

Take-Home Project: Long-Video Human Motion Analysis Agent

We're interested in how you **think about designing and building production-grade agentic systems** for long-form video understanding.

Imagine you're building a system that analyzes **videos of human physical activity** (sports training, posture correction, gymnastics etc.).

A user might ask:

- “Analyze the takeoff technique in this vault.”
- “Is this squat safe according to coaching standards?”

Your task is to **design and implement an agent** that can meaningfully answer questions like these.

This is intentionally **open-ended**. We care more about your **system design decisions** than about completeness or polish. Feel free to choose any type of human physical activity

What to Build (High Level)

- An **agentic system** that:
 - takes a long video (tens of minutes is fine),
 - accepts a natural-language question,
 - decides *what parts of the video to look at*,
 - extracts meaningful motion or pose information,
 - reasons about correctness / safety / technique,
 - and produces a grounded, explainable answer.
- You are **encouraged** to use:
 - open-source models
 - Agentic frameworks(OpenAI, Langchain,etc)
 - fine-tuned models
 - custom heuristics
 - LLMs, vision models, or hybrids

There is no “correct” architecture — we want to see *yours*.

What We're Looking For

We're especially interested in how you think about:

- **Separating perception from reasoning**
- Handling **long videos efficiently** (not brute force)
- Converting raw motion into **interpretable signals**
- Grounding answers in **evidence** (timestamps, metrics, references)
- Using **agents, tools, or planning** rather than a single monolithic model
- Making design tradeoffs explicit

You don't need to solve everything — depth in a few places beats shallow coverage everywhere.

Deliverables

- A **GitHub repository** with:
 - runnable code (local setup is fine),
 - a short README explaining your system and design choices.
- A **demo video** (5–10 minutes):
 - walk through the system,
 - show at least one example query,
 - explain *why* you made certain design decisions.

That's it. No strict API requirements, no UI expectations.