SMART INDIA HACKATHON 2025

- **Problem Statement ID –25242**
- **Problem Statement Title- Deep learning based ALM (Audio Language Model) which Listen, Think, and Understand the speech and non-speech Together**
- **Theme-** Smart Automation
- **PS Category- Software**
- **Team ID-59988**
- **Team Name (Registered on portal)-Quantum Quirks**

AUDIO LANGUAGE MODEL

🎧 Audio data is often unstructured and difficult to analyze.
🕓 Manual transcription and analysis are time-consuming and prone to human errors.
⚙Existing systems focus only on speech, ignoring environmental or contextual audio.
🌐 Lack of unified models that can handle multilingual, code-mixed, and noisy audio data.
🗣☐ Automatically transcribe speech from audio inputs.
🔊 Identify speech and non-speech segments in real-time.
☐ Extract meaningful insights and context from the processed audio (like emotion, environment, or activity).
🌍 Adapt to multilingual and real-world conditions.

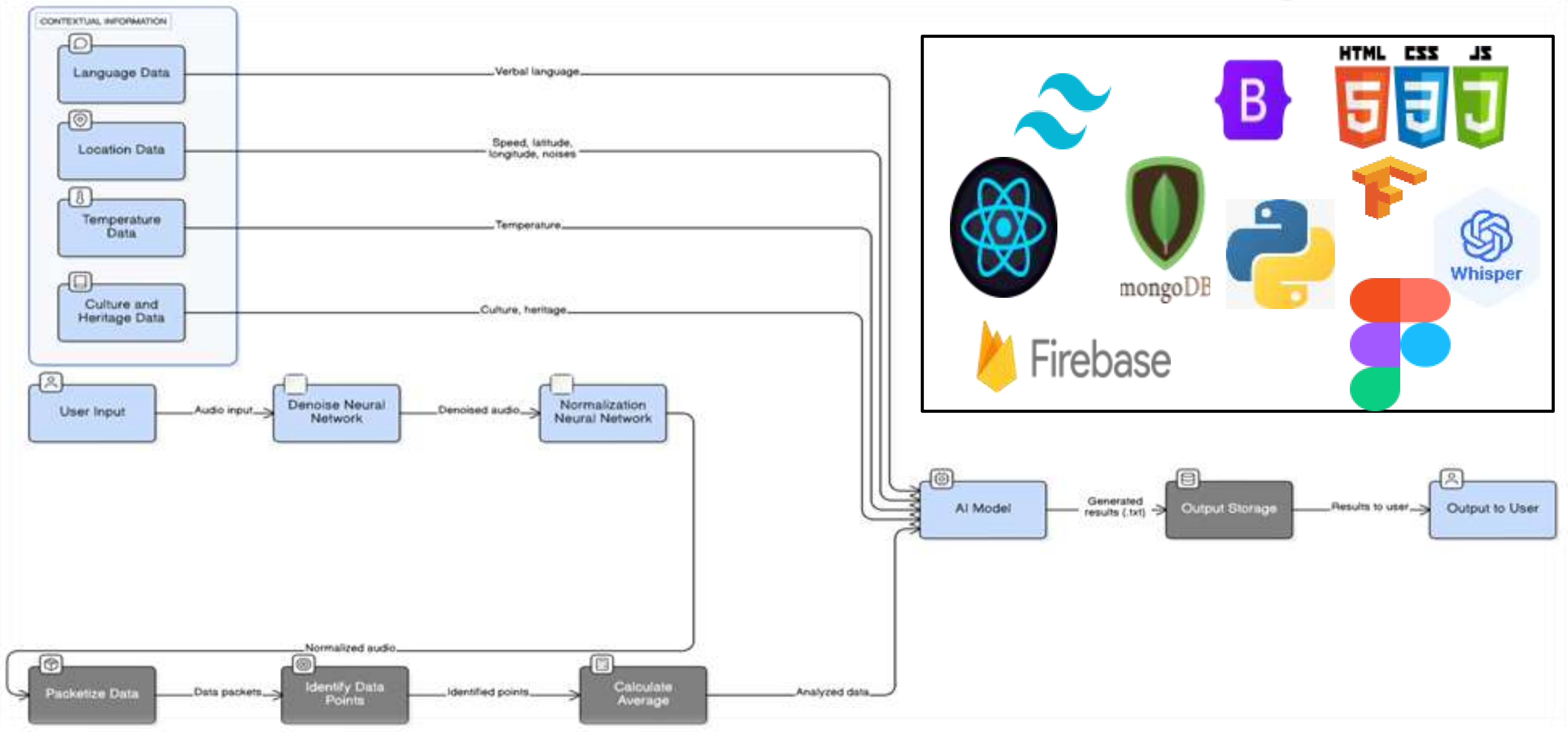
To build an Audio Language Model that:

❖ Imports and processes various audio file formats.
❖ Generates accurate text transcripts from speech.
❖ Combines both **acoustic** and **linguistic** information
❖ the model uses **deep neural networks** (like Transformers) to analyze patterns and make decisions..

Uniqueness: 1. Unified Understanding of Speech and Non-Speech Audio.
2. Multimodal Audio Context Awareness.
3. Cross-Domain Applicability.
4.Human-Like “Listen–Think–Respond” Loop

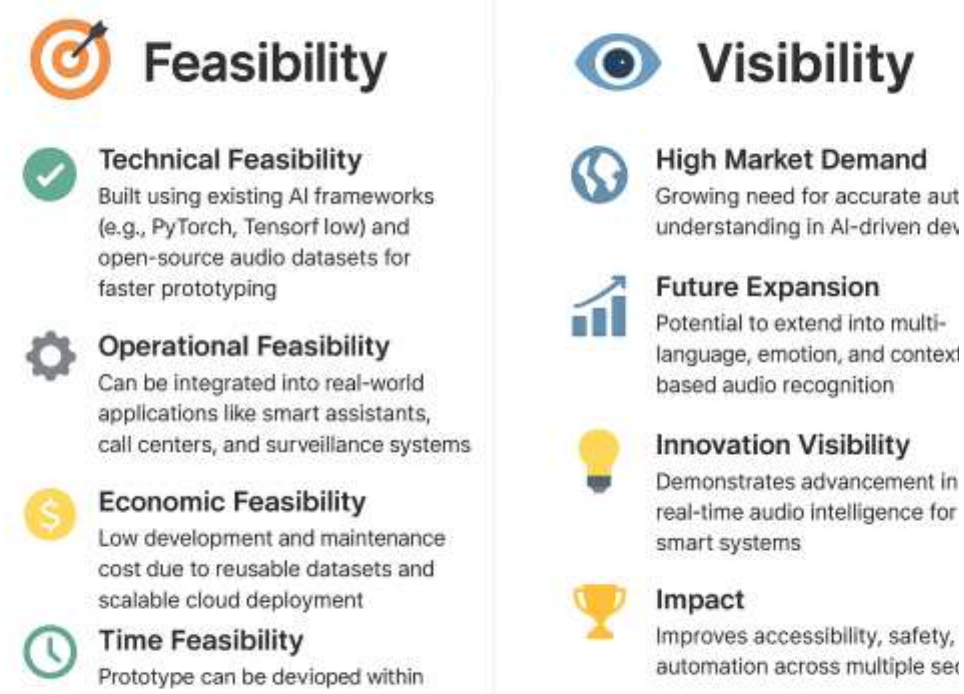
TECHNICAL APPROACH

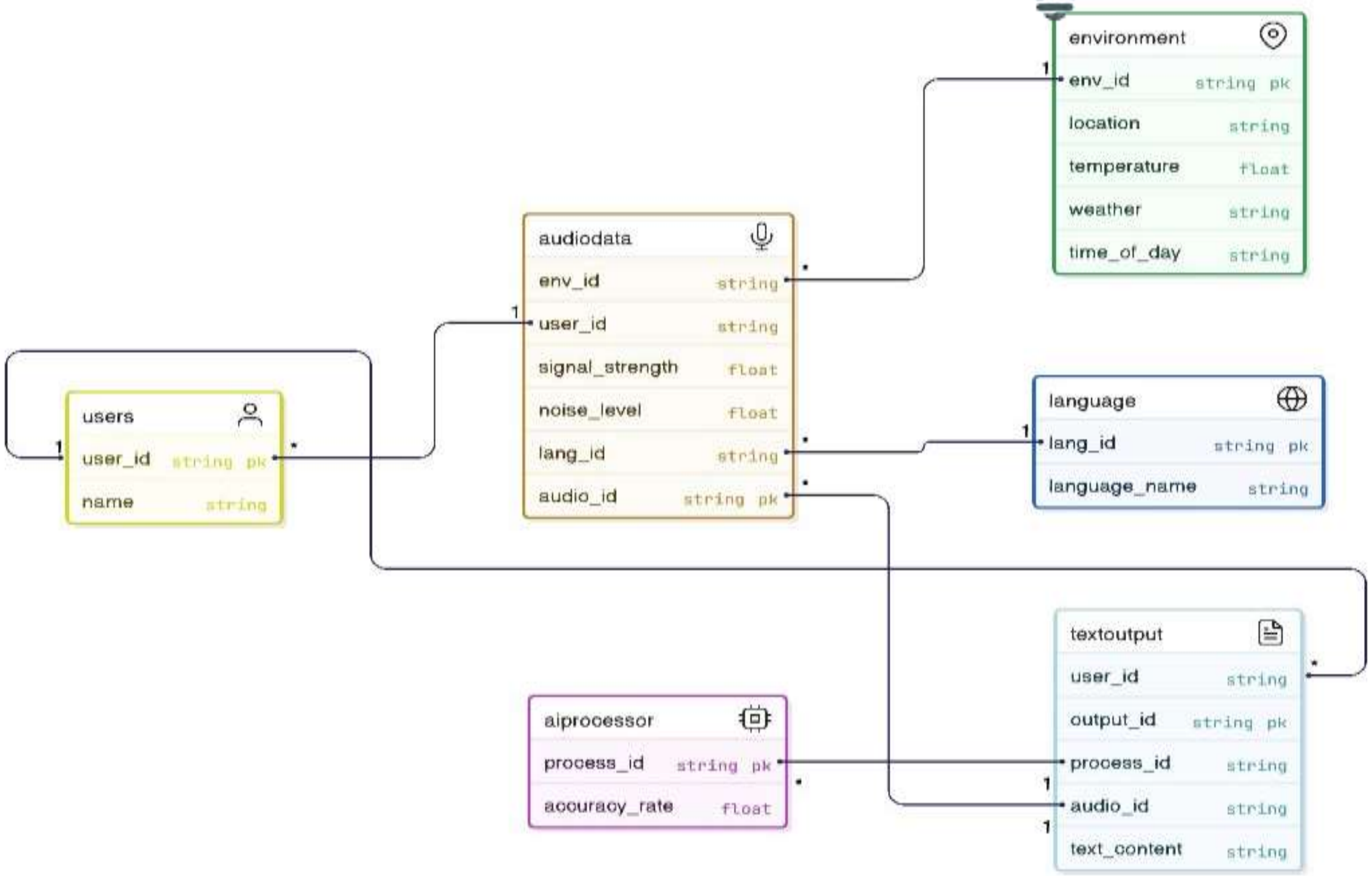
❖Develop a **unified Audio Language Model (ALM)** that understands both **speech and non-speech audio**.
❖Capture **raw audio input** from the user and **preprocess** it through denoising and normalization.
❖**Segment audio** into small data packets for efficient processing. identify **individual** and **average audio data points** for detailed analysis.
❖Collect **contextual information** such as
Fuse all contextual and audio data inside the **AI model** for interpretation.
❖Generate and store the analyzed output as a **text (.txt) file**.



Uniqueness:
🗣 **Objective:** Distinguishes between speech and non-speech audio in real-time.
💡 **Innovation:** Uses a custom-trained dataset for improved accuracy across environments.
⚙**Capability:** Powers applications like smart assistants, surveillance, and audio analytics.
🔊 **Efficiency:** Processes audio with low latency and high precision using optimized AI models.
🌐 **Scalability:** Easily adaptable to different languages, accents, and acoustic conditions.

FEASIBILITY AND VIABILITY





⚠ Potential Challenges

Challenge	💬 Description
🌐 **Multilingual & Code-Mixed Speech**	Accurately recognizing Indian regional and mixed-language speech.
🔊 **Non-Speech Sound Detection**	Distinguishing between speech, silence, and environmental sounds.
💾 **Data Collection & Annotation**	Building large, diverse, labeled datasets for training.

☐ **Overall Summary:** The Audio Language Model is technically strong, cost-efficient, and time-feasible — ensuring scalable, real-world deployment with high impact in multilingual AI applications.

IMPACT AND BENEFITS

Defence-Specific Impact

🚨 **Threat Detection & Surveillance**
The ALM can automatically identify **critical defence-related sounds** such as gunfire, explosions, distress calls, or unauthorized movements.

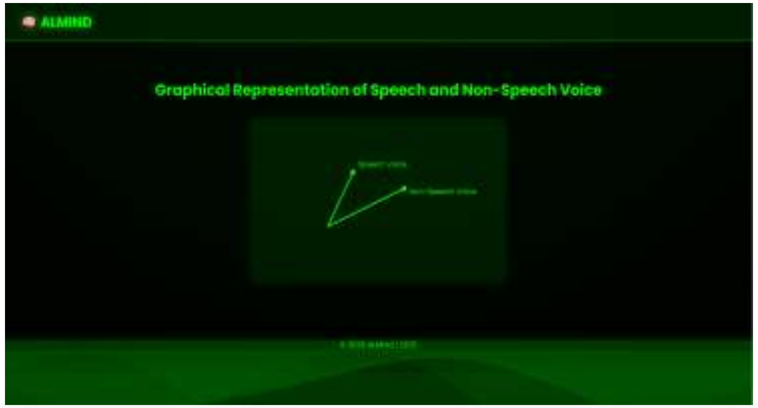
🎯 **Tactical Real-Time Alerts**
By processing live field audio, the system can **generate instant alerts** for suspicious activities or sounds.
This feature is critical for **battlefield awareness**, **base security**, and **emergency response coordination** among defence units.

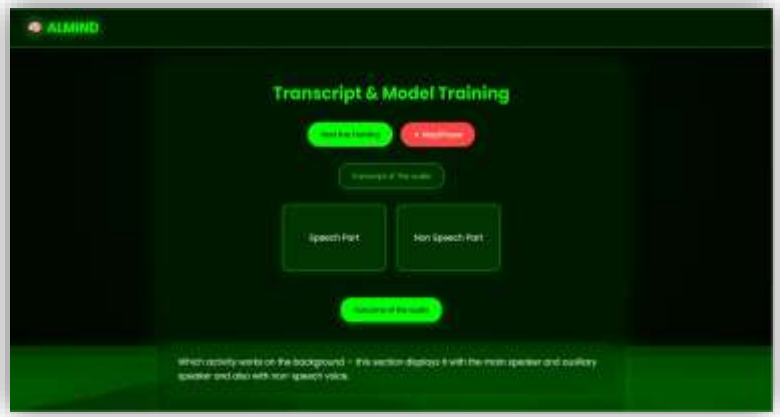
🔐 **Border and Coastal Security**
The ALM can detect **unusual sound patterns** near borders, coastlines, or restricted areas — such as vehicles, boats, or drone noises.
This helps strengthen **perimeter defence** and **prevents unauthorized crossings** or intrusions before they escalate.

☐ **Intelligence and Reconnaissance Support**
The model can analyze intercepted communications or ambient battlefield sounds to extract **strategic insights**.

☐ Future Scope – Defense Applications

🗣☐ **Secure Voice Commands** – Encrypted, hands-free control for weapons, drones & vehicles.
🌐 **Multilingual Communication** – Real-time translation between field units.
🎯 **Mission AI Assistant** – Instant voice-based tactical updates & intel access





💡 Key Benefits of the Audio Language Model (ALM)

🎧 **Automated Audio Understanding**
Transforms unstructured audio into meaningful information by **automatically detecting, classifying, and transcribing** both speech and non-speech sounds.

⚙ **Improved Efficiency & Accuracy**
Eliminates manual transcription errors and **enhances accuracy** in speech recognition and sound classification.

🌐 **Multilingual and Cross-Cultural Adaptability**
Supports multiple **languages, dialects, and regional accents**, enabling better communication across diverse communities or forces.

☐ **Context-Aware Insights**
Gathers **contextual data** (location, noise level, environment) to interpret sound meaningfully.

RESEARCH AND REFERENCES

📖 **References**

Research Papers

Tang, C. et al. (2023) – *SALMONN: Towards Generic Hearing Abilities for Large Language Models.* [arXiv:2310.13289]
→ Introduces the idea of AI with hearing capabilities for both speech & non-speech sounds.

Ardila, R. et al. (2019) – *Common Voice: A Massively-Multilingual Speech Corpus.* Mozilla Foundation.
→ Provides **diverse multilingual datasets** for training speech recognition systems.

Wu, J. et al. (2023) – *Speech-LLaMA: Decoder-Only Architecture for Speech and Language Model Integration.* [arXiv:2307.03917]
→ Shows how to integrate speech understanding into LLMs for contextual comprehension.

📀 **Datasets & Benchmarks**

AudioSet (Google Research, 2017) – Over 2 million labeled audio clips covering 600+ sound classes.

DCASE Challenge (2013–2023) – Annual benchmark for sound event detection & acoustic scene analysis.

🏠 **ReaLISED Dataset (MDPI, 2022)** – Real-world indoor sound event dataset for AI model training.

🌐 **FSDnoisy18k (Fonseca et al., 2019)** – Web audio dataset designed for training with noisy labels.

Defence & Strategic Research

🚨 **Gunshot & Blast Detection Systems** – IEEE papers on acoustic signal detection for defence and border safety.

📡 **AI-Driven Acoustic Surveillance** – Research integrating IoT sensors and machine learning for field monitoring.

🎯 **Sound-Based Intelligence Systems** – Defence studies on audio-driven situational awareness in combat zones.