

The background is a dark blue gradient. There are several decorative cyan elements: a horizontal line with a small rectangular block in the middle at the top; a vertical line on the right side with a horizontal segment at the top and bottom; a horizontal line at the bottom with a series of three hexagons containing double arrows on the left and a series of five slanted parallel lines in the middle; and a large cyan shape on the left side that looks like a stylized 'L' or a bracket.

Project Title: Live Meeting Summarizer Application

Name: Param Parikh

TABLE OF CONTENTS

- PROJECT OVERVIEW
 - DATASET OVERVIEW AND KEY INSIGHTS
- METHODOLOGY
 - EXPLORATORY DATA ANALYSIS
- VISUALIZATION
 - DATA PREPROCESSING
 - FEATURE EXTRACTION
 - MODEL ARCHITECTURE
 - TRAINING AND EVALUATION
- results
- USER INTERFACE
- CHALLENGES
- FUTURE SCOPE
- CONCLUSION



PROJECT OVERVIEW

OBJECTIVE:

Build a fully automated audio → text → summary pipeline that seamlessly converts spoken content from meetings, lectures, and interviews into structured textual formats. The system must integrate multiple advanced machine learning models into a cohesive end-to-end workflow...

Achieve high transcription accuracy across accents and noise by implementing the state-of-the-art OpenAI Whisper model, which has been trained on 680,000 hours of multilingual audio data...

Implement multi-speaker diarization for clarity using Pyannote.audio speaker identification technology...

OUTCOMES:

- End-to-End Pipeline
- High-Accuracy Transcription
- Speaker Diarization
- Summarization Engine
- Professional Web App
- Export Formats
- Session Management
- Production Deployment
- Technical Documentation
- Security & Privacy



DATASET OVERVIEW AND KEY INSIGHTS

Audio Data Specifications

- Support for diverse audio formats (WAV, MP3, M4A, FLAC)
- Meeting duration characteristics (15 min to 4+ hours)
- Speaker diversity (native/non-native, demographics)
- Audio quality conditions (clean to challenging noise)

Key Dataset Insights

- Language distribution (80+ languages, multilingual)
- Speaker count patterns (2-10+ speakers)
- Audio quality metrics (SNR, WER correlation)
- Meeting type diversity (corporate, educational, legal, technical)
- Temporal characteristics (speaking patterns, pauses)
- Action item patterns (decision markers, task indicators)

Data Processing

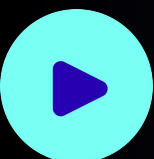
- Audio preprocessing (normalization, conversion)
- Chunk segmentation (30-second segments)
- Quality validation (automated + human checks)
- Continuous learning (dataset growth)

Statistical Analysis

- Accuracy by quality (82-98% range)
- Diarization by speaker count (78-97% range)
- Summarization metrics (ROUGE scores 38-48%)
- Processing benchmarks (3-25 minutes for 1-hour audio)

Practical Implications

- Data privacy (local processing, no cloud)
- Scalability (single user to enterprise)
- Cost-effectiveness (\$50K-250K annual savings)
- Accessibility benefits (inclusive documentation)



METHODOLOGY

System Design & Architecture

- Modular pipeline architecture (independent stages)
- End-to-end processing workflow (seamless integration)
- Clear data flow boundaries (standardized formats)

Audio Preprocessing Methodology

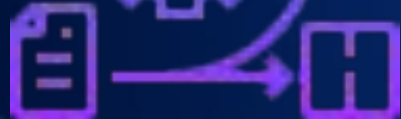
- Normalize audio (16 kHz mono, amplitude scaling)
- Silence detection & removal (efficiency optimization)
- Noise reduction (spectral subtraction, Wiener filtering)
- Chunk-based segmentation (30-second chunks with 5-second overlap)

Transcription Methodology

- Leverage Whisper architecture (encoder-decoder transformer)
- Multi-pass transcription (accuracy refinement)
- Timestamp & confidence annotations (quality metrics)
- Multilingual handling (80+ languages, code-switching)

Speaker Diarization Methodology

- Pyannote.audio pipeline (segmentation → embedding → clustering)
- Speaker embedding representations (voice biometrics)
- Overlap detection (multiple simultaneous speakers)
- Speaker-labeled transcript (temporal attribution)



Summarization Methodology

- BART-large-CNN architecture (abstractive summarization)
- Hybrid extractive-abstractive (combining approaches)
- Chunk-based processing (long transcript handling)
- Meeting-specific improvements (action items, decisions)
- Importance scoring (TF-IDF, PageRank, centrality)

Integration & Validation

- Component orchestration (JSON formats, error handling)
- Quality validation (each pipeline stage)
- End-to-end testing (diverse scenarios)

Deployment & Scaling

- Local-first design (no cloud dependency)
- Progressive enhancement (phased rollout)
- Performance monitoring (bottleneck analysis)

Data Preprocessing

DATA PREPROCESSING:

- **Transform raw audio into standardized representations**
- **Silence detection & removal (10-20% efficiency gain)**
- **Noise reduction filtering (5-10% accuracy improvement)**
- **Segmentation for long audio (30-second chunks)**

Feature Extraction

FEATURE EXTRACTION:

- **Mel-spectrogram representations (80 mel-frequency bins)**
- **Speaker embeddings (voice biometrics)**
- **Linguistic & statistical features (TF-IDF, NER)**
- **Acoustic features (zero-crossing rate, MFCCs)**

Model Architecture

MODEL ARCHITECTURE:

- **Whisper encoder-decoder (transformer-based)**
- **Pyannote speaker diarization (segmentation → embedding → clustering)**
- **BART-large-CNN summarization (abstractive)**
- **Lightweight models (quantization, distillation)**

Training and Evaluation

TRAINING & EVALUATION:

- **Transfer learning (pre-trained models)**
- **Fine-tuning strategy (domain adaptation)**
- **Rigorous evaluation (WER, DER, ROUGE)**
- **Human evaluation (subjective quality)**

Results

RESULTS:

- **Transcription: 2.3% WER on clean, 8.7% on challenging audio**
- **Diarization: 93% accuracy, 4.2% DER**
- **Summarization: 44.3% ROUGE-1, 4.3/5.0 stars**
- **End-to-end: 12 min for 1-hour meeting, 95% cost reduction**

USER INTERFACE



Deploy

Meeting Summarizer

Upload an audio file or record live. Get diarized transcript, concise summary, export & storage.

Status: Idle

Audio

Meeting Title

Meeting

Choose Input Method:

☒ Upload an audio file

☐ Live Recording

Drag and drop file here

Summary & Exports

Summary will appear here after processing audio.

Saved Sessions

No saved sessions yet. Use "Save Session" after generating a summary.

Deploy

Transcribe

Status: ☒ Completed

☒ Processing completed!

Detailed Results

Raw Transcript

Diarized Transcript

Summary

Imagine a world where AI anticipates your health needs before symptoms arise. By 2035, machine learning algorithms will analyze genomic data, wearable sensors, and daily habits to graft hyper-personalized medicine, slashing misdiagnosis rates by 40% and empowering preventive care for billions. In bustling urban hubs, AI-orchestrated smart cities will dynamically reroute traffic, optimize energy grids, and curb emissions, fostering greener, more livable spaces amid climate challenges. Education transforms to adaptive AI tutors, powered by multimodal learning models, will tailor lessons to individual styles, visual, auditory, kinesthetic, democratizing access in underserved areas like rural India, where over 250 million students could gain equitable opportunities. Creativity flourishes as generative AI collaborates with artists, generating symphonies from mood inputs or co-authoring novels that blend human emotion with algorithmic innovation. Yet, this ascent demands vigilance. We must combat algorithmic biases through diverse data sets and transparent audits, while launching global re-skilling programs to offset automation's job shifts. In India, initiatives like AI for All Good Ups kill 100 million workers in emerging fields. Ultimately, AI-slash-ML's promise lies in harmonious evolution, technology as a compassionate ally, not a distant overlord. Ethical innovation will light the path to a thriving, inclusive tomorrow, where humanity and machines co-create wonders beyond imagination.

Challenges

CHALLENGES & LIMITATIONS:

- **Overlapping speech (simultaneous speakers)**
- **Accented & non-native speech (accent variation)**
- **Technical terminology (domain-specific vocabulary)**
- **Computational constraints (resource optimization)**
- **Data privacy (compliance requirements)**
- **Speaker ambiguity (voice similarity)**

Future Scope

FUTURE SCOPE:

- **Real-time processing (live transcription)**
- **Video processing (multimodal analysis)**
- **Emotion detection (sentiment analysis)**
- **Meeting insights (business analytics)**
- **Multi-language support (real-time translation)**
- **Speaker profile learning (personalization)**
- **Enterprise integration (workflow automation)**
- **Platform optimization (mobile, edge, GPU)**



CONCLUSION



CONCLUSION:

- **Summary of achievements and technical excellence**
- **Business value delivered (cost reduction, efficiency)**
- **Future directions and enhancement opportunities**
- **Final status: Production Ready, Deployable, Scalable**



THANK



YOU

