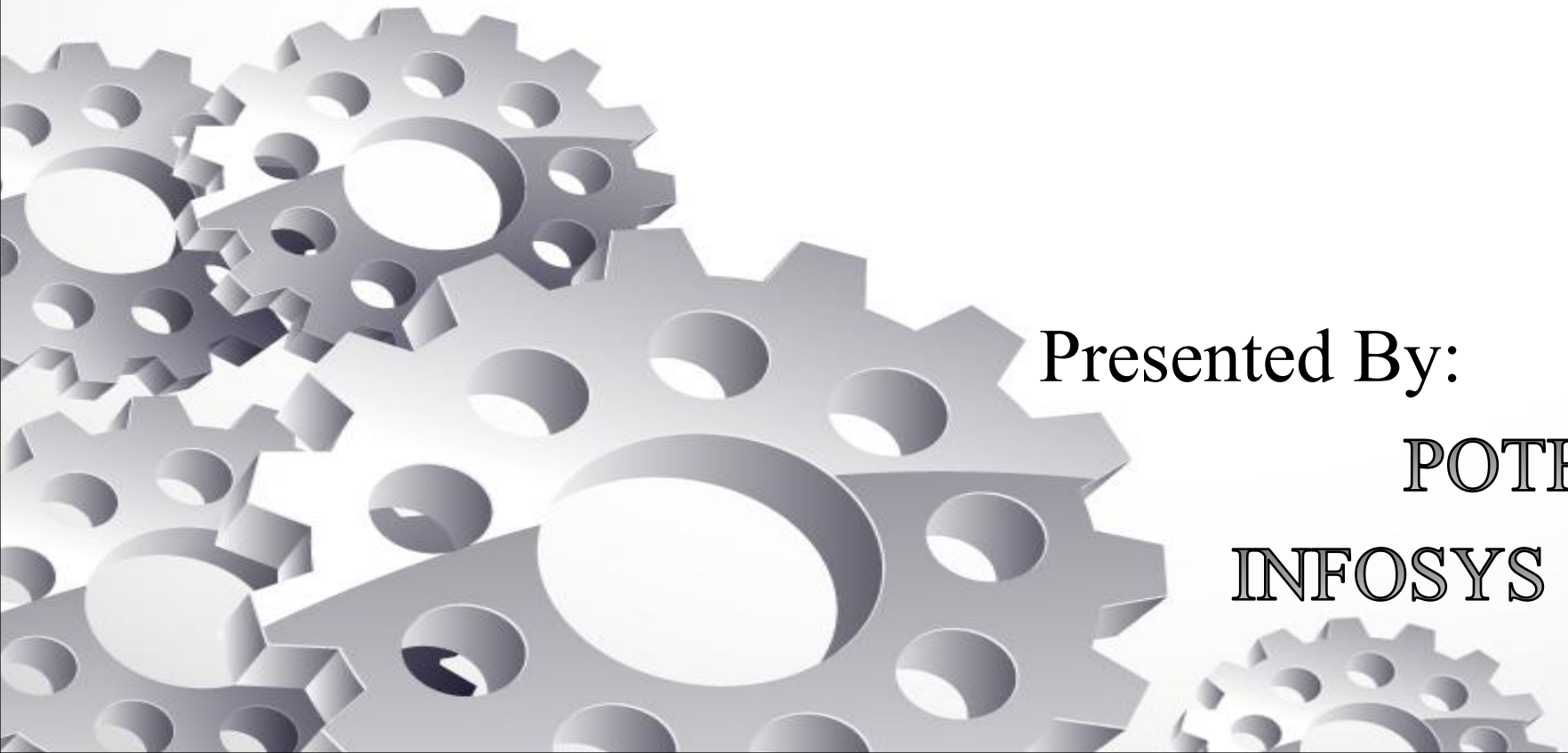
AudioMind: Automated Podcast Transcription & Insights

Transforming Raw Audio into Structured, Searchable Intelligence



Presented By:

POTHEESH VIGNESH K

INFOSYS SPRINGBOARD 6.0

Problem Statement

The Core Problem

- Audio content is unstructured and difficult to navigate
- Users must listen to entire files to find specific information
- Language barriers restrict accessibility.
- No structured insight extraction from long-form audio.
- Poor discoverability in educational & research contexts.

Real-World Impact

- Students struggle to revise lectures quickly
- Podcast creators spend hours creating show notes
- Hearing-impaired users lack structured transcript support
- Researchers cannot efficiently index audio archives

System Overview

- AudioMind converts raw audio into transcripted text.
- Full transcript
- Topic-based segmentation
- Abstractive summaries
- Context-aware titles
- Sentiment analysis
- Interactive timeline
- Translation & Romanization

Ai Flow:

[Audio Input] --> [Whisper Transcription]

[Whisper Transcription] --> [Semantic Segmentation]

[Semantic Segmentation] --> [Summarization & Sentiment]

[Summarization & Sentiment]--> [Translate& Localization]

[Translate& Localization] --> [Visualization & UI].

Pipeline:

Input → AI Processing → Intelligence Extraction → Visualization

Data Set

Sample	Genre	Purpose
I wanna be your(Music Track)	Lyrics + Background	Test transcription accuracy
Political Podcast	Multi-speaker	Test segmentation
Crowd Noise	No speech	Edge-case handling
TED Talk	Structured speech	Educational use-case
Audiobook	Narrative	Long-form storytelling

Formats:

- MP3
- WAV

Preprocessing:

- ffmpeg conversion
- Noise robustness (Whisper)
- File validation & sanitization

Methodology

1.Transcription:

- Whisper Small model (244M parameters).
- Auto language detection.
- Timestamp generation.

2.Topic Segmentation:

- Sentence splitting
- Embedding using MiniLM
- Cosine Similarity threshold = 0.82
- Dynamic boundary detection

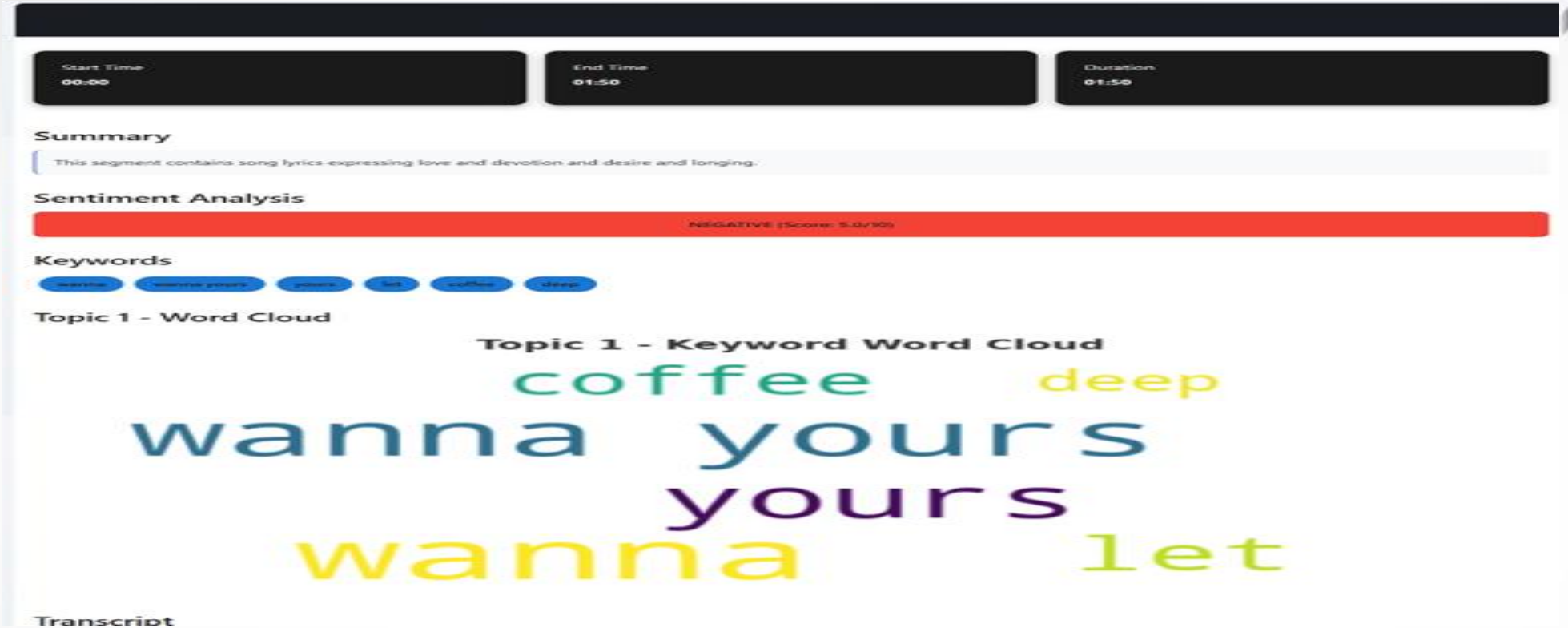
3.Summarization & Keyword:

- Flan-T5 for abstractive summaries
- Context-aware topic titles

4.Sentiment Analysis:

- TextBlob polarity scoring
- Converted to 1–10 scale

Visualization (Timeline & Keyword Cloud)



Interactive Timeline:

- Horizontal duration mapping
- Color-coded topic segments
- Quick structural overview

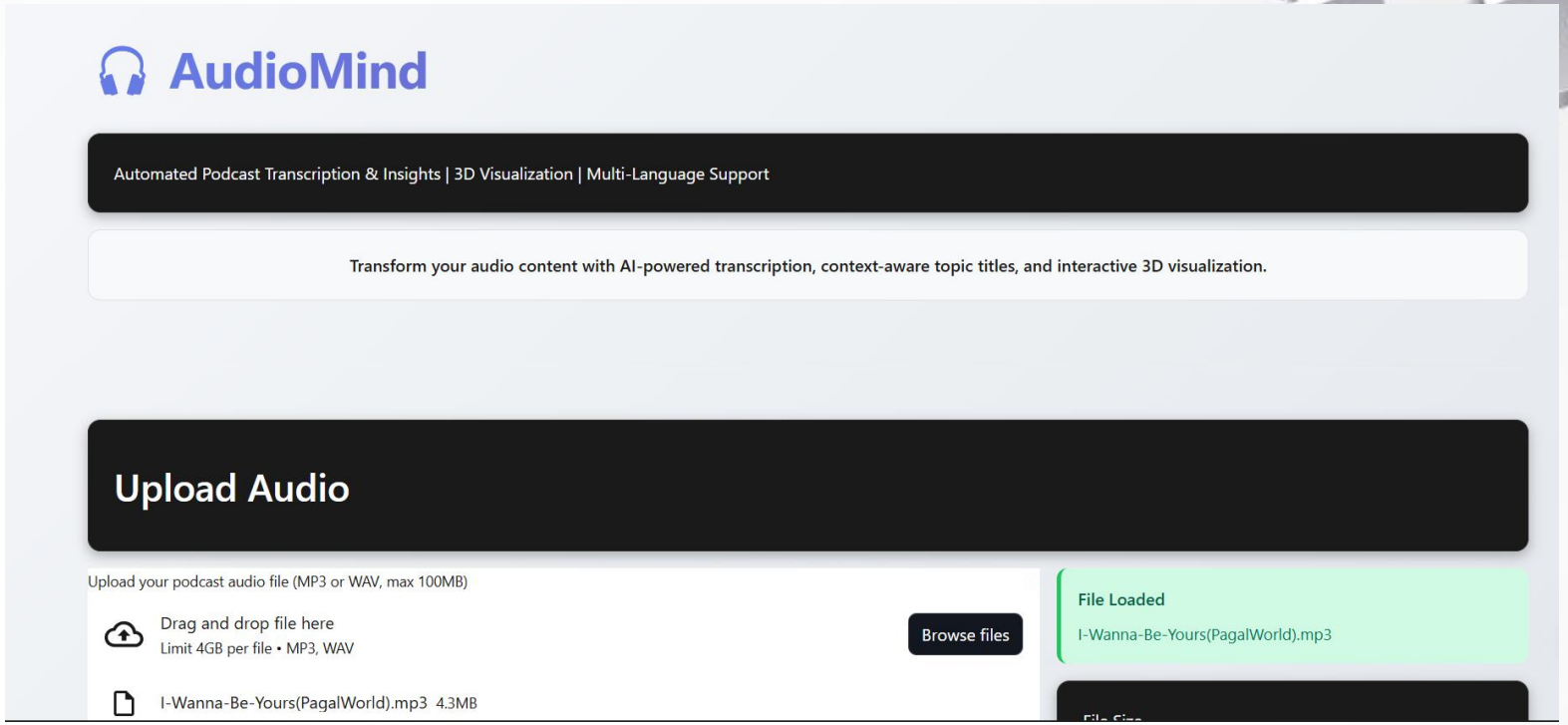
Keyword:

- Frequency-weighted keywords.
- Visual emphasis on dominant concepts.
- Rapid theme recognition.

User Interface Demonstration

Dashboard Sections:

- Upload Audio
- Process & Analyze
- Transcript View
- Topic Cards
- Timeline Visualization
- 3D Structural Animation



Dark-mode, high-contrast, clean layout.

Key Results

Podcast	Outcome	Score
Music	Accurate lyrics, 2 segments	9/10
Political Podcast	10 precise topics	10/10
Crowd Noise	Correctly detected no speech	N/A
TED Talk	3 structured topics	9/10
Audiobook	10 rich narrative segments	10/10

Strength Indicators:

- Strong semantic segmentation
- High-quality summaries
- Accurate edge-case handling
- Multi-language support functional

Testing & Feedback Summary

- **User Feedback:**

Review Problem --> Solved

- Low contrast issue → Implemented high-contrast dark UI
- Generic titles → LLM-based context-aware titles added
- Noise transcription attempt → Graceful handling implemented
- System improved through iterative testing.

Limitations & Future Work

Current Limitations:

- No speaker diarization
- Subjective segmentation boundaries
- Performance drops in poor audio
- Resource-heavy for long files

Future Enhancements:

- Speaker identification (Diarization)
- Real-time streaming transcription
- Cloud deployment
- Custom vocabulary injection
- Dedicated React frontend

THANK YOU