**Executive Summary / Cover Letter**

**Project:** Montage — AI Agent for Event Photo Curation & Instagram Publishing  
**Author:** Prem Kondru, BTech Engineering Physics, IIT Guwahati (Photography Club: Montage)

**Purpose & Manual Task**

At IIT Guwahati’s Photography Club (Montage), every event generates hundreds of photos that must be **deduplicated, clustered by moment/theme, captioned in a consistent club style, and packaged as Instagram carousels**. This is a repetitive, time-consuming manual workflow.  
**Montage** automates this end-to-end: **ingest → dedupe → cluster → caption → export**.

**How the Agent Reasons, Plans, and Executes**

* **Reasoning:** Uses CLIP embeddings to assess visual similarity, remove near-duplicates, and infer high-level labels (zero-shot).
* **Planning:** A lightweight **Supervisor (Planner)** orchestrates a fixed, reliable pipeline: Ingest → Embed → Dedupe → Categorize → Cluster → Caption → Export.
* **Execution:** Tool-style workers (Executors) run each step deterministically with clear inputs/outputs. The Streamlit UI allows human-in-the-loop edits (include/exclude photos, adjust labels) before export.

**Fine-Tuned / Parameter-Efficient Model (LoRA) — Choice & Integration**

* **Model:** BLIP-2 + Flan-T5 captioner with **LoRA** adapters (parameter-efficient finetuning).
* **Why this target:**
  + **Task specialization & style adaptation:** Our club prefers **abstract, mood-driven** captions that avoid factual claims and hashtags in body text. Full finetuning is heavy; **LoRA** lets us cheaply adapt tone and phrasing.
  + **Reliability:** We add guardrails (no proper nouns, optional *only event name* policy, length bounds), and fuse **per-image captions** into a **single cluster caption** using **common-word intersection**, which favors shared visual elements over outliers.
* **Integration:** The captioner runs in two modes—**template** (deterministic) and **blip2** (LoRA). In BLIP-2 mode we generate per-image captions in batches and **aggregate** them to one post caption. The agent also supports a **RAG** side input (nearest past captions) to keep tone/hashtags consistent.

**Evaluation: Quality & Reliability Metrics**

* **CLIPScore:** Measures caption–image alignment (per-image and cluster means).
* **Silhouette Score:** Quick proxy for clustering cohesion/separation.
* **Dedupe Rate:** % of near-duplicates removed.
* **Human Ratings (optional):** Abstractness, tone, IG readiness (Likert).
* **A/B Protocol:** Compare **template vs BLIP-2/LoRA** on the same clusters; track metrics per event and overall.  
  A full **Data Science Report (PDF)** and **Interaction Logs** (prompts + chat history) are generated to document method and outcomes.

**Optional Features (Bonus) Implemented**

* **Multi-agent collaboration:** Planner (Supervisor) + Executors (Ingestor, Embeddings, Deduper, Categorizer, Clusterer, Captioner, Exporter).
* **External integrations / custom tools:**
  + **RAG** from past\_captions.jsonl for style/hashtag hints.
  + CLIP-based tools for dedupe, zero-shot labeling, clustering, and CLIPScore.
* **User Interface:** A **Streamlit** app with image upload, runtime controls (label set, max images per post, captioner mode, preview zoom, event name), IG-style previews, thumbnail include/exclude toggles, and **JSON export** for an Instagram carousel publisher.

**Why This Meets the Objective**

* **Automates a real manual task:** Directly replaces a repetitive club workflow with an agent that **reasons** about image content, **plans** a reliable pipeline, and **executes** it end-to-end.
* **Uses a fine-tuned model:** Integrates **LoRA-adapted BLIP-2** for club-specific caption style and improved reliability.
* **Explains the fine-tuning choice:** LoRA targets **style specialization** while minimizing compute/storage; the abstract captioning policy reduces hallucinations.
* **Implements evaluation:** CLIPScore, silhouette, dedupe rate, and (optional) human ratings + A/B comparisons provide measurable evidence of quality.
* **Delivers bonus features:** Multi-agent design, RAG tooling, and a practical UI suitable for day-to-day use.

**What You Receive**

* **Working app** with a clean UI and exportable IG carousel payloads.
* **Architecture document** (components, flow, models, design rationale).
* **Data Science report** (fine-tuning setup: data, method, evaluation).
* **Interaction logs** (prompts and chat history) and a JSONL logging template.
* **Configurable controls**: label taxonomy, cluster cap, captioner mode, and event-name policy.

**In summary, Montage is a complete, defensible AI agent that automates a real university workflow, integrates a LoRA-tuned model for style-safe captioning, and ships with the metrics and documentation required to evaluate and maintain it.**