STATUS UPDATE - 10/01/2026

CAREMATE : AN AGENTIC AI – POWERED HOSPITAL ASSISTANT FOR PATIENT- CENTERED CARE

1. SPECIFIC STATUS

As of January 10th, 2026, the CareMate team has successfully completed the software part of Phase 1. This phase is concerned with patient interaction using speech.

- The speech input-output pipeline (STT to Translation to TTS) has also been entirely implemented and tested using the speech APIs of Sarvam AI.
- The complete pipeline for recording speech from patients, processing the speech into structured text, and then synthesizing speech responses has been proven to be functional.
- On the hardware side, the three main elements received on January 13, 2026, were the ESP32 development board, the I2S microphone module, and the I2S audio amplifier module.
- The remaining hardware components such as speakers, jumper wires, and other enclosure-related components will be purchased locally.

The project has therefore moved from the planning and experimentation phase to the development stage.

2. ADHERENCE TO PROJECT SCHEDULE

The work is going as planned, without any deviation from the submitted action plan.

- The phase 1 software development process regarding the speech input/output system was accomplished within the planned timeline.
- The acquisition of hardware encountered a slight delay in delivery; however, this did not affect the progress, since the software and hardware phases were designed to run in parallel.

- The shift from AI4Bharath's open-source models for speech to Sarvam's AI APIs was an optimization in the technical process, which in no way resulted in a delay in the schedule.
- This shift enhanced the stability of the system, the quality of the output, and facilitated quicker completion of Phase 1.

In conclusion, the whole project still fully meets the original action plan and timeline, while the next phases follow as scheduled.

3. TECHNICAL AND LOGISTICS CHALLENGE

Technical challenges encountered

- In the initial implementation phase, the project used open-source AI4Bharath models for the speech processing chain, which included the following models:
    - Speech-to-Text (STT): ai4bharat/indic-conformer-600m-multilingual
    - Translation: ai4bharat/indictrans2-indic-en-1B
    - Text-to-Speech (TTS): Initial experiments on the IndicF5 system, followed by IndicF5 Parlor.
- The STT and translation components functioned correctly and reliably, including support for conversational and code-mixed patient speech.
- However, challenges were identified primarily in the speech synthesis stage:
    - The speech synthesized was not clear and accurate, which is important when it comes to patient communication in a hospital setting.
    - For better audio clarity, the TTS module was later changed from the **'AI4Bharath models' to 'Svara TTS,'** which resulted in better outcomes.
    - Though this was an improvement, it was a resource-intensive process to have the whole speech processing pipeline run locally, which required constantly running servers.

- Running local models on a laptop/computer environment or a shared environment (such as Google Colab) led to an increase in the latency of the responses, thereby impacting the real-time conversation experience.
- It was therefore not cost effective to have local servers that are always on for the current project size.

Resolution and mitigation:

- For overcoming the above-mentioned challenges, the project used Sarvam AI speech APIs in the speech integration pipeline.
- The Sarvam APIs offer highly accurate speech recognition, translation, and speech synthesis capabilities with clear audio output.
- The API-based solution eliminates the need to have the local server always running, thus making the infrastructure simpler.
- The Sarvam APIs also provide low-latency responses, ensuring smooth real-time functionality for patient-oriented applications.
- In the case of the small-scale deployment that is currently underway, this strategy is cost-effective, stable, and reliable.
- On the other hand, the architecture of the system is made to be modular and scalable in the sense that:
  - The local AI4Bharath STT, translation, and TTS systems already tested are fully functional.
  - However, once the system is scaled to larger environments with their own infrastructure, self-hosted local models will become the best option with regards to long-term costs.

4. OBJECTIVES AND MILESTONES FOR THE UPCOMING MONTH (FEBRUARY 2026)

In February 2026, work on hardware integration and preparation of domain models will take place as indicated in the action plan, and agent-based system design will occur in Month 3.

Primary objectives:

- Hardware Integration:
    - ESP32 microphone input capture and testing
    - Playing audio through amplifier and speakers
    - End-to-end voice interaction testing between bedside device and backend system
- Start Phase 2: Development of Domain AI Models
    - Obtaining medical and patient interaction datasets
    - Handling and organization of filtered and generated patient inquiries
    - Early experiments & validation of the domain-specific language model for patient question answering
- Backend system readiness:
    - Developing and testing backend APIs related to patient inquiries and responses
    - Integration of the speech processing system with the domain model of AI

5. SUPPORT OR CLARIFICATION REQUIRED FROM FAER

At this point, the team does not need any technical help. But some guidance would be welcome in the following respects:

- Best Practices for Discussing Technical Trade-Offs (such as APIs vs. Open Source Tools) in Interim Evaluations.
- Clarification on the Expectations of Evaluation for Healthcare AI Prototypes in Simulated Settings.
- Guidelines on documentation level and organization for mid-project and final submissions.