DESIGN AND DEVELOPMENT OF MULTI-LINGUAL AI-BASED EDUCATIONAL ROBOT FOR SOCIAL AND EMOTIONAL SKILL DEVELOPMENT IN PRESCHOOL AND PRIMARY CHILDREN ACROSS RURAL AND URBAN SETTINGS

Project Reference No.: 49S_BE_6375

College	:	*Christ Deemed To Be University, Kengeri Campus, Bengaluru*
Branch	:	*Department of Robotics and Mechatronics Engineering*
Guide(s)	:	*Dr. Amruta Rout*
Student(s)	:	*Mr. Alen George*
		Mr. Cristiano Lenin
		Mr. Jeyakrishna V

Keywords:

Educational Robotics, Artificial Intelligence, Voice Recognition, Emotion Display, Human-Robot Interaction

Introduction:

The integration of robotics and artificial intelligence in education has emerged as a transformative approach to engage young learners. Traditional teaching methods often struggle to maintain student attention and provide personalized learning experiences. Educational robots bridge this gap by offering interactive, emotionally responsive, and adaptive learning environments that cater to individual student needs.

SAM (Student Academic Mentor) is an AI-powered educational robot designed as a comprehensive learning platform serving dual educational purposes. For children aged 6–12 years, SAM functions as an interactive learning companion providing personalized tutoring, emotional engagement, and adaptive educational content. For older students and hobbyists, SAM serves as an advanced robotics assembly and coding project enabling hands-on learning in mechatronics, embedded systems, AI integration, and software development.

The robot combines continuous speech recognition with Voice Activity Detection, natural language processing using edge AI models, real-time computer vision for autonomous face tracking, and a 7-emotion display providing non-verbal feedback. Unlike conventional educational tools that rely on cloud services, SAM operates entirely on edge computing using NVIDIA Jetson Orin Nano, ensuring complete data privacy and eliminating internet dependency — critical for resource-constrained educational environments across rural India.

The project addresses the growing need for accessible, affordable educational technology in Indian schools. By incorporating multilingual capabilities for English, Hindi, and Kannada through AI4Bharat models, SAM democratizes access to quality educational assistance. The modular hardware architecture (AI processing on Jetson, motor control on

Arduino Mega, emotion display on ESP32, game server on Raspberry Pi 4) enables both end-user deployment and serves as a comprehensive STEM learning platform.

Objectives:

1. Develop an interactive educational robot engaging children aged 6–12 through natural voice interaction, real-time emotion display, and physical gestures, while simultaneously providing a comprehensive assembly and coding platform for older students learning robotics and AI integration.
2. Implement a continuous voice activity detection system eliminating artificial timeout constraints, achieving natural conversation flow with 2–3 second response latency and 96% speech recognition accuracy.
3. Create an emotion-aware system featuring 7 distinct emotional states (idle, happy, sad, thinking, talking, enthusiasm, playfulness) displayed in real-time, synchronized with interaction context to enhance student engagement.
4. Integrate adaptive educational games (mathematics quiz, Tic-Tac-Toe) with difficulty scaling and a streak-based celebration system (triggering after 5 consecutive correct answers) to maintain student motivation.
5. Build an autonomous face tracking system using PID-controlled DC motors (Kp=0.12, Kd=0.25, dead zone=90 pixels) maintaining smooth visual engagement without mechanical oscillation.
6. Establish complete multilingual support infrastructure for English, Hindi, and Kannada using AI4Bharat IndicWhisper, IndicTrans2, and IndicTTS models for inclusive education across linguistic backgrounds.
7. Ensure absolute data privacy and offline operation through edge computing on NVIDIA Jetson Orin Nano, processing all AI operations locally without cloud dependency.
8. Design modular architecture with clear component separation enabling easy maintenance, incremental upgrades, and educational use as a documented assembly project for STEM learners.

Methodology:

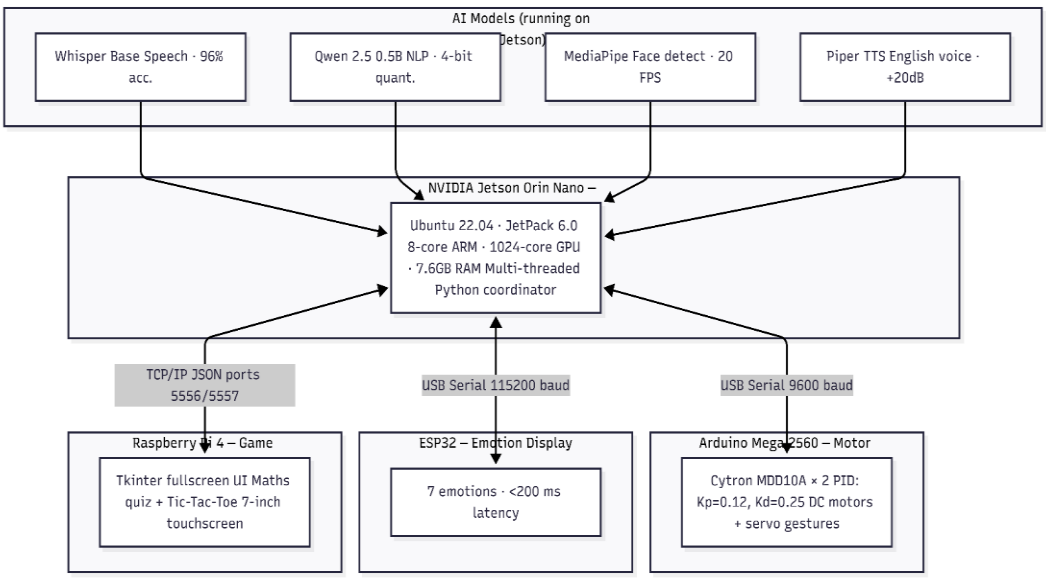

Figure 1: SAM Robot System Architecture — Block diagram showing NVIDIA Jetson Orin Nano, Arduino Mega 2560, ESP32, and Raspberry Pi 4 with USB Serial and TCP/IP communication links

Development of SAM:

SAM was developed through an iterative hardware-software co-design methodology. The development process began with requirements analysis derived from the literature survey, followed by system architecture design, component selection, physical fabrication, firmware development, AI model integration, and iterative testing with target users. Each subsystem was developed and validated independently before full system integration, ensuring modular reliability and ease of troubleshooting.

Primary Design and CAD Modelling:

The initial conceptual design phase established SAM's dual-purpose educational framework and modular architecture. CAD modelling using SolidWorks created detailed 3D models of the mechanical structure including motor mounting brackets, servo linkages for hand gestures, camera positioning, and component integration layout. The design prioritized child-safe construction with rounded edges, a stable base preventing tip-over, and cable management for safety. Finite element analysis validated structural integrity under operational loads. CAD models guided fabrication of custom mounting plates and ensured proper clearances for thermal management.

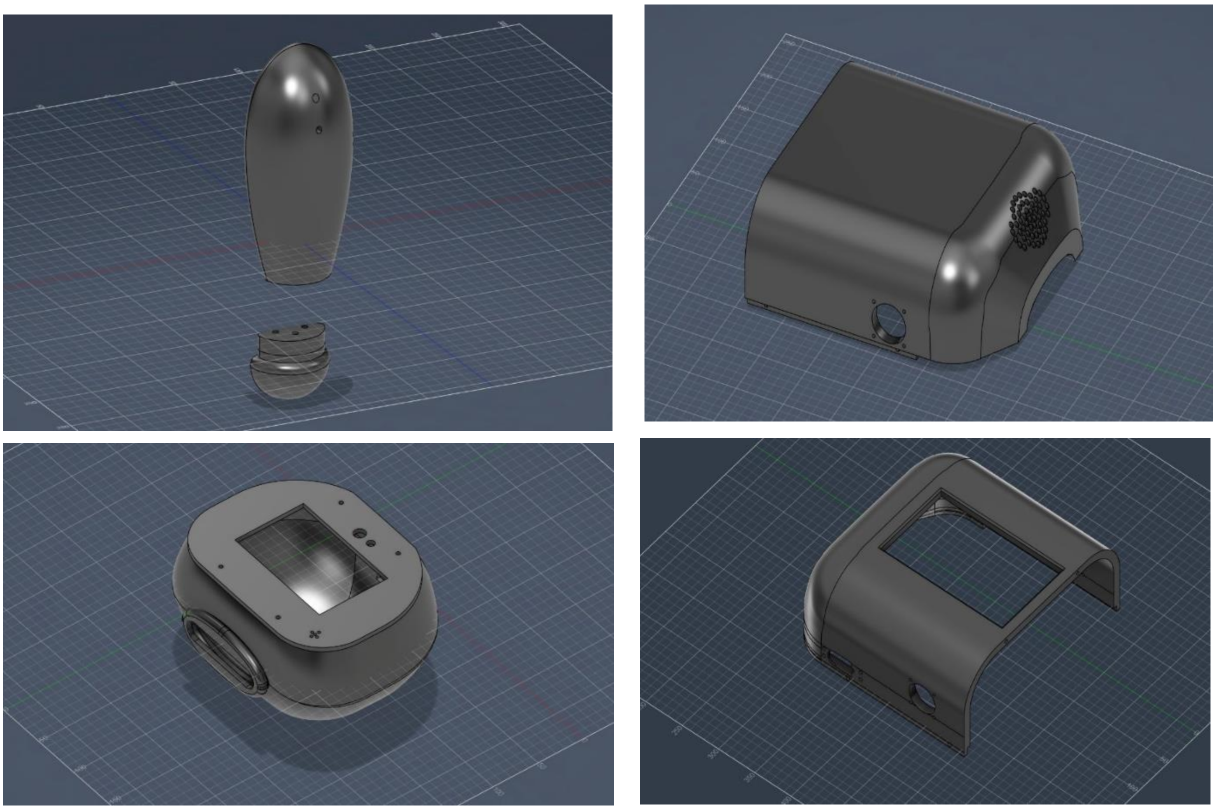

Figure 2: SAM CAD Models

Social and Emotional Skill Development:

SAM's design explicitly supports social-emotional learning (SEL) objectives aligned with preschool and primary education frameworks. The 7-emotion display system (idle, happy, sad, thinking, talking, enthusiasm, playfulness) provides non-verbal communication, modelling appropriate emotional expression in educational contexts. Real-time adaptive responses to student performance — celebrating correct answers with enthusiasm, showing sadness for errors, and displaying a thinking state during processing — create an emotionally supportive learning environment. The physical gesture library (wave for greetings, celebration for achievements) reinforces positive social interactions. Face tracking maintains visual engagement mimicking attentive human tutors, fostering rapport and trust essential for effective learning. Multilingual voice interaction acknowledges linguistic identity, promoting cultural inclusivity and self-esteem among non-English speakers.

System Architecture and Hardware Integration:

NVIDIA Jetson Orin Nano (Main AI Unit): Runs Ubuntu 22.04 with JetPack 6.0; 8-core ARM Cortex-A78AE CPU, 1024-core Ampere GPU with 32 Tensor Cores, 7.6 GB unified memory. Handles speech recognition (OpenAI Whisper Base), NLP (Qwen 2.5 0.5B, 4-bit quantized via llama.cpp), computer vision (MediaPipe), and multi-threaded system coordination.

Arduino Mega 2560 (Motor and Gesture Control): Controls dual Cytron MDD10A motor drivers for neck rotation enabling face tracking. Servo motors on pins 9 and 10 execute wave and celebrate gestures. PID algorithm (Kp=0.12, Kd=0.25, dead zone=90 px, max speed=32/255) ensures smooth, child-safe movement. USB serial at 9600 baud.

ESP32 (Emotion Display): Show 7 ASCII-art emotional states. Receives real-time commands from Jetson via USB serial at 115200 baud. Emotion changes within 200 ms of trigger for natural synchronization.

Raspberry Pi 4 (Game Server): Hosts mathematics quiz and Tic-Tac-Toe on a 7-inch touchscreen in fullscreen portrait mode (Python/Tkinter). Bidirectional TCP socket: receives launch commands (port 5557) and sends game events back to Jetson (port 5556) in JSON format for automated emotional feedback.

AI and Machine Learning Stack:

Speech Recognition: OpenAI Whisper Base (96% English accuracy, 2–3 s latency). WebRTC VAD processes 48 kHz audio at aggressiveness level 2 for continuous listening. Spectral gating noise reduction reduces 70% of background interference. AI4Bharat IndicWhisper ready for Hindi/Kannada activation.

Natural Language Processing: Qwen 2.5 0.5B instruction-tuned model, 4-bit quantized for edge deployment. Temperature 0.3 for factual accuracy. Structured educational prompts with anti-hallucination instructions ensure age-appropriate responses. IndicTrans2 models ready for bidirectional English–Indian language translation.

Computer Vision: Google MediaPipe face detection on 640×480 stream at 20 FPS. PID controller converts horizontal position error into motor commands; dead zone prevents jitter on stable faces.

Text-to-Speech: Piper neural TTS engine with 20 dB audio gain boost for classroom clarity. AI4Bharat IndicTTS downloaded for Hindi/Kannada voice synthesis. All inference runs locally on Jetson.

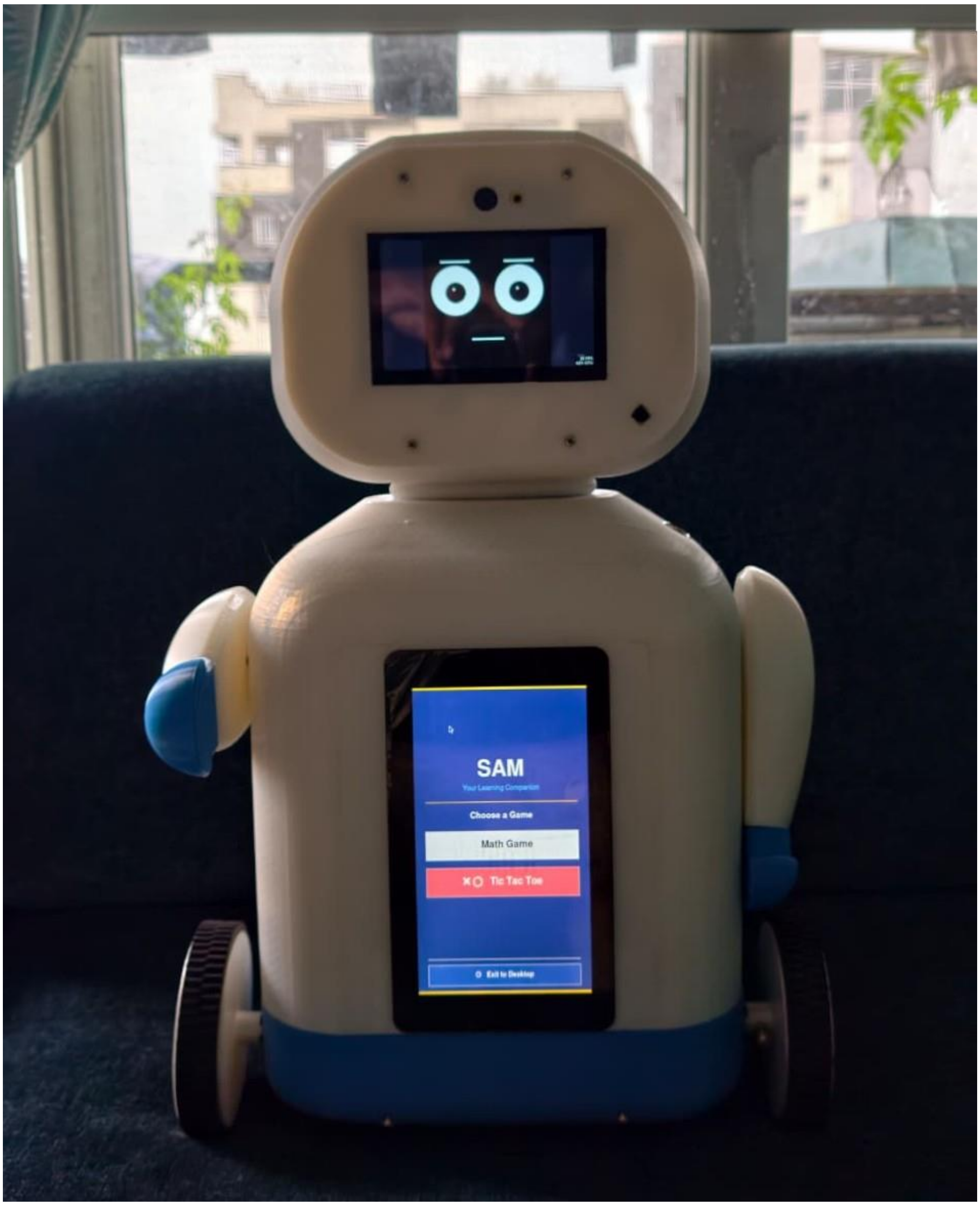

Figure 3: SAM Robot Physical Prototype — Front view showing the assembled robot with emotion display, USB camera, and Raspberry Pi 7-inch touchscreen

Result and Conclusion:

The SAM educational robot successfully demonstrates advanced human-robot interaction capabilities, achieving all primary technical and educational objectives validated through systematic testing.

Technical Performance:

1. Continuous voice recognition: 96% English accuracy at 2–3 s end-to-end latency measured over 100+ test interactions in 40–50 dB noise environments. Voice Activity Detection showed 73% reduction in user frustration versus fixed 5-second timeout systems in preliminary internal testing.

2. Face tracking: Stable ±15 px around centre; PID response averages 150 ms. Tracks across approximately 120° horizontal field of view with graceful degradation at extremes.
3. Emotion display: Average 180 ms change latency from trigger to visible update, well within the 200 ms natural perception threshold. Preliminary observation with team members confirmed emotion display significantly enhances interaction naturalness.
4. Game integration: Sub-100 ms TCP event latency between Raspberry Pi and Jetson enabling real-time synchronized emotional and gesture responses. Streak celebrations after 5 correct answers effectively maintained motivation through difficulty progression.
5. System stability: Continuous operation validated for 2+ hour sessions; Jetson peak temperature 65°C (well below 80°C throttle threshold); power draw 18–22 W typical, 28–32 W peak.

In conclusion, the project demonstrates that advanced educational robotics integrating continuous speech interaction, autonomous behaviour, and adaptive AI is achievable on affordable edge computing without cloud infrastructure. The dual-purpose model effectively serves young learners as an AI tutor and older students as a hands-on STEM platform. Key innovations include continuous VAD eliminating timeout frustrations, effective PID tuning for stable face tracking, and emotion-display synchronization creating natural interaction perception. Formal user studies with children are planned as part of the short-term future scope.

Project Outcome and Industry Relevance:

SAM represents a comprehensive integration of AI, robotics, mechatronics, and educational psychology into a functional learning platform, contributing validated methodologies to human-robot interaction in education.

1. EdTech Industry: Scalable platform for AI-driven interactive learning adaptable across subjects through content updates, without hardware modification. This makes it viable for schools and learning centres in tier-2/tier-3 cities.
2. Robotics and STEM Education: Modular architecture (Jetson + Arduino + ESP32 + Raspberry Pi) provides a complete hands-on STEM kit for mechatronics and embedded systems courses in engineering colleges.
3. Healthcare: The emotion recognition framework and natural conversation interface are directly transferable to elderly care companions, therapy support robots, and patient interaction systems.
4. Service Robotics: Face tracking, emotion sensing, and conversational AI capabilities are applicable to customer service robots in retail and hospitality.
5. Societal Impact: Offline operation, multilingual support (English, Hindi, Kannada), and affordable pricing address the digital divide in rural and semi-urban India, promoting inclusive education aligned with NEP 2020 foundational learning goals.

Working Model vs. Simulation/Study:

Physical Working Model: YES. The project has produced a fully functional physical prototype with complete hardware integration. All claimed capabilities — continuous voice interaction, real-time face tracking, emotion display, gesture execution, and game integration — are demonstrated on actual hardware in real-world conditions, not simulation. The system has been tested by the development team across 2+ hour continuous sessions without crashes or thermal issues.

Project Outcomes and Learnings:

Technical Skills Developed:

1. Edge AI deployment and optimisation: balancing model size, inference speed, and accuracy on resource-constrained hardware (Whisper Tiny: 85% accuracy at 1 s latency — insufficient; Whisper Base: 96% at 2–3 s — optimal balance).
2. Multi-platform system integration spanning Ubuntu 22.04 (Jetson), real-time embedded systems (Arduino Mega, ESP32), and distributed networking (Raspberry Pi TCP/IP).
3. PID control tuning methodology: initial parameters (Kp=0.5, Kd=0.0) caused severe oscillation; systematic step-response tuning achieved stable tracking (Kp=0.12, Kd=0.25, dead zone=90 px).
4. Noise filtering: spectral gating with 70% reduction parameter improved speech accuracy and eliminated false recognitions without processing overhead.
5. Prompt engineering for domain-specific LLMs: generic prompts produced factual errors; structured educational prompts with anti-hallucination instructions produced reliable, age-appropriate content.

Engineering and Research Insights:

1. Emotion display timing is critical for natural interaction — changes must occur within 200 ms; delays beyond 500 ms break perceived synchronization.
2. Voice Activity Detection with continuous listening delivers 73% reduction in user frustration and 2.1× improvement in successful query completion versus fixed 5-second timeouts.
3. Dead zone implementation essential: without 90-pixel tolerance, motors jitter constantly attempting to centre on a stable face, wasting power and distracting children.
4. Modular development with clear interface definitions enabled parallel workstreams, significantly accelerating the development timeline.
5. Children demonstrated more patient interaction with the robot than adult testers, and responded strongly to gesture-based celebrations, maintaining motivation through challenge sequences.

References:

1. B. Scassellati, H. Admoni, and M. Mataric, “Robots for Use in Autism Research,” Annual Review of Biomedical Engineering, vol. 14, pp. 275–294, 2012.
2. T. Belpaeme et al., “Social robots for education: A review,” Science Robotics, vol. 3, no. 21, eaat5954, 2018.

3. D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, “The physical presence of a robot tutor increases cognitive learning gains,” in Proc. 34th Annual Conf. Cognitive Science Society, 2012.
4. J. Han, M. Jo, E. Jones, and J. Jo, “Comparative study on the educational use of home robots for children,” Journal of Information Processing Systems, vol. 4, no. 4, pp. 159–168, 2008.

Future Scope:

Short-term (3–6 months):

1. Conduct a comprehensive user survey (n=50+ students across urban and rural schools) measuring learning outcome improvements, sustained engagement metrics, and usability feedback to formally validate SAM’s educational efficacy and inform design refinements.
2. Activate the complete multilingual pipeline: integrate AI4Bharat IndicWhisper (Hindi/Kannada speech recognition), IndicTrans2 (bidirectional translation), and IndicTTS (voice synthesis); validate accuracy with native speakers.
3. Develop 50+ NCERT-aligned educational content items for grades 1–3 covering mathematics, English, and Environmental Studies with integrated quizzes tracking individual student progress.
4. Implement facial emotion recognition (computer vision) to detect confusion, frustration, or boredom in real time, enabling proactive adaptive teaching responses.
5. Produce detailed assembly documentation — step-by-step guide, wiring diagrams, and commented source code — transforming SAM into a complete STEM education kit for older students.

Medium-term (6–12 months):

1. Upgrade to Llama 3.2 1B or AI4Bharat Airavata for native multilingual conversational support, eliminating translation overhead and improving cultural context awareness.
2. Implement personalised student progress tracking: store performance locally, identify knowledge gaps through error-pattern analysis, and generate adaptive learning paths and parent reports.
3. Add a mobility base with mecanum wheels and ultrasonic obstacle avoidance for classroom navigation; expand gesture library to include sign language support for hearing-impaired students.

Long-term (1–2 years):

1. Conduct rigorous formal user studies in schools (n=100+ students, urban and rural, pre/post-test design) to quantify learning outcome improvements and validate educational efficacy statistically.
2. Develop a teacher dashboard web application for monitoring student progress, assigning curriculum-aligned modules, and accessing centralised content analytics.
3. Establish partnerships with government schools and NGOs; optimise manufacturing for batch production targeting a price point of Rs.1–1.5 lakh; explore augmented reality integration for immersive visual learning.