



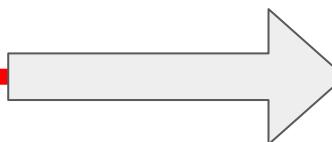
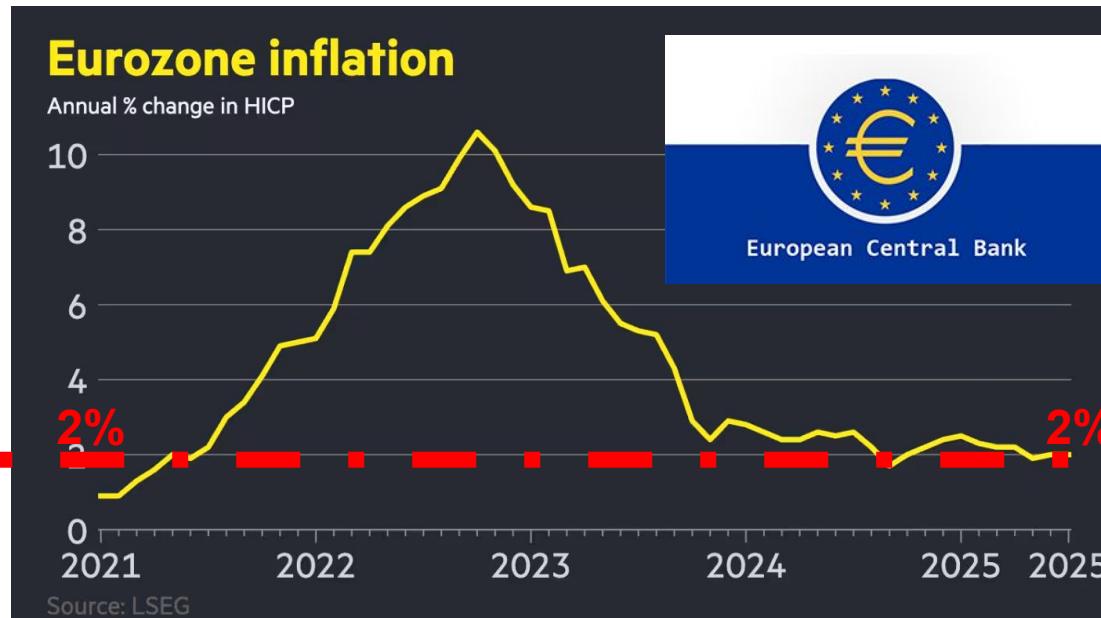
# AURA

RAG for VALUE INVESTING

ADRI

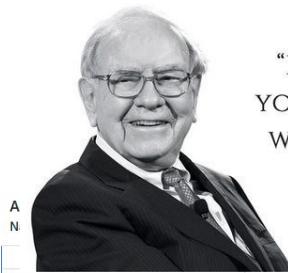
# WHY AURA?

AI tool designed for value investors



The **cumulative inflation** over the last 10 years was  
**21.9%**

# WHY AURA? AI tool designed for value investors



"PRICE IS WHAT  
YOU PAY. VALUE IS  
WHAT YOU GET."  
— WARREN BUFFETT —

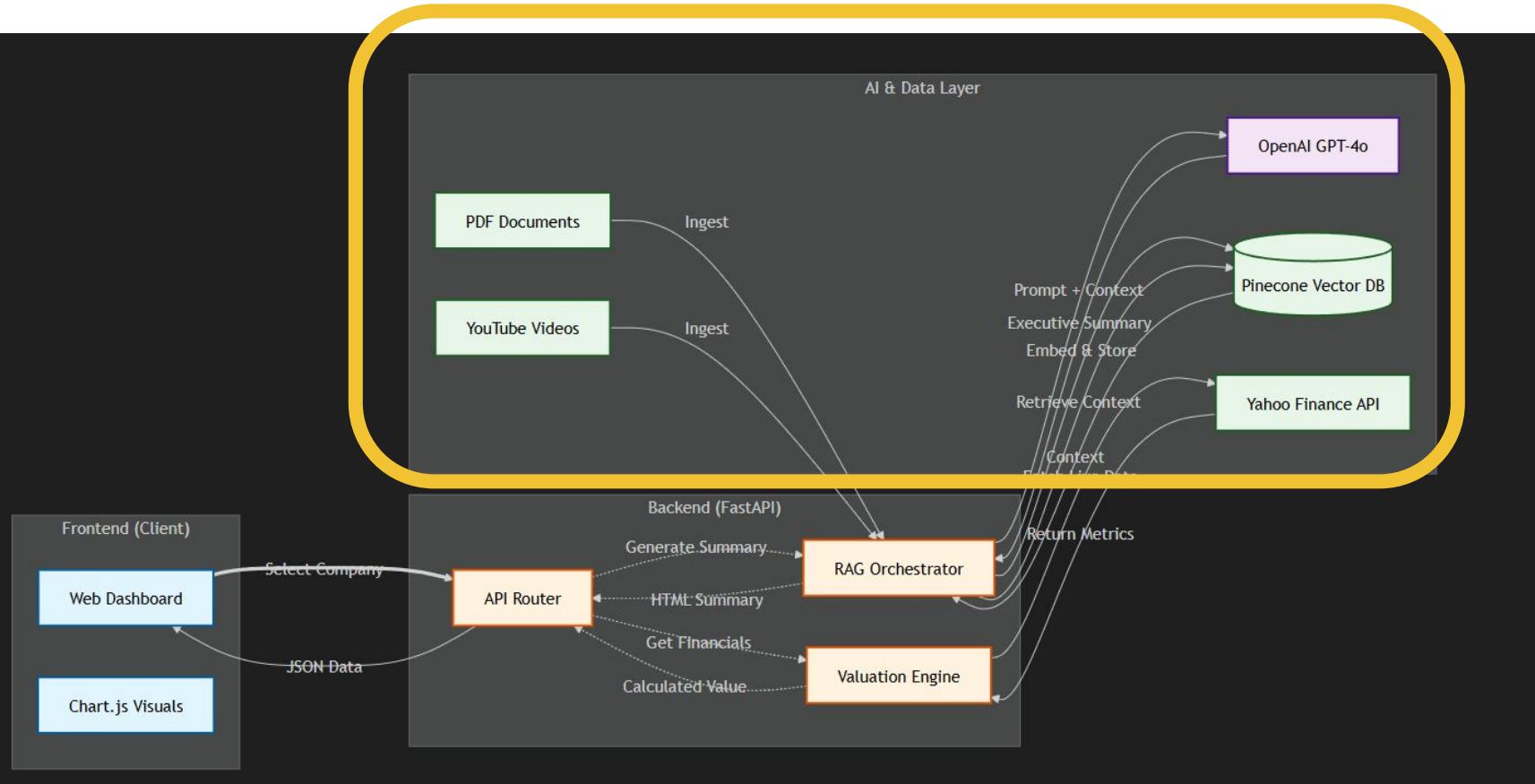


**Value investing** requires extensive time analyzing a company's fundamentals (income, balance..?) and understanding its competitive moat

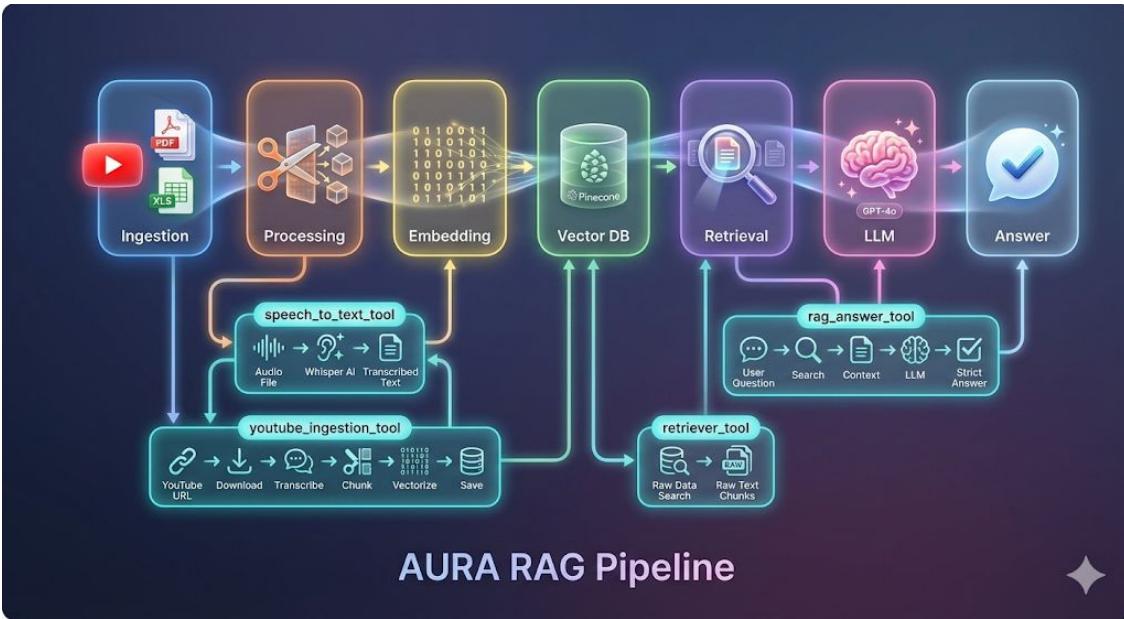


**AURA**  
turns deep financial analysis into simple, actionable insights—instantly

# System Architecture



# System Architecture: RAG



## Deployment:

- **Environment:** The application currently runs locally PARTIALLY on a development machine.
- **Server:** Hosted on localhost:8000 using **Uvicorn**.

## Step 1 — Ingestion (How AURA “reads”)

PDFs: HunyuanOCR

YT Videos( 20 videos of 15-30 min):  
OpenAI Whisper

## Step 2 — Indexing (How AURA “remembers”)

*Embedding: OpenAI  
ext-embedding-3-small*

*Storage: Pinecone*

## Step 3 — Retrieval (How AURA “thinks”)

retrieves the most relevant chunks using cosine similarity.

## Step 4 — Generation (How AURA “answers”)

We send the question and retrieved text to GPT-4o with strict instructions:

# DATASET

## Data Sources

AURA integrates **two major unstructured datasets** to understand a company both quantitatively and qualitatively:



### Financial Documents (PDFs)



### YouTube Videos

## Retrieval :

- **LangChain PyPDFLoader** → Extracts the main textual content reliably and efficiently.

## B. Video Processing

- **YoutubeLoader** → Retrieves subtitles directly *when available* (fastest method).
- **If subtitles are not available:**
  - **yt-dlp** downloads the audio track. (*Current bottleneck / issue to optimize.*)
  - **Whisper (local model)** transcribes the audio into high-quality text.

# DATASET

## Preprocessing & Embedding

### Preprocessing

- Cleaning: remove special characters/normalizing
- **Chunking:** RecursiveCharacterTextSplitter
  - Chunk size: **1000 chars**
  - Overlap: **100 chars**

### Embedding

- Model: **OpenAI text-embedding-3-small**
- Vector dimension: **1536**

### Storage in the Vector Database (Pinecone)

#### Index: youtube-and PDF rag-index

- Metric: **Cosine Similarity**
- Stored elements:

#### Metadata:

- **text** → raw chunk text
- **source** → filename or YouTube URL
- **page** → page number (PDFs)
- **title** → YouTube title
- **type** → "pdf" or "video"

# Q&A Agent

## General Architecture

AURA's Q&A system is built with **LangChain**.

- **LLM Engine:** OpenAI **GPT-4o**, Temperature = 0 (deterministic, factual) **MANY PROBLEMS WITH LIGHT MODELS**
- **Framework:** LangChain (orchestrator) + LangSmith (used for tracing and debugging)

## Agent Tools & Capabilities:

 **rag\_answer\_tool — The Core Brain**

 **youtube\_ingestion\_tool**

Ingests new YouTube videos directly from a URL and adds them to the knowledge base.

 **speech\_to\_text\_tool:** Transcribes uploaded audio or voice notes into text, enabling fully voice-based interaction.

 **retriever\_tool**

 **Yahoo Finance (yfinance):** Provides all real-time financial metrics

 **LangChain Framework**

 **LangSmith Tracing & Evaluation**

# Q&A Agent

## Memory System:

- **Memory Type:** ConversationBufferWindowMemory
- **Window Size (k = 5):** Stores last 5 interactions

## Prompt Engineering:

To prevent hallucinations and ensure the bot acts as a "Value Investing Expert"

```
# Strict System Prompt
template = """You are a specialized assistant for "Invertir Desde Cero" YouTube videos.
Answer the question strictly based on the following context from the video captions.
Do NOT analyze the video itself, only the captions provided in the context.
Do not use outside knowledge.

If the answer is not contained in the context, you must answer EXACTLY:
"This information does not appear in the Invertir Desde Cero videos."

Here are some examples of how you should answer:

Example 1:
Context: "In this video, we discuss that Apple's competitive advantage is its ecosystem."
Question: "What is Apple's competitive advantage?"
Answer: "Apple's competitive advantage is its ecosystem."

Example 2:
Context: "We are analyzing Tesla's production numbers for 2023."
Question: "Who is the president of France?"
Answer: "This information does not appear in the Invertir Desde Cero videos."

Example 3:
Context: "Warren Buffett advises to buy businesses with a moat."
Question: "What does Warren Buffett advise?"
Answer: "Warren Buffett advises to buy businesses that have a moat.

Now, answer the following question based ONLY on the context provided below.

Context:
{context}

Question:
{question}

Answer:"""
```

# Evaluation

To validate the quality of AURA's Q&A agent:

- **Judge Model:** GPT-4-Turbo
- **Goal:** Ensure answers are **faithful, correct**, and based on **relevant context** retrieved from YouTube transcripts.

**Results** from the Validation Run:

- Sample Size: 3 Test Cases.
- Average Latency: 5.86 seconds.
- Correctness Score: 0.33 (Strict).

VALIDATION AURA-2996500f		VALUE .../d32f7a	Compact	Full	Diff	Default	Group by	Display	Compare
Inputs	Reference Outputs	Outputs	Correctness	Doc_releva...	Faithfulness...	Latency	Completed		
What does Warren Buffett... #5d1e →	Warren Buffett advises to buy business...	Warren Buffett advises that if inve... ⏪	0.3333 AVG ↓	0.3333 AVG ↓	0.6667 AVG ↓	5.406 P50 ⏪			
Who is the president of F... #d868 →	This information does not appear in the...	The documents provided do not c... ⏪	1.00	0.00	1.00	7.19s	COMPLETED		
What is Apple's competit... #ecb0 →	Apple's competitive advantage is its ec...	The provided documents do not c... ⏪	0.00	0.00	1.00	5.41s	COMPLETED		

# Conclusion

Thank you very much

Muchas gracias

Muito obrigado

Moltes gràcies

Vielen Dank

Ευχαριστώ πολύ

(Efharistó polí)

Хвала вам пунно

(Hvala vam puno)

बहुत धन्यवाद

(Bahut dhanyavaad)

AURA turns complex financial analysis into simple, fast insights.

- In the future:

- I want to add a **psychological analysis** tool that measures market sentiment, including indicators like the **Fear Index and FOMO**.
- Provide information about all worldwide companies