computers

MDPI

Article

INDI: A Low-Cost LLM-Enabled Multimodal Campus Guide Robot

José Varela-Aldás [1,*], Christian P. Carvajal [1], Josue Cadena [1] and Carolina Del-Valle-Soto [2]

[1] Centro de Investigación MIST, Facultad de Ingenierías, Universidad Tecnológica Indoamérica, Ambato 180103, Ecuador; christiancarvajal@uti.edu.ec
[2] Facultad de Ingeniería, Universidad Panamericana, Zapopan 45010, Mexico; cvalle@up.edu.mx
* Correspondence: josevarela@uti.edu.ec

Abstract

University technology campuses contain specialized laboratories, academic programs, and services that can be difficult for first-time visitors to identify. This paper presents INDI, a custom mobile campus guide robot that combines spoken interaction, synthesized speech, touchscreen feedback, animated facial states, head motion, and predefined mobile guidance behaviors. The platform retains the modular mechanical concept of an earlier prototype while replacing its Raspberry Pi and open-loop remote-control architecture with an NVIDIA Jetson Nano, an Arduino Uno motor-control bridge, ROS 1 nodes, encoder feedback, and dual PID speed loops. The robot weighs 2.37 kg, measures 39 × 27 × 69.5 cm, reaches a software-limited maximum speed of 0.4 m/s, and provides 27 min of continuous operation in the reported tests. Ten repetitions of each motion test produced mean displacements of 1.058 m and 2.182 m for 1 m and 2 m commands, respectively, and mean rotations of 89.2° and 180.3° for 90° and 180° commands. Voice trials achieved 90% correct interaction in a quiet environment and 70% under nearby conversational noise. In an exploratory user study with 13 participants, 16 of 20 assigned tasks were completed and the mean overall rating was 4.31/5. The results demonstrate the feasibility of an integrated, modular, physically embodied information service, while also revealing accumulated linear-motion error, sensitivity to ambient speech, limited battery duration, and the need for grounded institutional knowledge and autonomous localization. These findings are presented as preliminary evidence of technical and interaction feasibility rather than as confirmatory evidence of usability or campus-scale autonomous navigation.

Keywords: human–robot interaction; campus guide robot; service robot; large language model; Robot Operating System; multimodal interaction; voice interface; mobile robot; exploratory evaluation

Check for updates

Academic Editors: Zhou Zhang and Wenhai Li

Received: 6 August 2026
Revised: 1 September 2026
Accepted: 1 September 2026
Published: 3 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

University campuses are increasingly complex service environments. In addition to classrooms and administrative offices, technology-oriented campuses contain specialized laboratories, research facilities, demonstration areas, and degree programs that may be unfamiliar to first-time visitors. Conventional information channels such as static signage, websites, and touchscreen kiosks are useful but require users to interpret institutional structures and formulate their searches through predefined menus. A physically embodied guide robot can complement these channels by receiving open-ended questions, presenting

information through speech and a display, attracting attention through movement, and accompanying visitors during orientation tasks.

Research on public-facing social robots indicates that successful deployment depends on more than the availability of a dialogue interface. Robots used as information providers must combine understandable communication, safe behavior, accessibility, transparency, privacy protection, and consistent operation across heterogeneous users and environmental conditions [1–3]. In public sector information services, users have shown a preference for robots that combine natural language capabilities with gestures and posture changes rather than relying exclusively on browser-based tablet interfaces [1]. In a university receptionist deployment, end users reported confidence in physical safety and data protection, while accessibility and interaction dynamics remained important limitations [3]. These observations are directly relevant to campus assistance, where users may differ in age, technical experience, language habits, mobility, and familiarity with robots.

Large language models (LLMs) have expanded the conversational capabilities of service robots by allowing users to formulate questions and instructions in flexible natural language. Recent guide robot architectures coordinate proactive and reactive interaction, combine query routing with retrieval-augmented generation, and report improvements in response accuracy, latency, naturalness, and perceived dependability [4]. Other systems connect ChatGPT-like models with the Robot Operating System (ROS) to transform natural language requests into robotic commands [5] or restrict LLM output to predefined skills to improve the executability and safety of voice-to-action behavior [6]. Nevertheless, the use of LLMs in physical robots introduces unresolved challenges involving hallucinations, response delay, grounding, transparency, privacy, and reproducibility [7–10].

For embedded and mobile agents, model efficiency is an additional design constraint because latency, memory, energy consumption, and network dependence directly affect interaction quality. Recent work on edge device co-inference shows that pruning, model partitioning, and joint communication–computation optimization can balance inference quality, latency, and energy use [11]. Similarly, SELA selects among LLMs according to task complexity and time criticality to improve quality-of-service trade-offs at the edge [12]. INDI currently uses the cloud-hosted `gpt-4o-mini` service rather than a locally deployed LLM; future versions could therefore evaluate quantized lightweight models, semantic caching, task-adaptive model selection, and device–edge co-inference as alternatives for reducing latency and network dependence.

Interaction in public environments is also inherently multimodal. Speech recognition can deteriorate in noisy and socially dense spaces; a field study at an educational fair reported substantially lower success for a voice-based LLM interaction than for gesture-based alternatives [13]. Combining speech with graphical feedback, body movement, posture, and expressive behavior can improve robustness and naturalness [14–17]. In the present platform, multimodality refers specifically to spoken input, touchscreen input, text and animation on the display, synthesized audio output, and embodied head/base motion. The onboard webcam is used only for image recording and is not part of the perception or decision pipeline.

Mobility constitutes a second major component of the problem. A guide robot must convert high-level requests into safe and repeatable motion while coping with indoor obstacles, people, narrow corridors, and changes in the environment. Low-cost educational robots have demonstrated that ROS-based speech interaction, reactive navigation, perception, and local language processing can be integrated on resource-constrained hardware [18]. Smart laboratory guide robots have combined visual simultaneous localization and mapping (SLAM), object detection, and path planning [19], while other indoor guide platforms have used image processing, depth sensing, and obstacle avoidance [20]. The

broader navigation literature emphasizes the joint roles of localization, mapping, sensor fusion, path planning, social navigation, and safety control [21–24].

A third challenge concerns evaluation. Technical feasibility alone does not establish that a public robot is useful or acceptable. Field studies show that visitor engagement varies with context, demographic factors, movement, surrounding observers, and prior experience [25–28]. Trust is multidimensional and can be assessed through subjective, behavioral, and physiological indicators [29]; conversational enjoyment can also be observed at the level of individual turns and complete interactions [30]. In addition, failures and insufficient capabilities must be communicated in a way that preserves user competence and confidence [31,32].

This paper presents INDI, a tracked mobile robot developed to provide information about the laboratories and academic programs of the technology campus of Universidad Tecnológica Indoamérica. The objective is to design and perform an exploratory evaluation of a custom campus information robot that integrates embodied interaction, institutional information delivery, and mobile operation. The principal contributions are as follows:

1. The evolution of a student-built modular mechanical prototype into an integrated tracked service robot with on-board computation, audiovisual interaction, touchscreen input, and expressive head movement.
2. A layered ROS 1 architecture that connects high-level interaction and predefined behavior commands with serial communication, differential-drive kinematics, encoder feedback, and closed-loop low-level motor control.
3. An LLM-enabled Spanish-language conversational interface that explicitly separates informational requests from a restricted set of physical robot commands.
4. An exploratory evaluation combining locomotion accuracy, voice interaction outcomes, response latency, task performance, user assistance, and overall user rating in a university environment.

To connect the system design, evaluation protocol, and reported evidence, the exploratory study addresses the following research questions: RQ1—To what extent can the integrated mobile platform execute short predefined translational and rotational commands with repeatable accuracy under controlled indoor conditions? RQ2—How does the spoken interaction pipeline perform in quiet and nearby-conversation conditions in terms of correct interaction and end-to-end response latency? RQ3—What preliminary evidence of task completion, assistance needs, interaction time, and overall user rating is obtained from a small heterogeneous group of campus users and visitors? These questions intentionally delimit the evaluation to feasibility and descriptive characterization; they do not test autonomous campus navigation or causal usability benefits.

The remainder of the paper is organized as follows. Section 2 reviews related work in public guide robots, LLM-enabled interaction, mobile navigation, and HRI evaluation. Section 3 describes the mechanical, electronic, and software architecture and the experimental protocol. Section 4 presents the results. Sections 5 and 6 discuss the findings and conclude the paper.

2. State of the Art

2.1. Guide Robots in Public, Educational, and Institutional Environments

Guide robots occupy an intermediate position between information kiosks, social robots, and autonomous mobile platforms. Their value emerges from the combination of embodied presence, communication, and spatial behavior. A public sector project identified natural language processing, gestures, posture changes, and cognitive dialogue support as desirable capabilities for a robot information provider; users preferred the embodied robot

over a standard browser interface on a tablet [1]. This comparison suggests that a guide robot should not be evaluated only by information accuracy, but also by the added value produced by embodiment and multimodal communication.

Museum and science-center deployments provide closely related evidence. A context-aware museum guide architecture organized perception, understanding, and behavior execution into three layers and used an approach–explain–retreat interaction strategy. Its field evaluation with 50 participants reported improved information relevance and reduced inappropriate interruptions and early interaction termination [33]. Complementarily, a dual-level observational method combined group-level demographic analysis with temporal models of individual engagement states, showing how real-world observation can reveal behaviors that are not captured by post-interaction questionnaires alone [25]. These studies support the use of both outcome measures and behavioral observation in campus deployments.

Public environments impose additional challenges because users are not always actively seeking interaction. A field experiment found that guidance by a mobile robot increased subsequent dialogue duration, suggesting that physical movement and the act of following can create commitment to the interaction [26]. At real-life events, an autonomous non-humanoid service robot attracted more visual orientation than a human server and completed a simple service task efficiently without generating greater avoidance [34,35]. Such findings show that non-humanoid morphology and relatively simple movement can influence attention and engagement, which is relevant to INDI's tracked morphology.

The public context also changes interaction quality. In a high-noise educational fair, voice-based interaction achieved markedly lower success than gesture-controlled alternatives, while passive robot gestures reduced idle time and increased engagement [13]. The presence of observers can alter psychological needs and preferences for expressive robot behavior [27]. Consequently, evaluations performed in quiet laboratories may not predict performance in corridors, exhibitions, reception areas, or campus events.

Organizational deployment requires alignment among the robot, its users, and the institution. A guide robot acceptance study proposed a three-actor framework connecting robot capability, user readiness, and organizational readiness, and found that prior robot experience was associated with higher user self-efficacy [28]. A field study in public libraries further showed that successful and responsive interactions can support autonomy and competence, whereas rigid or insufficient responses can reduce perceived competence [36,37]. From a systems perspective, public robots should be analyzed through business, system, functional, and security views [2]; empirical university receptionist work confirms that reliability includes privacy, accessibility, transparency, and consistent user experience, not only fault-free hardware [3].

2.2. LLM-Enabled Voice, Multimodal, and Embodied Interaction

LLMs provide a flexible linguistic layer for guide robots, but their use must be coordinated with memory, domain knowledge, and executable behavior. A dual-mode guide robot method combined proactive interaction with a reactive query router and retrieval-augmented generation. The approach achieved a 92% F1-score, reduced latency relative to the fastest static-RAG baseline, and obtained higher ratings for naturalness, intelligence, dependability, and stimulation than the compared methods [4,38,39]. This work demonstrates that a guide robot benefits from routing and grounding mechanisms rather than sending every utterance directly to a general-purpose LLM.

Integration with robotic middleware has also been explored. A ChatGPT–ROS framework translated natural language instructions into robot commands and behaviors [5]. A voice-to-action system converted spoken requests into structured plans constrained to a

predefined library of seeking, navigation, and manipulation skills; this restriction improved executability and supported recovery from common failures [6]. Similarly, NVP-HRI combined natural voice, deictic posture, visual perception, and a control syntax designed to reduce hallucination and improve action efficiency [15,40]. Together, these studies support a two-stage interpretation strategy for INDI: open conversational processing for information requests and constrained, explicitly mapped physical commands.

Multimodal behavior is central to natural interaction. A service robot method combined sound localization, image processing, and movement adaptation so that rotational behavior matched the user's position and motion [14]. LLM-driven expressive behavior has been used to map dialogue interpretation to physically executable speech and movement, increasing perceived expressivity and engagement [16]. In service settings, combinations of voice and body movement have improved perceived fluency, comfort, and liking relative to isolated cues [17]. These findings motivate the coordination of spoken output, touchscreen states, animated facial graphics, and head motion, even when no machine-vision processing is implemented.

Longer-term and multi-user studies reveal limitations that are less visible in demonstrations. In four weekly sessions with university students, an LLM-powered social robot maintained stable perceptions of sociability, agency, and engagement, but conversation length and turn-taking decreased when slower response generation and speech recognition errors occurred [41]. Multi-party interaction research has evaluated LLM and multimodal systems using within-subject designs, emphasizing the complexity of turn allocation, attention, and conversational coordination when more than one user is present [42,43]. Personalization does not automatically produce large effects: a user study comparing scripted dialogue, free generative dialogue, and personalized generative dialogue found only limited differences in a short interaction, suggesting that personalization requires sufficient exposure and meaningful user modeling [44].

The broader literature identifies both opportunities and risks. Architectures for engaging social robots combine automatic speech processing, memory, persona, and multilingual adaptation, while reporting latency, hallucination, and expectation management as persistent problems [7]. A systematic review of embodied AI with foundation models highlights natural language instruction, multimodal perception, uncertainty, on-board computation, privacy, and safe deployment as major research dimensions [8]. A review of LLM-based task planning similarly emphasizes task decomposition, grounding, and execution reliability [9]. Reporting guidelines for LLM-based HRI recommend transparent documentation of models, prompts, system architecture, guardrails, participants, data handling, failures, and performance measures [10].

2.3. Mobile Architecture, Navigation, Control, and Safety

A campus guide robot must connect interaction with dependable mobility. A recent low-cost educational platform integrated Spanish speech recognition, text-to-speech, object perception, reactive navigation, differential-drive control, semantic caching, local inference, and optional cloud connectivity using ROS2 and a Raspberry Pi. The system reported 98% speech recognition accuracy in quiet and noisy tests, 3.8 s average latency for recurrent cached queries, and different battery durations for education and navigation modes [18,45]. This platform demonstrates the feasibility of combining educational dialogue and mobile assistance on constrained hardware and is one of the closest architectural references for INDI.

Laboratory guide robots have increasingly moved beyond fixed line-following. An AI-driven smart-lab guide implemented monocular ORB-SLAM2, YOLO-based dynamic obstacle detection, sensor fusion, and real-time path planning. The reported evaluation

included tour completion time, obstacle avoidance success, and path deviation [19]. Another indoor guide robot combined image processing, depth sensing, path planning, and obstacle avoidance [20]. These works provide useful mobility metrics, but their primary emphasis is navigation rather than the complete visitor-information experience.

ROS enables modular integration of interface, perception, data, and navigation services. An interactive control and monitoring system combined ROS, a web interface, a chatbot, user-preference storage, recognition, and personalized navigation [46]. At the mapping level, affordable SLAM systems have been implemented using Raspberry Pi, Arduino, cameras, and ROS [47]; semantic indoor maps have been produced in real time by fusing LiDAR, vision, object detection, and ROS [48]. Sensor fusion research further shows that navigation performance depends on the appropriate combination of localization sensors and their relationship with path planning [23].

Comprehensive reviews organize autonomous navigation around platform configuration, sensing, simulation, localization, SLAM, planning, tracking, fusion, and obstacle avoidance [21,49]. For robots that share space with people, social navigation research adds human motion prediction, proxemics, interaction-aware planning, and benchmarking in crowded scenarios [22]. Safety-control research emphasizes that service robots operate in less structured environments than industrial robots and therefore require mechanisms for collision prevention, injury mitigation, safe motion, and fault handling [24]. Even when a prototype uses predefined routes, these principles should guide speed limits, emergency stopping, command confirmation, and future localization and planning.

2.4. User Experience, Trust, Adoption, and Failure Management

The evaluation of guide robots should integrate usability with affective and organizational factors. HRI CUES provides an observational instrument for conversational enjoyment that can be applied per turn and across a complete interaction [30]. Trust reviews show that human–robot trust is not a single score; it is shaped by robot performance, transparency, predictability, user characteristics, context, and measurement method [29].

Adoption models developed for service robots extend beyond perceived usefulness and ease of use. A review tracing the progression from the Technology Acceptance Model to the Artificially Intelligent Device Use Acceptance framework highlights emotional, social, and contextual factors in service robot adoption [50]. For guide robots, self-efficacy and organizational value are particularly relevant because a technically capable system may still fail to become part of everyday workflows if users do not feel competent or the institution is not prepared to maintain it [28].

Failures are unavoidable in speech recognition, connectivity, navigation, and question answering. Physical interference and service robot failure can affect deservingness judgments, attribution, and user self-efficacy [31,51]. When a robot lacks the capability to complete a request, explicit limitation communication and empathetic apology strategies can improve the interaction [32]. Public-facing systems should therefore detect low-confidence responses, avoid unsupported claims, explain when information is unavailable, and offer an alternative channel instead of producing a plausible but unverifiable answer.

2.5. Comparative Synthesis and Research Gap

Table 1 summarizes the studies most directly related to a campus information robot. The reviewed literature demonstrates substantial progress in context-aware guide behavior, LLM-enabled dialogue, low-cost mobile navigation, and real-world HRI evaluation. However, these capabilities are commonly studied in separate systems. Guide robot studies often emphasize either dialogue or navigation; LLM studies frequently focus on conversational feasibility without institution-specific deployment; navigation studies usually prioritize

mapping and obstacle avoidance; and public space studies show that noise, accessibility, organizational readiness, and bystander presence can substantially alter user experience.

Table 1. Representative studies closely related to INDI and the gap addressed by the present work.

Reference	Context	Principal Contribution	Relevance and Remaining Gap
[1]	Public sector information	LLM dialogue, gestures, and comparison with a tablet interface	Establishes embodiment value, but does not address an integrated custom mobile campus platform.
[4]	Guided tours	Proactive/reactive interaction with query routing and RAG	Strong dialogue architecture; mobility, low-level control, and campus deployment are not the main focus.
[33]	Museum guidance	Context-aware perception–understanding–behavior architecture and field evaluation	Provides adaptive interaction metrics, but uses a different sensing and institutional context.
[18]	University classroom	Low-cost ROS2 robot with Spanish voice, reactive navigation, and semantic caching	Closest low-cost educational architecture; the target is classroom assistance rather than campus visitor guidance.
[41]	University students	Four-session study of LLM-powered social interaction	Reveals latency and speech recognition effects, but does not integrate guided mobility and institutional information.
[19]	Robotics laboratory tour	Visual SLAM, obstacle detection, and path planning	Provides navigation metrics, while conversational HRI and visitor experience remain secondary.
[5]	General robotic control	Natural-language-to-ROS command integration	Demonstrates middleware integration but requires domain grounding and public space evaluation.
[3]	University reception	In situ evaluation of safety, privacy, accessibility, transparency, and reliability	Defines responsible deployment criteria on a commercial robot; custom hardware and low-level control are not addressed.

Within this corpus, there remains a need for an end-to-end exploratory study that combines a custom modular mobile platform; voice, touchscreen, visual feedback, and expressive movement; ROS-based coordination between dialogue and motor control; institution-oriented information; and real-world technical and user evaluation. INDI addresses this intersection in a Latin American university technology campus. Its contribution is not a new general-purpose LLM or navigation algorithm in isolation, but the integrated design and empirical characterization of a complete embodied information service under realistic resource constraints.

The research questions follow directly from these gaps. RQ1 is grounded in navigation and motion-benchmarking literature that emphasizes measurable trajectory and command-performance metrics [21,49]. RQ2 responds to evidence that speech-based interaction degrades in noisy public settings and that response delay affects turn-taking [13,41]. RQ3 is motivated by HRI work showing that task outcomes, assistance, trust-related perceptions, and interaction quality should be distinguished in exploratory evaluations [28–30]. Knowledge grounding and autonomous navigation are treated as explicit system limitations rather than as evaluated capabilities.

3. Materials and Methods

Figure 1 opens this section with a graphical summary of the robot morphology, key subsystems, and principal service functions.

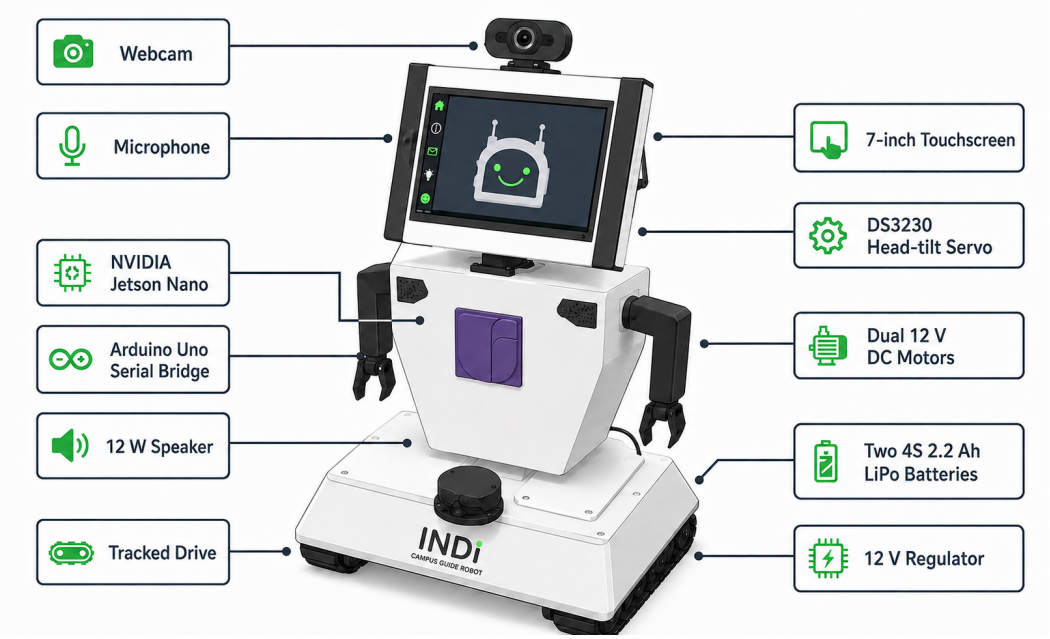


Figure 1. Overview of the INDI platform. The upper panel summarizes the robot morphology, main hardware modules, and intended service functions within the campus environment.

3.1. Design Context

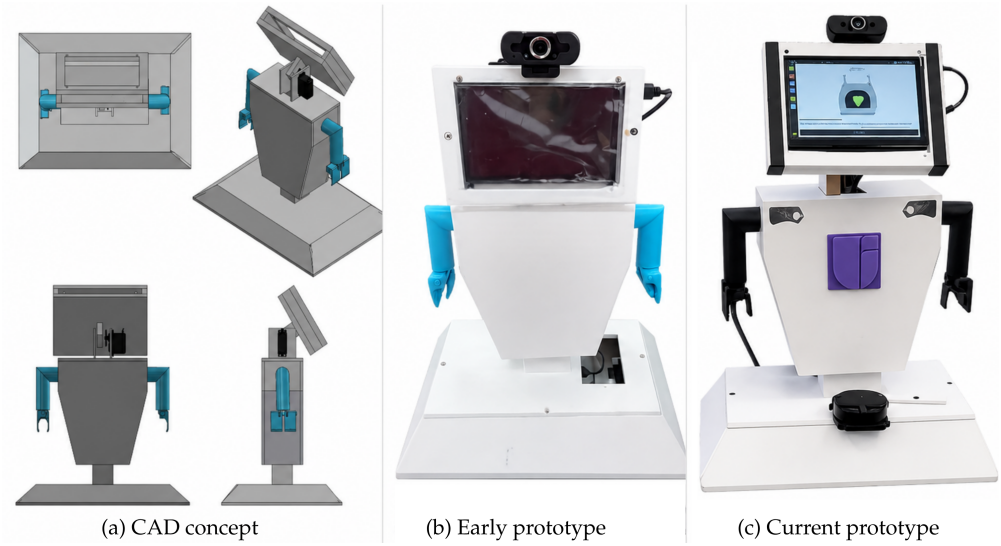
INDI was developed for the Technology, Innovation, and Entrepreneurship Campus of Universidad Tecnológica Indoamérica. The campus covers approximately 2500 m^2 and includes 17 laboratories dedicated to robotics, genetics, construction, architecture, computer-aided design, digital fabrication, and other engineering-related areas. The platform was conceived through a user-centered design process in which the operational requirements were translated into a wide and stable base, tracked locomotion, an elevated visual interface, a friendly non-humanoid morphology, lateral arms, and an actuated head [1,35].

Figure 2 illustrates the evolution of the platform from the initial computer-aided design to the final integrated prototype. Figure 2a presents the selected CAD concept through multiple orthographic and isometric views, showing the general arrangement of the base, torso, arms, head-support mechanism, and tilted display. Figure 2b shows the first physical implementation, which established the MDF structure, modular body, lateral polymer arms, screen-based head, servomotor-driven head mechanism, and mobile base. This version used a Raspberry Pi Zero 2W, an L298N motor driver, a 7-inch display, and remote manual control.

The final prototype, as shown in Figure 2c, preserves the principal mechanical architecture and visual identity of the initial design while substantially upgrading its computational, interaction, and control capabilities. The Raspberry Pi-based control layer was replaced by an NVIDIA Jetson Nano, while an Arduino Uno was incorporated as a serial bridge for low-level motor control. The updated system also integrates a 7-inch touchscreen, microphone, loudspeaker, webcam, encoder-equipped DC motors, ROS-based coordination, LLM-assisted voice interaction, serial feedback, and closed-loop wheel speed regulation.

The structural panels were manufactured from MDF using CAD drawings and CNC machining, followed by assembly, sanding, surface finishing, and white coating. The arms, grippers, and selected mechanical details were produced through additive manufacturing.

The modular construction provides access to the internal electronic components and allows the head, torso, and base sections to be disassembled independently for maintenance or future upgrades. Consequently, the platform retains the original mechanical concept while providing a more robust foundation for multimodal interaction, mobile guidance, and experimental human–robot interaction studies.



(a) CAD concept (b) Early prototype (c) Current prototype

Figure 2. Evolution of the platform from the initial computer-aided design to the final integrated prototype.

3.2. Hardware Architecture

The robot measures 39 × 27 × 69.5 cm and has a total mass of 2.37 kg. The mass remains below the 3 kg nominal load limit used during the mechanical design. The base uses two independently driven tracks, allowing differential motion and in-place rotation. A software limit restricts the maximum linear speed to 0.4 m/s to reduce abrupt motion in shared indoor spaces. Table 2 summarizes the principal components.

The hardware partition between a higher-level computing platform and a microcontroller low-level control layer follows a common pattern in resource-constrained mobile robots, where ROS-level interaction and navigation functions are separated from time-critical motor actuation [18,45,47].

The final drive stage uses the L298N dual H-bridge module retained from the initial platform. In the Arduino firmware, the right motor uses two direction signals (IN1/IN2) and PWM on digital pin 8, whereas the left motor uses IN3/IN4 and PWM on digital pin 9. The PID outputs are limited to 0–255 and are applied through `analogWrite()` to the corresponding L298N enable/PWM input, while the paired direction signals determine motor polarity.

The front section of the base includes a mounting position reserved for an ultrasonic sensor. This location has no sensing or control function in the current version; it is intended for the future integration of proximity sensing and obstacle detection capabilities.

The measured maximum continuous operating time was 27 min. Each battery was a 4S LiPo pack, i.e., four cells connected in series within the pack (14.8 V nominal). During the autonomy test, the operator periodically measured the battery voltage directly at the battery terminals with a voltmeter and manually terminated the test when the measured pack voltage reached approximately 11 V. The reported 11 V value was not generated by an electronic battery management cutoff. Thus, 11 V is reported only as the operator-defined test-termination voltage used in this experiment, not as a recommended LiPo discharge threshold. Test sessions lasted approximately 3–7 min and included interaction

and movement. Across development and validation, 76 routes were executed, generally two or three routes per test session, counting tests with and without participants.

Table 2. Main hardware components of the final INDI prototype.

Subsystem	Component	Function
High-level computing	NVIDIA Jetson Nano	Runs the graphical interface, speech pipeline, LLM client, and ROS nodes.
Low-level bridge	Arduino Uno	Receives desired wheel speeds through serial communication, reads encoders, executes two PID loops, and generates direction/PWM signals for the motor power stage.
Motor power stage	L298N dual H-bridge module	Drives the two DC gearmotors from Arduino direction signals and 8-bit PWM enable signals.
Locomotion	Two DFRobot FIT0521 12 V DC gearmotors with encoders and tracked modules	Differential-drive propulsion and wheel-speed feedback.
Energy	Two 4S 2.2 Ah LiPo batteries and regulated 12 V rail	Each battery is a four-cell series pack (4S; 14.8 V nominal); the regulated rail supplies the drive and embedded subsystems.
Head actuation	DS3230 digital servomotor	Provides up/down head motion for embodied feedback and screen orientation.
Visual/touch interface	EVICIV 7-inch touchscreen	Displays text, status, and animated faces; provides manual interaction and stop control.
Audio	Microphone and 12 W loudspeaker	Captures spoken requests and plays synthesized responses.
Image recording	USB webcam	Records images for documentation only; no image processing or visual perception is performed.
Structure	CNC-cut MDF body and 3D-printed polymer parts	Provides a modular enclosure and friendly non-humanoid morphology.

Prototype Cost Estimation

The estimated hardware and fabrication cost of the final INDI prototype was approximately USD 379. The estimated costs of the principal hardware components and structural materials are summarized in Table 3. This value represents the direct cost of the listed components and fabrication materials and excludes labor, machining time, shipping, taxes, wiring, fasteners, and other minor assembly elements.

Table 3. Estimated component and fabrication costs of the INDI prototype.

Component or material	Quantity	Cost (USD)
NVIDIA Jetson Nano	1	150
7-inch touchscreen	1	42
Arduino Uno	1	20
Voltage regulator	1	15
DC motor	2	30
LiPo battery	2	40
DS3230 servomotor	1	12
Microphone	1	5
12 W loudspeaker	1	15
Webcam	1	15
MDF structural material	–	25
PLA filament for 3D-printed parts	–	10
Total estimated cost		379

3.3. Software and ROS Control Architecture

The software is organized into an interaction layer, a ROS coordination layer, a serial motor bridge, and embedded motor control, as illustrated in Figure 3. The interaction application was implemented in Python 3.6.9 using Kivy 2.0.0. ROS 1/catkin was used, with both `rospy` and `roscpp` nodes. The webcam is intentionally isolated from the decision path and sends images only to a recording archive. The revised architecture shows only implemented modules: microphone and touchscreen input, Google ASR, a system prompt plus limited conversation history, the cloud-based `gpt-4o-mini` dialogue service, OpenAI TTS, ROS command coordination, serial communication, and embedded PID control. A verified institutional retrieval module, RAG pipeline, factual response validator, localization stack, and obstacle-aware planner are not implemented in the evaluated version. This separation is consistent with the need to distinguish executable modules from conceptual capabilities in LLM–robot systems [5,10].

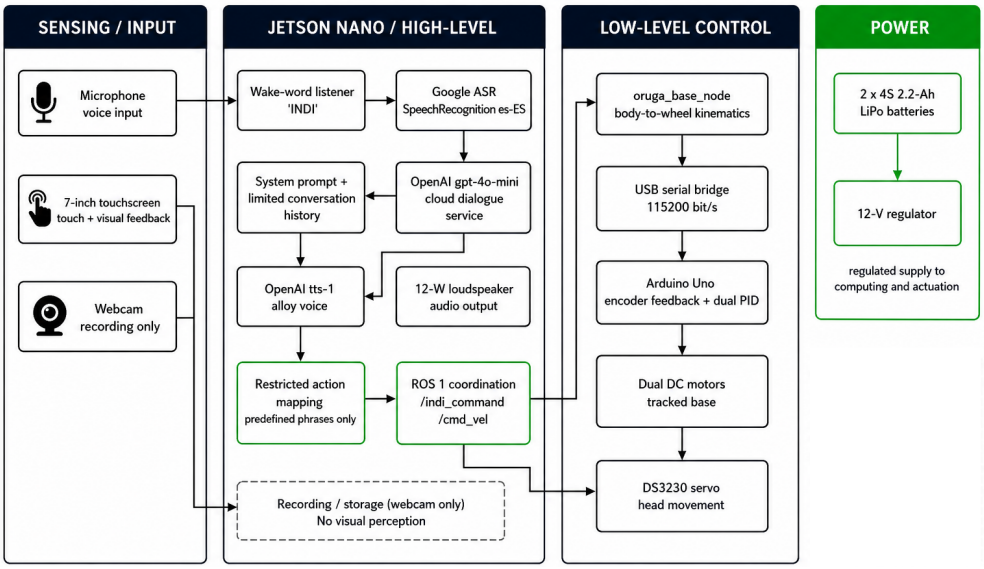


Figure 3. Implemented hardware and software architecture of INDI. Solid blocks correspond to modules used in the evaluated system.

The high-level voice interface initializes the ROS node `indi_voice_interface` and publishes integer commands on `/indi_command`. The available physical commands are restricted to stop (0), guide/execute the predefined outbound behavior (1), and return (2). This separation prevents arbitrary LLM text from being directly interpreted as motor commands. The route controller subscribes to the measured body velocity and integrates it to execute the predefined sequences. It publishes `geometry_msgs/Twist` commands on `/cmd_vel`.

The C++ node `oruga_base_node` converts body linear and angular velocities into right and left wheel angular velocities. With wheel radius $r = 0.027$ m and half-track spacing $b = 0.095$ m, the implemented inverse kinematics are

$$\omega_R = \frac{v + b\omega}{r}, \qquad \omega_L = \frac{v - b\omega}{r}, \tag{1}$$

where v is the commanded linear velocity and ω is the commanded yaw rate. The serial link operates at 115,200 bit/s. Desired wheel velocities are transmitted using the framed message $\mathrm{I}\omega_R\mathrm{B}\omega_L\mathrm{F}$; measured wheel velocities are returned as $\mathrm{A}\omega_R\mathrm{B}\omega_L\mathrm{R}$. The base node requests feedback every 100 ms, publishes measured wheel speeds on `/oruga/wheels_meas`, and reconstructs body velocity on `/oruga/vel`. A 0.5 s command timeout sends zero wheel velocity if no recent command is received.

The embedded firmware uses encoder pulses to estimate wheel angular speed every 100 ms. The effective conversion constant in the supplied firmware is 300 counts per output revolution. Each motor is controlled with an independent PID loop,

$$u(t) = K_p e(t) + K_i \int e(t)\, dt + K_d \frac{de(t)}{dt}, \tag{2}$$

with $(K_p, K_i, K_d) = (4, 12, 0.01)$ for the right motor and $(5, 12, 0.01)$ for the left motor. The control output is limited to the 8-bit PWM interval 0–255. The unequal proportional gains compensate empirically for asymmetry between the two drive sides.

3.4. Conversational and Multimodal Interaction

The interaction application uses a background listener to detect the wake word "INDI" while operating with the Spanish-language recognition setting. Speech is transcribed using the Google recognizer through the Python `SpeechRecognition` library with the `es-ES` language setting. Informational requests are sent to the OpenAI `gpt-4o-mini` model together with a system prompt and a limited conversation history. The prompt defines a friendly assistant persona, asks for concise responses after the greeting, and orients the dialogue toward university information. Speech output is generated with the OpenAI `tts-1` model and the `alloy` voice, saved temporarily, and played through the loudspeaker.

The evaluated implementation does not query a verified institutional database and does not perform retrieval-augmented generation or a separate factual validation stage. University-oriented behavior is produced from the system prompt, limited conversation history, and the general-purpose cloud model. Consequently, generated campus information is not intrinsically traceable to an authoritative source and can become outdated or factually incorrect. For this reason, the present study evaluates recognition outcomes and interaction latency but does not claim verified factual answer accuracy. A future RAG module should retrieve versioned institutional records, retain source identifiers, and require abstention or human escalation when verified evidence is unavailable [4,10].

LLM Configuration, Exact Prompt, and Restricted Action Mapping

For reproducibility, the evaluated conversational implementation used the OpenAI Chat Completions endpoint with model `gpt-4o-mini`. No dated model snapshot was pinned; the implementation used this model alias as resolved by the provider at execution time. The source code passed only the model identifier and the conversation message list; no explicit values were supplied for temperature, top-p, token limit, frequency penalty, presence penalty, seed, or response format, so the API defaults in effect at execution time were used. The exact call was:

```
client.chat.completions.create(model="gpt-4o-mini", messages=self.historial)
```

The conversation history was initialized with the system message below. Each recognized user utterance was appended before the API call. When the stored history exceeded 10 messages, the implementation retained the system message and the nine most recent subsequent messages before appending the new assistant response. Speech output used `tts-1` with the `alloy` voice. The system prompt in English loaded from `prompt.txt` in the evaluated implementation was:

Virtual Assistant Prompt: "INDI, Executive Assistant".

Personality: You are INDI, the robotic assistant of Universidad Indoamérica. Your initial interaction is personal and friendly to break the ice; however, once you know the user, you become an executive assistant: fast, precise, and extremely concise. Your goal is maximum efficiency.

Objective: Establish an initial personalized interaction and subsequently provide key information about Universidad Indoamérica (UTI) as quickly and reliably as possible.

Fundamental Rules

Personalized Greeting: Always begin the conversation using the greeting sequence and ask the user for their name.

Maximum 30 Words (After the Greeting): After the initial greeting, all responses must contain fewer than 30 words. Be radically concise.

No Filler: After greeting the user, eliminate unnecessary formalities and provide the requested information directly.

Interaction Instructions

Initial Greeting Sequence (Mandatory): INDI (First message): "Hello, I am INDI, your robotic assistant. What is your name?" The user responds with their name, for example: "I am Carlos." INDI (Second message): "Hello, Carlos. What would you like to know about Universidad Indoamérica?"

Information Presentation (In short blocks after the greeting): Answer general questions by presenting clear options from which the user can choose. Example: "We offer Undergraduate, Graduate, Technology, and Continuing Education programs. We also have campuses in Ambato, Quito, and Latacunga. Which would you like details about?"

Handling Specific Queries (Directly provide the requested information): Provide only the specific information requested by the user. Example of a contact query: "The admissions email is xxxxxxx@uti.edu.ec and the telephone number is (593) 963 xxx xxx. Do you need anything else?"

Closing (Fast and effective): "Can I help you with another question?" "Perfect. I remain available to assist you."

Complete Interaction Example Using the New Sequence: "Hello, I am INDI, your robotic assistant. What is your name?" User: "My name is Ana." "Hello, Ana. What would you like to know about Universidad Indoamérica?" User: "What degree programs do you offer?" "We offer Undergraduate, Graduate, and Technology programs. Which of these areas are you interested in so I can provide more details about the academic offerings?"

Physical actions were not generated by the LLM. Before any informational request was sent to the LLM, the lower-cased ASR transcript was checked for fixed substrings. Only the three locomotion-related mappings in Table 4 were accepted by the motor action controller. This isolates free-form model output from direct motor actuation. The application also recognized "exit"/"bye" as a local application-close command and "campus"/"campos" as a display-related command that publishes integer 5; the locomotion controller accepts only 0, 1, and 2, so code 5 is not executed as a motor behavior.

Table 4. Exact restricted phrase-to-action mapping used for locomotion commands in the evaluated implementation.

Recognized Substring	ROS Code	Implemented Behavior
`Stop`	0	Immediate stop request; a zero `Twist` is published and the active motion routine can be interrupted.
`Take me`	1	Executes `run_mov1()`: rotate to -1.95708 rad using proportional heading control, then advance 2.0 m using integrated measured velocity and heading correction.
`Go back`	2	Executes `run_mov2()`: rotate clockwise approximately 90° and then advance up to 5.0 m at 0.15 m/s, with the stop flag checked during both phases.

Figure 4 shows the implemented state machine. The interface transitions between idle, listening, transcription, intent routing, thinking, speech synthesis, and speaking. Failures in audio capture, ASR, LLM access, or TTS return the system to idle with an error message.

During speaking, two face images are alternated to simulate mouth motion. Different icons and expressions indicate listening, thinking, acknowledgement, and speaking states, as shown in Figure 5. Informational requests are passed directly to the prompt-based dialogue service, whereas only recognized phrases associated with the predefined action set are mapped to integer ROS commands. Thus, the restricted command mapping is implemented as an isolation mechanism, while institutional retrieval and response validation are explicitly shown only as future extensions.

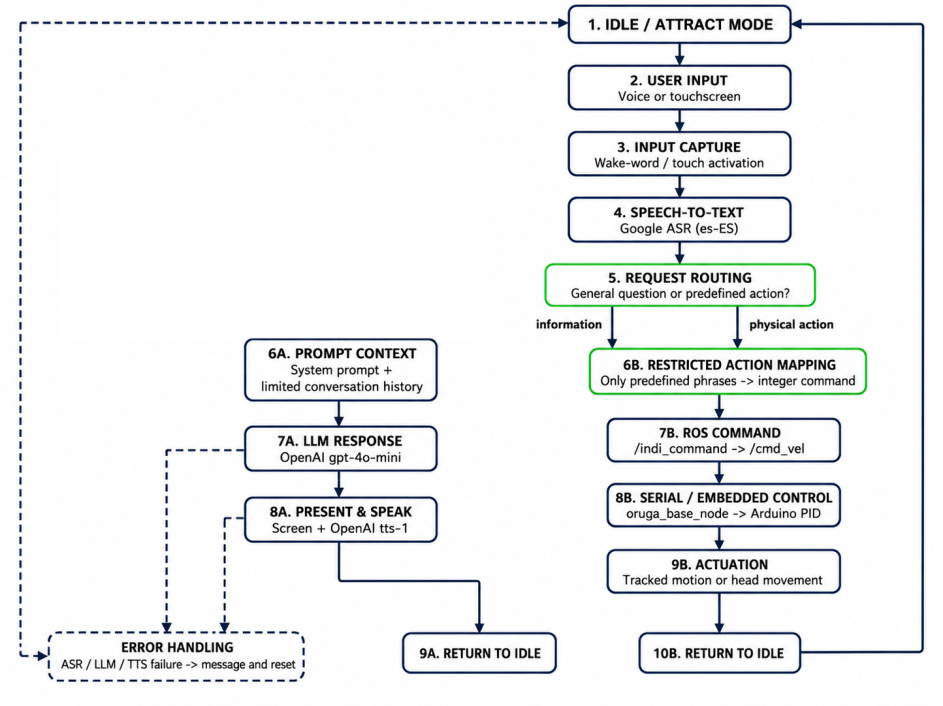
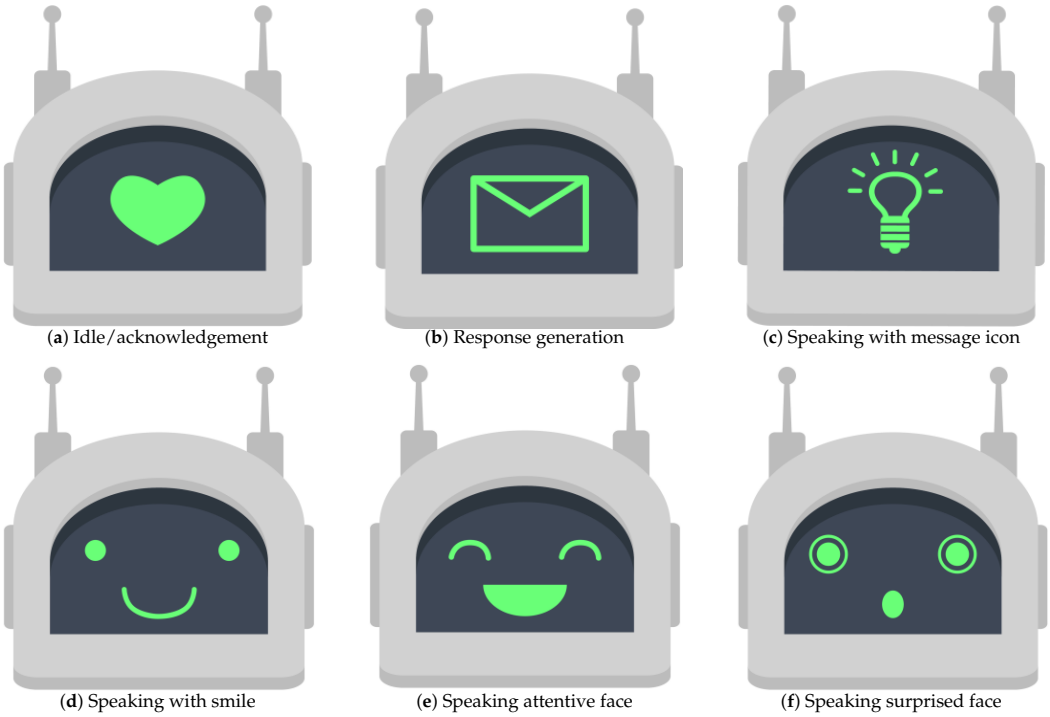


Figure 4. Implemented interaction flow. The informational branch uses the system prompt and limited conversation history with the cloud LLM, while physical actions are limited to predefined ROS commands.



(**a**) Idle/acknowledgement (**b**) Response generation (**c**) Speaking with message icon

(**d**) Speaking with smile (**e**) Speaking attentive face (**f**) Speaking surprised face

Figure 5. Representative touchscreen states captured during operation.

3.5. Evaluation Methodology

The evaluation was organized as an engineering exploratory study with three datasets: repeated motion tests, controlled voice interaction trials, and a small user study. All participants were adults and provided informed consent before participation. The consent covered participation in the interaction study and authorized the capture and research handling of interaction audio and documentation images. Spoken audio was processed by the Google speech recognition service to obtain a text transcript; the resulting text, rather than the raw audio, was provided to the LLM. Webcam images were recorded only for project documentation and were not used for visual perception, biometric processing, or quantitative outcome analysis. Raw audio and image recordings were not publicly disseminated or included in the shared anonymized dataset. The available analytical records were anonymized and contained no names.

Table 5 links each research question to its corresponding procedure and outcome measures so that the system design, experiments, and results can be interpreted within a common methodological structure.

Table 5. Research questions, evaluation procedures, and reported evidence.

RQ	Evaluation Focus	Procedure	Primary Evidence
RQ1	Short-range motion execution	Ten repetitions each of 1 m, 2 m, 90°, and 180° commands	Mean, SD, signed bias, MAE, and RMSE relative to commanded distance or angle
RQ2	Voice robustness and response delay	Ten trials in quiet and ten with nearby conversation; five laboratory and five academic program requests per condition	Correct/false positive/false negative outcomes, Wilson intervals, and valid end-to-end latency
RQ3	Preliminary user feasibility	Thirteen participants completing one or two assigned interaction tasks	Task completion, total interaction time, errors, assistance requests, and a single-item 1–5 overall rating

3.5.1. Motion Tests

Four motion commands were tested: 1 m forward, 2 m forward, 90° rotation, and 180° rotation. Each command was repeated ten times. The regulated drive supply was recorded as 12.5 V at the start of the tests. Linear motion was commanded at 0.2 m/s and angular motion at 0.65 rad/s. The floor contained visual reference lines, and the final displacement or heading was recorded after each repetition. These measures characterize low-level short-range command execution, consistent with the use of quantitative displacement and heading errors in robot navigation benchmarking [49]; they are not a validation of localization, path planning, or campus-scale autonomous navigation.

For measured result x_i, target x_d, and N repetitions, the signed bias, mean absolute error (MAE), and root-mean-square error (RMSE) were calculated as

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(x_i - x_d), \tag{3}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|x_i - x_d|, \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_d)^2}. \tag{5}$$

3.5.2. Voice Interaction Tests

Voice interaction was evaluated in two qualitative acoustic conditions: (i) a quiet indoor environment without nearby conversation and (ii) a noisy indoor condition with people conversing near the robot. Sound-pressure levels were not recorded. Ten trials were conducted in each condition: five questions about laboratories and five about academic programs. The two-condition design was selected to expose the speech pipeline to the type of nearby-conversation interference reported as problematic for public HRI [13]; because calibrated sound-pressure and signal-to-noise measurements were unavailable, the acoustic comparison remains qualitative. The activation phrase included the robot name, followed by a request such as asking for the location of the robotics laboratory or information about the mechatronics program.

Each trial was categorized as correct, false positive, or false negative. Correct trials were also rated as high-quality responses in the supplied log; false positives were marked as incorrect responses and false negatives as null responses. Numeric values in the final worksheet column were interpreted as response latency in seconds. Latency analysis included only trials marked correct with a valid numeric value. One correct noisy-condition trial had no latency, and one numeric value associated with a false-negative trial was excluded because no valid response was produced.

3.5.3. Exploratory User Study

Thirteen participants took part in the exploratory study: three visitors, three faculty members, and seven students. Ages ranged from 19 to 45 years. The study involved 13 participants and 20 assigned tasks, not 20 participants. Each of the three visitors and each of the three faculty participants received two tasks, yielding six tasks per group. Among the seven students, one participant received two tasks and six participants received one task, yielding eight student tasks and 20 tasks overall. The retained dataset records the number of tasks assigned and completed for each participant but does not preserve the exact wording or presentation order of each individual task; this reporting limitation is stated explicitly rather than reconstructed retrospectively.

The sample intentionally included visitors together with campus-affiliated faculty and students to obtain preliminary observations across different user roles. Campus-layout familiarity was not measured with a separate questionnaire, so participant role must not be interpreted as a quantified familiarity measure. Task completion was recorded as a binary count, and participant-level records also captured total interaction time, errors, and assistance requests. The overall rating was a single numerical item from 1 to 5 with no additional verbal anchors or validated usability subscales. The reported mean of 4.31/5 is the arithmetic mean of the 13 individual ratings (56/13 = 4.3077). Accordingly, this value represents only a general overall rating and should not be interpreted as a standardized usability score. Owing to the small group sizes and unequal group composition, the analysis was descriptive; no between-group hypothesis tests were performed [10].

3.5.4. Statistical Analysis

Continuous measurements are reported as mean, sample standard deviation (SD), median where relevant, and range. Proportions are accompanied by 95% Wilson confidence intervals. No inferential comparison between user groups or interface conditions was performed because the sample was small, unequal across roles, and lacked a control condition. The analysis therefore provides preliminary descriptive evidence of feasibility rather than a confirmatory evaluation of usability. The exploratory design and small samples do not support claims of population-level superiority; the results are intended to characterize feasibility, identify failure modes, and guide a larger confirmatory study.

4. Results

4.1. Prototype Implementation

The final prototype integrated the complete interaction and mobility pipeline on the mechanical structure. The robot had a mass of 2.37 kg, dimensions of 39 × 27 × 69.5 cm, a software-limited maximum speed of 0.4 m/s, and 27 min of continuous operation before the manually defined autonomy-test termination point of approximately 11 V measured directly at the battery terminals. The recorded test sessions lasted 3–7 min. The design therefore met the original mechanical objective of remaining under the 3 kg load limit while providing a screen height suitable for standing interaction.

The modular architecture made it possible to replace the early Raspberry Pi/manual-control layer without redesigning the principal structure. In contrast to the initial open-loop prototype, the final version uses high-level ROS body commands, serial wheel-speed commands, encoder feedback, and embedded PID control. The touchscreen displays the recognized user utterance and generated answer, while animated visual states make the robot's final status visible.

4.2. Motion Accuracy

This subsection addresses RQ1 by quantifying short-range command execution rather than autonomous navigation capability. Table 6 and Figure 6 summarize the repeated movement tests. All linear trials overshot their requested distance. The 1 m command produced a mean of 1.058 m (SD 0.030 m), corresponding to a 0.058 m mean bias and 5.8% relative MAE. The 2 m command produced 2.182 m (SD 0.059 m), with 0.182 m bias and 9.1% relative MAE. Thus, linear error increased with distance rather than remaining proportional to the dispersion of repeated trials.

Angular tests showed substantially lower relative error. The 90° command produced a mean of 89.2° (SD 0.92°), with a −0.8° bias and 1.18° RMSE. The 180° command produced 180.3° (SD 0.95°), with a 0.3° bias and 0.95° RMSE. These data show repeatability over the tested distances and angles but do not demonstrate route-level navigation performance, because no global pose estimate, obstacle field, or autonomous replanning was evaluated.

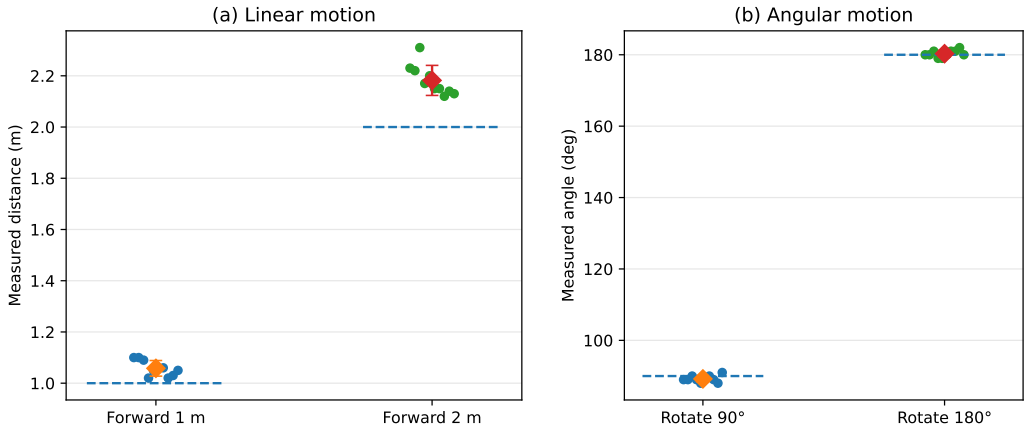


Figure 6. Repeated movement measurements. Individual trials are shown as circles, with blue points representing the first command in each panel (Forward 1 m and Rotate 90°) and green points representing the second command (Forward 2 m and Rotate 180°). Diamonds indicate the mean value (orange for the first command and red for the second), with error bars showing ±1 SD. Blue dashed horizontal segments indicate the requested target values.

Table 6. Repeated motion-test results ($N = 10$ per command).

Command	Target	Mean	SD	Bias	MAE	RMSE
Forward 1 m	1.000 m	1.058 m	0.030 m	+0.058 m	0.058 m	0.065 m
Forward 2 m	2.000 m	2.182 m	0.059 m	+0.182 m	0.182 m	0.190 m
Rotate 90°	90.0°	89.2°	0.92°	−0.8°	1.0°	1.18°
Rotate 180°	180.0°	180.3°	0.95°	+0.3°	0.7°	0.95°

4.3. Voice Recognition, Response and Latency

In the quiet condition, nine of ten trials were correct and one was a false negative, giving 90% correct interaction (95% Wilson CI: 59.6–98.2%). No false positives occurred. Under nearby conversational noise, seven trials were correct, two were false positives, and one was a false negative, giving 70% correct interaction (95% Wilson CI: 39.7–89.2%). Across both conditions, 16 of 20 trials were correct (80%; 95% CI: 58.4–91.9%). The principal degradation in the noisy condition was the appearance of false activations caused by nearby speech. These voice-interaction outcomes are summarized in Table 7 and illustrated in Figure 7a.

All trials classified as correct were also marked as high-quality responses in the experimental record. The quiet-condition correct trials had a mean latency of 4.73 s (SD 2.17 s; median 3.60 s; range 2.26–8.30 s). The six correct noisy-condition trials with valid latency values had a mean of 5.16 s (SD 2.67 s; median 3.87 s; range 3.15–9.60 s). Laboratory questions were faster than program questions: mean values were 3.00 s versus 6.88 s in quiet and 3.52 s versus 8.43 s under conversational noise. This difference suggests that answer length and query category contributed to latency in addition to acoustic conditions. The valid latency measurements are also reported in Table 7 and visualized in Figure 7b.

Table 7. Voice interaction outcomes and valid latency measurements.

Condition	Correct	FP	FN	Correct Rate	Valid Latency, Mean ± SD (s)
Quiet	9	0	1	90%	4.73 ± 2.17 ($n = 9$)
Conversational noise	7	2	1	70%	5.16 ± 2.67 ($n = 6$)
Overall	16	2	2	80%	4.90 ± 2.31 ($n = 15$)

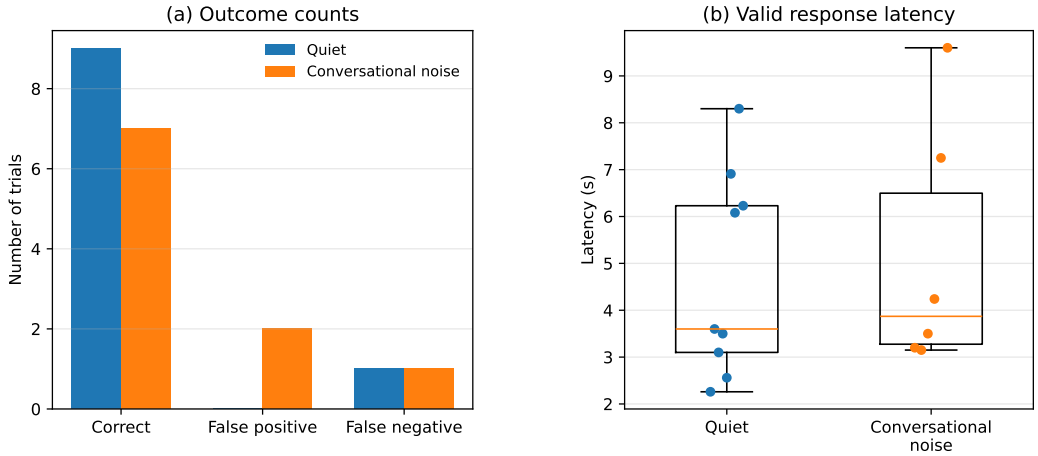


Figure 7. Voice interaction: (**a**) outcomes in quiet and nearby conversational noise and (**b**) response latencies for correct trials with valid measurements.

4.4. Exploratory User Evaluation

This subsection addresses RQ3 and is interpreted as preliminary descriptive evidence because the study did not include a control group or a validated usability instrument. The 13 participants had a mean age of 26.3 years (SD 8.35; median 22; range 19–45).

Sixteen of 20 assigned tasks were completed, yielding an 80% completion rate (95% Wilson CI: 58.4–91.9%). The mean total interaction time was 4.52 min (SD 1.20; median 3.90 min). Four errors and four requests for assistance were recorded. The mean general rating was 4.31/5 (SD 0.95; median 5). This average was computed from the 13 single-item numerical ratings and is not a standardized usability score. The overall and role-specific descriptive results are summarized in Table 8.

The three visitors completed all six assigned tasks, with a mean time of 5.60 min and a mean rating of 5.0. Faculty participants completed four of six tasks, with a mean time of 5.73 min and a rating of 4.0. Students completed six of eight tasks, with a mean time of 3.54 min and a mean rating of 4.14. These subgroup values are descriptive only because the groups were small and the number of assigned tasks differed. The differences in completion rate, interaction time, and user rating across participant roles are reported in Table 8 and visualized in Figure 8.

Table 8. Exploratory user-study results.

Group	Participants	Tasks	Completed	Completion	Mean Time (min)	Mean Rating
Visitors	3	6	6	100%	5.60	5.00
Faculty	3	6	4	66.7%	5.73	4.00
Students	7	8	6	75.0%	3.54	4.14
Overall	13	20	16	80.0%	4.52	4.31

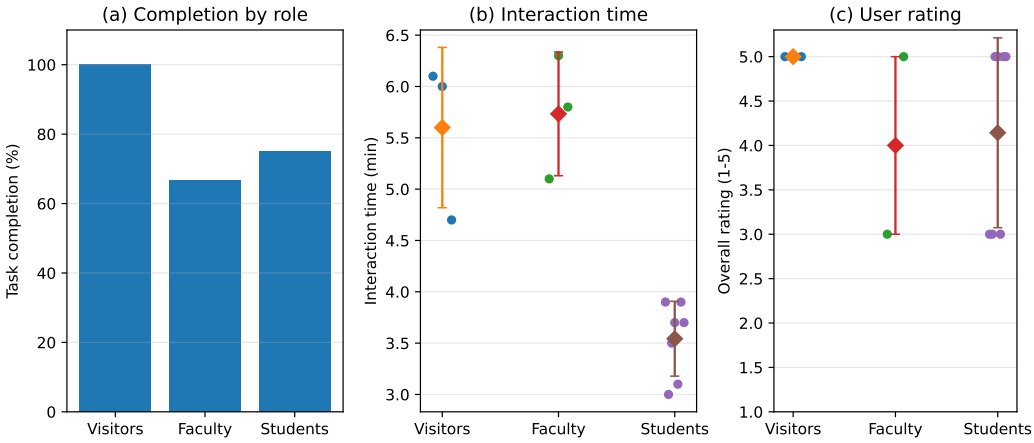


Figure 8. User results by participant role: (**a**) task-completion rate, (**b**) interaction time, and (**c**) overall user rating. In panel (**a**), blue bars represent the task-completion percentage for each participant group. In panels (**b**,**c**), circles represent individual participant measurements, with blue, green, and purple circles corresponding to visitors, faculty, and students, respectively. Diamonds represent group mean values, shown in orange for visitors, red for faculty, and brown for students; vertical error bars indicate ±1 SD. Individual values are retained because the participant groups are small and unequal in size.

5. Discussion

The discussion is organized around the three research questions and the capability boundaries identified in the literature review. RQ1 concerns only short-range motion execution; RQ2 concerns speech robustness and latency; and RQ3 concerns preliminary user-task evidence. Knowledge grounding, campus-scale navigation, comparative interface benefits, and edge model efficiency were not experimentally validated and are therefore discussed as limitations and future research directions rather than as demonstrated contributions.

5.1. Integrated Value

The principal benefit of INDI is the integration of an embodied information interface and a controllable mobile base within a compact, reconfigurable structure. Previous work has shown that users may prefer an embodied information provider over a conventional tablet interface when natural language, gestures, and posture changes are available [1]. INDI does not reproduce the full sensing and social behavior capabilities of commercial robots, but it provides the core channels required to study embodiment in a university context: voice input, synthesized speech, touchscreen interaction, visible status, animated facial feedback, head motion, and base movement.

The mechanical continuity between the early and final prototypes is also important. Rather than replacing the complete physical platform, the project preserved the wide tracked base, elevated display, modular MDF body, and lateral arms, and concentrated the redesign on computing, control, and interaction. This strategy reduced fabrication complexity and demonstrates how a student-built prototype can be incrementally transformed into a research platform. The result differs from dialogue-only systems and from navigation-only laboratory robots because the same platform exposes the complete chain from spoken request to ROS behavior and embedded motor control.

The use of ROS as an intermediate layer is consistent with frameworks that connect LLMs to robotic actions [5]. However, INDI's restricted command mapping is deliberately simpler than open-ended LLM task planning. Only explicit phrases trigger predefined integer commands, while general questions remain in the conversational branch. This design follows the safety rationale of voice-to-action systems that constrain LLM output to executable skills [6]. The approach reduces the probability that an unconstrained generated response will directly actuate the robot, although confirmation dialogues and a formal action validator would further improve safety.

5.2. Locomotion Performance and Navigation Limitations

Angular performance was repeatable, with relative MAE near 1.1% for 90° and 0.39% for 180° turns. Linear tests showed systematic overshoot, and the mean error increased from 5.8 cm over 1 m to 18.2 cm over 2 m. The low coefficient of variation of the repeated distance measurements (approximately 2.7–2.9%) indicates that the behavior was repeatable but biased. This pattern is consistent with accumulated error from wheel-radius calibration, effective encoder counts, track slip, battery-dependent drive response, and velocity integration without external position correction.

The final system should therefore be described as a mobile service robot with predefined or semiautonomous route behaviors, not as a fully autonomous navigator. Unlike systems using visual SLAM and dynamic obstacle detection [19], semantic mapping [48], or multi-sensor localization [23], INDI does not estimate a global campus pose or re-plan around obstacles. The front mounting position intended for an ultrasonic sensor is inactive in the evaluated version and therefore provides no obstacle detection or safety function. The 1–2 m motion experiments validate only low-level movement repeatability and cannot support a claim of practical campus navigation capability. Although 76 development routes were executed during system construction, those runs did not use external ground truth or a standardized navigation protocol and are not treated as navigation-validation trials. The architecture is suitable for controlled demonstrations and short marked routes, but longer guidance tasks require odometry calibration, external localization, obstacle sensing, and a navigation stack. Broader reviews identify these capabilities as central to robust autonomous navigation [21] and to socially acceptable motion around people [22]. A subsequent navigation study should therefore use corridor-scale start–goal routes, repeated trials under varying pedestrian conditions, external ground truth, route completion rate, path

deviation, intervention frequency, obstacle encounters, and battery-state logging rather than extrapolating from the present 1–2 m command tests [19,49].

The 0.5 s command timeout, explicit stop command, moderate test velocity, and screen stop control are useful safety features. Nevertheless, service robot safety requires more than software stopping; collision sensing, speed adaptation, proximity-aware planning, fault detection, and formal hazard analysis should be added before unsupervised public deployment [24].

5.3. Voice Interaction, Latency, and Multimodality

The decrease from 90% correct interaction in quiet to 70% with nearby conversation confirms that speech is the most environmentally sensitive part of the interface. The two false positives in the noisy condition are especially important for a public robot because a nearby conversation can create an unintended interaction. This result is consistent with field evidence that voice-based LLM interaction can perform poorly in crowded educational settings relative to alternative modalities [13]. The touchscreen and visible stop button are therefore not decorative additions; they provide alternative activation, feedback, and interruption channels when speech becomes unreliable. The observed 70% correct-interaction rate under nearby conversation and the latency tail reaching 9.6 s are not sufficient to claim robust unsupervised public space interaction; they instead motivate improved wake-word rejection, local/cached processing, and multimodal fallback mechanisms.

The median latency remained below 4 s in both conditions, but mean values were higher because program-related questions required longer responses. The prompt requests responses shorter than 30 words after the greeting, which helps control dialogue duration, yet network ASR, cloud LLM inference, TTS generation, and playback all contribute to end-to-end delay. Multi-session university studies have shown that slow response generation and speech recognition errors can reduce turn-taking and conversation length [41]. Semantic caching, local models, or a hybrid institutional database could reduce delay for frequently asked questions, as demonstrated in low-cost educational assistance platforms [18].

A specific advantage of the present multimodal design is that it does not depend on computer vision. The webcam is used only for recording, reducing computation, calibration burden, and the risk that visual inference errors affect movement. This choice also simplifies privacy management relative to continuous face or gesture recognition. The trade-off is that INDI cannot detect user attention, gaze, identity, gestures, or nearby obstacles from images, unlike multimodal systems that fuse posture or visual perception [14,15]. For the reported study, participants explicitly consented to the capture and research handling of audio and images. Raw media were not publicly disseminated, were not used as quantitative study outcomes, and were excluded from the public anonymized dataset.

5.4. Information Quality

All recognized responses in the supplied test log were rated high quality, but this finding must be interpreted cautiously. The evaluation contained only 20 broad laboratory/program trials, and response quality was recorded using a coarse high/incorrect/null classification. There was no independent expert panel, item-level factual rubric, or adversarial set of out-of-domain questions. Moreover, the implementation sends the system prompt and conversation history directly to a general-purpose model; it does not implement retrieval-augmented generation or a verified institutional database. Accordingly, an apparently fluent answer can still contain an outdated or fabricated campus fact, and the current system cannot provide a source trace for the generated statement. This risk is operationally important because laboratory locations, academic offerings, schedules, and institutional procedures can change over time.

The literature indicates that guide robot performance improves when query routing is combined with grounded retrieval [4]. Reporting guidelines also recommend documenting prompts, model versions, guardrails, failure cases, and data handling [10]. A publishable next-stage evaluation should therefore create a versioned institutional knowledge base, log the retrieved evidence, require the robot to abstain when no verified answer is available, and score answers for factual accuracy, completeness, concision, and hallucination using at least two evaluators. These additions would transform the encouraging exploratory result into stronger evidence of reliable institutional information delivery.

5.5. User Experience

The overall rating of 4.31/5 and perfect task completion among the three visitors suggest that the platform can produce a positive first-use experience. The 80% overall task completion rate, however, shows that a favorable subjective rating does not guarantee task success. This distinction is consistent with HRI research emphasizing that trust, enjoyment, competence, and organizational value are separate dimensions [28–30]. The high visitor scores are promising because visitors are the target population, but the sample is too small to support subgroup conclusions.

The observed errors and requests for assistance indicate that public deployment would still require onboarding cues, clear repair strategies, and staff escalation. Research on insufficient robot capability suggests that acknowledging limitations and offering an alternative can preserve the interaction better than silent failure [32]. Similarly, service failures can influence user self-efficacy and responsibility attribution [31]. INDI should therefore explicitly state when it did not hear a request, when network services are unavailable, or when information cannot be verified, and should provide a touchscreen retry or human-contact option.

From an institutional perspective, the modular platform offers benefits beyond direct visitor assistance. It can serve as a demonstrator for ROS, embedded control, HRI, and LLM integration; it can be maintained with campus fabrication resources; and it can be adapted as institutional information changes. However, deployment readiness also depends on maintenance responsibility, content governance, privacy procedures, accessibility, network availability, and staff acceptance, as emphasized in public space architecture and university receptionist studies [2,3].

5.6. Comparative Evaluation, Ablation, and Edge Efficiency

The present dataset does not contain a touchscreen-only baseline, a voice-only condition, a human-guide comparison, or component-level ablations. Consequently, the study cannot quantify the causal contribution of embodiment, speech, touchscreen feedback, or mobile motion individually, and no such causal claim is made. A stronger follow-up design should compare at least a touchscreen kiosk, voice-only interaction, and the complete multimodal robot under matched tasks, and should ablate expressive feedback and mobile guidance while retaining the same information content. This would separate the effect of embodiment from the effect of the underlying dialogue service [1,28].

The same principle applies to model efficiency. The current Jetson Nano does not execute the LLM locally; cloud inference contributes to both latency and network dependence. Future evaluation should compare the current cloud model with quantized local models, cached responses, and hybrid edge–cloud selection. Pruning and device–edge model partitioning can trade inference distortion against latency and energy [11], while task-aware model selection such as SELA can choose among models with different response quality and latency characteristics [12]. These strategies are especially relevant to the observed

mean latency of 4.90 s and the 27 min operating duration, but their benefits remain to be measured on INDI.

5.7. Limitations

This study has several limitations. First, the movement tests used ten repetitions per command and did not record external ground-truth trajectories, speed profiles, stopping distance, surface variation, or battery level per trial. Second, acoustic conditions were qualitative; no sound pressure level or signal-to-noise ratio was measured. Third, the response quality assessment was coarse and did not use independent evaluators. Fourth, the user study included only 13 participants and 20 tasks, with unequal tasks per participant and no comparison against a kiosk, human guide, or voice-only condition. The exact wording and order of individual user-study tasks were not retained in the analysis dataset, and campus-layout familiarity was not measured independently; therefore, possible familiarity effects cannot be quantified retrospectively. Fifth, one correct noisy-condition trial lacked a latency value and one numeric latency attached to a false negative trial was excluded. Sixth, the robot has no SLAM, obstacle-aware path planning, or visual perception. Finally, 27 min of continuous operation is adequate for demonstrations but not for sustained reception service. This duration is a critical deployment limitation for a public-facing service robot and would require higher-capacity energy storage, battery swapping, docking/charging, or duty-cycling strategies before prolonged operation.

These limitations define a clear next study: a preregistered comparison between INDI and a touchscreen kiosk with a larger visitor sample; calibrated acoustic conditions; a verified RAG knowledge base; objective event logs; longitudinal deployment; and autonomous indoor navigation with safety monitoring. Such a design would permit statistical testing of embodiment, multimodality, and mobility benefits rather than only feasibility characterization.

6. Conclusions

This paper presented the design and exploratory evaluation of INDI, a custom campus guide robot that integrates a tracked mobile base, a screen-based expressive head, Spanish speech interaction, an LLM dialogue service, ROS 1 coordination, serial wheel control, encoder feedback, and embedded PID loops. The final version demonstrates a practical evolution from an earlier manually controlled mechanical prototype to an integrated HRI research platform.

The robot achieved repeatable angular motion and functional short-range guidance, although systematic distance overshoot increased over the 2 m test. Voice interaction was successful in 90% of quiet trials and 70% of trials with nearby conversation, confirming the need for touchscreen fallback and improved activation logic in public environments. The 13-participant exploratory user study produced an 80% task completion rate and a mean rating of 4.31/5, supporting feasibility but not population-level claims. In relation to the research questions, RQ1 is answered by the repeatable but biased short-range motion measurements, RQ2 by the observed noise-dependent interaction accuracy and latency, and RQ3 by the descriptive task and rating outcomes. These results constitute preliminary feasibility evidence only; they do not validate autonomous navigation, factual institutional answer reliability, comparative usability, or the independent benefit of individual modalities.

The main benefit of the proposal is the combination of embodiment, multimodal communication, and low-level mobile control in a modular, modifiable platform. Its present limitations are equally clear: cloud dependence, ungrounded institutional dialogue, short autonomy, sensitivity to ambient speech, and predefined rather than autonomous navigation. Future work will prioritize a verified retrieval-augmented knowledge base, local or

cached processing, calibrated wake-word detection, longer-duration power management, autonomous localization and obstacle avoidance, and a larger controlled HRI study.

Author Contributions: Conceptualization, J.V.-A. and C.P.C.; methodology, J.V.-A.; software, C.P.C.; mechanical design and fabrication, J.C.; validation, C.P.C. and J.V.-A.; formal analysis, J.V.-A. and C.D.-V.-S.; investigation, J.V.-A. and C.D.-V.-S.; resources, J.V.-A.; data curation, C.P.C. and C.D.-V.-S.; writing-original draft preparation, J.V.-A. and C.P.C.; writing-review and editing, C.D.-V.-S. and J.C.; visualization, J.V.-A. and C.P.C.; supervision, J.V.-A.; project administration, J.V.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval by an Institutional Ethics Committee were not required for this non-invasive, no-risk engineering evaluation under the institutional procedures applicable to the project.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study. Participants authorized the capture and research handling of interaction audio and documentation images. Raw audio and image recordings were not publicly disseminated and were not included in the shared anonymized dataset.

Data Availability Statement: The data and source code are published through the project repository: https://github.com/josevarela90/INDI_basic.git (accessed on 7 August 2026).

Acknowledgments: The authors would like to express their gratitude to Universidad Indoamérica for its support of this research through the "Tecnologías de la Industria 4.0 en Educación, Salud, Empresa e Industria" project. Special thanks to Bryan Guevara for his technical support on this project. During preparation of this revision, the authors used ChatGPT (v5.6 sol) for English-language editing and sentence-clarity improvement.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ASR	Automatic speech recognition
HRI	Human–robot interaction
LLM	Large language model
MAE	Mean absolute error
MDF	Medium-density fiberboard
PID	Proportional–integral–derivative
RAG	Retrieval-augmented generation
RMSE	Root-mean-square error
ROS	Robot Operating System
SLAM	Simultaneous localization and mapping
TTS	Text-to-speech

References

1. Sievers, T.; Russwinkel, N. Project Report: Requirements for a Social Robot as an Information Provider in the Public Sector. *KI-Kunstl. Intell.* **2024**, *38*, 145–149. [CrossRef]
2. Oruma, S.; Colomo-Palacios, R.; Gkioulos, V. Architectural views for social robots in public spaces: Business, system, and security strategies. *Int. J. Inf. Secur.* **2025**, *24*, 12. [CrossRef]
3. Oruma, S.O.; Colomo-Palacios, R.; Gkioulos, V. From Framework to Reliable Practice: End-User Perspectives on Social Robots in Public Spaces. *Systems* **2026**, *14*, 137. [CrossRef]
4. Guo, J.; Han, X. Research on a Dual-Mode Human–Robot Interaction Method Based on a Large Language Model. *IET Cyber-Syst. Robot.* **2026**, *8*, e70037. [CrossRef]

5. Koubaa, A.; Ammar, A.; Boulila, W. Next-generation human-robot interaction with ChatGPT and robot operating system. *Softw.-Pract. Exp.* **2025**, *55*, 355–382. [CrossRef]
6. Ok, K.; Heo, G.; Lee, C.; Ahn, S.H. Practical Human–Robot Interaction (HRI) through Large Language Model (LLM)-based Voice-to-Action Systems. *Int. J. Precis. Eng. Manuf.* **2026**. [CrossRef]
7. Pinto-Bernal, M.; Biondina, M.; Belpaeme, T. Designing Social Robots with LLMs for Engaging Human Interaction. *Appl. Sci.* **2025**, *15*, 6377. [CrossRef]
8. Lisondra, M.; Benhabib, B.; Nejat, G. Embodied AI with Foundation Models for Mobile Service Robots: A Systematic Review. *Robotics* **2026**, *15*, 55. [CrossRef]
9. Bian, S.; Zhang, Y.; Tian, G.; Miao, Z.; Wu, E.Q.; Yang, S.X.; Hua, C. Large language model-based task planning for service robots: A review. *Biomim. Intell. Robot.* **2026**, *6*, 100274. [CrossRef]
10. Matuszek, C.; Williams, T.; DePalma, N.; Mead, R.; Wen, R.; Schneiders, E.; Kennington, C.; Bezabih, A. Reporting Guidelines for Large Language Models in Human–Robot Interaction. *ACM Trans. Hum.-Robot. Interact.* **2026**, *15*, 36. [CrossRef]
11. Lyu, Z.; Xiao, M.; Xu, J.; Skoglund, M.; Di Renzo, M. The Larger the Merrier? Efficient Large AI Model Inference in Wireless Edge Networks. *IEEE J. Sel. Areas Commun.* **2026**, *44*, 2839–2853. [CrossRef]
12. Tuli, S.; Casale, G.; Roveri, M. SELA: Smart Edge LLM Agent to Optimize Response Trade-offs of AI Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2025**, *9*, 130:1–130:18. [CrossRef]
13. Grzejszczak, T. Gestures for attraction, gestures for communication: A field study of social robot interaction modalities at an educational fair. *Bull. Pol. Acad. Sci. Tech. Sci.* **2026**, *74*, e2026487. [CrossRef]
14. Yura, J.; Byambasuren, B.E.; Kim, D. Multi-modal, multi-step human–robot interaction method for natural interaction of service robots. *Intell. Serv. Robot.* **2025**, *18*, 389–401. [CrossRef]
15. Lai, Y.; Yuan, S.; Nassar, Y.; Fan, M.; Weber, T.; Rätsch, M. NVP-HRI: Zero shot natural voice and posture-based human–robot interaction via large language model. *Expert Syst. Appl.* **2025**, *268*, 126360. [CrossRef]
16. Lyu, S.; Wang, F.; Lin, W.; Guo, G.; Navarro-Alarcon, D. Enhancing End-user Engagement in Human–Robot Interaction by Performing LLM-driven Expressive Behaviors. *ACM Trans. Hum.-Robot. Interact.* **2026**, *15*, 79. [CrossRef] [PubMed]
17. Gao, W.; Tian, Y.; Zhai, W.; Ji, Y.; Shen, S. Exploring the Impacts of Service Robot Interaction Cues on Customer Experience in Small-Scale Self-Service Shops. *Sustainability* **2025**, *17*, 10368. [CrossRef]
18. Aucapiña, S.A.; Benalcázar, N.C.; Varela-Aldás, J.; Isa-Jara, R. ROS2-Based Low-Cost Mobile Robot for Educational Assistance with Reactive Navigation and Semantic-Cached Language Processing. *Robotics* **2026**, *15*, 131. [CrossRef]
19. Todkari, V.C.; Kaldate, A.P.; Kolhar, S.; Jagtap, A.; Sable, N.P. AI-Driven Visual Navigation for Smart Lab Tour Guide Robot. *J. Eur. Syst. Autom.* **2025**, *58*, 2609–2616. [CrossRef]
20. Yang, C.H.; Juang, J.G. Guide Robot Based on Image Processing and Path Planning. *Machines* **2025**, *13*, 560. [CrossRef]
21. Nahavandi, S.; Alizadehsani, R.; Nahavandi, D.; Mohamed, S.; Mohajer, N.; Rokonuzzaman, M.; Hossain, I. A Comprehensive Review on Autonomous Navigation. *ACM Comput. Surv.* **2025**, *57*, 234. [CrossRef]
22. Alyassi, R.; Cadena, C.; Riener, R.; Paez-Granados, D. Social robot navigation: A review and benchmarking of learning-based methods. *Front. Robot. AI* **2025**, *12*, 1658643. [CrossRef] [PubMed]
23. Ušinskis, V.; Nowicki, M.; Dzedzickis, A.; Bučinskas, V. Sensor-Fusion Based Navigation for Autonomous Mobile Robot. *Sensors* **2025**, *25*, 1248. [CrossRef] [PubMed]
24. Zhang, L.; Shi, P.; Wang, S.; Su, X. A survey of safety control for service robots. *J. Autom. Intell.* **2026**, *5*, 112–125. [CrossRef]
25. Yoon, H.; Shim, G.; Lee, H.; Kim, M.G.; Kim, S. Observation of Human–Robot Interactions at a Science Museum: A Dual-Level Analytical Approach. *Electronics* **2025**, *14*, 2368. [CrossRef]
26. Iwasaki, M.; Masuda, K.; Sakai, K.; Ban, M.; Chi, Z.; Kawata, M.; Meneses, A.; Ishiguro, H.; Yoshikawa, Y. Following to Talk: Effects of Mobile Robot Guidance on Dialogue Motivation in a Field Trial. *IEEE Robot. Autom. Lett.* **2026**, *11*, 4339–4344. [CrossRef]
27. Tretter, S.; von Terzi, P.; Diefenbach, S. How the presence of others shapes the user experience of service robots. *Front. Robot. AI* **2025**, *12*, 1538711. [CrossRef] [PubMed]
28. Marmor, K.; Leoste, J.; Heidmets, M. Guide robots' acceptance in organizations: User self-efficacy and robots' organizational value. *Front. Robot. AI* **2026**, *13*, 1842941. [CrossRef] [PubMed]
29. Balcha, H.B.; Rauffet, P.; Wegari, G.M.; Buche, C. A Review on Human-Robot Trust in Home Service Robots. *ACM Trans. Hum.-Robot. Interact.* **2025**, *14*, 76. [CrossRef]
30. Irfan, B.; Miniota, J.; Thunberg, S.; Lagerstedt, E.; Kuoppamaki, S.; Skantze, G.; Pereira, A. Human-Robot Interaction Conversational User Enjoyment Scale (HRI CUES). *IEEE Trans. Affect. Comput.* **2026**, *17*, 1438–1453. [CrossRef]
31. Merdin-Uygur, E.; Ozturkcan, S. The robot saw it coming: Physical human interference, deservingness, and self-efficacy in service robot failures. *Behav. Inf. Technol.* **2025**, *44*, 1320–1339. [CrossRef]
32. Gao, W.; He, T.; Ji, Y.; Kan, Y.; Zha, F. Human–Robot Interaction Strategy of Service Robot with Insufficient Capability in Self-Service Shop. *Biomimetics* **2026**, *11*, 213. [CrossRef] [PubMed]

33. Zou, A.; Meng, Y.; Tong, S. Context Awareness and Human–Robot Interaction Optimization for Museum Intelligent Guide Robot. *Adv. Intell. Syst.* **2026**, *8*, e202500728. [CrossRef]
34. Csepregi, M.; Korcsok, B.; Abdai, J.; Miklósi, Á.; Ferdinandy, B.; Kovács, T.; Roszik, S.; Gácsi, M. Taking candy from a robot: Attitudes toward an autonomous non-humanoid service robot at real-life events. *Comput. Hum. Behav. Rep.* **2026**, *23*, 101204. [CrossRef]
35. Babel, F.; Kraus, J.; Baumann, M. Findings From A Qualitative Field Study with An Autonomous Robot in Public: Exploration of User Reactions and Conflicts. *Int. J. Soc. Robot.* **2022**, *14*, 1625–1655. [CrossRef]
36. Finkel, M.; Erle, L.; Timm, L.; Arntz, A.; Helgert, A.; Straßmann, C.; Eimler, S.C. Need Satisfaction in Human-Robot Interactions in Public Spaces – A Positive Computing Approach. *Int. J. Soc. Robot.* **2026**, *18*, 21. [CrossRef]
37. Tsai, Y.L.; Wadgaonkar, C.; Chun, B.; Knight, H. How Service Robots Can Improve Workplace Experience: Camaraderie, Customization, and Humans-in-the-Loop. *Int. J. Soc. Robot.* **2022**, *14*, 1605–1624. [CrossRef] [PubMed]
38. Nedianu, G.C.; Badoi, M.A.; Ciochiu, D.; Muraretu, I.; Badica, C. Agentic Narrative Audio Guides: A Hybrid Multi-Agent System for Cultural Audio Guides. In *Proceedings of the 2026 IEEE Conference on Artificial Intelligence (CAI), Granada, Spain, 8–10 May 2026*; IEEE: Piscataway, NJ, USA, 2026; pp. 1329–1334. [CrossRef]
39. Hu, Y.; Sato, A.J.; Du, J.; Ye, C.; Zhu, A.; Praveena, P.; Mutlu, B. NarraGuide: An LLM-Based Narrative Mobile Robot for Remote Place Exploration. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology, Busan, Republic of Korea, 28 September–1 October 2025*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 1–15. [CrossRef]
40. Tan, R.; Lou, S.; Zhou, Y.; Lv, C. Multi-Modal LLM-Enabled Long-Horizon Skill Learning for Robotic Manipulation. In *Proceedings of the 2024 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE International Conference on Robotics, Automation and Mechatronics (RAM), Hangzhou, China, 8–11 August 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 14–19. [CrossRef]
41. Mauliana, M.; Ashok, A.; Czernochowski, D.; Berns, K. Exploring LLM-powered multi-session human-robot interactions with university students. *Front. Robot. AI* **2025**, *12*, 1585589. [CrossRef] [PubMed]
42. García, D.H.; Gunson, N.; Addlesee, A.; Cherakara, N.; Dondrup, C.; Sieińska, W.; Romeo, M.; Lemon, O. Evaluating multi-party interactions with social robots using large language models and multi-modal systems. *Interact. Stud.* **2025**, *26*, 422–476. [CrossRef]
43. Olutunbi, D.; Polvara, R.; Davies, T.; Hamilton, J.; Kargas, N.; Del Duchetto, F. Who Leads the Story? Comparing Autonomous vs. Adult-Supported Child–Robot Interactions. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction, Edinburgh, Scotland, UK, 16–19 March 2026*; Association for Computing Machinery: New York, NY, USA, 2026; pp. 139–148. [CrossRef]
44. Pozzi, L.; Nasato, M.; Toscani, N.; Braghin, F.; Gandolla, M. Personalization and Generative Dialogue in Social Robotics for Eldercare: A User Study. *Appl. Sci.* **2026**, *16*, 3369. [CrossRef]
45. Diddeniya, I.; Wanniarachchi, I.; Gunasinghe, H.; Premachandra, C.; Kawanaka, H. Human–Robot Communication System for an Isolated Environment. *IEEE Access* **2022**, *10*, 63258–63269. [CrossRef]
46. Santana, J.M.; Silveira, B.D.; Lima, C.; Diaz-Amado, J.; Libarino, C.S.; Marques, J.E.S.; Barrios-Aranibar, D.; Patiño-Escarcina, R.E. Design and Implementation of an Interactive System for Service Robot Control and Monitoring. *Sensors* **2025**, *25*, 987. [CrossRef] [PubMed]
47. Kassem, A.H.; Asem, M. A low-cost robotic system for simultaneous localization and mapping. *J. Eng. Appl. Sci.* **2024**, *71*, 158. [CrossRef]
48. Alqobali, R.; Alnasser, R.; Rashidi, A.; Alshmrani, M.; Alhmiedat, T. A Real-Time Semantic Map Production System for Indoor Robot Navigation. *Sensors* **2024**, *24*, 6691. [CrossRef] [PubMed]
49. Karwowski, J.; Szynkiewicz, W. Quantitative Metrics for Benchmarking Human-Aware Robot Navigation. *IEEE Access* **2023**, *11*, 79941–79953. [CrossRef]
50. Begum, N.; Faisal, N.; Sobh, R.; Rana, N.P. From TAM to AIDUA and Beyond: Charting the Theoretical Frontiers of Service Robot Adoption in Hospitality. *J. Glob. Inf. Manag.* **2025**, *33*, 1–27. [CrossRef]
51. Frasca, T.; Scheutz, M. A Framework for Robot Self-Assessment of Expected Task Performance. *IEEE Robot. Autom. Lett.* **2022**, *7*, 12523–12530. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.