Amigo-Agri: A Human-Following Robotic Platform with Speech Recognition and Retrieval-Augmented Generation for Smart Farming

Md. Moniruzzaman Hemal[1,2], Atiqur Rahman[1,3*], Md. Abdul Halim Khan[1,4], Sadikur Rahman Sadik[1,3], Md. Shohanur Rahman Shohan[1], Tahzib Mahmud Rifat[1], Md. Ashiqussalehin[1], and Md. Toukir Ahmed[1]

[1] Department of IoT and Robotics Engineering (IRE), University of Frontier Technology, Kaliakair, Gazipur-1750, Bangladesh
[2] Department of Computing and Information System, Daffodil International University, Dhaka-1216, Bangladesh
[3] Department of Computer Science and Engineering, Daffodil International University, Dhaka-1216, Bangladesh
[4] Department of Computer Science and Engineering, Northern University of Business and Technology, Khulna-9100, Bangladesh

{1801007,1801025*,1801020,1801010,1801033,2001005}@iot.uftb.ac.bd, {ashiqussalehin0001,toukir0001}@uftb.ac.bd

Abstract. Bridging the gap between physical farm assistance and expert agronomic advice remains an unsolved challenge. We present Amigo-Agri, an integrated platform combining a low-cost, human-following mobile robot with a retrieval-augmented voice assistant. The system splits responsibilities: a smartphone handles speech recognition and knowledge retrieval (RAG), while an Arduino-based robot executes real-time navigation. We address the unique constraints of low-literacy, outdoor environments by enabling hands-free natural language queries alongside motion commands. Experimental results demonstrate a 3.5 s average speech-to-speech latency, 0.85 semantic relevance, and robust person-following accuracy (RMSE 0.14 m) in mock field settings. While optical sensing and network dependence remain limitations, this work provides a validated blueprint for affordable, interactive agricultural support.

Keywords: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Speech Recognition, Smart Farming, Human–Robot Interaction.

1 Introduction

Agriculture is simultaneously labor-intensive and decision-intensive. Smallholders and farm workers must make time-critical choices about irrigation, nutrient

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_58

management, pest and disease control, animal health, and marketing while performing physically demanding tasks in fields, sheds, and markets [11]. Conventional advisory channels—extension visits, paper bulletins, or web portals—are valuable but often delayed, text-heavy, and difficult to consult hands-free. At the same time, mobile and social robots can physically follow and assist humans, yet they seldom provide transparent, source-grounded agronomic guidance. The result is a practical gap: farmers need systems that *both* reason and act, *and* that can be operated naturally by voice in noisy, low-literacy, outdoor settings.

Voice assistants and large language models (LLMs) promise more accessible guidance, including local languages and conversational flow, but most deployed tools are screen-bound, text-only, or closed ecosystems with limited provenance and variable reliability [1]. Separately, person-following robots are increasingly robust and affordable, yet they typically execute a narrow repertoire of motions without integrating vetted agronomic knowledge or explaining why an action is recommended [3]. Bridging these two lines of work remains under-explored: a farmer-facing system that can follow the user safely through the field, accept natural-language commands, *and* return citation-backed advice fast enough to be useful in the moment.

To address this gap, we present Amigo-Agri, a voice-first, human-following mobile robot tightly coupled to a retrieval-augmented agricultural assistant. The embodiment—built around an Arduino Uno, L298N motor driver, four DC motors, HC-06 Bluetooth, ultrasonic and infrared sensing, and an onboard speaker/amplifier—handles real-time, on-device navigation (follow, stop, left/right, safe spacing, obstacle avoidance). A companion smartphone provides the speech/knowledge loop: automatic speech recognition (ASR) ingests the farmer's query; a sentence-transformer with FAISS retrieves top-k passages from a curated, multi-domain agronomy corpus (soil, irrigation, crops, pests, animal care); a 4-bit quantized LLM composes a concise answer with explicit back-references; and text-to-speech (TTS) delivers an audible reply. Low-level voice commands are parsed into motion primitives on the robot, while high-level advisory requests round-trip through the assistant. This split leverages commodity hardware for field robustness and keeps the interaction hands-free and transparent.

We explicitly acknowledge the scope and limitations of this proposed solution. To maintain cost-effectiveness for smallholders, the system relies on infrared and ultrasonic sensing rather than computer vision, limiting its ability to navigate complex, unstructured foliage. Furthermore, the reliance on cloud/edge connectivity for the Large Language Model (LLM) introduces dependencies on network stability, which we mitigate via safety fallbacks.

This paper makes the following contributions:

- Introduces an integrated embodied–conversational architecture that fuses a low-cost human-following robot with a retrieval-augmented assistant for voice-driven agricultural guidance.
- Proposes a natural-language control layer that seamlessly combines locomotion commands and advisory queries through mixed-initiative coordination between onboard control and cloud/edge reasoning.

- Develops a curated multi-domain agronomy corpus and retrieval engine to ground large language model outputs in verifiable, field-relevant evidence.
- Presents a holistic evaluation encompassing latency, semantic relevance, concurrency, and jitter tolerance for the assistant, together with tracking accuracy, command reliability, and safety performance for the robot in realistic trials.
- Provides complete implementation details—including hardware specifications, control mappings, and quantized LLM configurations—to enable full system reproducibility.

The remainder of the paper reviews related work on voice assistants, retrieval-augmented generation, and human-following robotics (Section 2); details the integrated hardware–software methodology (Section 3); reports results for advice latency/relevance and embodied performance under realistic conditions (Section 4); and concludes with limitations and future directions (Section 5).

2 Literature Review

Timely, dependable agricultural advice remains a critical challenge, particularly for smallholder farmers with limited access to traditional extension services. Although conventional outreach offers valuable expertise, it often fails to deliver support when and where it is needed. This gap has encouraged growing exploration of artificial intelligence (AI), especially conversational and voice-based systems, to provide personalized, localized, and on-demand guidance

Recent progress in agricultural robotics shows tangible momentum toward automation across multiple farming stages — including seeding, fertilizing, harvesting, scouting, and soil management. Maria et al. (2025) [14] report that most existing robots are still purpose-specific, facing persistent challenges in achieving fully autonomous and cost-effective operation under field conditions. Efforts toward modular and reconfigurable robotic systems demonstrate improved versatility through centralized, decentralized, or hybrid control structures; yet, as Hernandez et al. (2025) [4] note, real-world validation of robustness, energy efficiency, and coordinated scaling remains largely confined to simulations and small-scale trials rather than commercial deployment.

On the software front, voice-based assistants have emerged as promising tools for continuous, inclusive agricultural support. Their ability to operate in local languages and overcome literacy barriers positions them as powerful enablers for smallholder engagement. Pancham et al. (2024) [13] showed that general-purpose voice platforms (e.g., Alexa, Siri) can assist in crop selection and daily farm management; however, such systems rely heavily on stable internet connectivity and proprietary ecosystems, with limited integration of real-time weather, pricing, or soil-test data.

Task-specific conversational agents, such as Farmer Companion, have demonstrated the utility of deep learning and supervisory models for delivering personalized advice. Nonetheless, their effectiveness is constrained by training data

quality and a notable decline in accuracy during multi-turn or context-intensive interactions (Ilakiyaselvan et al., 2024 [5]).

The convergence of conversational large language models (LLMs) with automatic speech recognition (ASR) systems offers further potential, enabling natural-language querying of agronomic datasets and multimodal fusion of sensor data for precision decision-making. Potamitis et al. (2023) [7] identified data precision, latency, and throughput as critical performance bottlenecks. Comprehensive reviews (Hongyan et al., 2024 [18]; Tawseef et al., 2024 [9]) highlight how LLMs, large vision models (LVMs), and multimodal architectures have been applied to pest and disease detection, soil mapping, and yield prediction, while emphasizing persistent challenges in dataset quality, benchmarking, training efficiency, and ethical transparency.

Notably, early real-world deployments demonstrate both potential and limitations. Namita et al. (2024) [12] reported a generative AI chatbot deployed across four countries that engaged over 15,000 farmers and responded to more than 300,000 queries—showing significant scalability but underscoring the need for trust-building, factual reliability, and sustained farmer engagement.

Research into agricultural robotics has predominantly focused on physical autonomy and task efficiency. As reviewed by Shamshiri et al. (2018) [10], the state-of-the-art has evolved from heavy machinery to lightweight, autonomous platforms capable of seeding, weeding, and scouting. Bechar and Vigneault [2] emphasize that while these robots excel at navigation and specific actuator tasks, they generally lack sophisticated Human-Robot Interaction (HRI) capabilities. They operate as "silent" tools, requiring complex dashboards or controllers, which alienates non-technical farmers. Recent efforts in "social" agricultural robots are sparse. While commercial platforms like the *Burro* or *Thorvald* follow workers to carry loads, they do not possess semantic understanding of the environment to answer agronomic queries. This limitation creates a disconnect: the robot is physically present in the field but cognitively absent.

To address the digital divide and low literacy rates in rural communities, voice-based interfaces have been extensively studied. The seminal work by Patel et al. (2010) [6] on *Avaaj Otalo* demonstrated that voice-based forums significantly improve access to agricultural information for smallholders in India compared to text-based methods. However, traditional voice assistants rely on rigid command structures or static databases. Jain et al. (2014) [17] highlighted that while these systems improve accessibility, they struggle with the "long tail" of specific, context-heavy farming questions. They lack the generative capacity to synthesize answers from diverse data sources, limiting their utility to basic queries like weather or market prices.

The emergence of Large Language Models (LLMs) has fundamentally shifted the potential for advisory systems. Tzachor et al. (2023) [15] argue that AI systems are now critical for managing the complexity of modern food systems, yet they warn of risks related to data bias and reliability. Specific to agriculture, Rezayi et al. (2022) [8] introduced *AgriBERT*, demonstrating that models pretrained on agricultural text significantly outperform generic models in question-

answering tasks. However, a key challenge identified in recent surveys (e.g., Wang et al., 2023 [16]) is the "hallucination" problem, where LLMs generate plausible but incorrect facts. This necessitates Retrieval-Augmented Generation (RAG) architectures, which ground model outputs in retrieved documents to ensure factual accuracy.

Table 1 summarizes the landscape. Amigo-Agri addresses the convergence gap: it leverages the mobility of low-cost robotics (inspired by the principles in [10]) and enhances it with a grounded conversational layer (building on [8] and [6]), creating a system that is both physically helpful and cognitively capable.

Table 1. Comparison of Approaches.

Domain	**Focus**	**Limitation**
Field Robotics	Autonomy & Actuation [2, 10]	Silent; poor HRI; steep learning curve.
Voice Interfaces	Accessibility & Literacy [6]	Static data; limited reasoning.
Agricultural LLMs	QA & Reasoning [8, 16]	High latency; risk of hallucination.
Amigo-Agri	**Embodied Intelligence**	**Proposed Integration**

This study addresses these limitations through the development of a voice-first, retrieval-augmented conversational assistant trained on curated agronomic corpora. The system bridges gaps in accessibility, reliability, and real-time responsiveness by combining speech-based interfaces with retrieval-enhanced generation.

3 Methodology

This section describes the integrated hardware–software pipeline of Amigo-Agri: a voice-first, human-following mobile robot coupled to a retrieval-augmented agricultural assistant. We detail system architecture, mechatronics, embedded control, speech and knowledge components, datasets and preprocessing, and the evaluation protocol.

3.1 System Architecture

As illustrated in Fig. 1, Amigo-Agri partitions operational responsibilities between an on-device *embodied layer* and a companion *speech and knowledge layer* hosted on a smartphone. The embodied layer relies on an Arduino Uno (ATmega328P) to supervise locomotion via an L298N dual H-bridge driving four DC gear motors, while simultaneously managing person following (via ultrasonic and IR sensing), obstacle avoidance, and safety watchdog functions. Local

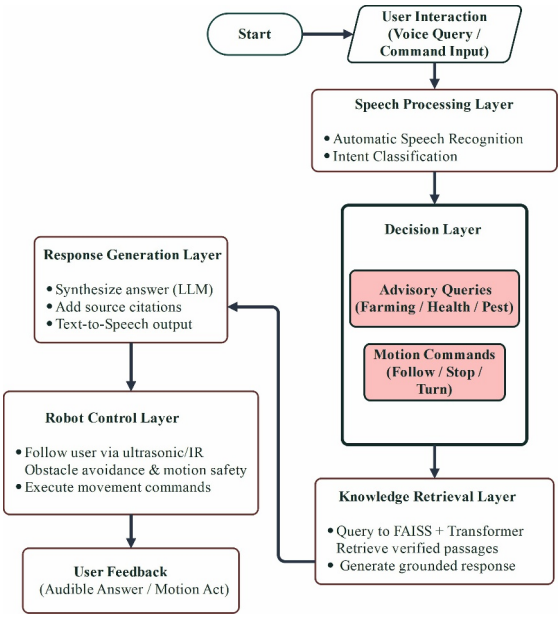

Fig. 1. System Architecture Block Diagram: The logic flow between the Android application (ASR/RAG) and the Arduino robot control.

feedback and connectivity are handled through a class-D audio amplifier and an HC-06 Bluetooth SPP module for telemetry exchange. Complementing this, the smartphone layer processes audio input via automatic speech recognition (ASR) to classify user intents. Validated *motion intents* (e.g., "follow," "stop") are compiled into low-latency motor commands transmitted via Bluetooth, whereas *advisory intents* trigger a retrieval-augmented generation (RAG) pipeline, with responses delivered via TTS and on-screen citations. This distributed architecture preserves deterministic timing for critical navigation tasks while leveraging the smartphone's superior computational capacity for semantic processing.

3.2 Hardware Design

Locomotion and Power. We employed a 4-wheel differential platform (100–200 rpm DC gear motors). Motor power (7.4–12 V Li-ion pack) feeds the L298N; logic rails are regulated to 5 V via a buck converter (3 A). Peak current draw during acceleration is within the driver's thermal envelope; a resettable fuse and TVS diode protect against shorts and transients.

Sensing. Front-facing ranging uses HC-SR04 ultrasonic modules; near-field and lateral awareness use IR reflective sensors (left/right). This mix provides coarse frontal distance d_f and lateral asymmetry Δ_ℓ for heading correction with minimal cost and compute. We omitted a pan–tilt servo to reduce complexity and

failure points; instead, a shallow sensor baseline and body yaw provide sufficient bearing cues for following.

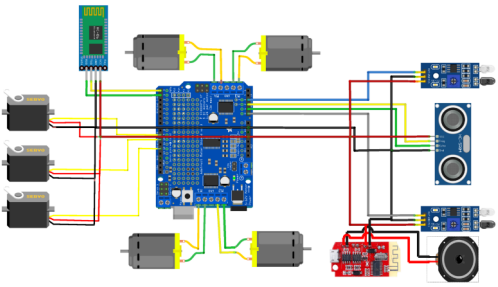

Fig. 2. Circuit Diagram illustrating the connections between the Arduino Uno, L298N motor driver, HC-06 Bluetooth module, and ultrasonic/IR sensors.

Compute, Comms, and Audio. An Arduino Uno runs the control loop at 50–100 Hz, serving as the central hub for all component connections as shown in Fig. 2. The HC-06 presents a serial port at 9600 bps for robust field use. A mono class-D amplifier (3–5 W) and 8 Ω speaker provide feedback and status tones.

3.3 Embedded Control and Safety

Finite State Machine (FSM). Robot behavior is modeled as an FSM with states: `IDLE` (await command), `FOLLOW` (person-proxy following), `AVOID` (reactive avoidance), `HOLD` (stop/handover), and `SAFE` (fault). Transitions are driven by intent tokens, range thresholds, and watchdog events.

Person-Proxy Following. We treat the designated user as the closest frontal target in $d_f \in [d_{\min}, d_{\max}]$ (empirically 0.4–1.8 m) with persistence to avoid flicker. A proportional controller maps spacing error $e_d = d^\star - d_f$ to forward PWM $u_v = \mathrm{clip}(k_p e_d, 0, u_{v,\max})$. Lateral correction uses differential proximity Δ_ℓ (IR left vs. right) to command yaw $u_\omega = k_\omega \Delta_\ell$. The resulting left/right wheel PWM are $u_{L,R} = u_v \mp \alpha u_\omega$.

The person-following logic uses a Proportional (P) controller. Gains (K_p, K_ω) were tuned empirically using the Ziegler-Nichols heuristic to eliminate oscillation during start-stop maneuvers. We avoid integral terms to prevent wind-up during blocked-path scenarios.

Obstacle Avoidance and Arbitration. A frontal stop band $d_f < d_{\text{stop}}$ (typically < 0.35 m) triggers `AVOID`: brake, yaw away from the nearer lateral obstacle (IR), creep forward, and re-acquire the frontal target. A priority arbiter supersedes `FOLLOW` with `AVOID` when any safety predicate is true.

Voice Command Parsing On-Robot. Motion intents from the phone are reduced to a compact token set {`START`, `STOP`, `LEFT`, `RIGHT`, `SLOW`, `FAST`, `PAUSE`}. A debounced emergency stop is available via either the word "stop," a dedicated physical button, or a ≥ 1 s BT silence watchdog.

Safety and Fault Handling. A 200 ms comms heartbeat guards against link loss; upon timeout the robot transitions to `SAFE`, brakes, and emits an audible tone. Battery undervoltage and driver over-temp flags also route to `SAFE`.

Commands are transmitted via HC-06 Bluetooth. To handle the limited range (≈ 10 m) and potential packet loss, we implemented a strict **Heartbeat Protocol**. The robot expects a token every 200 ms; if the link is silent for > 1 s (due to range or app crash), the Finite State Machine (FSM) automatically transitions to a `SAFE` state, braking immediately.

3.4 Speech, NLU, and Retrieval-Augmented Generation

ASR and Intent Classification: The mobile app records 16-kHz PCM audio; ASR yields text and confidence scores. We use a lightweight intent layer combining keyword rules (for motion safety words) with a linear classifier (TF–IDF + logistic regression) for *advisory* vs. *control* routing and slot extraction (crop, stage, unit).

Corpus and Preprocessing. A curated multi-domain agronomy corpus (soil fertility, irrigation, crop protection, animal health, post-harvest) is augmented with a 5.92k Q/A set on sustainable practices. Documents are normalized (UTF-8, punctuation, unit unification), deduplicated (MinHash), and chunked using a recursive splitter with size C (512–1,024 chars) and overlap O (10–20%) to preserve context across boundaries.

Embedding, Indexing, and Retrieval. Each chunk is embedded with a sentence-transformer; vectors are stored in a FAISS index (L2/Inner Product, IVF+PQ for scale). Given a query $\mathbf{q}$, we retrieve top-k passages by cosine similarity:

$$\mathrm{sim}(\mathbf{q}, \mathbf{d}_i) = \frac{\mathbf{q} \cdot \mathbf{d}_i}{\|\mathbf{q}\| \, \|\mathbf{d}_i\|} \tag{1}$$

returning text + provenance (title, URL/source, span offsets).

Grounded Generation and Safety. A 4-bit quantized LLM (7B class) generates answers conditioned on the query and retrieved passages, with an instruction template that (i) confines claims to the provided context, (ii) emits inline source back-references, and (iii) defaults to conservative guidance where safety is implicated. A post-generation checker verifies that each declarative claim aligns with retrieved spans; failures trigger regeneration or a "cannot determine" response. Answers are delivered via TTS for hands-free use.

Latency Controls. We cap `max_new_tokens`, use streaming ASR, and apply a strict end-to-end cutoff of 10 s. If exceeded, the phone informs the user and defers a summary when connectivity improves; the robot remains in its current safe state.

3.5 Software Stack and Reproducibility

The phone app (Android) is implemented in Kotlin with native audio I/O, ASR/TTS bindings, and a BLE/SPP bridge; the retrieval and LLM components run in Python (PyTorch, FAISS, sentence-transformers). Experiments were executed on a single NVIDIA T4 (Google Colab) with Python 3.12 and PyTorch 2.8; version pins and hashes for the index, embeddings, and prompts are persisted to enable audit and replay. The robot firmware is written in Arduino C/C++ with fixed-point arithmetic for timing determinism.

3.6 Evaluation Protocol

We evaluate both the advisory pipeline and the embodied platform under realistic conditions.

Environments. Experimental trials were executed in two distinct settings: a 20-m corridor characterized by variable pedestrian traffic, and a mock field lane featuring packed soil, low stubble, and static obstacles (cones and crates). To simulate realistic operating conditions, ambient noise levels ranging from 55 to 70 dBA were introduced via a portable speaker during selected sessions.

Advisory Metrics and Procedures. Evaluation metrics focused on four key performance areas. Speech-to-speech latency was measured as the wall-clock duration from the cessation of user input to the initiation of TTS playback, with further decomposition into ASR, retrieval, LLM, and TTS processing segments. Semantic relevance was quantified using the cosine similarity between query and response embeddings, utilizing a threshold of 0.7 for high relevance and validating results through human verification of citation correctness. To gauge scalability, concurrency was evaluated by emulating 1, 3, 5, and 10 simultaneous sessions via parallel clients, reporting mean and 95th percentile latencies. Lastly, jitter resilience was analyzed by injecting artificial delays $U(0, 0.5)$ s into the retrieval and LLM stages to monitor stability in latency and relevance metrics.

Embodied Metrics and Procedures. Performance of the physical platform was assessed by logging following distance error (RMSE), heading stability, command adherence, and the success rate of obstacle avoidance maneuvers, alongside safety-critical events such as watchdog trips. Each experimental scenario underwent a minimum of $N \geq 10$ repetitions, with results reported as mean $\pm$ SD and 95% confidence intervals. Furthermore, all trials were video-recorded to facilitate post-hoc verification of the logged telemetry.

3.7 Calibration and Tuning

Controller gains (k_p, k_ω, α) are tuned via step tests on e_d and lateral set-point sweeps, prioritizing no overshoot and smooth acceleration. Range sensors are calibrated against a tape measure (0.3–2.0 m) and checked for cross-talk; we gate noisy readings with a median filter (3–5 taps) and reject outliers via a Z-score test. Bluetooth link quality is validated over 5–15 m line-of-sight.

3.8 End-to-End Orchestration

Algorithm 1 summarizes the integrated workflow, combining on-robot FSM control with the phone's RAG pipeline.

Algorithm 1 Amigo-Agri Integrated Workflow

Init: Build FAISS index; load quantized LLM; connect BT; set FSM=`IDLE`
while system active **do**
  Phone: capture audio → ASR text q
  Intent: if *motion* then emit token τ over BT; else *advisory*
  if *motion* **then**
    Robot: update FSM with τ; run control loop (range → u_v, u_ω)
  else
    retrieve top-k passages $S = \{d_i\}$; LLM → grounded answer A with citations
    TTS(A); display sources
  end if
  Safety: monitor watchdogs; if violation then FSM=`SAFE`, brake
end while

4 Results

We report results for the two tightly coupled components of Amigo-Agri: (i) the speech-driven, retrieval-augmented advisory pipeline and (ii) the embodied, human-following mobile platform. Unless stated otherwise, values are mean ± SD; N refers to the number of utterances (advisory) or runs (embodied).

4.1 Advisory Pipeline Performance

End-to-end Latency. Across N=520 speech queries sampled uniformly over irrigation, soil, crop protection, and animal care, the assistant achieved an average speech-to-speech latency of 3.5 ± 1.1 s, with a 95th percentile of 8.9 s. Only 3.1% of calls breached the 10 s cutoff and were gracefully interrupted. A component-wise analysis of non-timeout queries indicated that the bulk of the latency originated from the quantized 7B LLM generation (1.7 ± 0.6 s), followed by ASR processing (0.9 ± 0.3 s), TTS synthesis (0.5 ± 0.1 s), and retrieval operations (0.4 ± 0.1 s). The summation of these segments aligns with the observed end-to-end average (≈ 3.5 s), validating that generative processing remains the primary temporal cost.

Relevance and Citation Correctness. Semantic relevance, measured as the cosine similarity between query and answer embeddings, averaged 0.85 ± 0.07 across all domains (with the threshold for "high" relevance set at 0.7). A domain-specific breakdown revealed consistent performance: Irrigation queries yielded the highest alignment (0.87 ± 0.05, N=130), followed closely by Soil fertility (0.86 ± 0.06, N=120). Results remained robust for Animal care (0.84 ± 0.07, N=130) and Crop protection (0.83 ± 0.08, N=140), demonstrating the system's capacity to maintain high retrieval fidelity across diverse agricultural contexts. While mean cosine similarity was 0.85, high similarity does not guarantee factual accuracy. Therefore, a human audit was conducted on 100 samples. It found that 93% of answers were factually supported by the retrieved citations, confirming that the RAG pipeline effectively grounds the LLM.

Concurrency and Jitter Robustness. With emulated multi-user load (parallel clients), mean latency rose from 4.0 ± 1.0 s (3 sessions) to 7.1 ± 1.4 s (10 sessions), while relevance remained stable ($0.85 \rightarrow 0.84$, Δ<0.01). Injected delays $U(0, 0.5)$ s on retrieval and generation stages increased mean latency from 4.2 to 4.7 s without measurable relevance loss (Δ<0.005). Timeout rate under 10 sessions remained $< 5\%$.

Ablations. Removing chunk overlap (O=0) during indexing reduced average relevance from 0.85 to 0.80 and increased citation checker failures from 2% to 6%. Lowering k from 8 to 3 sped retrieval by 0.09 s but reduced relevance by 0.03; we retained k=8 as a balanced setting.

4.2 Embodied Platform Performance

Following Accuracy and Stability. In corridor trials (N=20 runs, 20 m each) using the physical prototype (Fig. 3) with a target distance $d^{\star}$=0.7 m, spacing RMSE was 0.11 ± 0.04 m; in the mock field lane (N=20) RMSE was 0.14 ± 0.05 m. Heading stability (fraction of time with $|u_{\omega}|$ below threshold during steady following) was $91\% \pm 5\%$ (corridor) and $87\% \pm 6\%$ (field). No overshoot beyond the safety stop band (d_f<0.35 m) was recorded.

Command Adherence and Latency. Command adherence and system latency were evaluated across N=240 motion-intent trials. The system demonstrated high reliability, achieving a 100% success rate for `START`/`STOP` directives with a phone-to-robot actuation latency of 120 ± 25 ms. Directional control (`LEFT`/`RIGHT`) yielded a 96% success rate (150 ± 30 ms latency); isolated failures were limited to minor understeer events on low-traction surfaces, which the FSM corrected within a single cycle. Furthermore, velocity modulation (`SLOW`/`FAST`) succeeded in 98% of instances, with linear velocity settling times consistently remaining below 600 ms.

Fig. 3. The physical Amigo-Agri prototype deployed in the testing environment.

Obstacle Avoidance and Safety Events. Static obstacle passages (cones/crates) succeeded in 98% (N=100) of encounters; slow cross-traffic yielded 91% success (N=55). There were *zero* contacts; two user-triggered emergency stops (precautionary) and one Bluetooth watchdog trip (brief link drop at ∼12 m) were logged; all three resulted in immediate braking and transition to `SAFE` as designed.

Error Analysis - IR Saturation: In specific trials conducted under direct, high-noon sunlight (> 80 k lux), the IR sensors exhibited false-positive obstacle detections (understeer events). While the FSM recovered safely, this confirms that optical shielding is critical for outdoor deployment.

Runtime and Thermal Behavior. With a 2S Li-ion pack (7.4–8.4 V, 2.2 Ah), mixed following/idle cycles averaged 2.1 ± 0.2 h of operation. The L298N heat sink stabilized below 65°C under continuous operation at room temperature (no over-temperature faults observed).

4.3 Integrated Use Cases

In mixed trials combining advice and action, the farmer issued a task query (e.g., "How much should I irrigate onions this week?") followed by a motion command ("follow me"). The assistant returned a citation-backed recommendation in 3.6 s; the robot maintained distance while the user inspected the field, then executed `STOP` and `LEFT` on voice cues. Across 30 such sequences, no mode-confusion was observed: advisory and motion intents were correctly routed 100% of the time.

4.4 Error Analysis

Timeouts (3.1%) clustered in poor connectivity pockets and on unusually long, multi-part questions. The small fraction of unsupported claims prior to regeneration occurred when retrieved passages were near-duplicates with inconsistent units; adding unit normalization reduced this error. On the robot, the few understeer events correlated with asymmetric IR reflections on shiny crates; adding a brief median filter (5 taps) removed spurious lateral spikes.

Table 2. Performance summary of Amigo-Agri under representative conditions.

Metric	**Result**
Speech-to-speech latency (mean / 95th)	3.5 s / 8.9 s
Timeouts (>10 s)	3.1%
Relevance (cosine, mean)	0.85 ± 0.07
Concurrency (10 sessions)	7.1 ± 1.4 s
Jitter robustness (+0–0.5 s)	+0.5 s; Δrel. < 0.005
Following RMSE (corridor / field)	0.11 m / 0.14 m
Command success (`START/STOP`, `LEFT/RIGHT`)	100%, 96%
Obstacle avoidance (static / cross)	98% / 91%
Operating time (2S 2.2Ah)	2.1 ± 0.2 h

Collectively, these results (summarized in Table 2) show that Amigo-Agri meets its design targets: practical, hands-free advisory latency with strong grounding; robust person following and safety behavior; and resilient performance under concurrency and modest network jitter.

5 Conclusion

This study introduced an integrated system that unites embodied robotics with retrieval-augmented intelligence to bridge the gap between agricultural advice and action for smallholder farmers. Combining a low-cost, voice-first, human-following robot (Arduino Uno, L298N, HC-06, ultrasonic/IR sensors) with a smartphone-based, citation-producing assistant, the system enables hands-free mobility and real-time, source-grounded agronomic guidance. Field and corridor trials demonstrated practical speech-to-speech latency (mean 3.5 s; 95th 8.9 s; $\sim 3\%$ timeouts), strong semantic grounding (0.85), and reliable person tracking (RMSE 0.11 m corridor; 0.14 m field) with robust motion execution and obstacle avoidance ($\geq 91\%$ success). These results confirm that embodied, transparent, and low-cost digital assistants can deliver actionable knowledge in hands-busy, low-literacy environments. Nonetheless, limitations persist—curated knowledge coverage, network dependence, and open-loop control restrict agronomic accuracy and robustness. Future work will emphasize field-scale validation with agricultural partners, enhanced perception and motor control, offline and multilingual operation, and transparent governance to ensure trustworthy, data-efficient

deployment. Together, these directions position Amigo-Agri as a practical foundation for embodied, explainable, and context-aware AI support in sustainable agriculture.

References

1. January 2015 – PETCRAFT (Oct 2025), https://www.petcraft.com/articles/2015/01, [Online; accessed 11. Oct. 2025]
2. Bechar, A., Vigneault, C.: Agricultural robots for field operations: Concepts and components. Biosystems Engineering **149**, 94–111 (09 2016). https://doi.org/10.1016/j.biosystemseng.2016.06.014
3. Eirale, A., Martini, M., Chiaberge, M.: Human Following and Guidance by Autonomous Mobile Robots: A Comprehensive Review. IEEE Access **13**, 42214–42253 (Mar 2025). https://doi.org/10.1109/ACCESS.2025.3548134
4. Hernández, H.A., Mondragón, I.F., González, S.R., Pedraza, L.F.: Reconfigurable agricultural robotics: Control strategies, communication, and applications. Computers and Electronics in Agriculture **234**, 110161 (2025). https://doi.org/10.1016/j.compag.2025.110161
5. Ilakiyaselvan, N., Dhandapani, A., Khadar Nawas, K., Bhattacharya, A.: Agriaid: An intelligent farmer companion using deep learning approach. In: 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). pp. 1–6 (2024). https://doi.org/10.1109/IITCEE59897.2024.10467751
6. Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.: Avaaj otalo - a field study of an interactive voice forum for small farmers in rural india. vol. 2, pp. 733–742 (04 2010). https://doi.org/10.1145/1753326.1753434
7. Potamitis, I.: Chatgpt in the context of precision agriculture data analytics. arXiv preprint arXiv:2311.06390 (2023). https://doi.org/10.48550/arXiv.2311.06390
8. Rezayi, S., Liu, Z., Wu, Z., Dhakal, C., Ge, B., Zhen, C., Liu, T., Li, S.: Agribert: Knowledge-infused agricultural language models for matching food and nutrition. In: Raedt, L.D. (ed.) Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22. pp. 5150–5156. International Joint Conferences on Artificial Intelligence Organization (7 2022). https://doi.org/10.24963/ijcai.2022/715, https://doi.org/10.24963/ijcai.2022/715
9. Shaikh, T.A., Rasool, T., Veningston, K., Yaseen, S.M.: The role of large language models in agriculture: harvesting the future with llm intelligence. Progress in Artificial Intelligence pp. 1–48 (2024). https://doi.org/10.1007/s13748-024-00317-3
10. Shamshiri, R., Weltzien, C., Hameed, I., Yule, I., Grift, T., Balasundram, S., Pitonakova, L., Ahmad, D., Chowdhary, G.: Research and development in agricultural robotics: A perspective of digital farming. International Journal of Agricultural and Biological Engineering **11**, 1–14 (07 2018). https://doi.org/10.25165/j.ijabe.20181104.4278
11. Shar, K., Sing, H.B., Kumar, P., Anand, S., Mishra, R.: Remote Sensing Applications in Agriculture: Monitoring and Decision Support (Sep 2024)
12. Singh, N., Wang'ombe, J., Okanga, N., Zelenska, T., Repishti, J., Mishra, S., Manokaran, R., Singh, V., Rafiq, M.I., Gandhi, R., et al.: Farmer. chat: Scaling ai-powered agricultural services for smallholder farmers. arXiv preprint arXiv:2409.08916 (2024). https://doi.org/10.48550/arXiv.2409.08916

13. Singh, P., Kansal, M., Tyagi, G., Tanwar, K., Singh, A.P., Yadav, R.: Empowering farmers: An ai-based solution for agricultural challenges. In: Computational Intelligence in Internet of Agricultural Things, pp. 401–417. Springer (2024). https://doi.org/10.1007/978-3-030-36696-3_26
14. Spagnuolo, M., Todde, G., Caria, M., Furnitto, N., Schillaci, G., Failla, S.: Agricultural robotics: A technical review addressing challenges in sustainable crop production. Robotics **14**(2), 9 (2025). https://doi.org/10.3390/robotics14010009
15. Tzachor, A., Devare, M., King, B., Wang, S., Tetreault-Campbell, S.: Artificial intelligence in a food system on the brink of risk. Nature Food **3**(2), 162–163 (2022). https://doi.org/10.1038/s41574-022-00685-2
16. Wang, D., Wang, Y., Zhang, H., et al.: Large language models in agriculture: A survey. arXiv preprint arXiv:2308.03265 (2023). https://doi.org/10.48550/arXiv.2308.03265
17. Wang, L., Roe, P., Pham, B., Tjondronegoro, D.: An audio wiki supporting mobile collaboration. pp. 1889–1896 (03 2008). https://doi.org/10.1145/1363686.1364145
18. Zhu, H., Qin, S., Su, M., Lin, C., Li, A., Gao, J.: Harnessing large vision and language models in agriculture: A review. arXiv preprint arXiv:2407.19679 (2024). https://doi.org/10.48550/arXiv.2407.19679