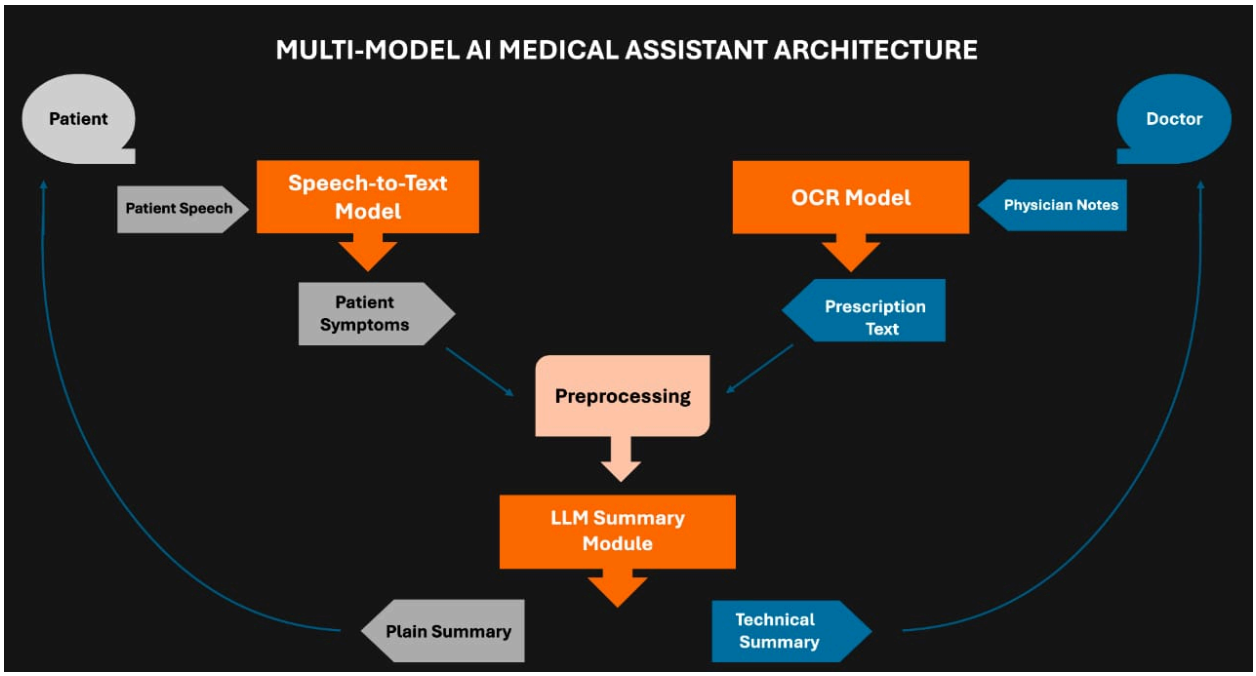
Milestone 3: Model Architecture



1. Input Modules (Speech-to-Text, OCR, Image Processing):

- **Modality Handling:** Medical data inherently comes in various forms (spoken dialogue, handwritten notes, images). Dedicated modules for Speech-to-Text, OCR, and Image Processing are crucial to accurately ingest and convert these raw, unstructured inputs into a format that downstream AI models can process. This addresses the real-world complexity of medical records.
- **Specialized Tools:** Using specialized tools (e.g., ASR for speech, OCR for handwriting) ensures higher accuracy for each specific data type compared to a general-purpose approach. These modules often incorporate domain-specific knowledge or are fine-tuned for medical contexts.

2. Data Integration & Preprocessing (Doctor Notes Aggregation):

- **Unification of Information:** Medical diagnoses and summaries require a holistic view of patient data. This module integrates information from disparate sources (transcriptions, digitized notes, image insights) into a coherent, structured representation.

- **Knowledge Extraction (NER, Relation Extraction):** Simply converting to text isn't enough. Medical information is rich with specific entities (diseases, drugs, body parts) and relationships between them (e.g., "drug X treats disease Y"). NER and Relation Extraction are vital for extracting this structured knowledge, which is essential for accurate and clinically relevant summarization.
- **Standardization (Text Normalization):** Medical terminology can vary. Normalization to standard ontologies ensures consistency and reduces ambiguity, improving the quality of the input to the summarization model.

3. LLM Summarization Module:

- **Advanced Language Understanding and Generation:** Large Language Models (LLMs) excel at comprehending complex text, identifying key information, and generating coherent, contextually relevant summaries. This makes them ideal for condensing extensive medical notes.
- **Adaptability to Summarization Tasks:** LLMs can be fine-tuned specifically for summarization tasks, allowing them to learn the nuances of medical discourse and the requirements for both concise and detailed summaries.
- **Multimodal Capability:** The optional multimodal integration allows the LLM to incorporate visual evidence directly into its understanding and summarization, leading to more comprehensive and accurate outputs, especially when images contain critical diagnostic information.
- **Dual Output (Plain Language & Technical):** This is a critical design choice driven by the need to serve different audiences. Patients require clear, jargon-free explanations, while medical professionals need precise, clinically rich details. An LLM can be prompted or fine-tuned to generate these distinct summaries from the same core information.

4. Post-processing & Output (Summary Refinement):

- **Quality Assurance:** Even advanced LLMs can hallucinate or produce inconsistent information. A refinement step adds a crucial layer of quality control, ensuring the summaries are factually accurate, consistent with medical guidelines, and appropriately toned for their intended audience (e.g., readability scores for patient summaries).

- **Safety and Compliance:** In medical applications, accuracy and safety are paramount. This step allows for a final check against potential errors or misleading information before dissemination.