

# Creating a Batch Image Description Process Tool Using Airflow and Claude Haiku 4.5

maría a. matienzo, Jason Raitz, Dan Schmidt,  
Lauren Sorensen, Anna Wilcox, Yuchai Zhou

UC Berkeley



## Meet your presenters



**Dan Schmidt**

**DevOps and Infrastructure Lead**  
[danschmidt5189@berkeley.edu](mailto:danschmidt5189@berkeley.edu)



**Jason Raitz**

**Application Developer**  
[raitz@berkeley.edu](mailto:raitz@berkeley.edu)

# Background

Identifying the problem

Exploring AI-backed solutions

UC Berkeley

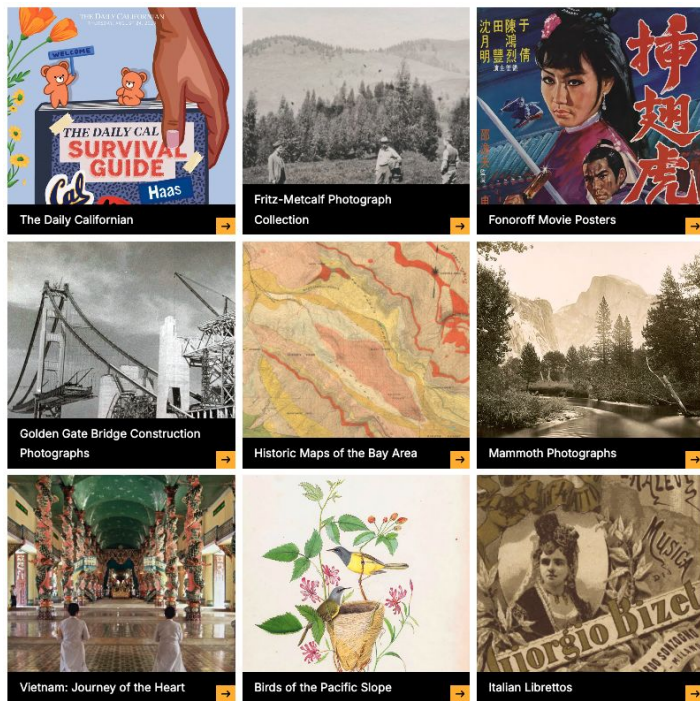


# Making Digital Collections Accessible

UC Berkeley Library's Digital Collections contain 1.3M images lacking descriptions for vision-impaired users.

We were charged with making those images accessible while ensuring:

- Descriptions are high-quality, unbiased, and comply with [WCAG 2.1 § 1.1.1 \(Non-Text Content\)](#).
- Metadata (MARC) updates are consistent and transparent.
- Workflows are maintainable at scale (cost-effective, observable, improvable, etc.).



<https://www.lib.berkeley.edu/find/digital-collections>

# What makes a good image description?

Descriptions ≠ alt-text

Good descriptions “paint a picture with words” while remaining concise, objective, and accurate.

Objectively measuring description quality is challenging (be forewarned).



<https://digicoll.lib.berkeley.edu/record/296153>

## What makes a good image description?

"A black-and-white photograph shows two men seated on folding chairs with their backs to the camera. They are positioned in the center of the frame against a dark background, with other people partially visible in the blurred distance behind them. Both chairs have large, white, sans-serif block lettering across the chair backs; the man on the left has "HOPPER" on his chair, while the man on the right has "CONNER." They are turned slightly toward each other, appearing to engage in conversation. A thin white border surrounds the photograph, and a color calibration strip is visible at the very bottom center of the mounting."



<https://digicoll.lib.berkeley.edu/record/296153>

# Pilot Project: Describing images with LLMs

Napkin math showed manual efforts couldn't scale:

1.3M images × 5 minutes / image  
≈ 50 years of staff time

So a small team was tasked with exploring LLMs:

- Assemble a highly varied test collection (landscapes, multilingual, text vs. non-text, ...)
- Write and validate acceptance criteria
- Experiment with three licensed models

Claude was chosen due to acceptable performance, reasonable cost, and platform availability.

| Collection filename             | Image | claude-sonnet-4.5                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | OpenAI gpt-5-chat                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | google gemini - gemini-2.5-flash                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| landscape<br>MSC_2024_A_001.jpg |       | <p>This is a black and white photograph showing a forest landscape with a stream or creek running through the center. The water flows from the foreground towards the background, flanked by dense trees and vegetation. The scene is captured from a low angle, looking down the length of the waterway.</p> <p>On both sides of the water are various types of vegetation. In the immediate foreground on the left are low grasses and shrubs. On the right side is a large bush or small tree. Dense foliage lines the banks of the stream are patches of grass and smaller plants.</p> <p>Two prominent tall trees with straight trunks frame the scene: one on the left side and one on the right side of the image. These appear to be cypresses, likely pine trees, with branches concentrated toward their upper portions.</p> <p>In the background, multiple coniferous trees of varying heights create a dense forest. The trees extend across the width of the image and back into the distance.</p> <p>On the far left, there appears to be a wooden fence or railing structure near the water's edge.</p> <p>The scene is lit gently, with the land rising slightly on both sides of the water course. The sky is visible in the upper portion of the image and appears overcast or hazy.</p> <p>The photograph has "2053" printed in white numbers in the bottom right corner. The entire image is mounted on a gray border or mat.</p> | <p>The image is a black and white photograph showing a natural outdoor scene. In the foreground, there is a stream or river with water flowing through the center of the image. Dense bushes and vegetation grow along the banks of the waterway. Tall coniferous trees, such as pines, stand throughout the scene, with some closer to the camera and others further away. In the left background, there is a wooden structure with a slanted roof, partially obscured by the trees. The ground is covered with some open areas and forested slopes extending toward the background, where more trees are visible. The image is framed with the number "2053" written in the bottom right corner.</p> | <p>This is a black and white photograph depicting a stream or river flowing through a forested landscape.</p> <p>The water occupies the center foreground and middle ground, appearing relatively shallow and rocky with a moderate flow creating a slightly blurred, dynamic effect. It is flanked by dense, low-lying shrubs on both banks in the foreground.</p> <p>The surrounding area is a coniferous forest, dominated by tall, thin pine and fir trees. Two particularly tall, large-diameter trees stand out in the background—one on the left and one on the right—acting as vertical anchors for the composition. The forest extends deeply into the background, where the more trees stand out at angle.</p> <p>On the far left, in the middle ground, there is a small, rustic wooden structure—possibly a cabin or shed—partially visible behind the brush and trees, near a small, open clearing. The ground slopes gently down toward the river.</p> <p>The sky is bright but featureless, contributing to a uniformity it, overcast, or slightly hazy appearance of the scene. A four-digit number, "2053", is barely visible in the bottom right corner of the photograph.</p>                                                      |
| landscape<br>MSC_2024_B_043.jpg |       | <p>This is a black and white photograph of a forest scene dominated by tall, straight tree trunks that extend vertically through the frame. In the center-right of the image, a large fallen tree lies on the forest floor. The view shows the massive exposed root system of the fallen tree, with tangled roots and soil attached. The trunk extends horizontally away from the root mass into the distance. A person is standing beside the fallen trunk near its base, providing scale to show the size of the tree. The ground is covered with leaf litter, small plants, and ferns. Light filters down between the tall trees, illuminating parts of the scene.</p> <p>The forest floor is covered with low vegetation and ferns. Multiple tree trunks of varying thickness are visible throughout the background, creating vertical lines across the composition. Light filters through the canopy, creating contrast between the darker tree trunks in the foreground and the brighter, hazier background foliage.</p> <p>A cleared path or dirt road runs through the lower portion of the image. The photograph appears to be taken in a forest with an old-growth forest with very large trees, possibly redwoods or sequoias based on their substantial diameter.</p>                                                                                                                                                                     | <p>The image is a black and white photograph of a forest with tall, straight trees, likely redwoods, extending vertically out of the frame. In the center-right of the image, a large fallen tree lies on the forest floor. The view shows the massive exposed root system of the fallen tree, with tangled roots and soil attached. The trunk extends horizontally away from the root mass into the distance. A person is standing beside the fallen trunk near its base, providing scale to show the size of the tree. The ground is covered with leaf litter, small plants, and ferns. Light filters down between the tall trees, illuminating parts of the scene.</p>                              | <p>This is a black and white photograph of a large, fallen tree in a forest of tall, straight trees, likely redwoods.</p> <p>The most prominent feature is the massive root ball of the fallen tree, which dominates the right side of the foreground. The root structure is irregular, densely tangled, and forms large, sculptural shapes. The trunk of the tree extends horizontally to the left from the root ball, partially covered by the dark tangle of roots and earth.</p> <p>A person is standing near the base of the fallen trunk, slightly to the left of the root ball. They appear relatively small next to the fallen tree, providing a sense of scale. The person is wearing light-colored pants and a shirt.</p> <p>The background is filled with a dense stand of very tall, slender tree trunks that rise straight up, creating a columnar effect. Sunlight streams down through the canopy and between the trees, creating bright patches of light on the forest floor, especially around the fallen tree. The forest floor is covered with leaf litter and some low-growing ferns or similar plants are visible in the foreground. The overall scene conveys a feeling of being deep within an ancient, monumental forest.</p> |

**Above:** Example output from the pilot study. Claude, GPT5, and Gemini results are seen left-to-right.

# Implementation

From pilot to production

UC Berkeley



# Assessing the infrastructure

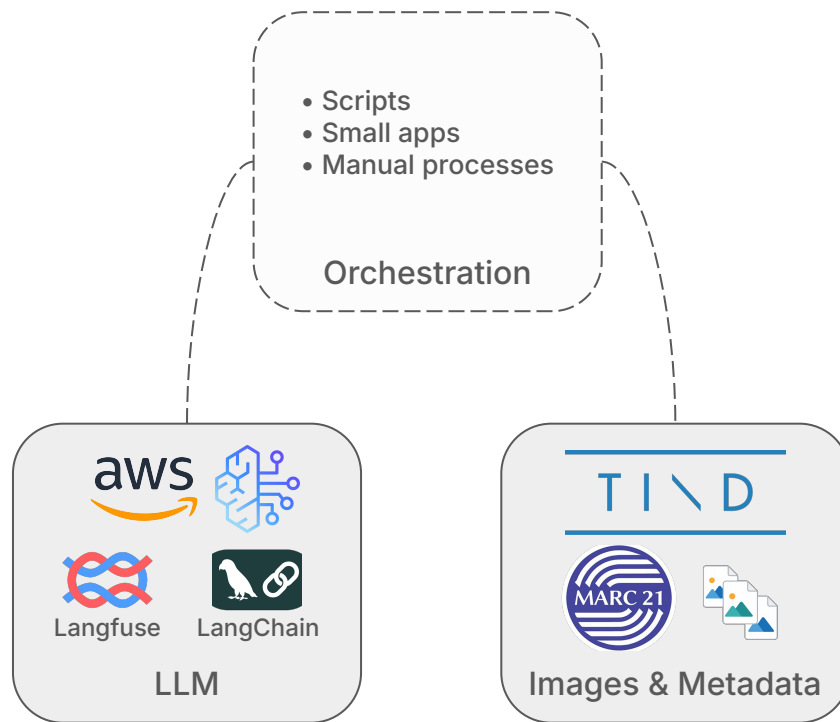
Some key components were in place and well understood:

- Digital Collections provided an API to our images and metadata (hosted by TIND).
- Prior AI projects provided an “LLM Stack”: Bedrock, LangChain, and Langfuse.

But tying it all together was a hodgepodge of scripts, small applications, and manual processes:

- Hard to integrate
- Poor observability
- Need to build the plumbing

Orchestration was missing—not only for AI.



# Introducing Apache Airflow

## The missing orchestrator:

- Airflow is a Python-based distributed system for running “workflows as code”
- Batteries included:
  - Web dashboard (Flask-based)
  - Sequencing and retry logic
  - Data source / sink
  - Triggers (ad-hoc, event-based, ...)
- Great [documentation](#) (or so it seemed)
- Flexible and extensible — would support many Library workflows

```
1 from datetime import datetime
2
3 from airflow.sdk import DAG, task
4 from airflow.providers.standard.operators.bash import BashOperator
5
6 # A Dag represents a workflow, a collection of tasks
7 with DAG(dag_id="demo", start_date=datetime(2022, 1, 1), schedule="@0 0 * * *") as dag:
8     # Tasks are represented as operators
9     hello = BashOperator(task_id="hello", bash_command="echo hello")
10
11
12 @task()
13 def airflow():
14     print("airflow")
15
16
17 # Set dependencies between tasks
18 hello >> airflow()
19
```

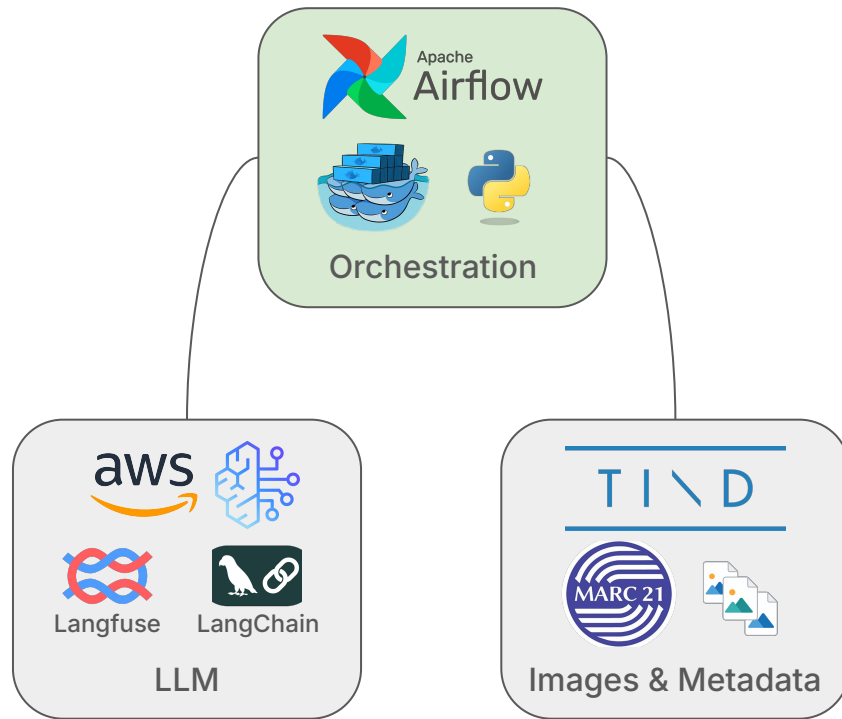


Above: A “hello world” workflow in AirFlow (from [the docs](#)).

## The Final Stack

- Airflow executes our workflows (with a GUI)
  - Existing Docker / Swarm infra (in VMware)
  - +100TB local storage capacity
- AI heavy lifting in the cloud
  - Considered [university GPU providers](#), but chose AWS Bedrock for ease of integration
  - Langfuse provides tracing and analysis tools
  - LangChain provides an SDK for model invocations with callbacks to LangFuse
- Digital Collections ([TIND.io](#))
  - REST API for querying, creating, and updating metadata records and associated images

Now we could implement the workflow!

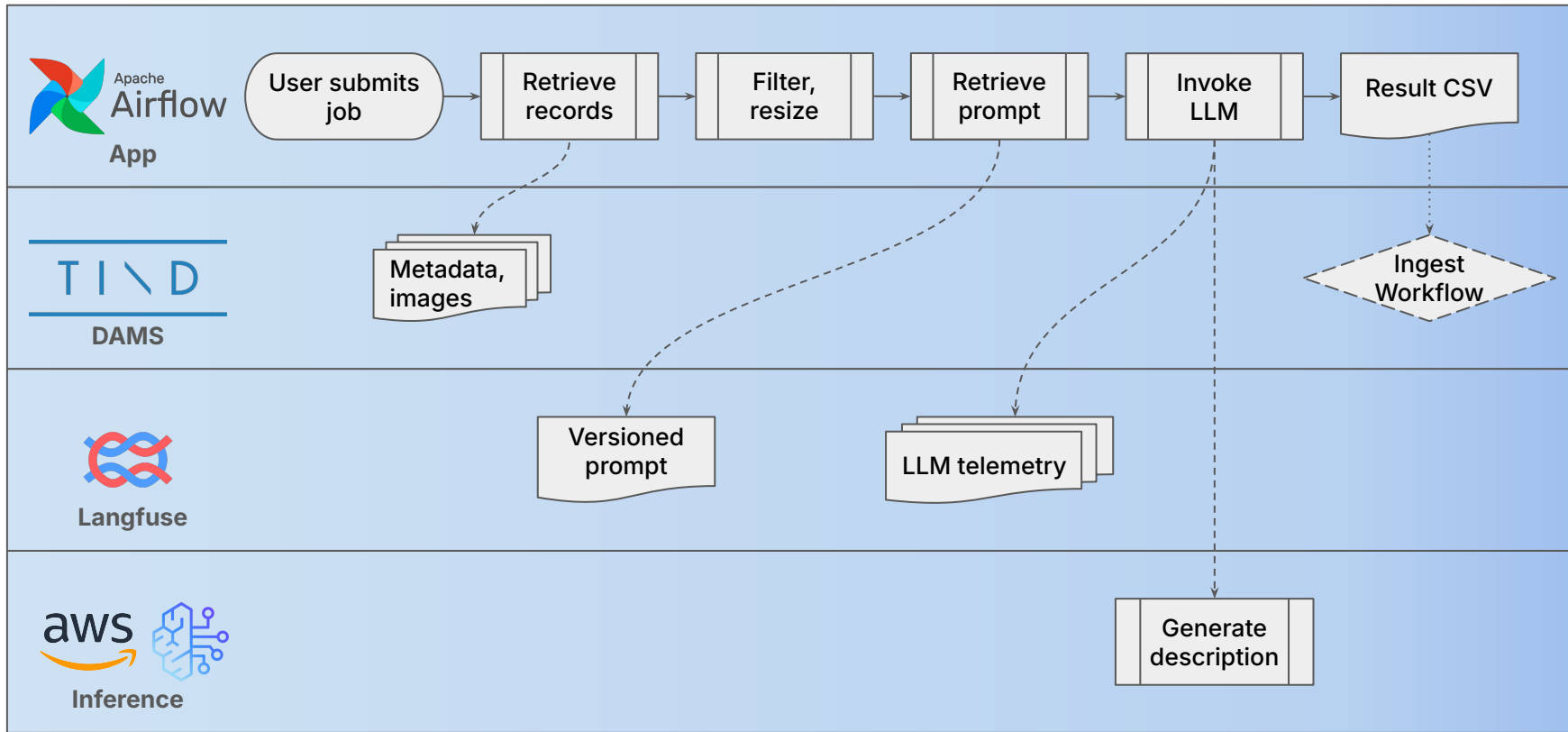


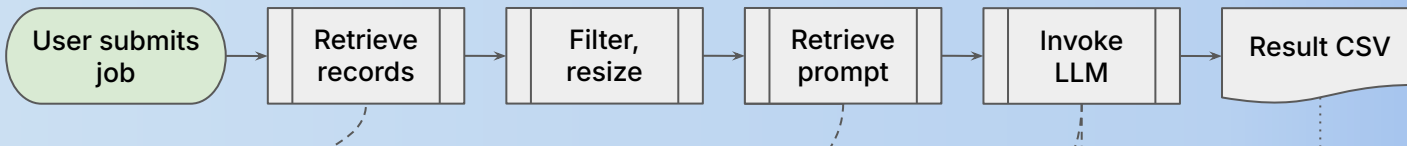
# Workflow Walkthrough

A stylized view of the image  
descriptions DAG

UC Berkeley



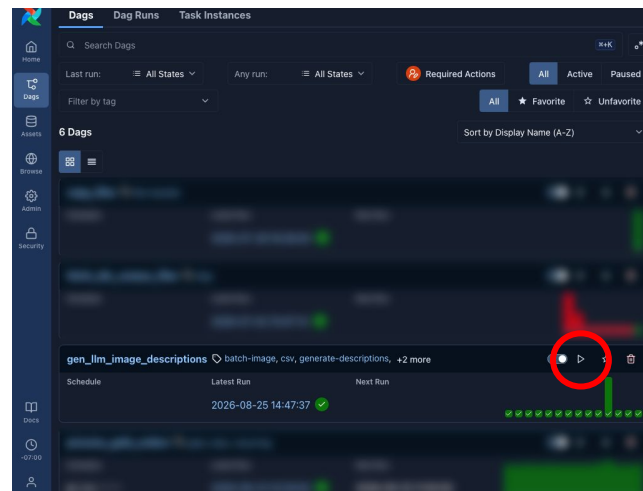




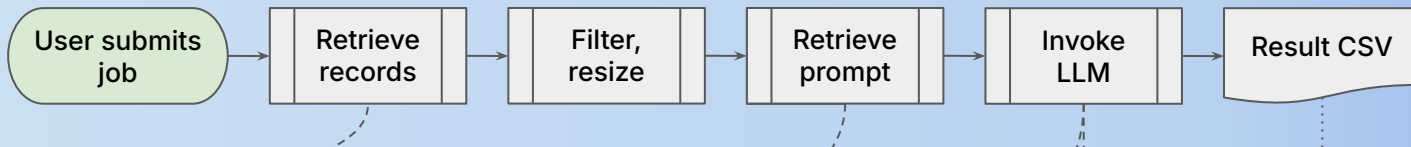
## Jobs are triggered manually

DAGs support per-run parameters:

- TIND Query (required)
- Prompt ID / Version (optional)
- Image max height / width (optional)
- "Advanced" options for AirFlow (e.g. annotating the run)



**Above:** Selecting from available DAGs.



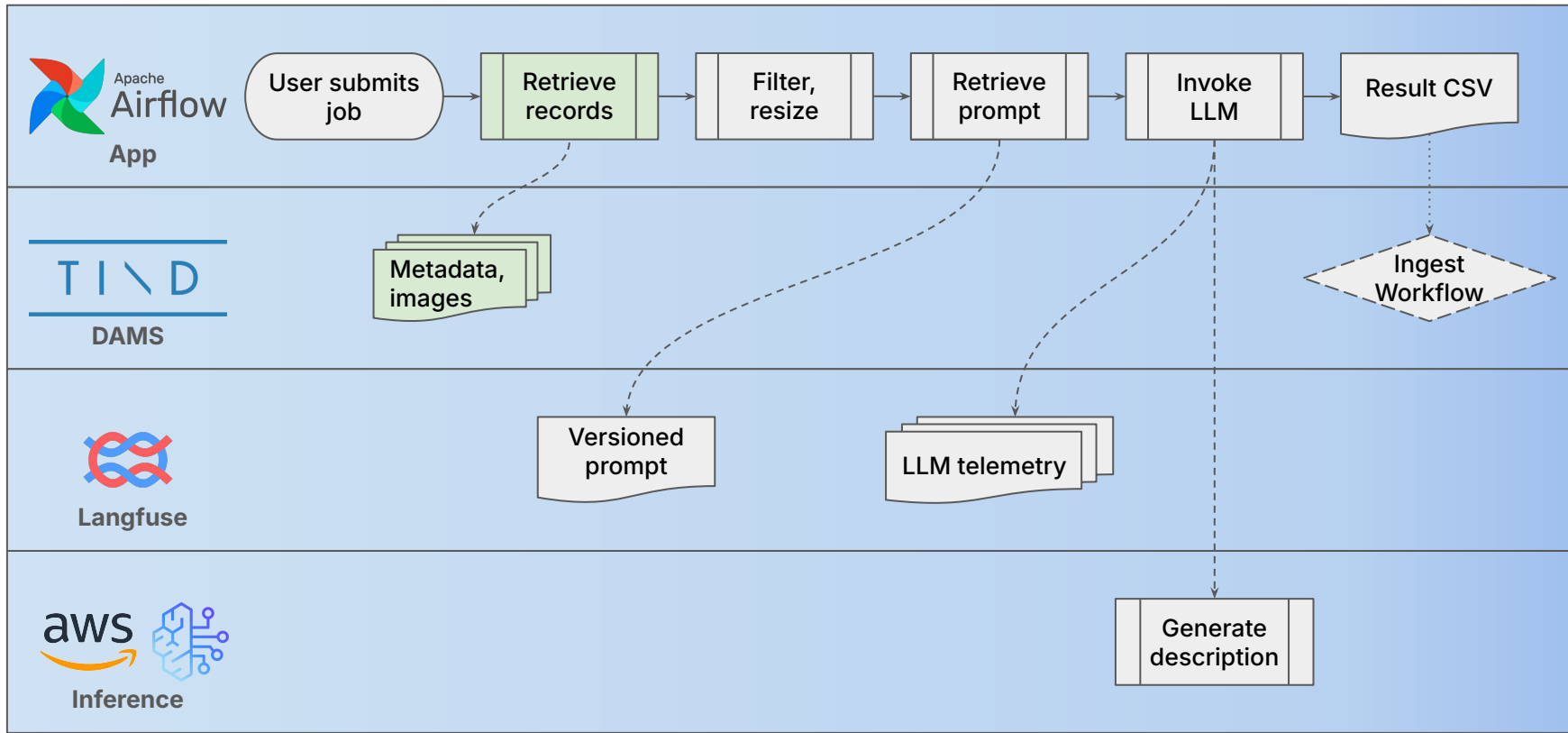
## Jobs are triggered manually

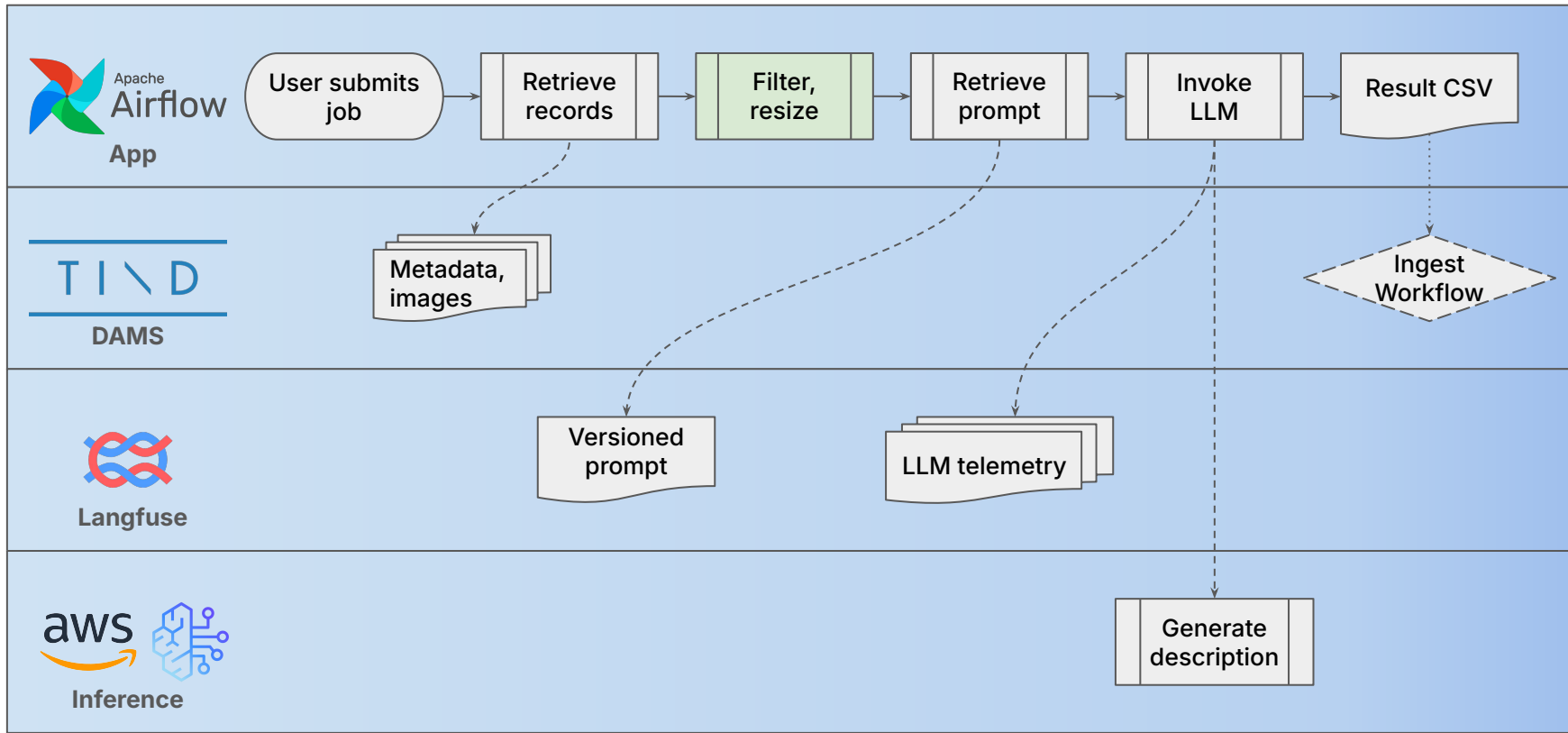
DAGs support per-run parameters:

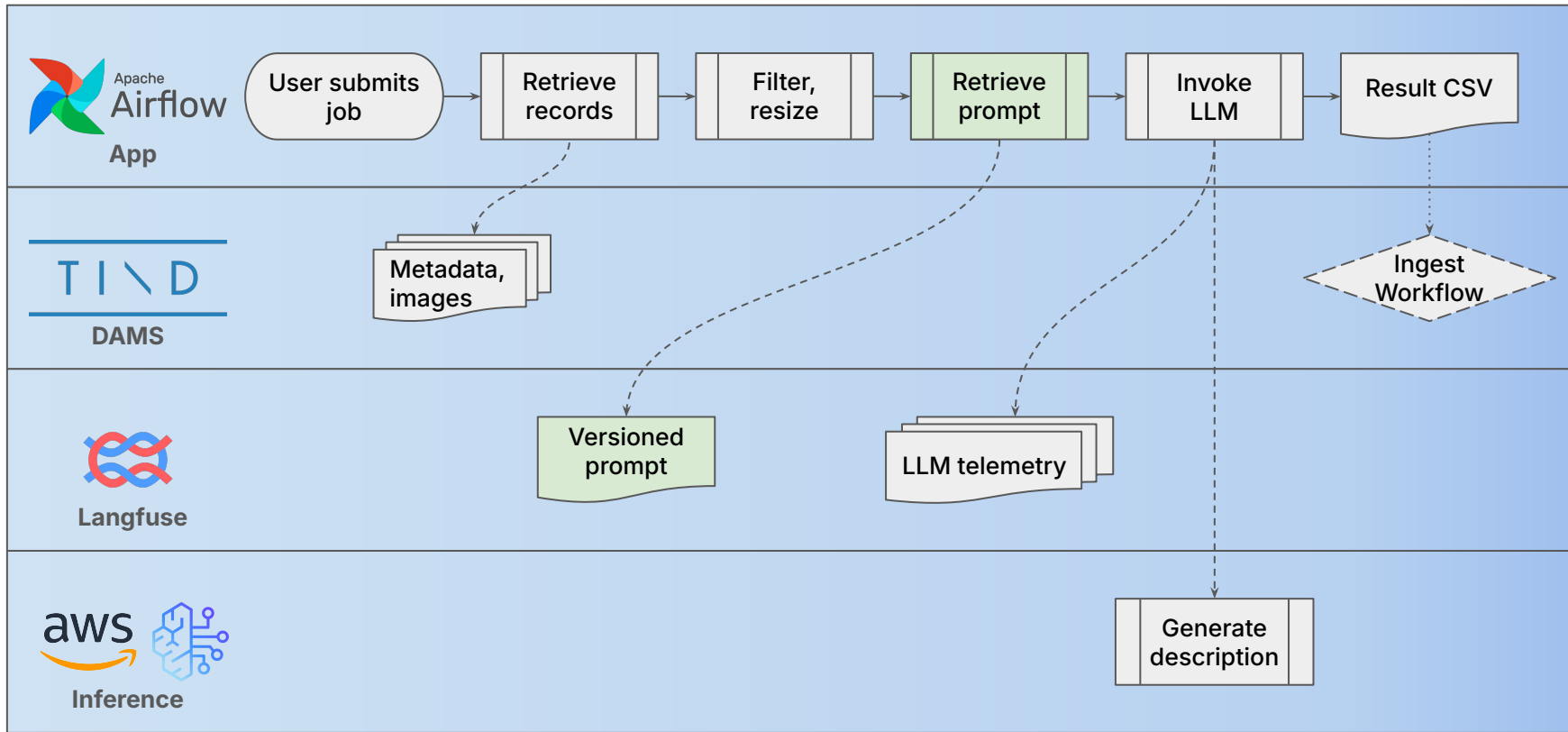
- TIND Query (required)
- Prompt ID / Version (optional)
- Image max height / width (optional)
- "Advanced" options for AirFlow (e.g. annotating the run)

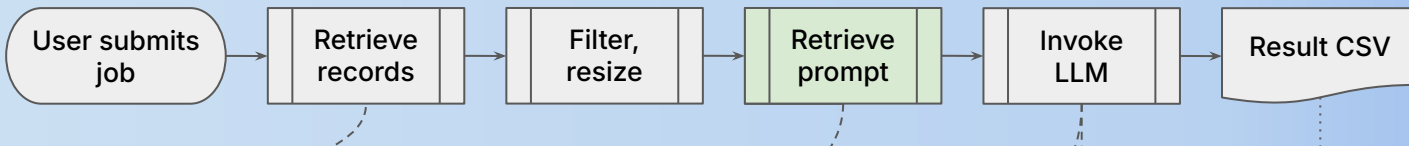
The screenshot shows the 'Trigger Dag - gen\_llm\_image\_descriptions' window. It has two tabs: 'Single Run' (selected) and 'Backfill'. The 'Single Run' tab has a description 'Trigger a single run of this Dag'. Below this is a 'Logical Date' field set to '09 / 10 / 2026, 01:46:15 AM'. Under the 'Run Parameters' section, there is a 'Tind query' field with the value 'farm mokelumne'. A tooltip for this field states: 'Search query for the Tind Search API. This is equivalent to the p (pattern) parameter in the Tind query syntax.' Below this are expandable sections for 'Prompt configuration', 'Image fetcher configuration', and 'Advanced Options'. A 'Trigger' button is at the bottom right.

**Above:** Configuring the DAG run.



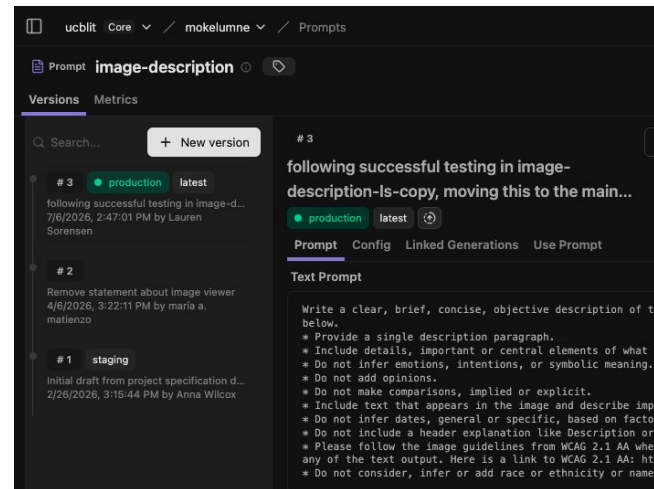




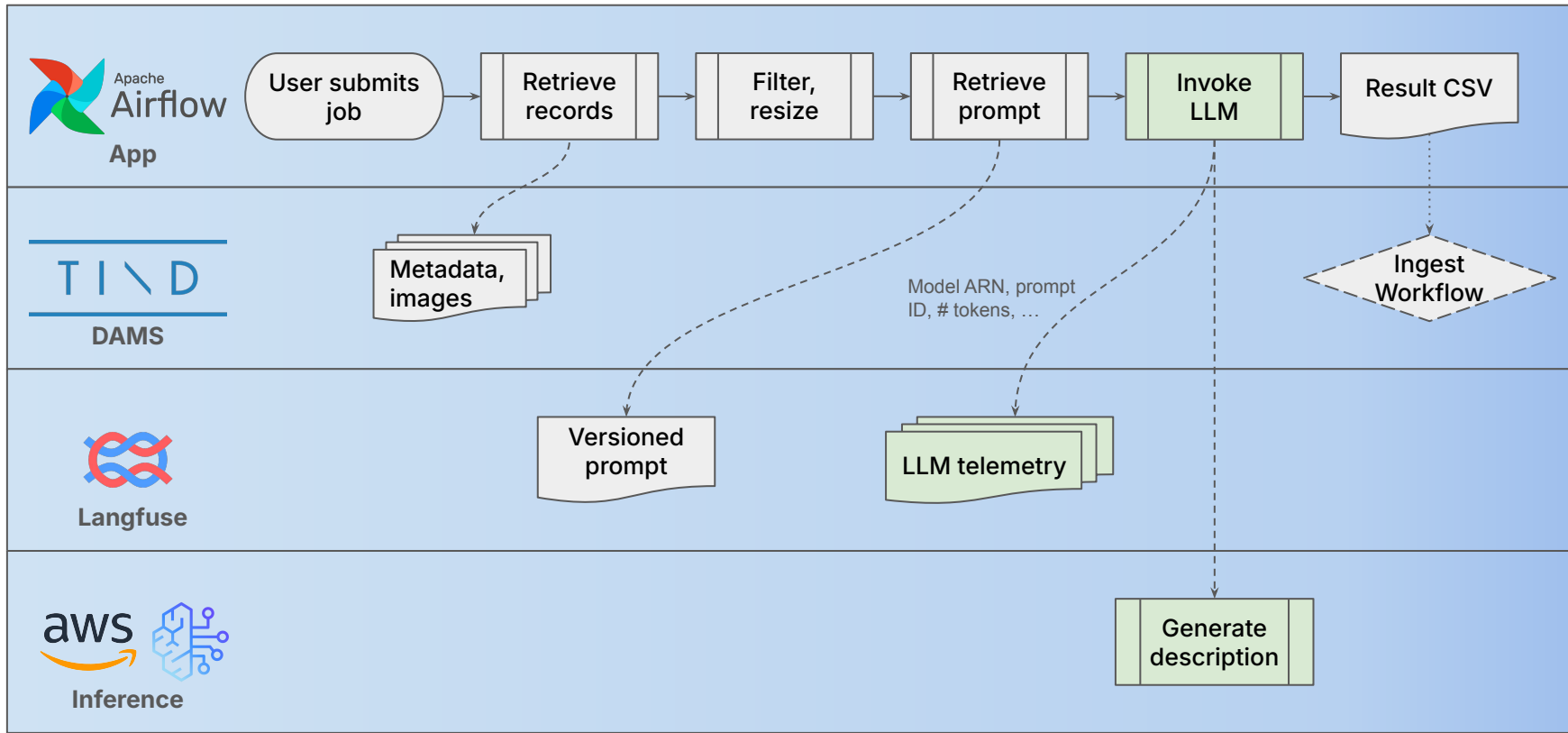


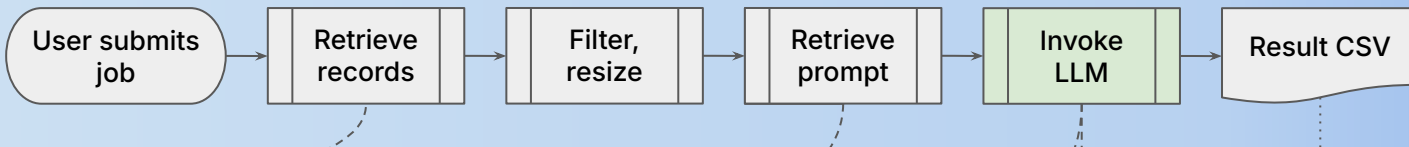
## LangFuse for prompts

- LangFuse provides prompt version control
- The DAG supports selecting a prompt type and specific version
- Built-in A/B testing support (but we have not yet used this)



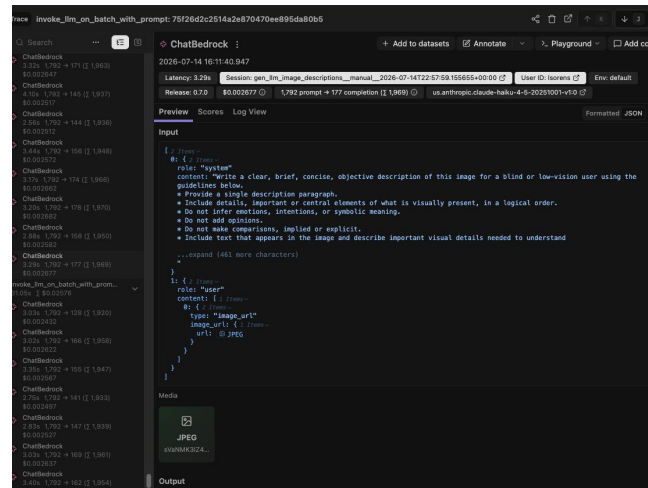
**Above:** Prompt versioning in LangFuse.



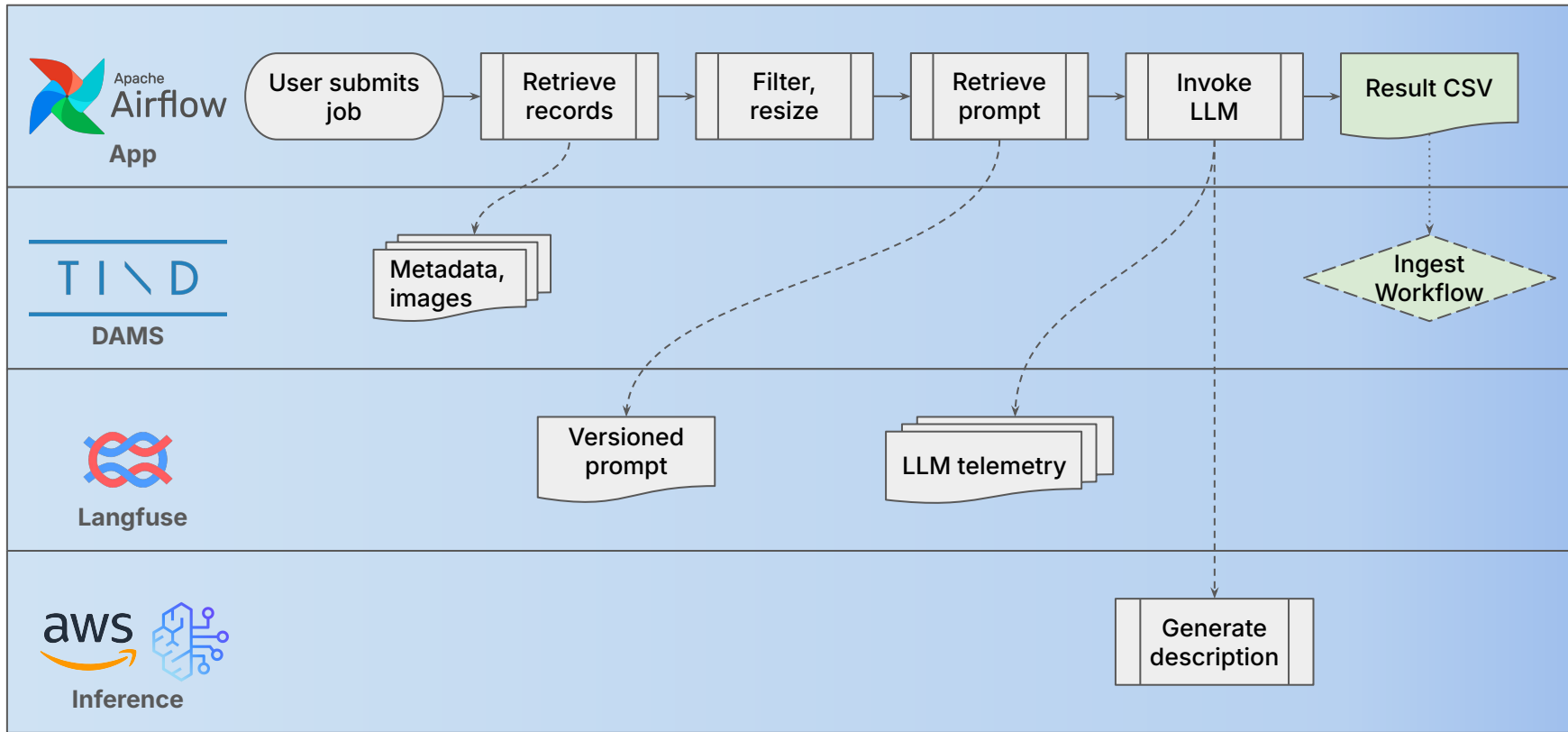


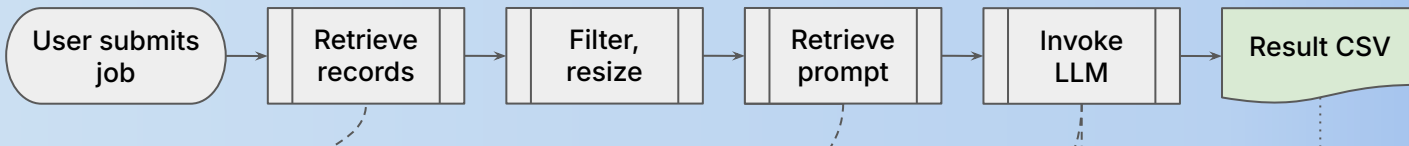
## LangFuse for tracing

- Sent full I/O and key-val metadata
- Many reports are available out-of-the-box (e.g. per-session cost estimation)
- Sent separately from Bedrock calls (LangFuse is not a proxy)
- Current free plan provides 50k traces/month



Above: Tracing in LangFuse.





## DAG outputs multiple CSVs

- Submitter downloads “processed” CSV:
  - Format supports QA + ingestion to TIND
  - Basis for A/B testing / prompt scoring
- Additional “fetched” and “skipped” CSVs assist with debugging
- AirFlow notifies user via email and provides a summary in the UI

Photograph Album] OR collection:[California Scenic Line] OR collection:[Fricot (D) Grass Valley] OR collection:[Ghirardelli Co. Chocolate Manufacturing]

[723 images processed](#)

723 succeeded, 0 failed

[723 images fetched](#)

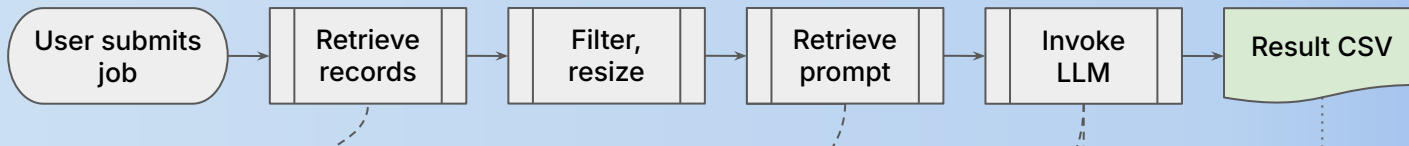
723 succeeded, 0 failed

[0 records skipped](#)

imagedesc\_processed\_20260714\_2312

| 035_a                            | 5880_a                                                                          | Status  | 520_a-1                                                                                                                                                                          | 983_a-1                                                                             | 983_d-1    |
|----------------------------------|---------------------------------------------------------------------------------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------------|
| calher_cubanc.25<br>_36_00139346 | Image description generated by AI (Claude Haiku 4.5) and reviewed on [MM/YYYY]. | success | This black and white photograph shows the interior of an elegant room with exposed wooden beam ceiling and ornate wooden wall paneling. Large multi-pane windows line the        | mokelumne-image-description[us.anthropic.claude-haiku-4-5-20251001-v1:0]mokelumne/3 | 2026-07-14 |
| calher_cubanc.25<br>_36_00139357 | Image description generated by AI (Claude Haiku 4.5) and reviewed on [MM/YYYY]. | success | This black and white photograph shows an ornate interior staircase viewed from above. The staircase features a curved wooden handrail with decorative turned balusters and newel | mokelumne-image-description[us.anthropic.claude-haiku-4-5-20251001-v1:0]mokelumne/3 | 2026-07-14 |
| calher_cubanc.25<br>_36_00139324 | Image description generated by AI (Claude Haiku 4.5) and                        | success | This sepia-toned photograph shows a                                                                                                                                              | mokelumne-image-description[us.anthropic.claude-haiku-4-5-20251001-v1:0]mokelumne/3 | 2026-07-14 |

**Above:** Summary output & processed CSV snapshot



## "Old Emperor Norton in 1876"

|         |                                                                                                                                                                                                                |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 035_a   | calher_cubanc_25_132_00180697                                                                                                                                                                                  |
| 5880_a  | Image description generated by AI (Claude Haiku 4.5) and reviewed on [05/2026].                                                                                                                                |
| Status  | success                                                                                                                                                                                                        |
| 520_a-1 | This is a sepia-toned historical photograph showing a bearded man in military uniform seated on a bicycle in an alleyway. The man wears a dark jacket with distinctive shoulder epaulettes (military decora... |
| 983_a-1 | mokelumne-image-description us.anthropic.claude-haiku-4-5-20251001-v1:0 mokelumne/1                                                                                                                            |
| 983_d-1 | 2026-05-19                                                                                                                                                                                                     |
| 983_t-1 | 520                                                                                                                                                                                                            |



# Takeaways

Metrics

Challenges

What's next

UC Berkeley



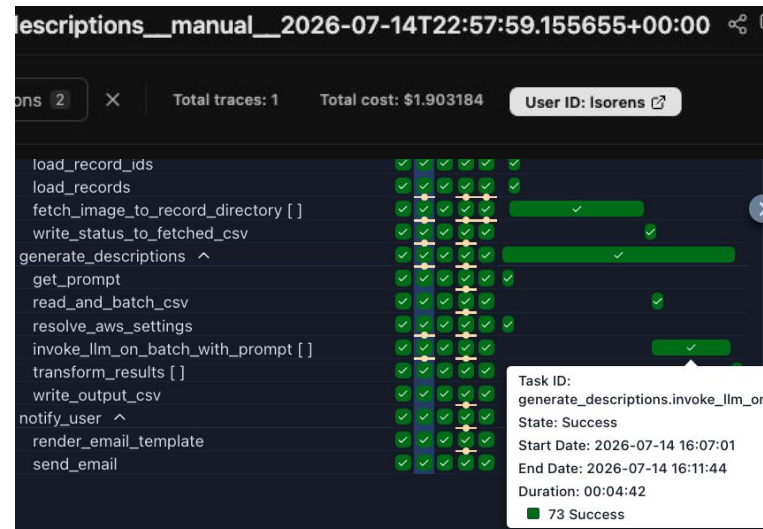
## Metrics at a glance

For a query returning 727 images:

- \$1.90 inference costs (\$0.0026/image)
- 14m10s total runtime (50 images/min)
  - 8m retrieval
  - 4m inference

Quality isn't formally quantified (yet).

AirFlow and LangFuse both provide useful metrics.



**Above:** Cost estimates (LangFuse, top) and task times (AirFlow, bottom) for a large DAG run.

# Challenges

- Integrating with LLMs is easy; iterating on prompts is not
- Bottleneck moved from generation to assessment
- Earning trust across the Library is the key challenge
- Launching, learning, and testing AirFlow (esp. v3.1+)

## What's Next

- Digital repository metadata updates are underway
- Assessment workflows are under active development (automated and manual)
- Airflow is hugely valuable even outside of AI-specific workloads. We've implemented six DAGs already and have many more planned!

# Get in touch

- [#airflow](#) in the Code4Lib Slack
- [BerkeleyLibrary/mokelumne](#) on GitHub
- Email us!
  - [matienzo@berkeley.edu](mailto:matienzo@berkeley.edu)
  - [raitz@berkeley.edu](mailto:raitz@berkeley.edu)
  - [danschmidt5189@berkeley.edu](mailto:danschmidt5189@berkeley.edu)
  - [lsorens@berkeley.edu](mailto:lsorens@berkeley.edu)
  - [awilfox@berkeley.edu](mailto:awilfox@berkeley.edu)
  - [zhouyu@berkeley.edu](mailto:zhouyu@berkeley.edu)

UC Berkeley



# Appendix

Example images, descriptions, prompts,  
and cost estimates.



## Comparing models

- Google models are generally cheaper, especially in one-shot conditions where reasoning efficiency is irrelevant.
- The original pilot was limited to only a handful of approved and available models.
- Reviewers generally liked Claude and Gemini's descriptions, but noted more and more significant hallucinations from Gemini (see next few slides).

| Model                             | Input (\$/M)        | Output (\$/M) |
|-----------------------------------|---------------------|---------------|
| <a href="#">gemini-2.5-flash</a>  | \$0.30<br>\$0.03**  | \$2.50        |
| <a href="#">claude-haiku-4.5*</a> | \$1.00<br>\$0.10**  | \$5.00        |
| <a href="#">gpt-5-chat</a>        | \$1.25<br>\$0.125** | \$10.00       |
| <a href="#">gemini-3.5-flash*</a> | \$1.50<br>\$0.15**  | \$9.00        |
| <a href="#">claude-sonnet-4.5</a> | \$3.00<br>\$0.30**  | \$15.00       |

**Above:** Model pricing as of Sep. 2026. Asterisk (\*) indicates the model was unavailable for the pilot. Double-asterisk (\*\*) indicates cached token price.

# Token Counts

To generate a description of this image of the Chabot Observatory using Claude Haiku 4.5:

- 1,792 input tokens (168 words + image)
- 177 output tokens (136 words)
- \$0.003 estimated cost

Image tokenization **is more involved** than for text and may include server-side downscaling.

Max context window varies; Haiku's is 200k (typical prompt is <1%).

Data is sourced from LangFuse tracing and checked for reasonableness against AWS stats and Anthropic docs.



**Above:** The Chabot Observatory. The JPEG image submitted to Claude was scaled to 1024×1536 (~200kb).

## Comparing Descriptions

### *Fallen Redwood // claude-sonnet-4.5*

#### claude-sonnet-4.5

"This is a black and white photograph of a forest scene dominated by tall, straight tree trunks that extend vertically through the frame. In the center-right of the image is a massive upturned tree root system - a large, complex tangle of roots and soil that has been exposed, creating a wall-like structure. The root mass appears dark and textured with irregular, twisted forms.

A single human figure stands near the fallen trunk on the left side of the root system, **providing scale - the root mass towers several times higher than the person [clearly describes scale]**. The individual appears to be looking up at or examining the roots.

The forest floor is covered with low vegetation and ferns. Multiple tree trunks of varying thickness are visible throughout the background, creating vertical lines across the composition. Light filters through the canopy, creating contrast between the darker tree trunks in the foreground and the brighter, hazier background foliage.

A cleared path or dirt road runs through the lower portion of the image. The photograph appears to be taken in what looks like an old-growth forest with very large trees, possibly redwoods or sequoias based on their substantial diameter."



**Above:** Emanuel Fritz photograph of a large, fallen redwood. (biosphotos MISC\_Box\_B\_043.jpg)

**Left:** Claude's response is clear and descriptive, making for easy and compelling reading.

## Comparing Descriptions

### *Fallen Redwood // gpt-5-chat*

#### **gpt-5-chat**

"The image is a black-and-white photograph of a forest with tall, straight trees, likely redwoods, extending vertically out of the ground. In the center-right of the image, a large fallen tree lies on the forest floor. The view shows the massive exposed root system of the fallen tree, with tangled roots and soil attached. The trunk extends horizontally away from the root mass into the distance. A person is standing beside the fallen trunk near its base, **providing scale to show the size of the tree [what scale?]**. The ground is **covered [too strong]** with leaf litter, small plants, and ferns. Light filters down between the tall trees, illuminating parts of the scene."



**Left:** This description is characteristic of GPT-5-Chat's overly-short descriptions given the prompt(s) used in the pilot.

## Comparing Descriptions

### *Fallen Redwood // gemini-2.5-flash*

#### gemini-2.5-flash

"This is a black and white photograph of a large, fallen tree in a forest of tall, straight trees, likely Redwoods.

The most prominent feature is the massive root ball of the fallen tree, which dominates the right side of the foreground. This root structure is irregular, densely tangled, and looms large, suggesting the immense size of the tree. The tree's trunk extends horizontally to the left from the root ball, partially covered by the dark tangle of roots and earth.

A person is standing near the base of the fallen trunk, slightly to the left of the root ball. They appear relatively small next to the fallen tree, providing a sense of scale. The person is wearing light-colored pants and a shirt.

The background is filled with a dense stand of very tall, slender tree trunks that rise straight up, creating a colonnade effect. Sunlight streams down through the canopy and between the trees, creating bright patches of light on the forest floor, especially around the fallen tree. The forest floor is covered with leaf litter and some low-growing ferns or similar plants are visible in the foreground. **The overall scene conveys a feeling of being deep within an ancient, monumental forest."**



***Left:** Gemini's description is similar to Sonnet's at 1/10th the nominal (no-cache) cost, but reviewers preferred Claude.*

## Comparing Descriptions

Angela Davis, 1972 // claude-sonnet-4.5

### claude-sonnet-4.5

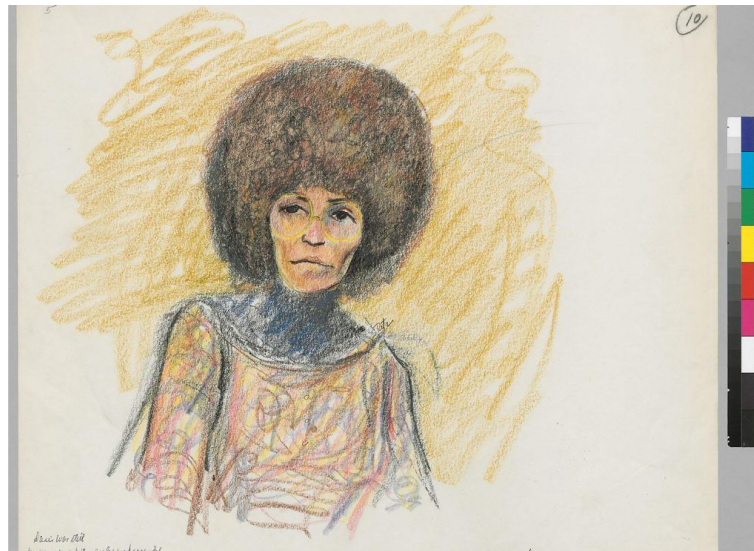
"This is a courtroom sketch drawn in colored pencil or pastel on white paper. The image shows a portrait of a person with a large, voluminous afro hairstyle that is dark brown or black in color. The subject's face has peachy-tan skin tones and they are depicted from roughly the chest up.

The person is wearing what appears to be a turtleneck or high-necked garment in gray or blue tones, with a lighter colored outer garment or shirt that has been rendered with loose, sketchy strokes in multiple colors including pink, yellow, and orange.

The background is filled with diagonal yellow-gold shading that creates a halo effect around the figure. The drawing style is loose and gestural, with visible pencil or pastel strokes throughout.

At the bottom of the image, there is handwritten text that appears to be a date and additional notes **[does not transcribe the text]**. In the upper right corner, there is a **circled number "19"** **[it's 10]**. On the right edge of the image, there is a color calibration strip showing various color swatches.

The paper appears to have a perforated edge at the top, suggesting it came from a spiral-bound sketchbook. The overall composition is a straightforward portrait study typical of courtroom sketch art."



**Above:** [Courtroom sketch of Angela Davis, 1972.](#)

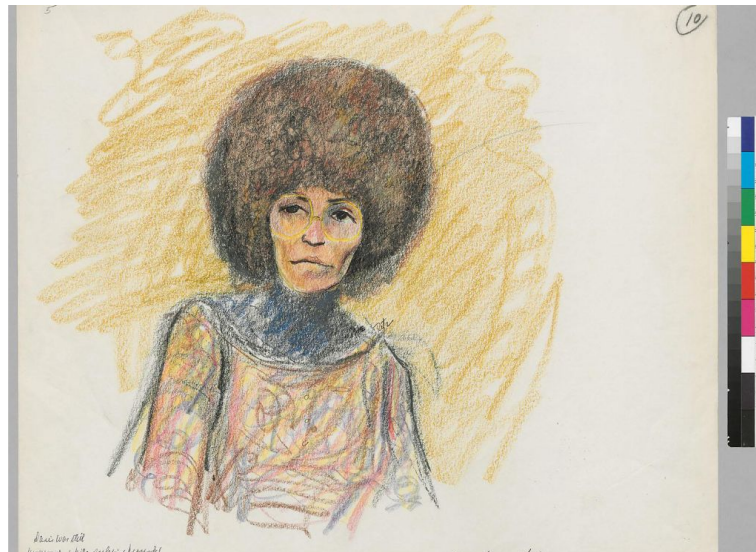
**Left:** Claude mistranscribed the text, even failing to mention the name "Angela Davis" although it appears in script.

## Comparing Descriptions

Angela Davis, 1972 // gpt-5-chat

### gpt-5-chat

"The image is a courtroom sketch created with colored pencil and pastel on white paper. It shows the back and shoulders of a person wearing a multicolored garment with intersecting lines of red, yellow, blue, and purple. The person's hair is dark brown with lighter highlights and drawn with a textured, rounded shape. The background is filled with a large area of yellow shading. Handwritten text appears along the bottom left and right edges of the paper, **noting a name [misses name and other text]** and date. The sheet has a perforated edge along the top and a vertical color calibration strip on the right side."



**Above:** gpt-5-chat writes a characteristically too-short description and misses key information.

## Comparing Descriptions

Angela Davis, 1972 // gemini-2.5-flash

### gemini-2.5-flash

"This is a color portrait sketch, drawn with ink lines and color pencils, on a sheet of white paper with a perforated top edge.

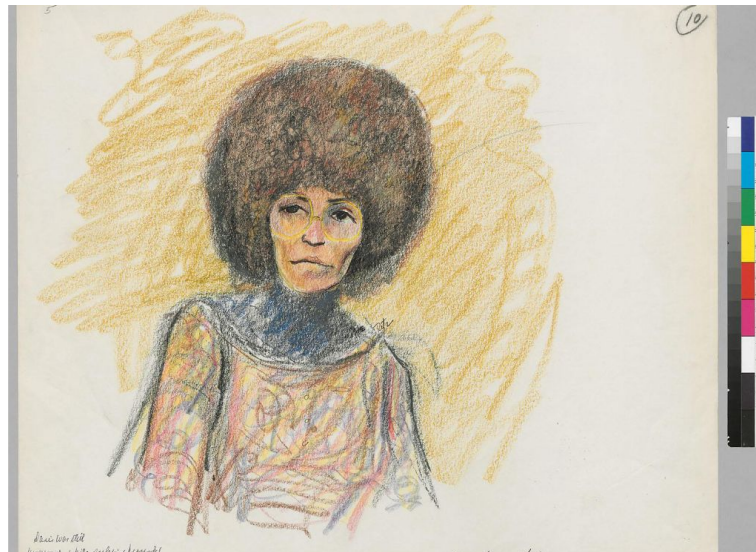
The central subject is a woman depicted from the shoulders up, dominating the center of the frame. She has a large, rounded, dark brown afro hairstyle that extends out widely around her head. She is wearing light-rimmed eyeglasses and looking directly at the viewer. Her clothing consists of a dark blue or black turtleneck underneath a patterned outer garment. The outer garment features a complex design with colors including pink, yellow, and red.

The background is filled with heavy, scribbled strokes of ochre and yellow-brown color pencil, creating a textured field behind her head and shoulders that strongly contrasts with her dark hair.

Handwritten text appears along the bottom of the sketch:

In the lower left corner, there is a note in small script: **"Davis wore old 'hippieman' white sweater - however I have seen it on."** [Hallucinated, see DigiColl: **"Davis was still imprisoned while preliminary hearings were going on."**] In the lower right corner, there is a signature and date: "Angela Davis 1/26/72".

The number "10" is circled in the upper right corner."



**Left:** Gemini's low-cost model picks up on most of the text but badly misreads the text in the lower-left corner. Reviewers consistently noted this problem with Gemini and GPT.

## Prompt Evolution

### Pilot

Describe the image for a blind person to see what a sighted person sees without implying emotion and feelings

## Prompt Evolution

### Production Version 1

Write a clear, brief, concise, objective description of this image for a blind or low-vision user using the guidelines below.

- \* Provide a single description paragraph.
- \* Include details, important or central elements of what is visually present, in a logical order.
- \* Do not infer emotions, intentions, or symbolic meaning.
- \* Do not add opinions.
- \* Do not make comparisons, implied or explicit.
- \* Include text that appears in the image and describe important visual details needed to understand the image.
- \* Do not infer dates, general or specific, based on factors such as if the image is black and white or not.
- \* Do not include a header explanation like Description or any variation of that.
- \* Please follow the image guidelines from WCAG 2.1 AA when providing the description, but do not mention the standard in any of the text output. Here is a link to WCAG 2.1 AA: <https://www.w3.org/TR/WCAG21/>.
- \* Do not consider, infer or add race or ethnicity or names.

## Prompt Evolution

### Production Beta Version (pseudo-XML, in testing)

<role>You write accessible image descriptions for blind and low-vision users.</role>

<task>Write a clear, brief, concise, and objective description of the provided image, following WCAG 2.1 Level AA image guidance: <https://www.w3.org/TR/WCAG21/>.</task>

<requirements>

- Return exactly one paragraph of plain text.
- Describe the central elements and important visible details in a logical order, beginning with the most important information.
- Include all relevant legible text visible in the image.
- Describe technical features such as color bars and archival markers as briefly as possible.

</requirements>

<constraints>

- Report only what is visually evident.
- Use neutral language without opinions or explicit or implied comparisons.
- Do not infer emotions, intentions, symbolic meaning, or dates.
- Do not infer a date from visual characteristics such as whether the image is black and white.
- Do not infer or assign race, ethnicity, gender, or sexuality.
- Do not identify or name people. A name may appear only when reproducing text that is visibly present in the image.
- Do not mention WCAG or accessibility standards in the output.
- Do not add a heading or introductory label such as "Description" or "Image description."

</constraints>

<output>Return only the description paragraph.</output>