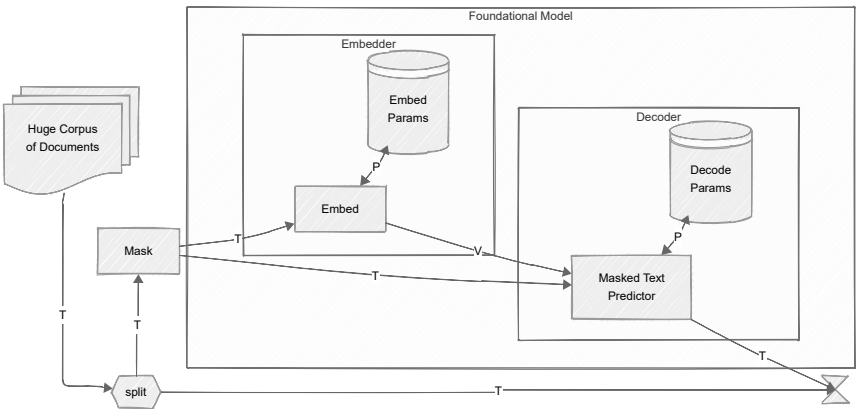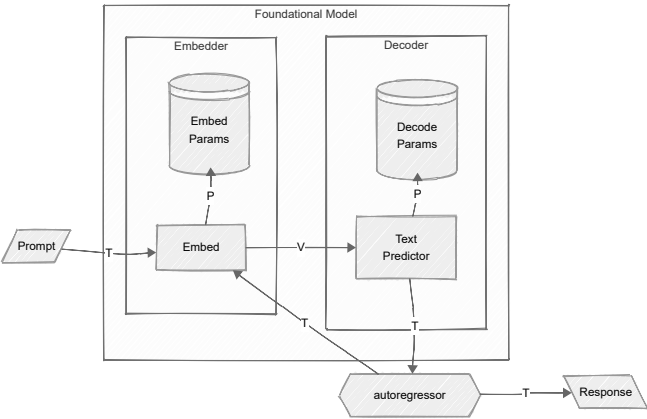# BUSI/COMP 488 Class12: Foundation Models & Retrieval Augmented Generation (RAG)
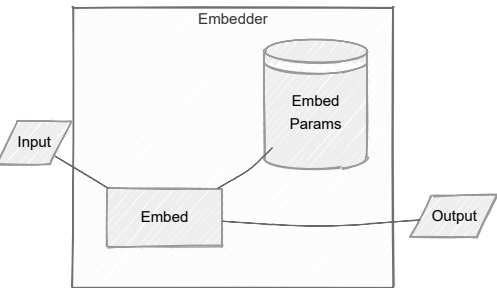
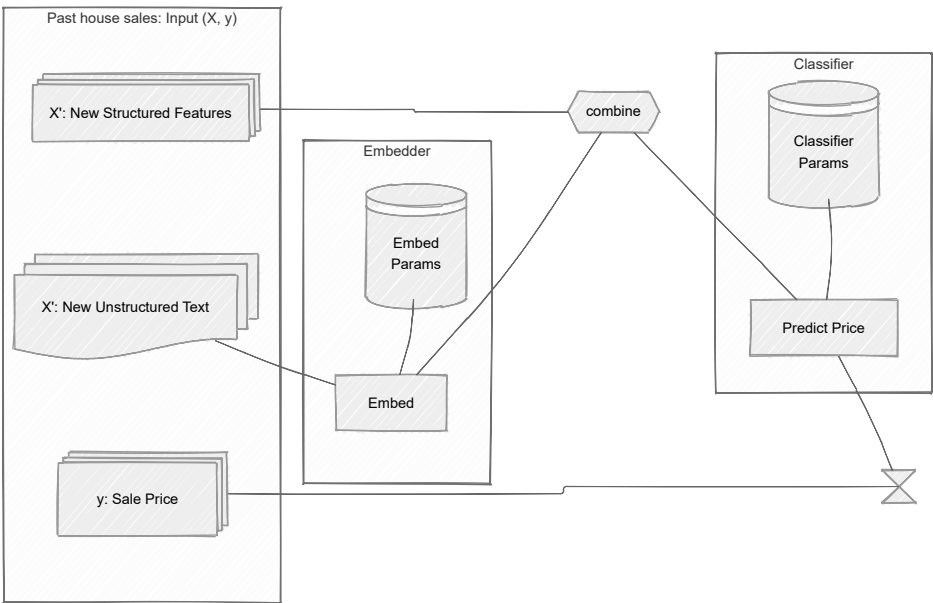## 3.3.1 Foundation Model training

3.3.3 FM generation

abel the lines and draw in arrowheads to identify how data flows in production use (after training).
* Use type abbreviations: T=text/token, V=vector, P=model parameters, or write out types.
* Add arrowheads where data flows to double headed arrows are possible.
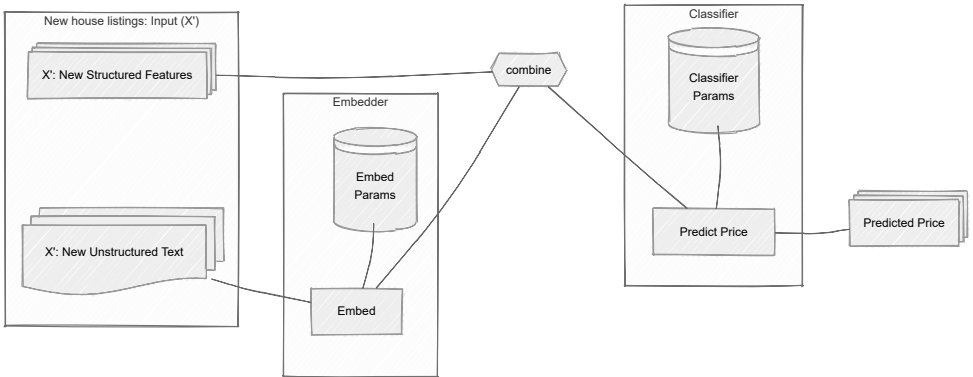* After you've added your labels, click (▶) in the notebook to compare with mine.
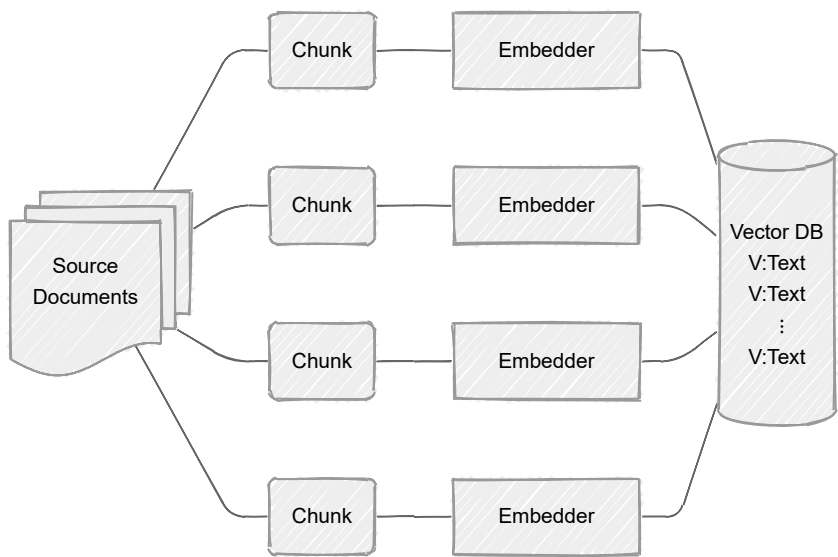
3.3.2   mbedder from FM
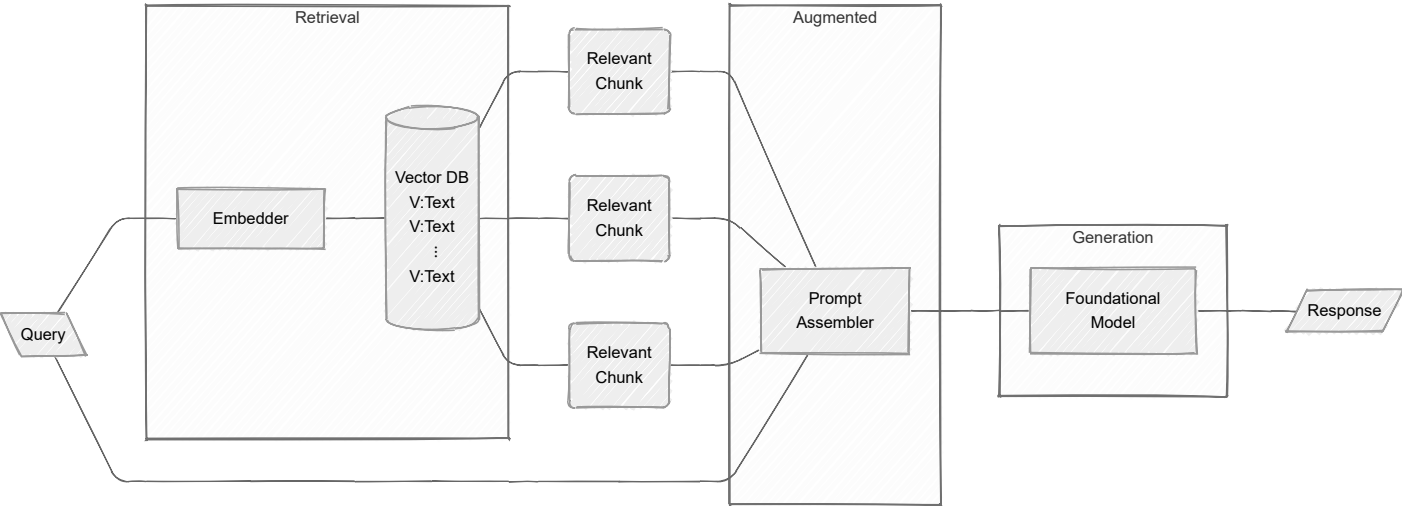

3.3.4 House Price prediction training

3.3.4 (cont) House Price Prediction

4.2.1 RAG preparation: Add arrow directions and labels (T=text/token, V=vector, P=model parameter)

Source Documents → Chunk → Embedder → Vector DB V:Text V:Text ⋮ V:Text

Chunk → Embedder

Chunk → Embedder

Chunk → Embedder

4.2.2 RAG generation

Retrieval

Query → Embedder → Vector DB V:Text V:Text ⋮ V:Text → Relevant Chunk

Relevant Chunk

Relevant Chunk

Augmented

Prompt Assembler

Generation

Foundational Model → Response

4.4 CrewAI example of AI Agents

Query

Prompt: scrape relevant text return insights as narratives

Prompt: Find relevant URLs since 2024

ResearchAgent

Search builder Fnd Model

Tool: URLS from DuckDuckGo

AnalystAgent

Web page gleaner FM

Tool: Scrape from URL

Prompt: Write article

JournalistAgent Fnd Model

Prompt: linked-in post

BloggerAgent Fnd Model