Design and Implementation of a Greeting Robot for the AIML Department

Ishan Kulkarni
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
ishanproj@gmail.com

Sharada Pujari
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
sharadapujari7709@gmail.com

Aditi Patil
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
adi.patil2329@gmail.com

Sourabh Gurav
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
gsourabh369@gmail.com

Girish Desai
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
girish.desai@aol.com

Uma Gurav
CSE (AIML)
KIT's College of Engineering
*Kolhapur*Maharashtra, India
gurav.uma@kitcoek.in

***Abstract*—This paper presents the design and implementation of a budget-friendly greeting robot for the Artificial Intelligence and Machine Learning (AIML) department at KIT's College of Engineering. The robot serves as an interactive assistant, providing navigational guidance and answering visitor queries. Built using a Raspberry Pi 5, Arduino Uno (for LED-based user feedback), and a custom-trained Mistral LLM hosted via Ollama, the system integrates pyttsx3 for text-to-speech and speech_recognition for speech-to-text. Designed for cost-effectiveness, the robot leverages open-source tools and off-the-shelf hardware, achieving a total hardware cost under $300—significantly lower than commercial alternatives. Preliminary results demonstrate 85% speech recognition accuracy and 90% LLM response relevance, validated through user feedback. This project highlights the feasibility of deploying AI-driven assistive robots in educational settings with constrained budgets.**

Index Terms: Human-Robot Interaction, Low-Cost Robotics, Conversational AI, Educational Assistants.

I. Introduction

Imagine stepping into the AIML department corridor, a bustling hub of students, faculty, and visitors, and being greeted by a friendly robotic assistant ready to assist. This vision sparked our project: a greeting robot designed to serve as a reliable companion, offering directions, answering queries, and easing the load on human staff. Where traditional aids like signs and personnel falter—limited by availability and scale—our robot steps in, powered by cutting-edge AI to deliver real-time, personalized support.

The system marries Ollama's hosting of Samantha-Mistral, a sophisticated LLM, with pyttsx3 for speech synthesis, speech_recognition for audio input processing, and pyaudio for seamless audio handling. This paper unfolds the story of its creation, structured as follows: Section II explores prior work that inspired us, Section III narrates the design and implementation journey, Section IV shares early outcomes, Section V reflects on insights and future paths, and Section VI wraps up with our contributions and vision.

This paper is structured as follows: Section II reviews related work on greeting robots, conversational agents, and relevant technologies. Section III outlines the methodology, detailing the system's design and implementation. Section IV presents preliminary results, while Section V discusses findings and future directions. Section VI concludes with the project's contributions and implications.

II. Literature Review

Our robot's story draws from a rich tapestry of research. Kanda et al. [1] brought robots to shopping malls, guiding visitors with spoken dialogue, while Burgard et al. [2] crafted a museum guide robot that spoke contextual tales. In education, Kerly et al. [3] and Fryer et al. [4] showcased chatbots easing tutoring and inquiry burdens, inspiring our shift to a physical form. LLMs like GPT-3 [5] and their fine-tuned successors [6] fueled Samantha-Mistral's conversational prowess. Speech technologies—neural TTS [7] and deep learning SR [8]—set the stage, though noise challenges linger [9]. HRI principles from Fong et al. [10], Dautenhahn [11], and Norman [12], alongside works on robotics and NLP [13]-[25], guided our design, blending past lessons into a fresh narrative.

III. Related Works

The development of our greeting robot builds upon a rich body of research in robotics, conversational agents, and human-robot interaction (HRI), particularly in public and educational settings. Robots designed for public spaces have demonstrated significant potential in assisting users through natural language interaction. For instance, Kanda et al. [1] introduced a communication robot deployed in a shopping mall, capable of guiding visitors and providing information. Their work highlighted how robots can enhance user engagement through intuitive spoken dialogue. Similarly, Burgard et al. [2] developed a museum guide robot that used speech to deliver contextual information, emphasizing the need for adaptability in dynamic, unpredictable environments. These studies inform our project by demonstrating the effectiveness of robots in navigation and information dissemination, key functionalities of our greeting robot in the AIML department.

In educational contexts, conversational agents have gained traction as tools to support students and staff. Kerly et al. [3] explored chatbots in tutoring systems, noting their ability to deliver personalized learning experiences. Fryer et al. [4] evaluated a university chatbot designed to address student inquiries, reporting benefits such as reduced staff workload and improved satisfaction due to timely responses. While these systems are predominantly virtual, our project distinguishes itself by integrating such capabilities into a physical robot. This embodied approach aims to leverage the advantages of physical presence, such as increased user trust and engagement, which virtual agents may lack.

Advancements in large language models (LLMs) have significantly enhanced the capabilities of conversational systems. Brown et al. [5] introduced GPT-3, a groundbreaking LLM capable of generating coherent, human-like responses across diverse topics. Radford et al. [6] further showcased the adaptability of LLMs in chatbot applications through domain-specific fine-tuning. Our project employs Samantha-Mistral, a variant of such models, fine-tuned for the AIML department's needs, enabling contextually relevant and accurate responses to user queries.

Speech technologies form the backbone of our robot's interaction capabilities. Zen et al. [7] reviewed neural text-to-speech (TTS) models, which offer superior naturalness compared to traditional systems like pyttsx3, used in our implementation. Despite the availability of advanced options, we selected pyttsx3 for its offline functionality and ease of integration, aligning with our project's practical constraints. On the speech recognition (SR) front, Hinton et al. [8] advanced the field with deep learning techniques, improving transcription accuracy across various conditions. However, challenges remain in noisy environments [9], a critical consideration for our corridor deployment, where ambient noise could impact performance.

Human-robot interaction research provides essential design principles for our system. Fong et al. [10] outlined the characteristics of socially interactive robots, emphasizing responsiveness and friendliness—core attributes of our greeting robot. Dautenhahn [11] stressed the importance of adaptive behaviors in service robots, guiding our approach to handling diverse user queries. Additionally, Norman's [12] principles of user experience design influenced our iterative testing process, incorporating user feedback to refine the system.

Complementary studies further enrich our project's foundation. Siegwart et al. [13] explored autonomous mobile robots, offering insights into potential future enhancements like mobility, though our current design remains stationary. Yamagishi et al. [14] investigated robust TTS systems, relevant for upgrading our speech output in subsequent iterations. Nielsen's [15] usability engineering framework informed our evaluation metrics, ensuring a user-centric design.

In summary, our greeting robot synthesizes insights from public-space robotics, educational conversational agents, and cutting-edge AI technologies. By combining a physical presence with advanced natural language processing (NLP), this project carves a unique niche, addressing specific needs within the AIML department corridor while building on established research.

IV. Proposed Work

The greeting robot is designed to function as an interactive assistant stationed in the AIML department corridor, offering greetings, navigational guidance, and responses to queries from parents, staff, and students. This section details the system's hardware and software components, integration approach, and the innovations and challenges addressed in its development.

Hardware Components
The robot's hardware is carefully selected to ensure reliable performance in a corridor environment:

Microphone: A high-sensitivity omnidirectional microphone captures user speech effectively, even amidst moderate background noise. Its omnidirectional design allows the robot to detect input from multiple directions, accommodating users approaching from various angles.
Speaker: A compact, high-quality speaker delivers clear audio responses, optimized for human voice reproduction to ensure natural and audible output in the corridor setting.
Chassis: The robot is housed in a stationary chassis, prioritizing stability and approachability. While future versions may explore mobility, the current design focuses on providing a consistent, accessible interaction point.

Software Architecture

The software stack integrates several components to enable seamless spoken interaction:
Speech Recognition (SR): The speech_recognition library, paired with pyaudio for audio stream management, transcribes spoken input into text. We utilize the Google Web Speech API for transcription, striking a balance between accuracy and computational efficiency.
Large Language Model (LLM): Samantha-Mistral, hosted on the Ollama platform, processes user queries and generates responses. Fine-tuned on a custom dataset tailored to the AIML department, this LLM ensures contextually appropriate answers.
Text-to-Speech (TTS): Pyttsx3 converts text responses into speech. Chosen for its offline capabilities and straightforward integration, it supports the robot's operation without constant internet reliance.
Audio Engine: Pyaudio synchronizes input capture and output playback, ensuring smooth audio interactions.
System Integration

The interaction process follows a streamlined flow:
Input Capture: The microphone listens continuously, activating upon detecting speech and passing the audio to the SR module.
Transcription: The SR module converts the audio into text via the Google Web Speech API.
Query Processing: The transcribed text is sent to Samantha-Mistral through the Ollama API, which generates a text response.
Response Generation: Pyttsx3 synthesizes the text into speech.
Output: The speaker plays the synthesized response, completing the interaction cycle.

LLM Training
To optimize Samantha-Mistral for the AIML department, we

curated a specialized dataset including:
Room location descriptions (e.g., "The machine learning lab is on the second floor, room 203.")
Staff details (e.g., "Dr. Smith is the head of the department, located in room 101.")
Common queries (e.g., "When are Professor Johnson's office hours?")
Navigational instructions (e.g., "Take the elevator to the third floor and turn left for the seminar hall.")
This dataset, sourced from department resources and manually crafted Q&A pairs, was used to fine-tune the LLM via transfer learning, enhancing its domain-specific knowledge while preserving general language proficiency.

Innovations and Challenges
Our system introduces several innovations:

Embodied Interaction: By integrating advanced NLP into a physical robot, we enhance user engagement in an educational context, distinguishing it from virtual assistants.
Contextual Awareness: Fine-tuning Samantha-Mistral ensures responses are tailored to the AIML department's unique needs.
Key challenges include:

Noise Handling: Corridor noise can degrade SR accuracy. We implemented a wake-word system to activate listening only when addressed, mitigating false triggers.
LLM Limitations: Samantha-Mistral may falter with ambiguous or out-of-domain queries. Fallback responses and rephrasing prompts help address this.
User Experience: A minimalist design and friendly voice were chosen to make the robot approachable, validated through user feedback.
This proposed work delivers a practical, AI-driven solution tailored to the AIML department, overcoming technical hurdles while prioritizing usability.

V. METHODOLGY

The tale of our greeting robot unfolds through its design—a harmonious blend of hardware and software crafted to thrive in the corridor's lively chaos.

A. Hardware Design

Raspberry Pi 5: Central controller for speech processing and system coordination .
Arduino Uno: Dedicated LED control for visual cues (e.g., blinking when listening, solid when responding).
PC Server: Hosts the Mistral LLM via Ollama (using existing lab infrastructure).
Shotgun Microphone: Directional audio capture to mitigate corridor noise .
Speaker: Basic but clear audio output.
Custom Chassis: 3D-printed enclosure.
Total Hardware Cost: ~$130 (excluding PC, using existing resources).

B. Software Architecture

The robot's voice comes alive through a carefully woven software stack:

Speech Recognition: Google Web Speech API (free tier) via speech_recognition.

- **LLM**: **Mistral-7B** (open-weight, fine-tuned on department-specific Q&A).
- **TTS**: pyttsx3 (offline, no API costs).
- **LED Control**: Custom Arduino firmware for real-time visual feedback.

C. System Workflow

1. **Continuous Listening**: No wake-word trigger—microphone remains active (optimized for quick responses).
2. **Visual Feedback**: Arduino-driven LEDs indicate system states:
   - *Blue*: Idle/Ready
   - *Pulsing Yellow*: Processing Query
   - *Green*: Speaking Response
3. **Cost Optimization**:
   - Avoided premium components (e.g., commercial TTS/ASR APIs).
   - Repurposed existing lab PCs for LLM hosting.

D. LLM Training

To make Samantha-Mistral a true AIML insider, we fed it a custom diet: room maps ("Room 203, second floor"), staff bios ("Dr. Smith, room 101"), and common queries ("When's Professor Johnson free?"). Transfer learning honed its skills, blending general fluency with department savvy.

E. Innovations and Challenges

Our robot shines with innovations like its physical presence and tailored responses, but challenges emerged:

- **Noise**: The shotgun mic battles corridor clamor, though positioning matters. User prompts help.
- **Networking**: The PC server split demands a steady router link; we built in error handling for hiccups.
- **Vision**: The camera hints at future tricks—recognizing faces or waves—ripe for exploration.

This journey birthed a robot ready to serve, with lessons paving the way for growth.

VI. RESULTS AND DISCUSSIONS

Early tests tell a promising tale. Speech recognition hit 85% accuracy across 50 queries, stumbling on noise and accents—proof the shotgun mic helps but isn't perfect. Samantha-Mistral nailed 90% of 100 AIML-specific questions, though tricky ones tripped her up. Ten users rated it: 4.2 for helpfulness, 4.5 for friendliness, 4.0 for ease, lauding its charm but noting noise woes. The camera's potential sparked excitement for what's next.

Our robot weaves AI and robotics into education's fabric, tackling noise [9], LLM quirks [6], and design finesse [12]. It's a working prototype lifting visitor support, with paths to sharper SR [16], richer training [17], and camera-driven flair [18] ahead.

VII. Conclusion

In the AIML corridor, our greeting robot stands as a beacon of AI and robotics united. Samantha-Mistral's wit, paired with speech tech, welcomes all with directions and answers. Early praise for its performance and demeanor fuels its promise. Fixed yet flexible, it's a model for broader use, with room to grow—taming noise, enriching its voice, or seeing with its camera. This robot is both a solution today and a glimpse of tomorrow's educational allies.

VI. References

[1] T. Kanda et al., "A Communication Robot in a Shopping Mall," *IEEE Trans. Robot.*, vol. 27, no. 5, pp. 897-908, Oct. 2011.

[2] W. Burgard et al., "A Museum Guide Robot," *Proc. IEEE Int. Conf. Robot. Autom.*, 1998, pp. 224-229.

[3] A. Kerly et al., "Conversational Agents in E-Learning," *Int. J. Artif. Intell. Educ.*, vol. 18, no. 2, pp. 89-112, 2008.

[4] L. Fryer et al., "Chatbots in Education," *Educ. Technol. Soc.*, vol. 22, no. 3, pp. 56-68, 2019.

[5] T. Brown et al., "Language Models are Few-Shot Learners," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 1877-1901, 2020.

[6] A. Radford et al., "Improving Language Understanding by Generative Pre-Training," arXiv:1801.06146, 2018.

[7] H. Zen et al., "Statistical Parametric Speech Synthesis," *Speech Commun.*, vol. 51, no. 11, pp. 1039-1064, 2009.

[8] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82-97, 2012.

[9] L. Deng, "Speech Recognition in Adverse Environments," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 5, pp. 925-936, 2013.

[10] T. Fong et al., "A Survey of Socially Interactive Robots," *Robot. Auton. Syst.*, vol. 42, no. 3-4, pp. 143-166, 2003.

[11] K. Dautenhahn, "Socially Intelligent Robots," *Robot. Auton. Syst.*, vol. 55, no. 3, pp. 199-214, 2007.

[12] D. Norman, *The Design of Everyday Things*, New York: Basic Books, 2013.

[13] R. Siegwart et al., "Autonomous Mobile Robots," *IEEE Robot. Autom. Mag.*, vol. 11, no. 1, pp. 12-20, 2004.

[14] J. Yamagishi et al., "Robust Speaker-Adaptive HMM-Based TTS," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 6, pp. 1208-1230, 2009.

[15] J. Nielsen, "Usability Engineering," *ACM Comput. Surv.*, vol. 26, no. 1, pp. 65-68, 1994.

[16] Y. Gong, "Speech Recognition in Noisy Environments," *Speech Commun.*, vol. 16, no. 3, pp. 261-291, 1995.

[17] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers," *Proc. NAACL*, 2019, pp. 4171-4186.

[18] C. Breazeal, "Toward Sociable Robots," *Robot. Auton. Syst.*, vol. 42, no. 3-4, pp. 167-175, 2003.

[19] P. Liu et al., "Fine-Tuning LLMs for Specific Domains," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3456-3468, 2021.

[20] S. Young et al., "The Hidden Markov Model Toolkit," *Proc. Eurospeech*, 1997, pp. 133-136.

[21] R. Brooks, "A Robust Layered Control System for a Mobile Robot," *IEEE J. Robot. Autom.*, vol. 2, no. 1, pp. 14-23, 1986.

[22] M. Matarić, "The Robotics Primer," *IEEE Robot. Autom. Mag.*, vol. 14, no. 3, pp. 20-21, 2007.

[23] A. Vaswani et al., "Attention is All You Need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998-6008, 2017.

[24] B. Zhang et al., "Real-Time Speech Recognition Systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 2145-2157, 2020.

[25] J. Pineau et al., "Reinforcement Learning for Adaptive Robots," *IEEE Trans. Syst. Man Cybern. B*, vol. 36, no. 6, pp. 1295-1307, 2006.