# Real-Time CNBC Audio Parsing Pipeline (Future Reference)

## ✅ Overview

Goal: Monitor live CNBC audio during market hours, transcribe it, and feed it into MarketPlayground's news parser pipeline.

## Step 1: Live TV Stream Access (CNBC)

- Requirement: Legal access to CNBC’s stream (via Sling TV, YouTube TV, etc.)  
- Tool: Use `ffmpeg` to extract audio from live video streams in real time.

## Step 2: Real-Time Audio Transcription

- Tool Options:  
 • Open-source: Whisper (by OpenAI) — high quality, runs locally  
 • Paid APIs: AssemblyAI, Deepgram, Google Speech-to-Text — offer streaming support  
- Feed audio from ffmpeg to the transcription engine in real time.

## Step 3: Text Cleanup + Parsing

- Remove filler words, banter, and non-market commentary  
- Use regex or LLMs to chunk text into meaningful sections  
- Prepare each chunk as a belief input

## Step 4: Feed Into MarketPlayground News Parser

- Treat each text chunk like a news article:  
 • Detect tickers, sentiment, themes  
 • Pass into belief-to-strategy pipeline  
 • Optionally weight confidence by speaker importance

## 🧠 Optional Enhancements

- Speaker diarization (identify individual speakers)  
- Emotion/tone detection in voice  
- Visual + audio (multimodal) AI integration

## ⚠️ Legal & Technical Notes

- Make sure you have legal rights to access and process the stream  
- Real-time transcription requires GPU or paid API access  
- CNBC contains a lot of noise — strong filtering will be critical

Save this for the future. When ready, revisit to prototype it.