# 📱 Product Screenshots Document – Montage Agent

Dashboard / Home Screen

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

Central entry point for the app.  
- Upload new images or select existing datasets.  
- Quick navigation to settings and configurations

## Image Upload & Ingestion

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

Drag-and-drop or select folders of images.  
- Previews thumbnails for quick verification.

## Deduplication & Quality Check

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

Highlights near-duplicates flagged by CLIP embeddings  
- Current accuracy: ~100% at similarity threshold 0.8.

## Clustering, Captioning View

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

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)

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

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)

🖼️ A screenshot of a computer

AI-generated content may be incorrect.

CLIPScore → semantic alignment

## Metrics & Evaluation (Silhouette)

🖼️

A screenshot of a computer

AI-generated content may be incorrect.

**silhouette** → **clustering quality:** cohesion vs. separation of image clusters; range −1 → 1 (higher = better).

## Publishing Preview

🖼️ A screenshot of a computer

AI-generated content may be incorrect.

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, …