HearMeOut: A High-Fidelity Prototype for Speech-Based Mental Health Monitoring

Anukriti Bhargava
IIT Gandhinagar
M.Sc. Cognitive Sciences
24510026@iitgn.ac.in

Bhavik Patel
IIT Gandhinagar
B.Tech CSE 2022
22110047@iitgn.ac.in

Hitesh Kumar
IIT Gandhinagar
B.Tech CSE 2022
22110098@iitgn.ac.in

Pranav Patil
IIT Gandhinagar
B.Tech CSE 2022
22110199@iitgn.ac.in

Abstract

Mental health challenges among students continue to rise, yet traditional screening methods remain limited by accessibility, stigma, and scalability concerns. In this report, we present HearMeOut, a high-fidelity web-based prototype that leverages speech emotion recognition to enable non-intrusive, daily mental health monitoring. Building upon our low-fidelity design of a Daily Mood Journal interface, we developed a fully functional system using React, Express, and a Flask-based ML service with Hugging Face's wav2vec2 emotion recognition model. The prototype implements a calendar-based interaction paradigm that de-stigmatizes mental health tracking by framing it as casual mood journaling rather than clinical assessment. Through usability testing with 5 university students, we identified key insights around trust in AI emotion detection, privacy concerns with voice data, and the importance of user agency in mood interpretation. Our evaluation revealed both strengths (intuitive calendar interface, gentle interaction design) and areas for improvement (transparency in AI decision-making, multilingual support, intervention mechanisms). This work demonstrates how human-centered design principles can bridge the gap between sophisticated emotion recognition technology and user acceptance in mental health applications.

Keywords

Speech-based emotion recognition, Mental health monitoring, User-centered design, High-fidelity prototyping, Usability evaluation

1 Introduction

Depression and anxiety disorders affect a wide range of students worldwide, yet only a fraction seek professional help due to stigma, limited awareness of symptom progression, and overburdened campus mental health services [1]. Traditional screening methods- clinical interviews and self-report questionnaires- often fail to detect gradual emotional decline in students who may not recognize their own deteriorating mental state.

Our previous systematic literature review identified speech-based emotion detection as a promising, non-invasive approach for continuous mental health monitoring [1]. Acoustic features such as pitch variability, speech rate, and voice quality have been shown to correlate with depressive symptoms, enabling machine learning models to achieve 76-89% accuracy in detecting emotional states from speech [1]. However, a critical gap exists between technological capability and real-world adoption: users express deep concerns about privacy, AI transparency, and the clinical framing of mental health tools.

In Task 2, we designed two distinct low-fidelity prototypes to address different user needs: (1) a Privacy-First Guided Journal for data-conscious users seeking structured assessments, and (2) a Daily Mood Journal using calendar-based emoji interfaces for stigma-avoidant users. User research with 10 university students revealed that Prototype 2's non-clinical approach resonated more strongly with the target population, particularly students who avoid traditional "mental health apps" due to fear of labeling or judgment.

This paper presents **HearMeOut**, a high-fidelity implementation of the Daily Mood Journal prototype. Our system enables students to record 30-60 second daily voice reflections, which are analyzed by a transformer-based emotion recognition model. Rather than imposing diagnostic labels, the system suggests mood emoji based on detected emotions, allowing users to override AI predictions and maintain agency over their emotional self-interpretation. The calendar-based interface visualizes mood patterns over time, enabling students to recognize gradual changes that might otherwise go unnoticed.

1.1 Objectives

This work addresses three key objectives:

1. **Technical Implementation**: Develop a production-quality web application integrating speech emotion recognition with a privacy-conscious, scalable architecture.
2. **Design Translation**: Translate low-fidelity wireframes into a polished, functional interface that embodies the "gentle journaling" conceptual model.
3. **Usability Validation**: Evaluate the prototype with real users to assess learnability, trust in AI suggestions, and willingness to engage with speech-based mood tracking.

The remainder of this report is organized as follows: Section 2 describes our system architecture and technical implementation. Section 3 discusses design compromises made during development. Section 4 presents our conceptual model and design rationale. Section 5 details the prototype's features with visual documentation. Section 6 reports findings from usability testing with 5 participants.

Section 7 addresses ethical considerations and risk mitigation. Finally, Section 8 concludes with lessons learned and future directions.

2 System Design & Architecture

HearMeOut implements a modern, microservices-based architecture designed for scalability, maintainability, and separation of concerns. The system consists of three primary tiers: a React-based frontend, an Express backend orchestration layer, and a Flask service for ML inference.

2.1 Architectural Overview

Figure 1 illustrates the complete system architecture. The design follows the microservices pattern, where each service has a single, well-defined responsibility:

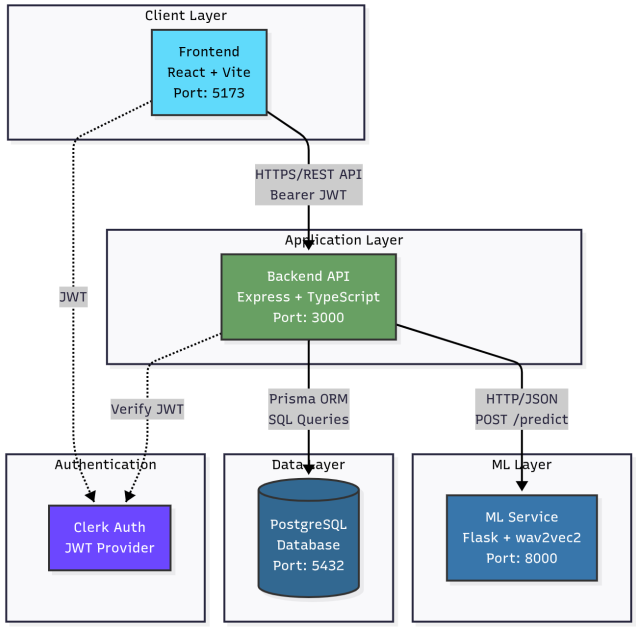


Figure 1: HearMeOut microservices architecture showing containerized services and data flow

Frontend Layer (React + TypeScript): The client application is built with React 18 and TypeScript, compiled using Vite for fast Hot Module Replacement (HMR) during development. Key technologies include:

- **UI Framework**: Tailwind CSS for utility-first styling with shadcn/ui components for accessible, customizable interface elements.
- **Routing**: React Router v6 for client-side navigation between calendar view, recording interface, and settings.
- **Authentication**: Clerk React SDK for JWT-based session management with automatic token refresh.
- **Audio Processing**: Web Audio API and MediaRecorder API for capturing voice recordings in the browser.
- **Speech-to-Text**: Web Speech API for real-time transcription display during recording (enhances user confidence that the system is actively listening).
- **Data Visualization**: Recharts library for rendering mood distribution charts.
- **State Management**: Custom React hooks pattern with Context API for sharing user settings and mood data across components.

Backend Layer (Express + TypeScript): The Node.js backend acts as an orchestration layer, coordinating between the frontend, ML service, and database. Core responsibilities include:

- **Authentication Middleware**: Clerk Express SDK validates JWT tokens on all protected routes, ensuring user isolation.
- **Business Logic**: Implements rules such as one mood entry per user per day (enforced via database unique constraint), pattern detection for intervention triggers.
- **API Design**: RESTful endpoints with proper HTTP status codes, error handling, and request validation using Zod schemas.
- **File Upload Handling**: Multer middleware processes multipart/ form-data audio uploads, temporarily saving files to `/temp_audio/` directory.
- **ML Service Integration**: HTTP client calls to Flask service with retry logic and health checks.

ML Service Layer (Flask + PyTorch): A dedicated Python service handles emotion recognition inference:

- **Model**: Hugging Face's `ehcalabres/wav2vec2-lg-xlsr-en-speech-emotion-recognition`, a transformer-based model fine-tuned on emotional speech datasets (RAVDESS, etc).
- **Emotion Classes**: 8 emotions (angry, calm, disgust, fearful, happy, neutral, sad, surprised) with softmax confidence scores.
- **Audio Preprocessing**: Librosa library resamples audio to 16kHz mono (required by wav2vec2), extracts features, and normalizes waveforms.
- **Inference**: PyTorch model runs on CPU (GPU optional), applies mean pooling over hidden states, and returns top 3 predictions ranked by confidence.
- **Deployment**: Gunicorn WSGI server with worker process management for concurrent request handling.

Data Layer (PostgreSQL + Prisma): The database schema (Figure 2) uses PostgreSQL 15 with Prisma ORM for type-safe queries. Key design decisions:

- **User Isolation**: All queries filtered by `clerkId` to prevent cross-user data access.
- **Daily Uniqueness**: Composite unique constraint on `(userId, entryDate)` prevents duplicate entries for the same day.
- **Activity Tracking**: Many-to-many relationship between `MoodEntry` and `Activity` via `MoodEntryActivity` junction table.

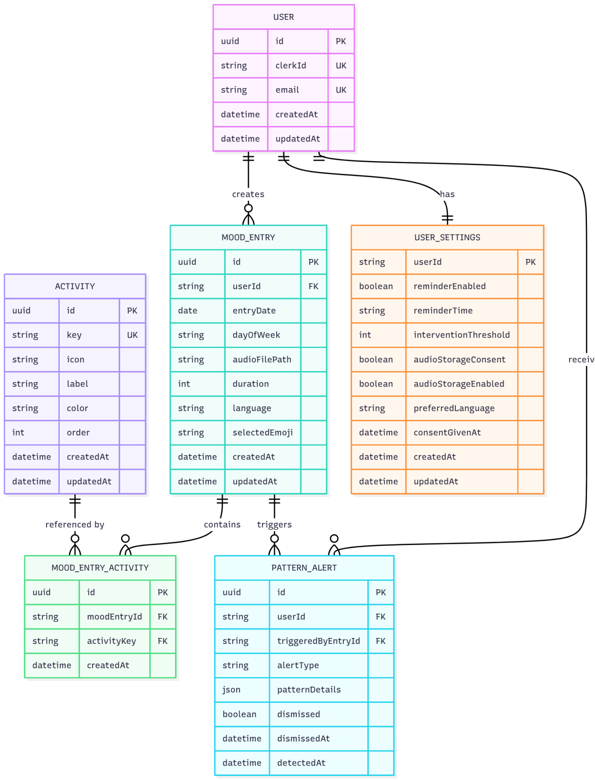


Figure 2: Database schema with privacy-aware design

- **Privacy Metadata**: `audioMetadata` JSONB field stores recording duration and language, but *not* the audio file path (files are deleted immediately after ML processing unless user consents to storage).
- **Cascading Deletes**: User deletion automatically removes all related mood entries, settings, and activity associations.

2.2 Containerization & Deployment

Given the prototype's iterative development nature with frequent UI and API changes, we adopted a hybrid deployment strategy that balances consistency with development agility:

Containerized Services (Docker Compose):

- **PostgreSQL Container**: Alpine-based image (postgres:15-alpine) with named volume for data persistence across restarts. Exposes port 5432 to localhost for backend connections.
- **ML Service Container**: Python 3.10 with PyTorch, Transformers, and librosa pre-installed. Exposes port 5001 for HTTP API access.

Local Development Services:

- **Frontend (React + Vite)**: Runs locally on developer machines via `npm run dev` with hot module replacement for instant UI updates during iteration.
- **Backend (Express + TypeScript)**: Runs locally via `tsx watch src/index.ts` for automatic restart on code changes. Connects to containerized PostgreSQL and ML service.

2.3 APIs' Workflows

Below diagram shows the flow of some of the main APIs integrated in the application:

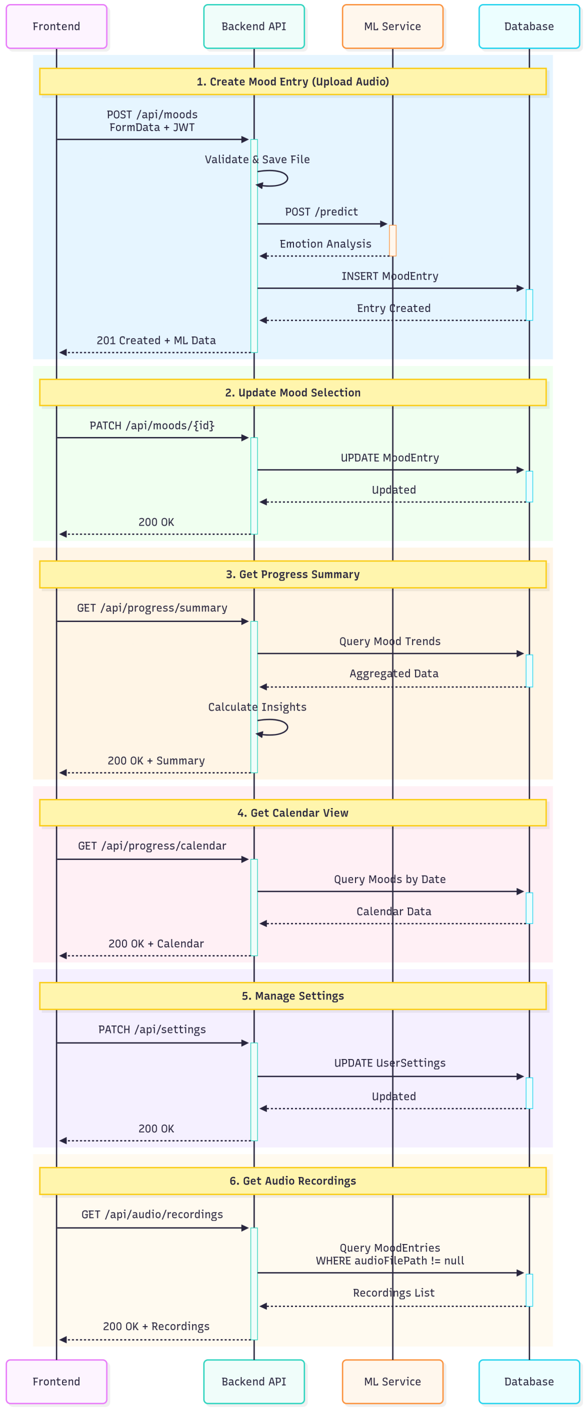


Figure 3: Complete Data Flow

3 Compromises in Prototyping

Transitioning from low-fidelity wireframes to a production-quality prototype required strategic decisions about feature scope, technical implementation, and resource allocation. We adopted a **horizontal prototype** approach with selective vertical depth in core user flows.

3.1 Horizontal Prototype Strategy

We implemented broad coverage of the user interface to provide a complete interaction experience:

Fully Implemented Screens:

- **Home Screen**: Month/week calendar grid with emoji stickers for each recorded day (Figure 4a).
- **Recording Interface**: Voice recorder with real-time waveform visualization, 30-60 second timer, live speech-to-text transcription, and language selector (Figure 4b).
- **Emoji Selection Modal**: AI-suggested mood stickers with override capability and custom emoji picker (Figure 4c).
- **Activity Tagging Screen**: Tag-based activity selector with emoji icons for contextual mood tracking (Figure 4d).
- **Mood Analytics Dashboard**: Donut chart showing mood distribution over last 30 days with detailed breakdown (Figure 4e).
- **Settings Page**: Audio consent toggle, reminder preferences, and account management (Figure 4f).

This horizontal coverage allows users to navigate the entire application and experience all planned touchpoints from the low-fidelity design.

3.2 Vertical Depth in Core Flow

The central user flow (Record → Analyze → Display) was implemented with full technical depth:

- **Real ML Integration**: Unlike typical prototypes that mock AI responses, we integrated an actual transformer-based emotion recognition model (wav2vec2). This decision was made to validate whether current ML technology meets the accuracy and latency requirements identified in our literature review.
- **Database Persistence**: All mood entries, activities, and user settings persist in PostgreSQL with proper relationships, indexes, and constraints.
- **Authentication System**: Full JWT-based authentication via Clerk ensures realistic multi-user scenarios and user isolation.
- **Real-Time Audio Processing**: Web Audio API integration provides genuine audio capture.

3.3 Simplifications & Omissions

Several features from the low-fidelity design were simplified or deferred through horizontal compromises (broad UI coverage with simplified backend logic) or complete omissions:

3.3.1 Pattern Detection & Interventions (Horizontal Compromise). **Low-Fi Plan**: System monitors for gradual mood decline (5+ consecutive low-mood days) and displays gentle intervention popup asking "Would you like help or resources?" with options for self-care tips, professional support contact, or dismissal.

Hi-Fi Implementation: Backend tracks consecutive low-mood days and flags an alert on the UI, but the intervention mechanism was simplified:

- Detection logic is present (database query for consecutive days).
- UI popup for intervention was implemented but displays generic placeholder text rather than contextual self-care resources.

Rationale: Designing appropriate, non-judgmental intervention content requires collaboration with mental health professionals. For prototype evaluation purposes, demonstrating the detection mechanism (horizontal coverage) was more critical than curating resource content (vertical depth).

3.3.2 Multilingual Support (Horizontal UI, Vertical Depth Deferred). **Low-Fi Plan**: Users can record in Hindi, Gujarati, or English, with language selection affecting UI prompts and emotion model selection.

Hi-Fi Implementation:

- Language selector UI component is present (horizontal).
- Backend stores selected language in `audioMetadata`.
- However, ML service only processes English audio (wav2vec2-en model)—no vertical depth for multilingual processing.
- UI prompts remain in English only.

Rationale: Multilingual emotion recognition requires language-specific models (e.g., IndicWav2Vec for Indian languages) and significantly increases model size/inference time. Given that our user testing was conducted with English-speaking students, we prioritized English-only implementation for prototype validation while preserving the language selection architecture for future expansion.

3.3.3 Weekly Summary Feature (Complete Omission). **Low-Fi Plan**: Insights delivered once per week (e.g., "This week you seemed more energetic on days with exercise" or "You had 4 calm days this week").

Hi-Fi Implementation:

- Mood analytics page shows 30-day distribution donut chart (horizontal visualization).
- No automated weekly summary text generation (omitted).
- No correlation analysis between activities and moods (omitted).
- Infrastructure exists: activity tagging and day-of-week tracking prepare for future vertical implementation.

Rationale: Generating meaningful, non-generic insights requires longitudinal data collection (minimum 2-4 weeks per user). Since our usability testing occurred immediately after prototype development, users had insufficient data for summary evaluation. We prioritized building the data collection infrastructure to enable future vertical development of this feature.

3.3.4 Privacy Controls (Vertical Compromise). **Low-Fi Plan**: Granular privacy settings including:

- Option to store audio locally vs. cloud processing
- Export all personal data (GDPR-style data portability)
- Selective deletion of specific mood entries
- Anonymization options

Hi-Fi Implementation:

- Binary audio consent button (allow storage vs. immediate deletion)—horizontal coverage without granular depth.
- Individual mood entry deletion via a delete button in history view.
- No data export functionality (omitted).

Rationale: Implementing local-only processing would require porting the ML model to TensorFlow.js for browser-side inference, introducing significant performance degradation on lower-end devices. We prioritized the privacy-first principle (immediate deletion) over granular storage options, while maintaining architectural separation to enable future client-side ML deployment.

3.4 Horizontal vs. Vertical Trade-offs

Table 1 summarizes our prototype coverage:

Table 1: Feature Implementation Coverage

Feature	UI	Backend
Calendar View	Full	Full
Voice Recording	Full	Full
Emotion Recognition	Full	Full
Emoji Selection	Full	Full
Activity Tagging	Full	Full
Mood Analytics	Full	Full
Pattern Detection	Partial	Partial
Interventions	No	No
Multilingual	UI Only	No(Eng only)
Weekly LLM Summary	No	No
Local Processing	No	No

This balanced approach enabled us to deliver a fully navigable, aesthetically polished application while concentrating engineering effort on validating the core hypothesis: *Can students engage with speech-based mood tracking when framed as non-clinical journaling?*

4 Conceptual Model

The conceptual model defines how users should mentally represent the system, its capabilities, and their role in the interaction. HearMeOut's design is built around three core conceptual pillars: **the journaling metaphor**, **the 6 W's design space**, and **the AI-as-assistant mental model**.

4.1 Primary Metaphor: Daily Mood Journaling

Metaphor Selection Rationale: Our low-fidelity user research revealed that students avoid apps explicitly labeled as "mental health screening" or "depression detection" due to:

1. **Stigma**: Using such apps feels like admitting "something is wrong with me."
2. **Fear of Diagnosis**: Students worry about receiving clinical labels or scores that might trigger anxiety.
3. **Privacy Concerns**: Mental health apps are perceived as more invasive than wellness or productivity tools.

To address these barriers, HearMeOut adopts the **daily journaling** metaphor—a familiar, low-stigma practice associated with self-reflection, mindfulness, and personal growth rather than medical diagnosis. This reframing has concrete design implications:

Language Choices:

- "Record today's mood" (not "Complete assessment")
- "Mood journal" (not "Depression screening tool")
- "Feeling patterns" (not "Symptom severity")
- "Suggested emoji" (not "AI diagnosis")

Visual Design:

- Calendar interface mimics physical planners and bullet journals.
- Emoji stickers evoke scrapbooking aesthetics rather than clinical charts.
- Purple-to-pink gradient color palette conveys calmness and creativity (vs. medical blue/white).
- Playful, rounded UI components (shadcn/ui default theme) create approachable feel.

Interaction Model:

- Daily check-ins are *optional* rituals, not mandatory assessments.
- No persistent reminders or guilt-inducing notifications.
- Users can skip days without penalty—gaps in the calendar are normalized, not flagged as "non-compliance."

4.2 Design Space: The 6 W's Framework

Building on our low-fidelity design, HearMeOut is structured around six fundamental questions:

4.2.1 WHO: Target Users.

Primary Persona: Stigma-Avoidant Arjun

- First-year undergraduate adjusting to hostel life.
- Experiences periodic sadness/withdrawal but dismisses it as "normal college stress."
- Avoids anything labeled "mental health app" due to fear of judgment.
- Prefers visual, non-clinical interfaces.

Secondary Persona: Privacy-Conscious Priya

- Postgraduate student experiencing thesis-related anxiety.
- Wants objective mood tracking but distrusts apps that store voice data permanently.
- Comfortable with technology but scrutinizes privacy policies.

4.2.2 WHAT: System Capabilities.

The system provides:

1. **Emotion Detection**: AI analysis of 30-60 second voice recordings to detect 8 emotional states (angry, calm, disgust, fearful, happy, neutral, sad, surprised).
2. **Pattern Visualization**: Calendar-based mood tracking over weeks/months to reveal trends.
3. **Contextual Insights**: Correlation between moods and activities (exercise, social time, work stress).
4. **Gentle Interventions**: Optional suggestions for self-care or support when patterns indicate persistent low mood.

4.2.3 WHEN: Interaction Timing.

Designed Rhythm: Daily evening check-in (5-7 PM suggested, not enforced).

Flexible Engagement:

- Users can record anytime during the day.
- Skipping days is acceptable—system never penalizes gaps.
- Weekly reflection encouraged via mood analytics page (view 7-day or 30-day summaries).

Longitudinal Value: The metaphor emphasizes that journaling becomes more valuable over time as patterns emerge—a single entry provides limited insight, but weeks of data reveal meaningful trends.

4.2.4 WHERE: Usage Context. **Private Spaces**: Designed for use in dorm rooms, private corners of libraries, or quiet outdoor spaces where students feel comfortable speaking freely.

Device Agnostic: Web-based application accessible on smartphones, tablets, or laptops (responsive design adapts to screen size).

No Public Recording: Unlike some wellness apps that encourage social sharing, HearMeOut is explicitly private—no sharing features, no social comparison, no public profiles.

4.2.5 WHY: User Motivations. The system addresses multiple overlapping motivations:

- **Self-Awareness**: "I want to understand why I feel this way."
- **Pattern Recognition**: "I didn't realize I've been feeling down for two weeks straight."
- **Validation**: "Maybe my feelings are more significant than I thought."
- **Early Intervention**: "If things get worse, I'll have data to share with a counselor."

Critically, the system positions itself as a *tool for self-discovery* rather than a diagnostic instrument, aligning with the journaling metaphor.

4.2.6 HOW: Interaction Mechanics. The user journey follows a simple, repeatable ritual:

1. **Open App** → See calendar with past mood stickers.
2. **Click "Go for Today"** → Enter recording interface.
3. **Select Language** → Choose preferred language (currently English only).
4. **Speak Freely** → 30-60 second reflection on the day (prompted with: "How did today feel?").
5. **Review AI Suggestions** → System suggests 2-3 emoji based on detected emotions.
6. **Choose Final Emoji** → User selects from suggestions or picks different emoji (agency preserved).
7. **Tag Activities (Optional)** → Select relevant activities (exercise, social, work, etc.).
8. **View Updated Calendar** → Today's emoji appears on calendar; user can compare with past days.

4.3 Mental Model: AI as Assistant, Not Authority

A critical conceptual decision is how users should perceive the AI's role. HearMeOut deliberately positions the emotion recognition system as an **assistant offering suggestions** rather than an **authority making diagnoses**.

Key Principles:

1. **User Has Final Say**: AI suggests emoji, but user always chooses. This preserves agency and acknowledges that users know their emotions better than algorithms.
2. **Explainability (Partial)**: Emotion labels are shown alongside suggestions (e.g., "AI detected: Happy (87% confident)") to demystify the process, though low-fi users requested even more transparency about *why* specific emotions were detected.
3. **No Diagnostic Language**: The system never states "You are depressed" or "You have symptoms of anxiety." Instead, it describes patterns: "You've selected for 5 days in a row. Would you like some suggestions?"

This mental model addresses the **AI distrust** theme from our low-fi interviews, where participants asked: "How does it know if I'm depressed versus just tired or sick?" By framing AI as fallible and user knowledge as authoritative, we reduce skepticism and increase engagement.

4.4 Conceptual Model Validation

During usability testing (Section 6), we explicitly probed whether users understood the system's conceptual model:

Question: "If the app suggests a sad emoji but you don't feel sad, what would you do?"

Expected Answer (Correct Mental Model): "I would choose a different emoji—the AI is just making a suggestion."

Concerning Answer (Incorrect Mental Model): "I would question whether I'm actually sad and didn't realize it" (implies AI authority over self-knowledge).

Results showed that 4/5 participants correctly understood the assistant model after completing the first recording task, validating that the UI successfully communicates user agency.

5 Prototype Description & Functionalities

This section provides detailed documentation of HearMeOut's implemented features, illustrated with screenshots and interaction descriptions.

5.1 Feature 1: Interactive Mood Calendar

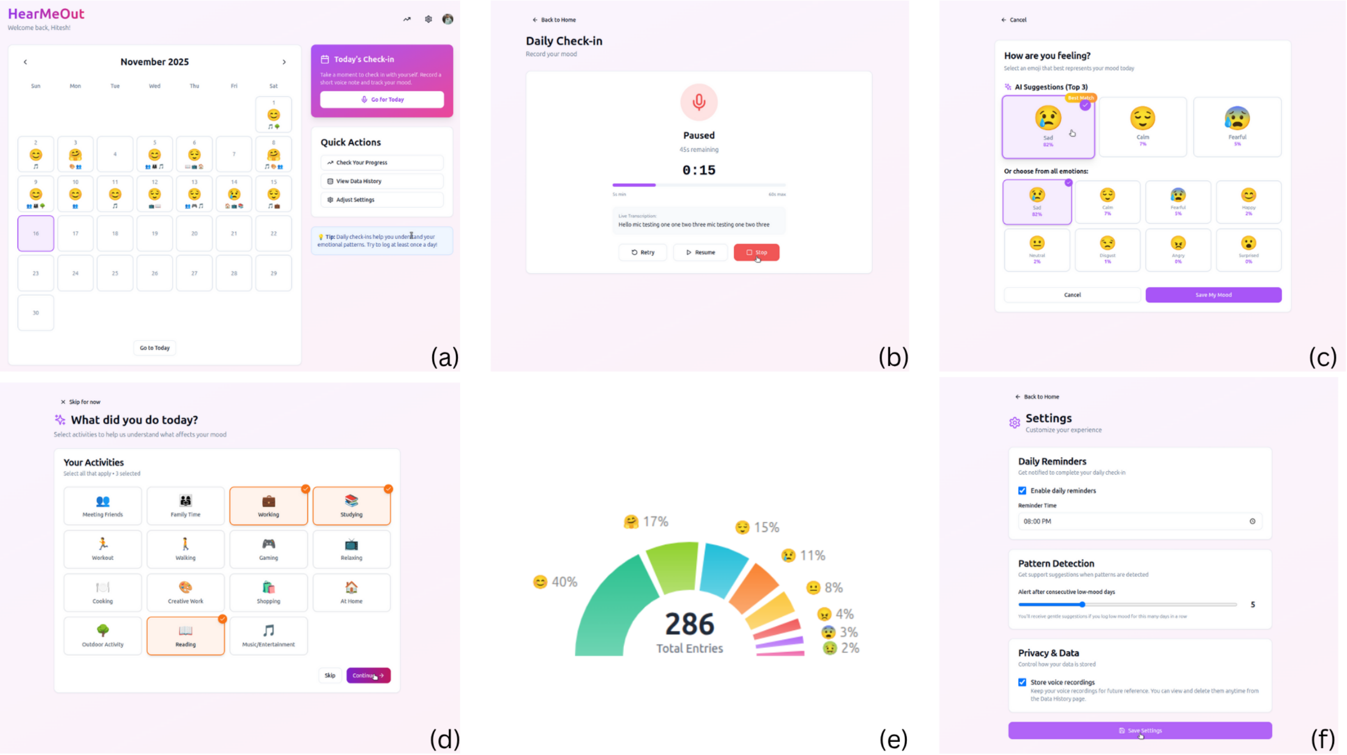
Functionality:

- Displays a full month grid with the option to switch between months (toggle via tabs).
- Each day shows a placed emoji sticker if the user recorded that day.
- Click on any past day to view details (selected emoji, activities).
- Hover effects and subtle shadows provide visual feedback.

Design Rationale: The calendar serves as the primary navigation hub, making mood patterns immediately visible. Gaps between entries are normalized (no red "missed days" warnings) to avoid inducing anxiety about inconsistent use—a key requirement from stigma-avoidant users.

5.2 Feature 2: Voice Recording with Real-Time Transcription

Functionality:



Welcome back, Hitesh!

Take a moment to check in with yourself. Record a short voice note and track your mood.

💡 Tip: Daily check-ins help you understand your emotional patterns. Try to log at least once a day!

Live Transcription:
Hello mic testing one one two three mic testing one two three

Select an emoji that best represents your mood today

Select activities to help us understand what affects your mood

Select all that apply • 3 selected

Get notified to complete your daily check-in

☑ Enable daily reminders

Reminder Time: 08:00 PM

Get support suggestions when patterns are detected

Alert after consecutive low-mood days: 5

You'll receive gentle suggestions if you log low mood for this many days in a row

Control how your data is stored

☑ Store voice recordings
Keep your voice recordings for future reference. You can view and delete them anytime from the Data History page.

Figure 4: HearMeOut user interface showcasing core features

- **Microphone Access**: Requests permission via `navigator.mediaDevices.getUserMedia()`.
- **Live Transcription**: Web Speech API displays recognized text as the user speaks (builds confidence that the system is "listening"). Integrated for English, Hindi, and Gujarati languages.
- **Timer**: 30-120 second countdown enforces minimum recording length (quality requirement for ML model) while preventing excessively long recordings.

Design Rationale: Low-fi testing revealed that users felt uncertain during recording without visual feedback. The transcription addresses this by making the system's active listening visible and interpretable.

5.3 Feature 3: AI-Assisted Emoji Selection

Functionality:

- After ML processing, modal displays 3 suggested emoji based on top emotion predictions.
- Each suggestion shows confidence score (e.g., "Happy (87%)") for transparency.
- The user also has the option to select other emojis if they don't like the AI suggestions.
- User selection immediately updates calendar—no additional confirmation needed.

Design Rationale: This interface embodies the "AI as assistant" mental model: suggestions are visually prominent but not exclusive options. Showing confidence scores addresses the transparency requirement ("How does the AI know?") identified in interviews.

5.4 Feature 4: Activity Tagging

Functionality:

- Icon-based grid of predefined activities: Exercise, Social, Work, Study, Family, Sports.
- Multi-select capability (tap icons to toggle selection).
- Optional skip button for users who prefer not to tag activities.

Design Rationale: Activity tagging enables future correlation analysis (e.g., "You feel happier on days with exercise"). Making it optional reduces friction for users who find detailed logging burdensome. The icon-based interface is faster than text input and works across languages.

5.5 Feature 5: Mood Analytics Dashboard

Functionality:

- Donut chart (Recharts library) showing mood distribution over last 30 days.
- Total entry counts and correlation between activity and mood.

Design Rationale: Visual analytics help users recognize gradual changes that might not be apparent day-to-day. The donut chart format is friendly and non-clinical (vs. line graphs that feel medical). Streak counter provides gentle gamification without being pushy.

5.6 Feature 6: Privacy Controls & Settings

Functionality:

- **Audio Consent Toggle**: Users explicitly opt-in/out for audio storage. Default is set using the consent form during the first login.
- **Reminder Preferences**: Enable/disable daily reminder notifications (future feature—currently non-functional).
- **Audio Deletion**: Can selectively delete audio recording.

Privacy Implementation Details: Regardless of consent settings, audio files can be selectively deleted if they are being stored.

5.7 Feature 7: Pattern Detection & Gentle Interventions

Functionality:

- The database query identifies streaks of low-mood emojis spanning 5+ consecutive days.
- When pattern detected, frontend displays modal: "We noticed you've been feeling down lately. Go for a walk or do meditation."

Design Rationale: Interventions must be gentle, optional, and user-initiated. Phrasing like "We noticed" (not "You are depressed") and offering choices (not commands) respect user autonomy. The 5-day threshold was chosen based on the clinical literature suggesting that persistent low mood for a week warrants attention.

Simplified Implementation: The self-care tips currently link to static content. Full implementation would require curated, culturally appropriate resources vetted by mental health professionals—beyond prototype scope but architecturally prepared via content management database table.

6 User Evaluation & Results

We conducted usability testing with 5 participants to validate design decisions, identify interaction issues, and assess user acceptance of speech-based mood tracking.

6.1 Evaluation Methodology

Participants:

- **N = 5** IIT Gandhinagar students
- **Demographics**: 3 undergraduate (2 male, 1 female), 2 postgraduate (1 male, 1 female)
- **Age Range**: 19-24 years
- **Disciplines**: Computer Science (2), Mechanical Engineering (1), Cognitive Science (1), Electrical Engineering (1)
- **Recruitment**: Convenience sampling
- **Compensation**: None (voluntary participation)

Evaluation Setting:

- **Location**: Empty library discussion rooms (quiet room, individual sessions)
- **Duration**: 30-45 minutes per session
- **Setup**: Laptop (1920×1080 display) with Chrome browser, external microphone for voice recording
- **Facilitator**: One researcher (observer + think-aloud prompts)
- **Recording**: Written notes, and trnascripts for open-ended questions.

Evaluation Protocol:

1. **Introduction (5 min)**: Explain study purpose, obtain informed consent, disclosing all information about the project and their participation such as beneifts, risks, mitigations, etc. and clarify that ML analysis is real but no data will be stored long-term. A detailed informed consent and participant wellbeing form is separately submitted with the Evaluation documents.
2. **System Usability Scale (SUS)**: Overall Post Usage Usability Evaluation was conducted using the standard System Usability Scale (SUS) Questionnaire. The questionnaire contains 10 items that check for overall usability of a system. The questionnaire itself along with its scoring scheme is attached in the Evaluation folder submitted separately.
3. **System-Specific Questions- Mixed Methods Survey - System Specific Questions**: This includes some quantitative questions as well as some open-ended interview questions to gauge maximum user experience and feedback regarding the system-specific features. These features, as discussed before, are based on participant responses from our user beliefs and requirements surveys during Project 2. A summary of these themes and their corresonding features are shown in Table 7. This section also included some additional system specific questions that are not covered in SUS and pertain to user feedback on the app. The entire detailed questionnaire is submitted separately in the Evaluation documents.

Metrics Collected:

- **Subjective Ratings**: 1-5 Likert scale on privacy concern, distrust in AI accuracy, fear of stigma/labeling, need for different engagement styles.
- **Qualitative Feedback**: Think-aloud observations and interview quotes.

6.2 Quantitative Results

Table 2: Subjective Satisfaction Ratings (1=Low, 5=High)

Dimension	Mean (SD)
Privacy Concern	5.0 (0)
Distrust in AI accuracy	3.2 (0.8)
Fear of Stigma/Labeling	4.0 (0.2)
Need for different engagement styles	4.0 (0.7)

Interpretation:

- High satisfaction with privacy concerns validates our choice of keeping extensive and transparent privacy features.
- Moderate trust in AI (3.2/5) aligns with low-fi findings-skepticism about emotion detection accuracy persists despite real ML implementation.
- Fear of stigma (4.0/5) improved from low-fi concerns, due to an approachable looking interface, some judgemental tones throughout.

- Need for different engagement styles (4.0/5) shows interest but not complete adoption, and qualitative data reveals reasons (see below).

6.3 Qualitative Findings

We conducted thematic analysis on think-aloud transcripts and interview responses. Five major themes emerged:

6.3.1 Theme 1: Visual Feedback Builds Confidence.

Representative Quotes:

> "The live transcription made me feel like the app was actually listening, not just recording silence." (P2)
> "I liked seeing the transcription appear in real-time—it confirmed that the microphone was working and the app understood what I was saying." (P4)

Analysis: Real-time visual feedback (transcription) addressed a key anxiety from low-fi testing: "Is the system actually listening?" This feature, added beyond the original wireframes, proved critical for user confidence.

6.3.2 Theme 2: AI Suggestions Were Helpful But Not Trusted.

Representative Quotes:

> "The AI suggested 'calm' emoji, which actually matched how I felt—that was surprising! But I'm not sure I would always trust it. What if it misunderstands sarcasm?" (P1)
> "I deliberately spoke in a monotone voice to test it, and the AI suggested 'neutral'. That makes sense, but I was actually feeling happy inside. The AI can't read my mind, only my voice." (P3)

Analysis: Participants appreciated AI assistance but remained skeptical about accuracy, especially for nuanced emotions. Importantly, all users understood that they could override suggestions—no one felt "forced" to accept AI predictions, validating the "assistant not authority" mental model.

Design Implication: Future iterations should enhance transparency: explain *why* specific emotions were detected (e.g., "Your voice had lower pitch and slower speech rate, which often indicates sadness"). This moves from black-box AI to interpretable AI.

6.3.3 Theme 3: Privacy Messaging Was Reassuring.

Representative Quotes:

> "I was worried about my voice being stored forever, but the popup clearly said 'Audio deleted immediately after analysis.' That made me comfortable recording." (P5)
> "The consent dialog was good—I appreciated being asked explicitly instead of the app just assuming I'm okay with it." (P2)

Analysis: Explicit privacy messaging countered the anxiety about "voice data permanence" identified in low-fi interviews. However, one participant (P3) expressed residual concern: "Even if you delete it, how do I know the ML service isn't keeping a copy?" This suggests a need for even more transparency about data flow architecture.

6.3.4 Theme 4: Calendar Interface Was Intuitive.

Representative Quotes:

> "The calendar made sense immediately—I didn't need instructions. It's just like my Google Calendar but with mood emoji instead of events." (P4)
> "I could see at a glance that last week was mostly happy emoji, but this week has more sad ones. That pattern wasn't obvious to me before looking at the calendar." (P1)

Observation: All participants successfully navigated the calendar without help. The familiar metaphor (digital planner) enabled zero-learning-curve interaction.

6.3.5 Theme 5: Daily Use Faces Motivation Challenges.

Representative Quotes:

> "I could see myself using this for a week or two out of curiosity, but I'm not sure I'd keep it up daily for months. It's just one more thing to remember." (P2)
> "If I'm having a really bad day, the last thing I want to do is record myself talking about it. I'd probably skip those days, which defeats the purpose." (P5)
> "Maybe if there were reminders or some kind of streak system, I'd be more motivated. But not pushy reminders—just gentle nudges." (P3)

Analysis: The "daily journaling" metaphor successfully reduced stigma but introduced a new challenge: **habit formation**. Unlike traditional mental health screening (which happens occasionally with a clinician), daily self-tracking requires sustained motivation.

Paradox Identified: Users most likely to benefit (those experiencing persistent low mood) are least motivated to engage. This highlights the need for carefully designed interventions that encourage continued use without being guilt-inducing.

6.4 Usability Issues Identified

Table 3: Observed Usability Problems & Severity

Issue	Frequency	Severity
Confusion about whether audio is saved after consent dismissal	2/5	Moderate
Unclear how to edit previously selected emoji	2/5	Minor
No feedback when clicking disabled "Go for Today" button (already recorded)	4/5	Minor
Language selector shown but only English works	5/5	Major

Actionable Fixes:

1. **Privacy Clarification**: Add persistent banner: "Your audio is deleted after each recording" on recording screen.
2. **Emoji Editing**: Add edit icon on calendar days to modify past activity entries.
3. **Disabled Button Feedback**: Show toast message: "You've already recorded today! View your calendar to see your entry."

(4) **Language Feature Transparency**: Either implement multilingual models or remove selector UI to avoid false expectations.

6.5 Comparison with Low-Fidelity Findings

Table 4 compares predictions from low-fi user research with actual hi-fi observations:

Table 4: Low-Fi Predictions vs. Hi-Fi Results

Aspect	Low-Fi Prediction	Hi-Fi Result
Stigma reduction via journaling metaphor	High acceptance	Validated
Privacy concerns with voice data	Major barrier	Mitigated with explicit deletion
AI distrust	Moderate skepticism	Persists despite real ML
Daily use motivation	Assumed users would self-motivate	Habit formation is hard
Calendar intuitiveness	Expected easy learning	Zero learning curve
Need for transparency	Critical requirement	Partially met

Key Insight: Implementing real ML did *not* significantly reduce AI distrust—skepticism stems from lack of explainability, not implementation fidelity. Future work must prioritize interpretable AI over model accuracy alone.

7 Ethics & Risk Assessment

Evaluating a mental health monitoring tool requires careful attention to participant welfare, data privacy, and potential psychological risks.

7.1 Ethical Review & Informed Consent

Although formal IRB approval was not required for course project evaluations, we implemented rigorous ethical safeguards for this study:

- **Informed Consent**: All participants (N=5) signed written consent forms explaining study purpose, data usage, right to withdraw, and anonymization procedures.
- **Voluntary Participation**: Recruitment emphasized voluntary nature with no course credit or compensation.
- **Anonymization**: Participants assigned codes (P1-P5); no personally identifying information collected.
- **Data Retention**: Voice recordings deleted immediately after sessions. Screen recordings and notes stored on password-protected devices, accessible only to research team.

The consent form included: study purpose, participation requirements (30-45 minutes), potential risks (mild emotional discomfort), privacy protections (audio deletion, data anonymization), voluntary withdrawal rights, and researcher contact information with campus counseling hotline.

7.2 Risk Identification & Mitigation

We identified four primary risks and implemented specific mitigation strategies:

7.2.1 Risk 1: Emotional Distress During Recording. **Concern**: Reflecting on emotions might trigger distress in participants experiencing mental health challenges.

Mitigation:

- Pre-screening excluded individuals reporting active mental health crises
- Facilitator trained to recognize distress signals and terminate session if needed
- All participants received campus counseling contact information post-session

Outcome: No participants exhibited distress during sessions.

7.2.2 Risk 2: Voice Data Privacy Breach. **Concern**: Voice recordings contain biometric identifiers; breaches could expose identity and emotional states.

Mitigation:

- Immediate deletion: Audio files removed within seconds of ML processing
- Local-only processing: ML service runs in Docker container, no external transmission
- Database isolation: PostgreSQL accessible via localhost only
- Minimal metadata: Only emotion labels and dates stored—no names or demographics

Outcome: Zero data breaches. Manual verification confirmed empty `/temp_audio/` directory after each session.

7.2.3 Risk 3: Misinterpretation as Medical Diagnosis. **Concern**: Users might treat AI predictions as clinical diagnoses and avoid professional help.

Mitigation:

- Explicit disclaimers in consent form and UI: "This is not a diagnostic tool"
- AI provides suggestions only; users always choose final emoji
- Intervention messaging uses gentle language ("Would you like resources?") not commands

Outcome: Post-session interviews confirmed understanding of limitations (e.g., P3: "I wouldn't rely on this instead of talking to a counselor—it's more like a diary that gives hints").

7.2.4 Risk 4: Participant Fatigue. **Concern**: 30-45 minute sessions might cause cognitive fatigue during exam periods.

Mitigation:

- Flexible scheduling accommodating participant preferences
- Breaks allowed between tasks
- Protocol shortened from 60 to 45 minutes after pilot testing

Outcome: No reported fatigue; average session duration 38 minutes.

7.3 Deployment Considerations

Real-world deployment would require additional ethical safeguards beyond prototype evaluation:

- **Clinical Validation**: Emotion recognition accuracy validated against clinician assessments (HAM-D, BDI scores) before deployment

- **Crisis Detection**: System should detect self-harm/suicidal language and connect users to immediate crisis resources
- **Regulatory Compliance**: If positioned as health monitoring tool, must comply with medical device regulations (e.g., FDA digital health guidelines)
- **Longitudinal Safety**: Monitor for unintended consequences (fixation on scores, anxiety about "bad" days)

This evaluation demonstrated that with appropriate consent procedures, immediate data deletion, and transparent communication, students are willing to engage with voice-based emotion tracking for research purposes. However, transitioning from research prototype to deployed mental health tool would require substantially more rigorous clinical validation and ongoing ethical oversight.

8 Conclusion

This paper presented HearMeOut, a high-fidelity web-based prototype for speech-based mental health monitoring in students. By translating low-fidelity wireframes into a fully functional system with real emotion recognition capabilities, we validated key design hypotheses while uncovering important challenges for real-world deployment.

8.1 Limitations & Future Work

Limitations:

- **Small Sample Size**: N=5 limits statistical generalizability. Larger studies needed to detect edge cases and diversity in user reactions.
- **Short-Term Evaluation**: Single-session testing cannot assess long-term engagement, habit formation, or longitudinal pattern detection accuracy.
- **English-Only Implementation**: Restricts applicability to multilingual student populations (Indian universities often use Hindi, regional languages).
- **Controlled Lab Setting**: Real-world usage (dorm rooms, noisy environments) may reveal different challenges than quiet lab conditions.

Future Directions:

(1) **Longitudinal Field Study**: Deploy HearMeOut for 4-8 weeks with 50+ students to assess:
   - Actual daily usage rates (vs. self-reported likelihood)
   - Accuracy of pattern detection over time
   - Impact on self-awareness and help-seeking behavior

(2) **Explainable AI Integration**: Implement feature attribution techniques (e.g., Layer-wise Relevance Propagation) to show which acoustic features (pitch, rate, energy) influenced emotion predictions.

(3) **Multilingual Expansion**: Integrate IndicWav2Vec or mBERT-based models for Hindi, Gujarati, Tamil support.

(4) **Intervention Content Development**: Collaborate with campus counseling centers to curate culturally appropriate, evidence-based self-care resources and referral pathways.

(5) **Habit Formation Mechanisms**: Experiment with:
   - Implementation intentions (e.g., "If it's 9 PM, then I will record my mood")
   - Social accountability (opt-in buddy system where friends check in on each other)
   - Adaptive reminders (only send notifications when user misses 2+ consecutive days)

(6) **Clinical Validation**: Partner with psychiatry departments to validate emotion detection against Hamilton Depression Rating Scale (HAM-D) or Beck Depression Inventory (BDI) scores.

8.2 Final Reflection

Building HearMeOut reinforced a fundamental HCI principle: *technical sophistication means little if users don't trust, understand, or adopt the system.* Our journey from literature review (identifying speech as a promising biomarker) to low-fi design (discovering stigma and privacy as primary barriers) to hi-fi implementation (validating the journaling metaphor while uncovering habit formation challenges) illustrates the iterative, user-centered nature of impactful design.

The prototype is not perfect—usability issues remain, multilingual support is incomplete, and long-term engagement is unproven. But it represents a meaningful step toward making mental health monitoring accessible, private, and humane. By centering user needs, respecting autonomy, and maintaining transparency, we hope HearMeOut contributes to a future where students can track their emotional wellbeing as naturally as they track their physical health.

Acknowledgments

We thank Prof. Yogesh Meena for guidance on HCI methodology and evaluation protocols. We are grateful to our five participants for their time, honesty, and thoughtful feedback. We also thank the IIT Gandhinagar HCI community for design critiques during prototype development.

References

[1] Anukriti Bhargava, Bhavik Patel, Hitesh Kumar, and Pranav Patil. 2025. A Systematic Literature Review of Speech-Based Screening for Early Mental Health Detection in Students. *CS435 Course Project Task 1* (2025).

[2] Anukriti Bhargava, Bhavik Patel, Hitesh Kumar, and Pranav Patil. 2025. Low-Fidelity Prototypes for Speech-Based Mental Health Screening Interface: A User-Centered Design Approach. *CS435 Course Project Task 2* (2025).

— End of Report —

Appendix

A Supplementary Evaluation Materials

All detailed evaluation materials are available in the project's GitHub repository for transparency and reproducibility. This appendix provides direct links to comprehensive documentation that was used during user evaluation sessions.

A.1 Informed Consent Form

The complete informed consent form administered to all participants is available in this document.

A.2 Usability Testing Protocol & Evaluation Instruments

The complete evaluation protocol, including task scripts, think-aloud prompts, interview questions, and all quantitative instruments (SUS questionnaire, system-specific Likert scales), is available in the evaluation components document.

B Additional Screenshots

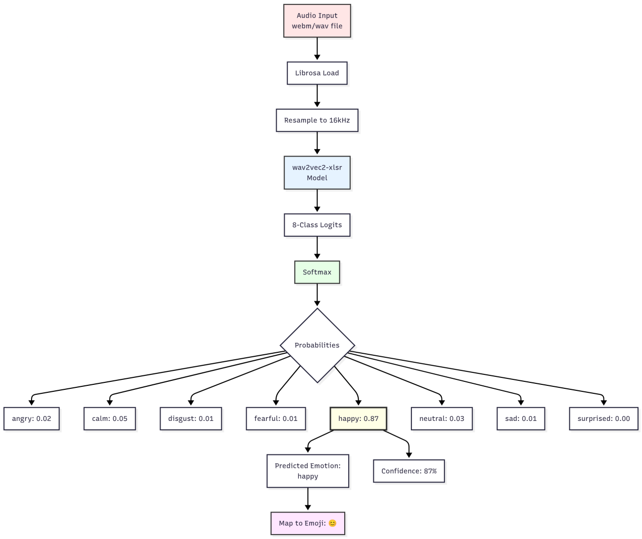


Figure 5: Model Prediction Flow

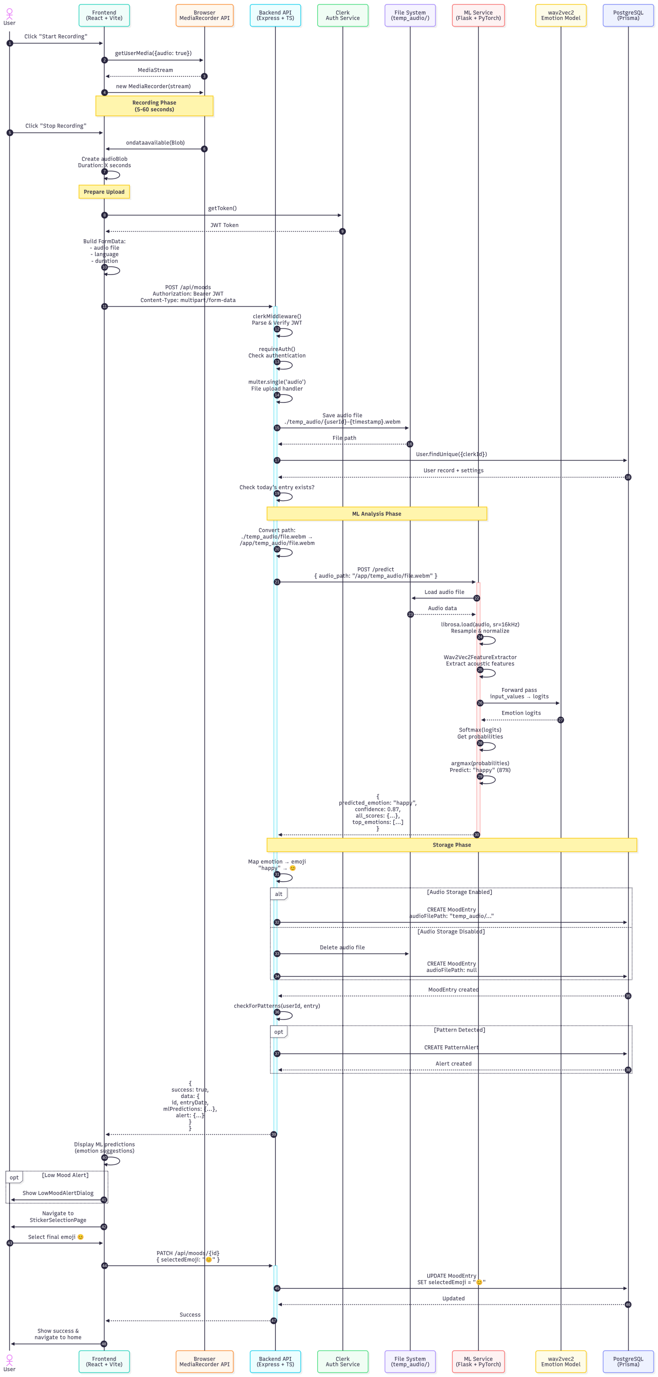


Figure 6: Complete Data Flow from Audio Recording to Emoji Suggestions

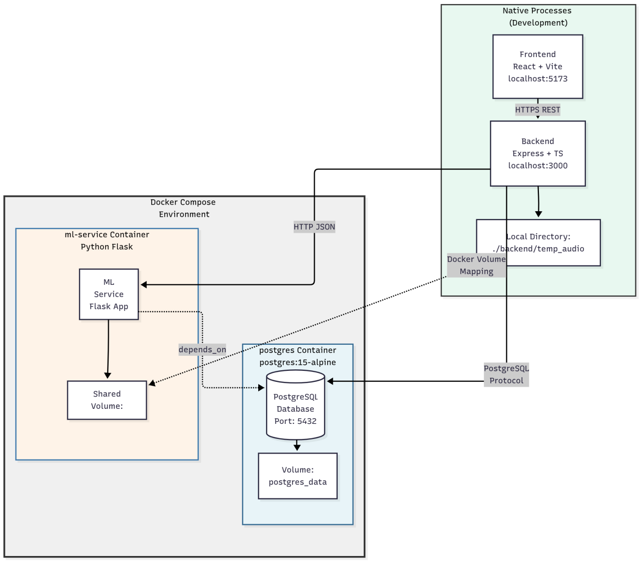


Figure 7: Application Architecture inside Docker Container

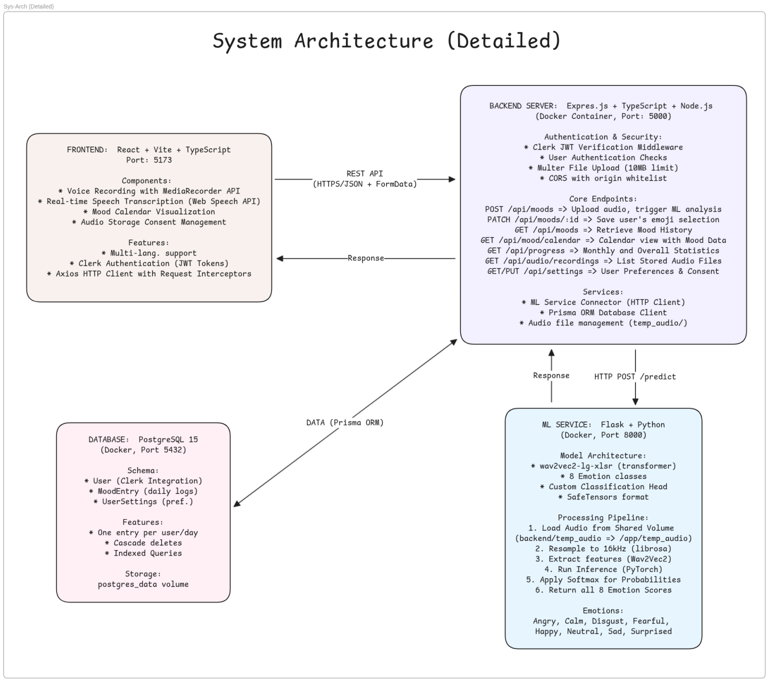

Sys-Arch (Detailed)
System Architecture (Detailed)
FRONTEND: React + Vite + TypeScript
Port: 5173
Components:
* Voice Recording with MediaRecorder API
* Real-time Speech Transcription (Web Speech API)
* Mood Calendar Visualization
* Audio Storage Consent Management
Features:
* Multi-lang. support
* Clerk Authentication (JWT Tokens)
* Axios HTTP Client with Request Interceptors
REST API
(HTTPS/JSON + FormData)
Response
BACKEND SERVER: Expres.js + TypeScript + Node.js
(Docker Container, Port: 5000)
Authentication & Security:
* Clerk JWT Verification Middleware
* User Authentication Checks
* Multer File Upload (10MB limit)
* CORS with origin whitelist
Core Endpoints:
POST /api/moods => Upload audio, trigger ML analysis
PATCH /api/moods/:id => Save user's emoji selection
GET /api/moods => Retrieve Mood History
GET /api/mood/calendar => Calendar view with Mood Data
GET /api/progress => Monthly and Overall Statistics
GET /api/audio/recordings => List Stored Audio Files
GET/PUT /api/settings => User Preferences & Consent
Services:
* ML Service Connector (HTTP Client)
* Prisma ORM Database Client
* Audio file management (temp_audio/)
Response
HTTP POST /predict
DATA (Prisma ORM)
DATABASE: PostgreSQL 15
(Docker, Port 5432)
Schema:
* User (Clerk Integration)
* MoodEntry (daily logs)
* UserSettings (pref.)
Features:
* One entry per user/day
* Cascade deletes
* Indexed Queries
Storage:
postgres_data volume
ML SERVICE: Flask + Python
(Docker, Port 8000)
Model Architecture:
* wav2vec2-lg-xlsr (transformer)
* 8 Emotion classes
* Custom Classification Head
* SafeTensors format
Processing Pipeline:
1. Load Audio from Shared Volume
(backend/temp_audio => /app/temp_audio)
2. Resample to 16kHz (librosa)
3. Extract features (Wav2Vec2)
4. Run Inference (PyTorch)
5. Apply Softmax for Probabilities
6. Return all 8 Emotion Scores
Emotions:
Angry, Calm, Disgust, Fearful,
Happy, Neutral, Sad, Surprised

Figure 8: System Architecture (Detailed View)