Project title: Adaptive Swedish Language Tutor with Affective Feedback

Babitha Mannaravalappil Kuvvakkattayil Pooja Ravindra Nagalikar Revathy Unnikrishna pillai

Course: 1MD032, HT2025

I. INTRODUCTION

We will build a Furhat-based Swedish language tutor for beginner/intermediate learners (A1–A2). The system detects user affect from webcam video (confused/frustrated vs. neutral vs. engaged/happy) and adapts dialogue difficulty, prosody, and nonverbal behaviours to keep learners on track.

II. ELEVATOR PITCH

An affect-aware Swedish tutor: Furhat listens to your Swedish, senses confusion or engagement from your face, and adjusts difficulty, speed, and hints in real time.

III. OBJECTIVES

Overall objective: deliver a coherent, adaptive tutoring demo with measurable affect detection and interaction quality.

1) Objective 1: Real-time affect detection (webcam → face detection → DiffusionFER fine-tuned emotion head) mapped to engagement buckets; tools: PyTorch, ONNXRuntime, MediaPipe/RetinaFace.
2) Objective 2: Adaptive dialogue manager with Swedish ASR/TTS and Furhat behaviours; tools: Whisper small (sv), rule-based state machine, Furhat SDK, SSML/voice controls.
3) Objective 3: Evaluation and robustness: F1/accuracy + latency for perception; scripted sessions rated for adaptation correctness; ablations (smoothing vs. none, self-data vs. none).

IV. DELIVERABLES

We will deliver an end-to-end affect-adaptive tutor, with code, model weights, and demo.

1) Deliverable 1: Perception module (trained weights + ONNX; inference script).
2) Deliverable 2: Interaction subsystem (dialogue manager, ASR/TTS integration, behaviours, configs).
3) Deliverable 3: Integration demo pipeline (webcam loop, on-screen overlays, logs).
4) Presentation/Demo: Live or recorded demo showing confused vs. engaged adaptations.
5) Final Report: PDF covering design, data, training, evaluation, limitations, future work.

V. SUCCESS METRICS

- Metric 1: Perception F1/accuracy on held-out set; target ≥ 0.75 macro-F1 across buckets.
- Metric 2: Per-frame latency; target $<$ 150–200 ms on webcam loop (including detection + model).
- Metric 3: Interaction appropriateness; target $\geq 80\%$ turns judged “appropriate adaptation” in scripted sessions (N=5–10).
- Progress Tracking: weekly checkpoints; log training runs, evaluation tables, and integration tests; maintain Kanban for tasks/risks.

VI. POTENTIAL ISSUES

- Issue 1: Class imbalance for confusion/frustration; mitigation: self-recorded clips, oversampling.
- Issue 2: ASR errors/noisy environment; mitigation: push-to-talk or VAD gating, repeat prompts, simpler phrasing.
- Issue 3: Latency spikes on CPU-only; mitigation: smaller models, downsampled crops, process every other frame.
- Time-Intensive Tasks: Fine-tuning and evaluation of perception model; integration and testing of ASR/TTS + behaviours.

VII. PROJECT BREAKDOWN

Deadline	Task Description	Assigned To	Notes/Dependencies
Nov 26	Submit project specification	All	Uses this document
Dec 2	Perception baseline (detection + pretrained emotion head)	Babitha and Pooja	Needs DiffusionFER prep
Dec 9	Plenary Feedback Session	All	Slides + baseline metrics
Dec 12	Fine-tune emotion head; add smoothing; ONNX export	Revathy	Requires GPU time
Dec 18	Individual Feedback Session	All	Show updated metrics/demo stub
Dec 22	Dialogue manager + ASR/TTS integration	Pooja	Needs perception output schema
Jan 5	Behaviours + adaptation policies wired; logging	Babitha	Depends on dialogue manager
Jan 10	Full integration + scripted session runs	All	Stable pipeline required
Jan 14	Project Presentation	All	Demo videos + slides
Jan 16	Report Submission	All	Integrate results/figures

Gantt Chart (edit as needed):

Task	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11
Spec & plan	X										
Perception baseline	X	X									
Fine-tune & smoothing		X	X								
Dialogue + ASR/TTS			X	X							
Behaviours & policy				X	X						
Integration & tests					X	X					
Demo prep						X	X				
Presentation								X			
Report									X		

VIII. AIMED GRADE

Aim: Grade 5. Justification: end-to-end functioning system with clear affect-driven adaptation, measurable perception performance and latency, interaction evaluation, robustness considerations (fallbacks, noise), polished demo with overlays, and concise report .