

# ***Stampli ML Engineer Home assignment***

**Background**

Disney wants to leverage AI to enable customer experience teams to query customer reviews in natural language. Your role is to design and implement a minimal, production-oriented prototype that demonstrates how this could work in practice. You are provided with a dataset of customer reviews about their parks (disney-reviews.zip).

**Task**

**1. System Design & Architecture**

* Propose an end-to-end architecture for the solution.
* Cover data flow, model serving (LLM or RAG), and performance considerations.
* Include monitoring components (logging, metrics, alerts).
* **Deliverable**: a diagram + short explanation of tradeoffs.

**2. Prototype Implementation**

* Build a minimal service (FastAPI/Flask) that answers natural language questions about the dataset (Use both review text and metadata).
* Leverage the provided 4o-mini LLM.
* Add simple logging (e.g., request counts, latency).

**3. Monitoring & Validation**

* Define key metrics to track (system latency, throughput, error rate; model drift/quality).
* Implement at least one automated test.
* Propose how you would evaluate quality and handle scale.

**Example Questions**

Your system should be able to handle queries like:  
- What do visitors from Australia say about Disneyland in Hong Kong?  
- Is spring a good time to visit Disneyland?  
- Is Disneyland California usually crowded in June?  
- Is the staff in Paris friendly?

**Deliverables**

* Code for the prototype (API + minimal tests).
* Architecture diagram + explanation.
* Short writeup or slides describing architecture, performance, and monitoring.

**Evaluation Criteria**

We will assess candidates based on:

* Architecture Thinking.
* Implementation Skills.
* Performance & Monitoring Awareness.
* Validation Strategy.
* Communication.