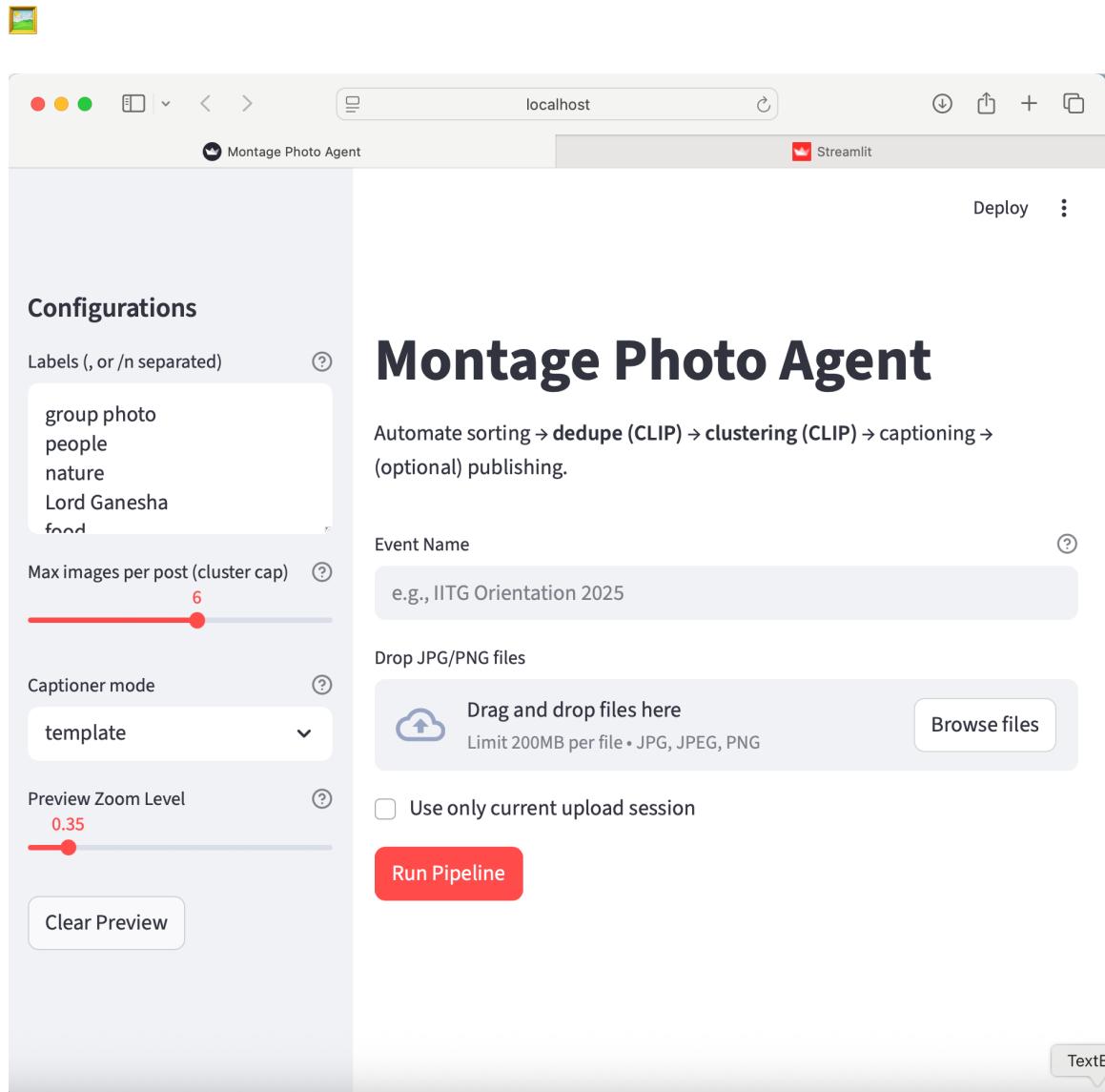


Product Screenshots Document – Montage Agent

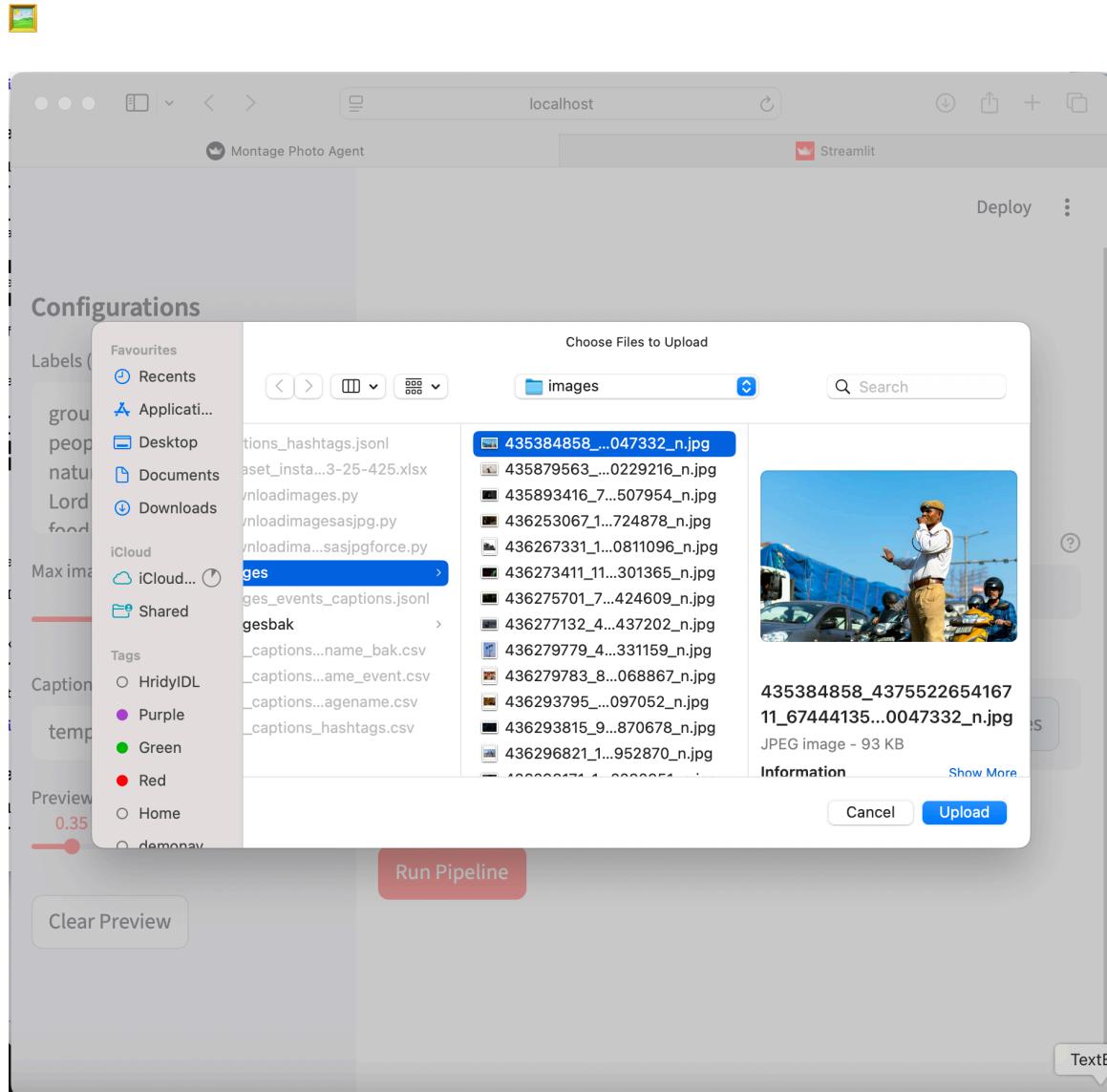
Dashboard / Home Screen



Central entry point for the app.

- Upload new images or select existing datasets.
- Quick navigation to settings and configurations

Image Upload & Ingestion



Drag-and-drop or select folders of images.

- Previews thumbnails for quick verification.

Deduplication & Quality Check

The screenshot shows a Streamlit application window titled "Montage Photo Agent" running on "localhost". The interface is divided into two main sections: "Configurations" on the left and "Pipeline step outputs" on the right.

Configurations:

- Labels (, or /n separated):** group photo, people, nature, Lord Ganesha, food
- Max images per post (cluster cap):** 6 (set via a slider)
- Captioner mode:** blip2 (selected via a dropdown)
- Preview Zoom Level:** 0.25 (set via a slider)
- Clear Preview:** A button to clear the preview area.

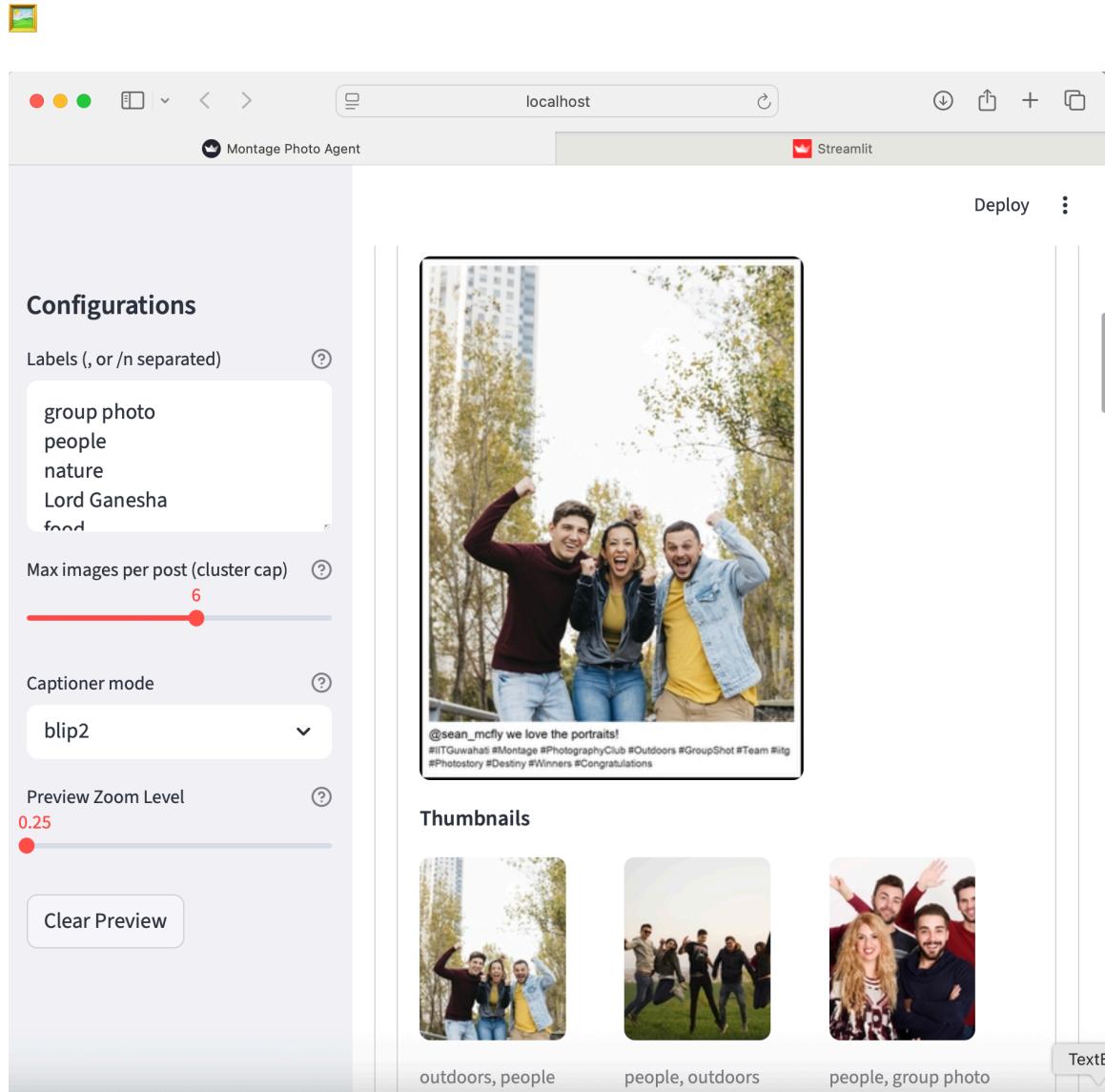
Pipeline step outputs:

- ingest:** { "n_items": 144 }
- dedupe_quality:** { "n_items": 59 }
- categorize:** { "n_items": 59 }
- cluster:** { "n_clusters": 5, "k": 5, "fused": false, "silhouette": 0.11800000071525574 }

Highlights near-duplicates flagged by CLIP embeddings

- Current accuracy: ~100% at similarity threshold 0.8.

Clustering, Captioning View



Groups images into cohesive event/story clusters.

- Displays silhouette score for cluster quality (-1 → 1).
- User can adjust K or use auto-detection.

Captioning (blip2 / Template Mode)

The screenshot shows the 'Montage Photo Agent' application running in a web browser on 'localhost'. The interface has a sidebar on the left labeled 'Configurations' containing settings for 'Labels', 'Max images per post (cluster cap)', 'Captioner mode', and a 'Clear Preview' button. The main area displays the title 'Montage Photo Agent' and a brief description: 'Automate sorting → dedupe (CLIP) → clustering (CLIP) → captioning → (optional) publishing.' It includes fields for 'Event Name' (set to 'e.g., IITG Orientation 2025') and a file upload section for 'Drop JPG/PNG files' with a limit of 200MB per file. Three files are listed: '436275701_7871263979592430_46224248622124246...', '436273411_1178651726461514_56150813931823013...', and '436267331_1987876951609257_3428810220700811...'. A message at the bottom indicates 'Showing page 1 of 3'. A green notification bar at the bottom states 'Saved 7 new file(s) to /Users/srisys/Code/MontageAgent/data/events/upload_session_175803855'.

Template Mode:

- Deterministic captions with safe, consistent wording.
- Suitable for announcements or campaign posts.
- Example: "Highlights from the Photo Exhibition."

Fine-tuned BLIP-2 + LoRA captioner Mode.

- Produces abstract, creative, Instagram-style captions.
- Example: "Frames alive with stories woven in light."
- Human reviewers rated ~4.3/5 (vs ~3.2/5 for Template).

Metrics & Evaluation (Clipscore)

The screenshot shows a Streamlit application window titled "Montage Photo Agent" running on "localhost". The interface is divided into two main sections: "Configurations" on the left and "Pipeline step outputs" on the right.

Configurations (Left Side):

- Labels (, or /n separated):** group photo, people, nature, Lord Ganesha, food
- Max images per post (cluster cap):** A slider set to 6.
- Captioner mode:** blip2
- Preview Zoom Level:** A slider set to 0.25.
- Clear Preview:** A button.

Pipeline step outputs (Right Side):

- ingest:** {...}
- dedupe_quality:** {...}
- categorize:** {...} (with a copy icon)
- cluster:** {...}
- captioner:** {...}
- clipscore:** {
 - "n_posts": 5
 - "posts": [...]
 - "label_index": {...}
 - "clipscore_mean": 0.1903
 - "clipscore_median": 0.1839
 - "clipscore_min": 0.1701
 - "clipscore_max": 0.2135}

CLIPScore → semantic alignment

Metrics & Evaluation (Silhouette)

The screenshot shows the Montage Photo Agent configuration interface running on a Streamlit server at localhost. The left sidebar contains configuration options like labels, cluster cap, captioner mode, and preview zoom level. The right panel displays the 'Pipeline step outputs' section, which includes the 'ingest' step (n_items: 144), 'dedupe_quality' step (n_items: 59), 'categorize' step (n_items: 59), and the 'cluster' step (n_clusters: 5, k: 5, fused: false, silhouette: 0.11800000071525574). A tooltip for the silhouette value indicates it represents clustering quality.

Configurations

Labels (, or /n separated) ?

- group photo
- people
- nature
- Lord Ganesha
- food

Max images per post (cluster cap) ? 6

Captioner mode ? blip2

Preview Zoom Level ? 0.25

Clear Preview

Pipeline step outputs

ingest

```
{ "n_items": 144 }
```

dedupe_quality

```
{ "n_items": 59 }
```

categorize

```
{ "n_items": 59 }
```

cluster

```
{ "n_clusters": 5, "k": 5, "fused": false, "silhouette": 0.11800000071525574 }
```

silhouette → clustering quality: cohesion vs. separation of image clusters; range $-1 \rightarrow 1$ (higher = better).

Publishing Preview

The screenshot shows the Streamlit interface for the Montage Photo Agent. On the left, a sidebar titled "Configurations" contains the following settings:

- Labels (, or /n separated):
 - group photo
 - people
 - nature
 - Lord Ganesha
 - food
- Max images per post (cluster cap): A slider set to 6.
- Captioner mode: A dropdown menu set to "blip2".
- Preview Zoom Level: A slider set to 0.25.

On the right, a grid of six images is displayed, each with a caption and an "Include" checkbox:

Image Category	Caption	Include Status
Row 1, Column 1	landscape, nature	<input checked="" type="checkbox"/> Include
Row 1, Column 2	landscape, outdoors	<input checked="" type="checkbox"/> Include
Row 1, Column 3	outdoors, nature	<input checked="" type="checkbox"/> Include
Row 2, Column 1	outdoors, person	<input checked="" type="checkbox"/> Include
Row 2, Column 2	landscape, nature	<input checked="" type="checkbox"/> Include
Row 2, Column 3	landscape, nature	<input checked="" type="checkbox"/> Include

A "Deploy" button is located at the top right of the main area. At the bottom, there is a button labeled "EXPORT: Download Instagram Carousel JSON".

Final assembled posts ready for export.

- IG-style layout (image + caption + hashtags block).
- Supports single or multi-image posts.

Posts Export



```
{  
  "generated_at": 1758038674,  
  "posts": [  
    {  
      "caption": "@sean_mcfly we love the portraits!",  
      "hashtags": [  
        "#IITGuwahati",  
        "#Montage",  
        "#PhotographyClub",  
        "#Outdoors",  
        "#GroupShot",  
        "#Team",  
        "#iitg",  
        "#Photostory",  
        "#Destiny",  
        "#Winners",  
        "#Congratulations"  
      ],  
      "images": [  
        "data/events/upload_session_1757759101/multiracial-friends-screaming-happily.jpg",  
        "data/events/upload_session_1757759101/group-friends-jumping-nature.jpg",  
        "data/events/upload_session_1757759101/young-happy-people-dressed-with-winter-clothes.jpg",  
        "data/events/upload_session_1757817240/multiracial-friends-screaming-happily.jpg",  
  
        "data/events/upload_session_1757964520/436279783_858869759394044_4088594490777068867_n.jpg"  
      ]  
    },  
  ],  
}
```

Export posts as JSONL, CSV, or ready-to-share images with captions.

- Store deliverables in /deliverables/ for reports or uploads.

Settings & Config



```

pipeline: [ingest, dedupe_quality, categorize, cluster, captioner, publisher]
clipscore:
  enabled: true
  model: "ViT-B-32"
  pretrained: "laion2b_s34b_b79k"
eval:
  generate: true
  eval_subset_size: 512
  metric_to_monitor: "rougeL"
  early_stopping_patience: 2
  num_beams: 3
  max_new_tokens: 32
embeddings: {backend: openclip, model: ViT-B-32, pretrained: laion2b_s34b_b79k,
device: cpu}
dedupe: {method: clip, clip_threshold: 0.800, near_dup_threshold: 5}
categorize:
  #labels: ["candid","portrait","group
photo","stage","audience","speaker","award","sports","food","night","landscape",
,"architecture","indoors","outdoors"]
  labels: ["group photo","people","nature","Lord
Ganesha","food","night","landscape","indoors","outdoors","person"]
  threshold: 0.40  # raise for stricter tags (e.g., 0.32–0.36)
  topk: 3
cluster:
  use_clip: true
  k: auto
  max_images_per_post: 6
  use_label_features: false      # ← enable fusion
  label_weight: 0.35            # ← scales label scores before concat
  label_norm: zscore            # ← zscore | none
  fuse_normalize: true          # ← L2-normalize fused vectors
  selection: {k: 48}
captioner:
  mode: "template"           # "template" (now) | "blip2" (optional future)
  montage_max_tiles: 9        # how many images to pack into the grid
  montage_tile_px: 384         # tile size (px) before packing
  adapter_path: "checkpoints/lora_blip2_montage/best"
  max_hashtags: 15
  include_swipe_hint: false
  calc_clipscore: true       # compute cosine(img, caption) scores
openers:
  - "Highlights from"
  - "Moments from"
  - "Scenes from"

```

```
- "Snapshots of"
base_hashtags: ["#IITGuwahati", "#Montage", "#PhotographyClub"]
label_hashtags:           # extend as you like
  "portrait": ["#Portrait", "#Faces"]
  "group photo": ["#GroupShot", "#Team"]
  "stage": ["#Stage", "#Concert", "#Live"]
  "audience": ["#Audience", "#Vibes"]
  "speaker": ["#Speaker", "#Talk"]
  "award": ["#Awards", "#Winners"]
  "sports": ["#Sports", "#Action"]
  "food": ["#Food", "#Treats"]
  "night": ["#Night", "#LowLight"]
  "indoors": ["#Indoors"]
  "outdoors": ["#Outdoors"]
publisher: {enabled: false, dry_run: true}
```

Load or modify pipeline YAML configs.

- Choose captioning mode (Template / BLIP-2 / Hybrid).
- Adjust thresholds for deduplication, clustering, ...