



National University
Of Computer and Emerging Sciences

NATIONAL UNIVERSITY OF COMPUTER AND
EMERGING SCIENCES
FAST-NUCES

Meeting Insight Generator:
An AI-Powered System for Meeting Transcription and
Analysis

A Project Report
In partial fulfillment of the requirements for the course of

Natural Language Processing

Submitted By:

Hafsa Imtiaz	(i220959)
Umer Farooq	(i221007)
Areeba Riaz	(i221244)
Zayyam Hassan	(i221247)
Muhammad Rayyan	(i221022)
Abeer Jawad	(i221041)

Instructor:
Sir Owais Idrees

28 December 2025

1 Introduction

Meetings consume over 18 hours of professional time weekly, yet ineffective documentation causes up to 50% productivity loss. Manual note-taking is time-consuming, error-prone, and fails to capture critical decisions and action items. This project presents the **Meeting Insight Generator**, an AI-powered system that automatically converts meeting recordings into structured, searchable insights.

1.1 Objectives

The system aims to achieve accurate transcription with speaker diarization, automatic extraction of topics, decisions, action items, sentiment, and summaries, semantic search across historical meetings, and scalable modular architecture with $>80\%$ extraction accuracy.

2 Literature Review

Transformer-based models have revolutionized speech recognition. OpenAI Whisper demonstrates robust performance across diverse audio conditions with on-premises deployment capability. Large language models excel at understanding conversational context for extracting structured information from informal dialogue. Semantic search using vector embeddings (FAISS, sentence-transformers) enables contextual retrieval beyond keyword matching. Multi-agent architectures improve modularity by distributing tasks among specialized agents.

3 System Design and Methodology

The system follows a three-tier architecture: React frontend for user interaction, FastAPI backend orchestrating transcription and analysis, and hybrid storage (PostgreSQL for structured data, FAISS for vectors, filesystem for audio). The processing pipeline comprises six stages: (1) file upload with validation, (2) Whisper transcription with PyAnnote speaker diarization, (3) sentence-transformers generating 384-dimensional embeddings, (4) five specialized agents processing in parallel (Topic, Decision, Action Item, Sentiment, Summary), (5) PostgreSQL storage with vector indexing, and (6) JSON API output with semantic search.

The multi-agent architecture uses LangChain orchestration with 60-second timeouts, error handling with three retries using exponential backoff, and rule-based fallbacks ensuring graceful degradation.

Technology Stack: FastAPI, Whisper large-v3, PyAnnote, LangChain, Mistral AI/Groq, sentence-transformers, FAISS, PostgreSQL, React/TypeScript, Docker.

4 Results and Evaluation

A 30-minute meeting processed in 14 minutes extracted 6 topics, 4 decisions with participants, 12 action items (8 with assignees, 4 with deadlines), sentiment analysis (68% positive/24% neutral/8% negative), and a 4-paragraph summary.

Table 1: Performance Metrics - Achieved vs. Targets

Metric	Target	Achieved	Performance
Transcription WER	<15%	8.3%	45% better
Decision Extraction	>80%	85% recall	Exceeds
Action Item Extraction	>80%	87% recall	Exceeds
Agent Execution	<30s	18.5s avg	38% faster
Search Query	<500ms	180ms	64% faster
Test Coverage	>80%	82%	Meets

Agent Quality: Topic Agent achieved 89% boundary accuracy, Decision Agent 85% recall/82% precision, Action Item Agent 87% recall/84% precision, Sentiment Agent 79% human agreement, Summary Agent 4.2/5 rating.

System Performance: Parallel execution processes at 35 seconds per audio minute (31% faster than sequential). Peak CPU 78%, memory 6.2GB. Successfully handled 5 concurrent meetings with 15% time increase.

Testing: Achieved 81.15% code coverage across 1,777 statements with 211 tests passing. Pyramid approach included unit tests (90%+ agent coverage), integration tests, end-to-end workflows, and performance tests.

5 Limitations

The system supports batch processing and English only. Transcription accuracy depends on audio quality. External LLM API reliance introduces availability constraints. Real-time processing and mobile support are outside current scope.

6 Conclusion

The Meeting Insight Generator successfully delivers an integrated AI system transforming meeting recordings into actionable insights. Key achievements include 8.3% WER with 91% speaker accuracy, >80% extraction precision for decisions and actions, 92% search relevance, and 82% test coverage. The system eliminates manual documentation (15-30 min/meeting) through full automation while enabling semantic search across meeting history. The modular multi-agent design ensures scalability, robustness, and extensibility for modern organizations.

References: Radford et al. (2022) Whisper; Jiang et al. (2023) Mistral 7B; Johnson et al. (2019) FAISS; Reimers & Gurevych (2019) Sentence-BERT; Lewis et al. (2020) RAG.