**Thesis Structure**

**1. Introduction**

**├─ 1.1 Background & Motivation**

**├─ 1.2 Problem Statement**

**├─ 1.3 Research Questions**

**├─ 1.4 Scope & Contribution**

**└─ 1.5 Thesis Structure Overview**

**2. Theoretical Background**

**├─ 2.1 Retrieval-Augmented Generation (RAG)**

**│ ├─ RAG Architecture**

**│ ├─ Embedding vs Generation**

**│ └─ Indexing and Retrieval Flow**

**├─ 2.2 Large Language Models (LLMs)**

**│ ├─ GPT-3 / 3.5 / 4 / 4o**

**│ └─ Token windows & cost models**

**├─ 2.3 Chunking Strategies**

**│ ├─ Fixed-length**

**│ ├─ Recursive**

**│ ├─ Document-based**

**│ └─ Semantic**

**├─ 2.4 Vector Databases**

**│ ├─ Pinecone**

**│ └─ PostgreSQL (pgvector)**

**├─ 2.5 LangChain Overview**

**└─ 2.6 Speech-to-Text (STT)**

**├─ Whisper (OpenAI)**

**├─ Google STT API**

**├─ Vosk / Mozilla DeepSpeech**

**└─ Use Cases in RAG**

**3. Literature Review**

**├─ 3.1 RAG Systems in Academia**

**├─ 3.2 Hybrid Systems Using Multi-Modal Inputs (HTML and Speech)**

**├─ 3.3 Comparative LLM Performance**

**└─ 3.4 Metadata & Conversational Context Handling**

**4. Methodology**

**├─ 4.1 System Design Overview**

**├─ 4.2 PDF and HTML Document Pipeline**

**├─ 4.3 Audio-to-Text Pipeline (STT)**

**│ ├─ Audio Cleaning & Preprocessing**

**│ ├─ Transcription & Timestamping**

**│ └─ Chunking of Transcript**

**├─ 4.4 Embedding & Vector Storage (Postgres)**

**├─ 4.5 RAG Query Flow**

**└─ 4.6 Prompt Design & Conversation Memory**

**5. Implementation**

**├─ 5.1 Backend Stack (Python + LangChain)**

**├─ 5.2 STT Module**

**├─ 5.3 PostgreSQL + pgvector Setup**

**├─ 5.4 Document & Transcript Upload Interface**

**└─ 5.5 Query and Answer Flow**

**6. Evaluation & Results**

**├─ 6.1 Accuracy on Document-Based Q&A**

**├─ 6.2 Accuracy on Transcribed Speech Q&A**

**├─ 6.3 Time-to-Response & Relevance**

**├─ 6.4 Comparison: STT+RAG vs Text-only RAG**

**└─ 6.5 Expert Feedback / Usability**

**7. Discussion**

**├─ 7.1 Challenges in Audio/Document Chunking**

**├─ 7.2 Cost vs Accuracy Trade-offs**

**├─ 7.3 Conversational Context Handling**

**└─ 7.4 Potential for Multimodal Integration**

**8. Conclusion & Future Work**

**├─ 8.1 Summary of Findings**

**├─ 8.2 Lessons Learned**

**└─ 8.3 Future Enhancements**

**├─ Real-time STT Integration**

**├─ TTS / Speech Output**

**├─ Multi-Language Support**

**└─ Legal/Medical Domain Adaptation**

**9. References**

**10. Appendices**

**├─ A. Architecture Diagrams**

**├─ B. Prompt Templates**

**├─ C. Sample Transcripts**

**└─ D. Evaluation Logs**