Electrify | Applied AI Engineer | Take-Home Task

Goal

Build a simple agentic workflow that **learns what high-performing video titles look like** (for a given YouTube channel) and then generates **3–5 titles for a new video idea**, each with **brief, data-grounded reasoning**.

Deliverables

1. A **FastAPI endpoint** (runnable locally) with the following request parameters:
    a. `channel_id` (str): The unique identifier of the YouTube channel.
    b. `idea` (str): A single-sentence summary of the YouTube video to generate titles for.

    We'll leave the exact response structure up to you, but the endpoint should return **3–5 titles** and **per-title reasoning** grounded in discovered patterns.

2. A **README** with setup & run instructions and a short "Architecture" section (which provides the high-level design of your workflow).

Training data

Link to training data

The training data is a CSV containing video-level metadata & performance data for 3 YouTube channels, with the following columns:

- `channel_id` (str): The unique identifier of the YouTube channel
- `video_id` (str): The unique identifier of the YouTube video
- `title` (str): The title of the YouTube video
- `summary` (str): The single-sentence summary of the YouTube video
- `views_in_period` (int): The number of views accrued per video in the last 28 days (this is metric you should use to measure the performance of a particular title)

You should use this data to identify the features of high-performing titles; how you do this is up to you.

Evaluation

You will have **1 week to work on this task** and it should take **no longer than 4 hours of development time**; if you reduce the scope of your solution, note any trade-offs.

You will present your solution to Electrify's Director of Data & Technology & Senior Technical Advisor in a 1 hour meeting. During this meeting, we will provide you with **three unseen test cases (one for each of the channels included in the training data)** which correspond to new video ideas; you'll use your solution to provide new title suggestions for these three unseen test cases.

We will also ask you to **add a small feature to your solution in real-time.**

Considerations

- Your approach should **generalise to any channel** as long as we have similar performance data. Please avoid channel-specific hardcoding.
- Your solution should be **runnable by anyone (locally on their machine)**. Please bear this in mind when setting up your project.
- **Do not use external data/scraping.** Use only the provided CSV.
- You can use either **external LLM APIs** (e.g. OpenAI, Anthropic, etc) or **locally run LLMs.**
- Please ensure your code is **pushed to Github prior to the meeting.**