an earlier subject without linkage, or 4) closed a topic prematurely. While related to coherence, topic shifts capture disruptive changes in the semantic flow and direction of the conversation, whereas coherence addresses whether consecutive turns of the conversation flow logically. In this way, *most topic shifts* can be *classified as coherence issues*, but *not all coherence issues are topic shifts*.

Two evaluators (co-authors on this paper, who were not involved in the design or implementation of the Enhanced NAO) developed the coding schema before viewing any videos, to minimize expectation bias. All recordings were then coded by a primary researcher, with 20% independently coded by a secondary researcher to assess reliability. Although evaluators were aware of the condition assignments, all ratings followed standardized criteria to minimize bias.

IV. Results

Table II summarizes the hardware and software improvements of the Enhanced NAO compared to the NAO AI.

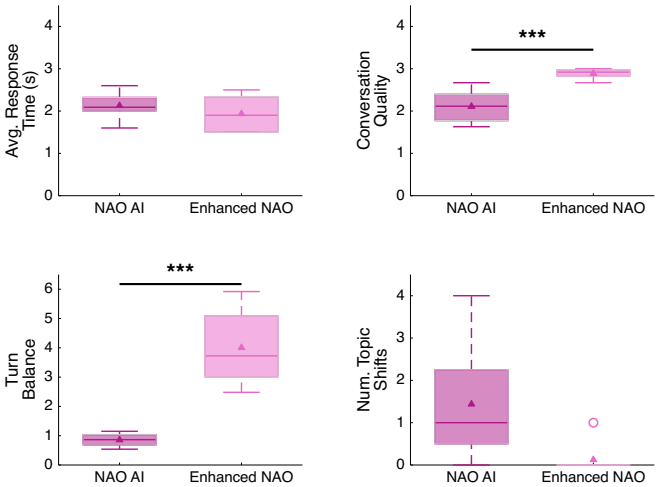

Fig. 6. A comparison of average response times, conversation quality, turn balance, and number of topic shifts between the NAO AI (dark pink) and the Enhanced NAO (light pink). Boxes show the interquartile range (25th-75th percentiles), with the centerline representing the median and triangles marking the mean. Whiskers extend to the most extreme values not considered outliers. Circles indicate outliers. Black lines with asterisks indicate statistically significant differences after Bonferroni correction.

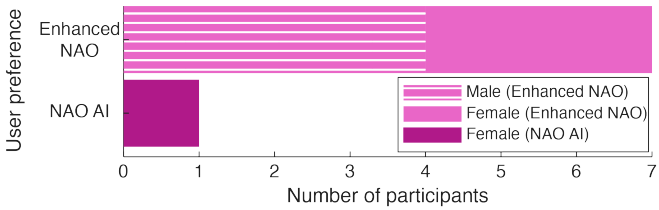

Fig. 7. Participant preferences in the pilot study for interactions with the NAO AI (dark pink) versus the Enhanced NAO (light pink). Shading indicates participant gender: stripes indicate male, solid indicates female.

A. Hardware

In addition to the NAO's two head-mounted cameras, we integrated a depth and thermal camera, expanding multimodal vision with synchronized RGB-D and thermal streams. The Hailo-8L accelerator allows real-time inference with custom models on these streams, a capability unavailable on the NAO AI. The ReSpeaker handles audio capture, rather than the NAO v6's built-in microphones, bypassing NAOqi's restrictions. NAOqi provides audio only through ALAudioDevice in either 16 kHz or 48 kHz in 170 ms chunks [1], without support for beam steering or noise suppression, limiting custom ASR pipelines.

B. Software

We restructured the conversational pipeline by replacing the NAO's intent-driven dialogue system with real-time, cloud-based transcription and diarization (Deepgram [36]). The Enhanced NAO uses an LLM, rather than fixed templates, for response generation, which enables dynamic, free-flowing conversation. To produce natural-sounding voices, we use Amazon Polly [38] for text-to-speech (TTS) generation. These software modifications are part of the system's functional improvements, which directly contribute to the Enhanced NAO's observed gains in conversational quality and motivated the selection of our evaluation metrics.

Feature	**NAO AI** [31]	**Enhanced NAO** [1] with Hardware/Software Upgrades
Cameras	*2 front-facing cameras* (OV5640); 1920×1080 up to 30 FPS; 2560×1920 at 15 FPS Frame rates may be impacted by NAO v6 compute resources.	*In addition to the base NAO camera features:* Orbbec Gemini 2 L *depth camera*; 1280×800 at 30 FPS Seek *Thermal camera*; 320×240 at 27 FPS. Support for 2 additional CSI Cameras.
Microphones	*Four omnidirectional microphones*; four channels at 48kHz or one channel at 16kHz. Accessible via NAOqi ALAudioDevice in 170ms chunks [1]. No algorithms to support ASR pipelines.	*Replaces existing NAO microphones* Seeed Studio ReSpeaker *4 Mic Array* [13] *Six channels* at 16kHz (4 raw, ASR, AEC) Additional LED indicators.
Speech (TTS)	Converts via NAOqi ALTextToSpeech. Supports pitch, speed, volume, pause, emphasis control, and voice effects. [40]	Flexible with NAO's default TTS engine, local TTS engines, or cloud-based TTS services. Tested implementation uses Amazon Polly [38].
Speech (ASR)	Via NAOqi ALSpeechRecognition. *Only recognizes predefined keywords.* Requires a set vocabulary and returns matches with confidence scores. Supports word spotting within longer speech. [41]	*Flexible transcription* using a dedicated ASR channel from the ReSpeaker. Tested implementation uses Deepgram to facilitate real-time speech-to-text.
Conversational Capabilities	*Static, intent-driven* system focused on matching predefined intents with spoken keywords that trigger pre-determined responses. TTS engine has limited voices and lacks dynamic range.	Supports *open-ended dialogue* with Deepgram [36], uses prompt engineering to tune an LLM (GPT-4o) [37] to generate relevant responses, and supports dynamic TTS voice generation via Amazon Polly [38].

TABLE II

COMPARISON OF NAO AI CAPABILITIES AND ENHANCED NAO SYSTEM FEATURES.

C. Conversation

As described in Sec. III, all conversations were independently coded by two researchers. Inter-rater reliability [42], calculated using a two-way random intraclass correlation coefficient (ICC[2,1]), indicated excellent agreement between coders (ICC $= 0.9897$).

For each conversation attribute (average response time, conversational quality, turn balance, and number of topic shifts), we performed a paired *t*-test with a Bonferroni correction ($\alpha = 0.0125$) to account for multiple comparisons (Fig. 6). We report effect sizes using Cohen's d. *Average response time* did not differ significantly between the Enhanced NAO ($M = 1.94, SD = 0.42$) and the NAO AI ($M = 2.13, SD = 0.30$), $t(14) = 1.05, p = 0.3105, d = 0.53$. *Conversational quality* was significantly higher for the Enhanced NAO ($M = 2.89, SD = 0.11$) compared to the NAO AI ($M = 2.11, SD = 0.37$), $t(14) = -5.69, p < 0.001, d = -2.85$. *Turn balance*, was also significantly higher for the Enhanced NAO ($M = 4.00, SD = 1.19$) than the NAO AI ($M = 0.86, SD = 0.21$), $t(14) = -6.89, p < 0.001, d = -3.44$, indicating that the Enhanced NAO produced substantially more words relative to the user. Additionally, the Enhanced NAO exhibited fewer *topic shifts* ($M = 0.12, SD = 0.35$) than the NAO AI ($M = 1.44, SD = 1.35$), $t(14) = 2.66, p = 0.0185, d = 1.33$, which approached significance after the alpha correction.

Finally, user preference data (Fig. 7) shows that 87.5% of users (seven out of eight) preferred the Enhanced NAO over the NAO AI, citing more natural and engaging conversation.

V. DISCUSSION

The hardware and software upgrades of the Enhanced NAO transformed it into a more capable and sustainable platform by adding sensing modalities, real-time inference, and low-level audio access. These enhancements move beyond the NAO AI's capabilities and limitations to demonstrate how modern compute and perception capabilities can be layered onto legacy hardware. Although we did not benchmark the new camera and microphone system against the NAO's integrated sensors, the additional RGB-D and thermal streams, together with advanced audio processing, provide valuable new data sources and greater developer control. Importantly, the upgrades were delivered in a fully self-contained, untethered package, demonstrating a platform-agnostic path for extending the lifespan and usability of unsupported robots.

These significant hardware upgrades directly improved the *conversational quality* of the Enhanced NAO. Cleaner, lower-latency audio from the ReSpeaker and the LLM-driven conversational pipeline generated more natural, context-aware responses. In contrast, the NAO AI's rigid intent-based models produced repetitive, unnatural, and disruptive replies, reducing flow. *Response times* did not differ significantly between conditions, showing that richer interactions can be delivered without added latency, despite additional sensing, processing, and cloud services. Interestingly, *turn balance* for the Enhanced NAO exceeded our desired range, yet participants still preferred it. This suggests that turn balance may be less critical to user preference than we initially assumed, or that our acceptable range was too conservative. Nevertheless, avoiding overly long or monologic robot responses is still necessary for maintaining enjoyable, user-centered conversation. Additionally, the Enhanced

NAO exhibited fewer *topic shifts*, suggesting conversations stayed more coherent and user-driven.

While the primary focus of this work was technical enhancement, we also acknowledge that the visible mechanical additions may have influenced user perception. The NAO's humanoid, friendly appearance supports comfort and approachability, especially among children [43]–[45]. Exposing additional hardware elements may make the robot appear more mechanical and potentially reduce approachability. Future adaptations should therefore consider visual and tactile design elements. Prior research on social and affective robots highlights the benefits of softness and warmth for comfort during social-physical interactions [19], [46]–[50]. Inspired by this line of work, future iterations of the Enhanced NAO could incorporate a soft cover or hat over exposed components to preserve the NAO's approachability while maintaining the improved functionality.

A. Limitations

This work was a pilot validation study to demonstrate the feasibility of our integrated system pipeline, rather than providing definitive evidence of the Enhanced NAO's impact. The participant pool was necessarily small (eight participants), limiting the statistical power and generalizability of results. Nevertheless, we performed statistical analyses and reported effect sizes alongside significance tests, presenting results as transparently as possible and allowing readers to interpret the findings in context. Importantly, the large observed effects indicate clear promise and motivate follow-up studies with fully powered designs. For example, detecting a medium effect size in a within-subjects design comparing two robot conditions across two gendered populations would require approximately 36 participants, which we plan for future work.

A second limitation involved a technical inconsistency in one session, where a participant interacted with the Enhanced NAO using a male TTS voice instead of the female voice presented in all other sessions. This mismatch may have influenced the system's performance and the participant's perception of the robot's gender, and notably, this was the only participant who preferred the NAO AI. The deviation could also have interacted with turn-balance patterns. However, it also revealed the flexibility of our pipeline to support multiple voices. We did not analyze male vs. female voice as an experimental factor in this pilot, but future work should explicitly test how voice characteristics influence conversational quality and user preference. Future experiments will also implement stricter quality control to ensure consistent conditions across participants.

Additionally, participants were not blinded to the experimental condition. To mitigate bias, we used a counterbalanced condition order and asked participants to evaluate the conversation quality rather than the robot itself. Future studies could incorporate blinded or partially blinded setups to minimize potential bias due to visual differences between conditions.

Forth, because this pilot emphasized system integration over conversational design, prompt engineering and pretesting were not fully optimized, occasionally producing overly long or interview-like responses. These refinements are straightforward to implement and will be incorporated in future studies with larger participant pools, standardized conditions, and improved dialogue strategies.

Finally, although the new multimodal camera system was successfully integrated, we did not formally evaluate its perceptual or interaction benefits in this pilot. Integration was the primary goal; follow-up work will characterize the contributions of these sensing modalities in depth. Taken together, these limitations reflect the preliminary nature of the study, but also underscore its success in validating the technical feasibility of our integrated framework and establishing the groundwork for more comprehensive evaluation.

VI. Conclusion and future work

This work demonstrates a pathway for revitalizing legacy social robots by integrating modern sensing, speech, and dialogue capabilities, demonstrated with the NAO. Our system restored usability while delivering higher conversational quality and stronger user preferences, without added latency, compared to the NAO AI, highlighting the value of extending rather than retiring such platforms.

Future work will address the remaining technical challenges to further improve interaction quality. In particular, we will migrate our speech-to-text and text-generation pipelines from cloud services to local processing using compact, high-performance models such as Gemma 3n [51] or Qwen 3 [52]. Running these models locally is expected to further reduce latency, strengthen privacy, and increase system autonomy. We will also leverage multimodal visual perception to detect and localize users prior to speech, enabling natural dialogue initiation. In multi-party settings, fused visual and thermal data will support more accurate speaker separation by disambiguating overlapping voices and tracking active participants, further improving responsiveness.

Ultimately, we envision extending this framework beyond the NAO to demonstrate the platform-agnostic strategy for sustaining unsupported social robots. We are already deploying the framework to other legacy platforms like the Baxter and the TurtleBot 2, showcasing the system's flexibility for humanoid and mobile robots. More broadly, this work highlights how thoughtful system integration can transform aging robots into modern research tools, extending their lifespan and ensuring they continue to advance human-robot interaction.

Acknowledgments

The authors thank the Ohio Space Grant Consortium for their generous support of the first author.

References

[1] Aldebaran, "NAO v6." [Online]. Available: https://aldebaran.com/en/nao6/

[2] C. Strathearn and E. Sobolewska, "Universities face getting stuck with thousands of obsolete robots – here's how to avoid a research calamity," *The Conversation*, May 2025. [Online]. Available: http://theconversation.com/universities-face-getting-stuck-with-thousands-of-obsolete-robots-heres-how-to-avoid-a-research-calamity-256829

[3] S. Nussey, "Softbank shrinks robotics business, stops pepper production- sources," *Reuters*, June 2021. [Online]. Available: https://www.reuters.com/technology/exclusive-softbank-shrinks-robotics-business-stops-pepper-production-sources-2021-06-28/

[4] C. Lawrence, "Rise and fall of rethink robotics," *The Conversation*, April 2019. [Online]. Available: https://www.asme.org/topics-resources/content/rise-fall-of-rethink-robotics

[5] R. Robotics, "Baxter robot," 2012.

[6] W. Garage, "Turtlebot 1 and 2," 2011.

[7] ——, "Pr2 robot," 2012.

[8] J. D'Onfro, "How a billionaire who wrote google's original code created a robot revolution," *Business Insider*, February 2016. [Online]. Available: https://www.businessinsider.com/a-look-back-at-willow-garage-2016-2

[9] M. Mattamala, G. Olave, C. González, N. Hasbún, and J. Ruiz-del Solar, "The nao backpack: An open-hardware add-on for fast software development with the nao robot," 2017. [Online]. Available: https://github.com/uchile-robotics/nao-backpack

[10] A. Bono, K. Brameld, L. D'Alfonso, and G. Fedele, "Open access nao (oan): a ros2-based software framework for hri applications with the nao robot," 2024. [Online]. Available: https://arxiv.org/abs/2403.13960

[11] E. Gestrin, "Naochat," https://github.com/ElliotGestrin/NAOChat, 2024.

[12] R. Pi, "Raspberry pi 5," *Raspberry Pi*, 2025. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-5/

[13] S. Studio, "ReSpeaker 4 Mic Array v2.0," 2023. [Online]. Available: https://wiki.seeedstudio.com/ReSpeaker_Mic_Array_v2.0/

[14] K. Chatzilygeroudis, "NAO Backpack and Helmet," May 2014. [Online]. Available: https://www.thingiverse.com/costashatz/designs

[15] Y. Bergeon, V. Křivánek, and J. Motsch, "Raspberry pi as an interface for a hardware abstraction layer: Structure of software and extension of the turtlebot 2–kobuki protocol," in *International Conference on Military Technologies (ICMT)*. IEEE, 2019, pp. 1–6.

[16] J. Avalos and O. E. Ramos, "Real-time teleoperation with the baxter robot and the kinect sensor," in *IEEE Colombian conference on automatic control (CCAC)*. IEEE, 2017, pp. 1–4.

[17] P. J. da Cruz Lino, "The control of baxter robot, and its interaction with objects using force sensitive ar10 hands, guided by kinect." Master's thesis, Universidade de Coimbra (Portugal), 2019.

[18] S. K. Paul, M. T. Chowdhury, M. Nicolescu, M. Nicolescu, and D. Feil-Seifer, "Object detection and pose estimation from rgb and depth data for real-time, adaptive robotic grasping," in *Advances in Computer Vision and Computational Biology: Proceedings from IPCV'20, HIMS'20, BIOCOMP'20, and BIOENG'20*. Springer, 2021, pp. 121–142.

[19] A. E. Block and K. J. Kuchenbecker, "Softness, warmth, and responsiveness improve robot hugs," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 49–64, 2019.

[20] H. Song, J. Tan, Y. Xing, and G. Hou, "Communication effiency and user experience analysis of visual and audio feedback cues in human and service robot voice interaction cycle," in *IEEE WRC Symposium on Advanced Robotics and Automation (WRC SARA)*. IEEE, 2019, pp. 215–221.

[21] L. R. Macias, J. E. Aleman-Gallegos, U. Orozco-Rosas, and K. Picos, "Map based localization using an rgb-d camera and a 2d lidar for autonomous mobile robot navigation," in *Optics and Photonics for Information Processing XVII*, vol. 12673. SPIE, 2023, pp. 112–122.

[22] S. Devine, K. Rafferty, and S. Ferguson, "Real time robotic arm control using hand gestures with multiple end effectors," in *International Conference on Control (CONTROL)*. IEEE, 2016, pp. 1–5.

[23] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, T. Darrell, and K. J. Kuchenbecker, "Robotic learning of haptic adjectives through physical interaction," *Robotics and Autonomous Systems*, vol. 63, pp. 279–292, 2015.

[24] NaoDevils, "NAOImage." [Online]. Available: https://github.com/NaoDevils/NaoImage

[25] O. Kermorgant, "Ros2 bridge for baxter," https://github.com/CentraleNantesRobotics/baxter_common_ros2, 2025.

[26] A. Kumar, A. Sahasrabudhe, C. Perugu, S. Nirgude, and A. Murugan, "Kinematics & dynamics library for baxter arm," *arXiv preprint arXiv:2409.00867*, 2024.

[27] X. Wang and M.-A. Williams, "Pyride: An interactive development environment for pr2 robot," *arXiv preprint arXiv:1605.09089*, 2016.

[28] N. Ramezani and M.-A. Williams, "Smooth robot motion with an optimal redundancy resolution for pr2 robot based on an analytic inverse kinematic solution," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 338–345.

[29] J. Alas Escobar and A. Holm, "Controlling a turtlebot 2 through a web interface," 2016.

[30] N. T. Fitter and K. J. Kuchenbecker, "Equipping the baxter robot with human-inspired hand-clapping skills," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 105–112.

[31] RobotLab, "NAO AI Edition." [Online]. Available: https://www.robotlab.com/store/nao-ai-edition

[32] H. T. LTD, "Hailo-8l entry-level ai accelerator," *Hailo Technologies LTD*, 2025.

[33] Orbbec, "Gemini 2 L." [Online]. Available: https://www.orbbec.com/products/stereo-vision-camera/gemini-2l/

[34] Seek Thermal, "S314SPX," 2025. [Online]. Available: https://shop.thermal.com/Mosaic-Core-Starter-Kit-320x240-57HFOV-FF

[35] F. Grondin and F. Michaud, "Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations," *arXiv preprint arXiv:1812.00115*, 2018.

[36] Deepgram, "Deepgram real-time speech-to-text," *Deepgram*, 2025. [Online]. Available: https://deepgram.com/product/speech-to-text

[37] OpenAI, "Gpt-4o system card," *arXiv preprint arXiv:2410.21276*, 2024.

[38] "Amazon polly." [Online]. Available: https://aws.amazon.com/polly/

[39] J. K. Miller and W. Tang, "Evaluating llm metrics through real-world capabilities," *arXiv preprint arXiv:2505.08253*, 2025.

[40] Aldebaran, "ALTextToSpeech – NAOqi 2.8 Documentation." [Online]. Available: http://doc.aldebaran.com/2-8/naoqi/audio/altexttospeech.html#altexttospeech

[41] ——, "ALSpeechRecognition – NAOqi 2.8 Documentation." [Online]. Available: http://doc.aldebaran.com/2-8/naoqi/audio/alspeechrecognition.html

[42] K. A. Hallgren, "Computing inter-rater reliability for observational data: an overview and tutorial," *Tutorials in quantitative methods for psychology*, vol. 8, no. 1, p. 23, 2012.

[43] S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. Zahari, "Humanoid robot nao interacting with autistic children of moderately impaired intelligence to augment communication skills," *Procedia Engineering*, vol. 41, pp. 1533–1538, 2012.

[44] A. Amirova, N. Rakhymbayeva, E. Yadollahi, A. Sandygulova, and W. Johal, "10 years of human-nao interaction research: A scoping review," *Frontiers in Robotics and AI*, vol. 8, p. 744526, 2021.

[45] S. Rossi, S. J. Santini, D. Di Genova, G. Maggi, A. Verrotti, G. Farello, R. Romualdi, A. Alisi, A. E. Tozzi, and C. Balsano, "Using the social robot nao for emotional support to children at a pediatric emergency department: randomized clinical trial," *Journal of Medical Internet Research*, vol. 24, no. 1, p. e29656, 2022.

[46] A. E. Block, "How should robots hug?" Ph.D. dissertation, University of Pennsylvania, 2017.

[47] A. E. Block and K. J. Kuchenbecker, "Emotionally supporting humans through robot hugs," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 293–294.

[48] A. E. Block, S. Christen, R. Gassert, O. Hilliges, and K. J. Kuchenbecker, "The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception," in *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, 2021, pp. 380–388.

[49] A. E. Block, H. Seifi, O. Hilliges, R. Gassert, and K. J. Kuchenbecker, "In the arms of a robot: Designing autonomous hugging robots with intra-hug gestures," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 2, pp. 1–49, 2023.

[50] A. E. Block, "Huggiebot: An interactive hugging robot with visual and haptic perception," Ph.D. dissertation, ETH Zurich, 2021.

[51] G. Team, "Gemma 3n," 2025. [Online]. Available: https://ai.google.dev/gemma/docs/gemma-3n

[52] Qwen Team, "Qwen3 Technical Report," 2025. [Online]. Available: https://arxiv.org/abs/2505.09388