HRI / 99P Labs Final Presentation

SDS 410 Capstone

Xinxin Zhang, Nora Zhou,
Yujie Gong, Ramsha Rauf

Spring 2025

Overview

- **Tool for researchers and data practitioners for Vehicle Risk and Safety assessment using Dashcam Images and Videos**
- **Explore the capabilities of multimodal AI systems to generate captions and analysis for images and videos**
- **Develop a workflow pipeline for the tool**

Contents

02 Data

- **57 dashcam images and 1 video**
  - PNG or JPEG
  - Screenshot from youtube dashcam videos
  - Personal collections
  - Include a variety of road conditions (congested highway, rain-soaked pond, animal, etc.)

Images from Dataset

Model Exploration

CLIP

OpenAI

Not for captioning

BLIP

salesforce

30/57

52.6% Accuracy

LLaVA

Microsoft

Computation Heavy

Gemini 2.0 Flash

Google

54/57

94.7% Accuracy

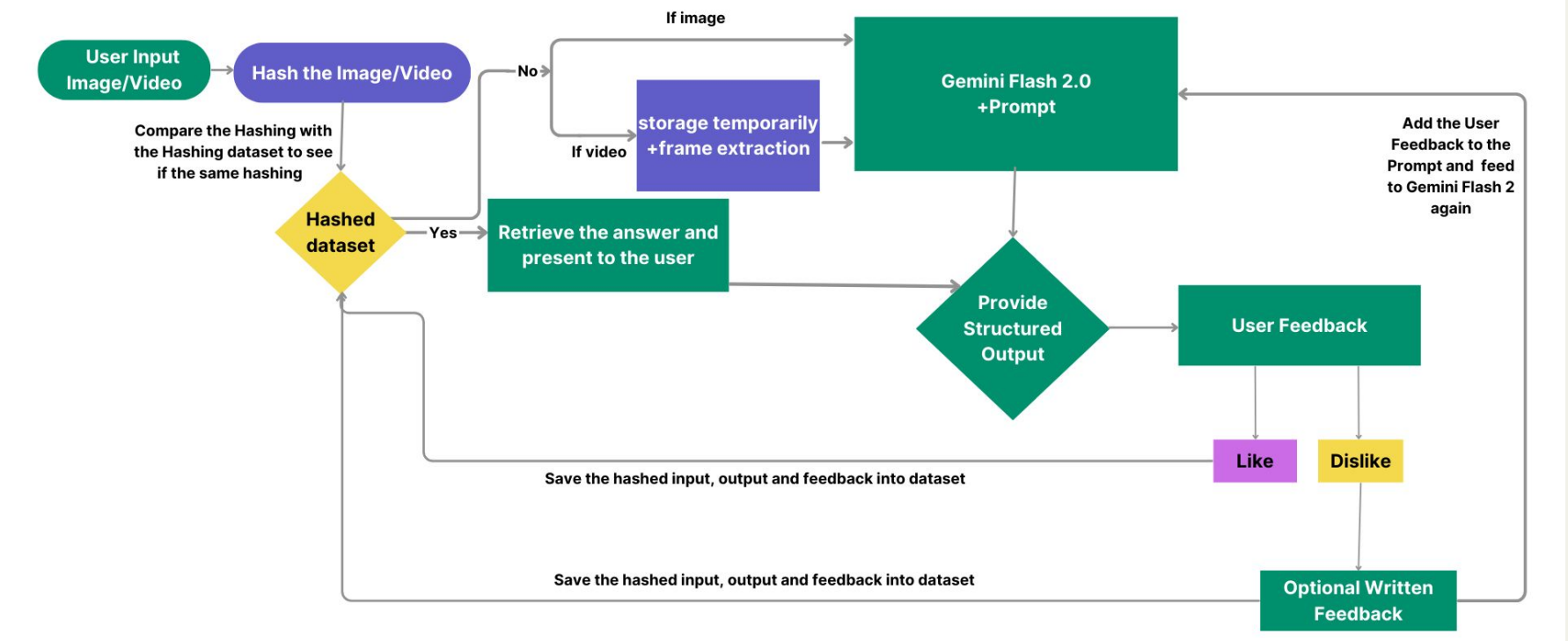
ChatGPT 4o-mini

OpenAI

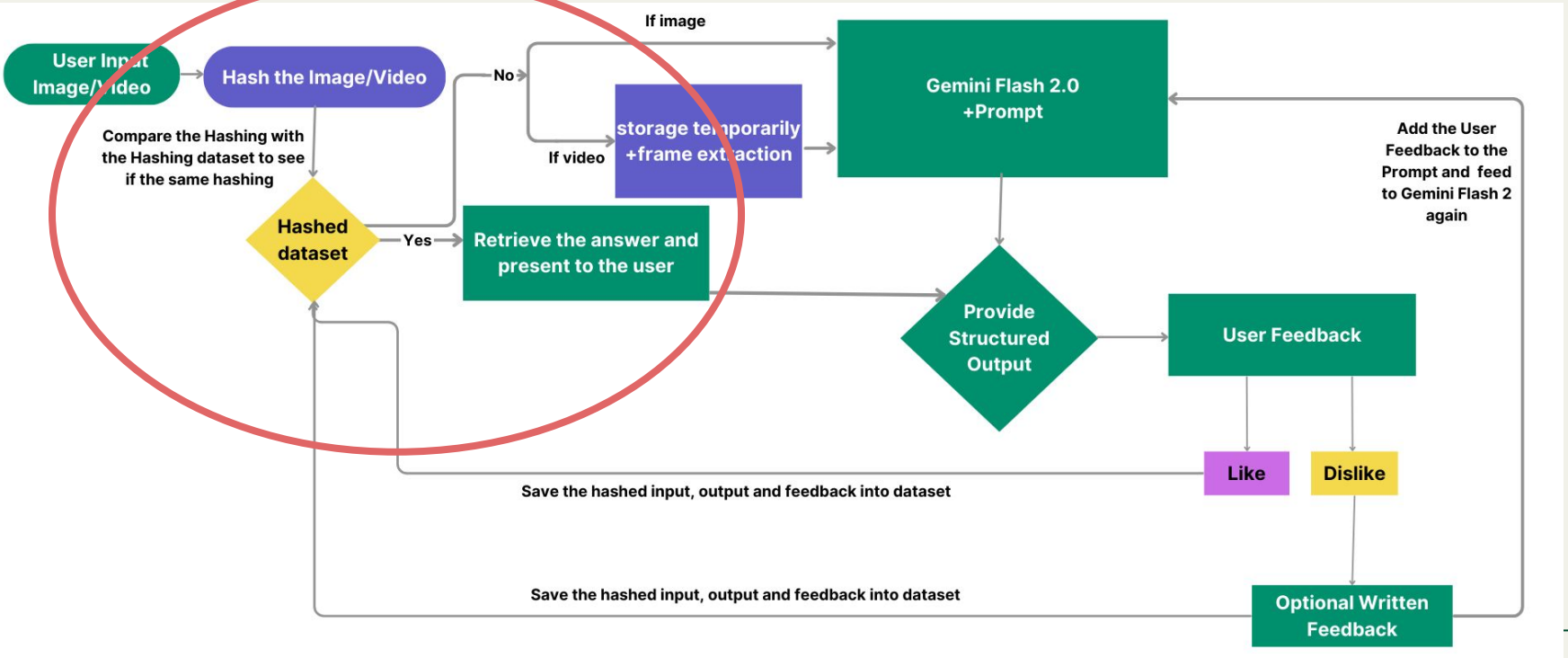
52/57

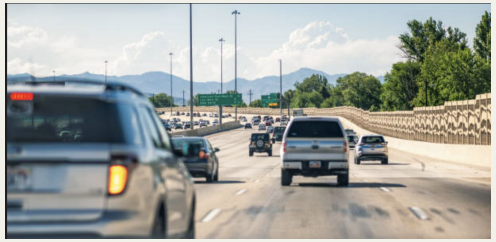
91.2% Accuracy

Workflow


User Input Image/Video
Hash the Image/Video
Compare the Hashing with the Hashing dataset to see if the same hashing
Hashed dataset
No
Yes
If image
If video
storage temporarily +frame extraction
Gemini Flash 2.0 +Prompt
Retrieve the answer and present to the user
Provide Structured Output
User Feedback
Like
Dislike
Add the User Feedback to the Prompt and feed to Gemini Flash 2 again
Save the hashed input, output and feedback into dataset
Save the hashed input, output and feedback into dataset
Optional Written Feedback

Workflow


User Input Image/Video
Hash the Image/Video
Compare the Hashing with the Hashing dataset to see if the same hashing
Hashed dataset
No
Yes
If image
If video
storage temporarily +frame extraction
Gemini Flash 2.0 +Prompt
Retrieve the answer and present to the user
Provide Structured Output
User Feedback
Like
Dislike
Add the User Feedback to the Prompt and feed to Gemini Flash 2 again
Save the hashed input, output and feedback into dataset
Save the hashed input, output and feedback into dataset
Optional Written Feedback



File to File Bytes

File Bytes

```
\x02d\x03\x01\x11\x00\x02\x11\x01\x03\x11\x01\xff\xc4\x00\x1b\x00\x00\
x02\x03\x01\x01\x01\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x01\x02
\x03\x04\x05\x06\x07\xff\xc4\x00\x1a\x01\x01\x01\x01\x01\x01\x01\x01\x0
0\x00\x00\x00\x00\x00\x00\x00\x00\x01\x02\x03\x04\x05\x06\xff\xda\x
00\x0c\x03\x01\x00\x02\x10\x03\x10\x00\x00\x01\xf4:\xc0\x001\xc0*\x04\x
80\x00\x00\x00\x00\x00\x05\x10\xc6\x032i\xa2,\x80@\x02\xa0\x00\x01\x10\
x00\x80\x02\x90\x00\x81\x00\x01\x00\x00\x00\x80\x00\x00\x00\x04\x00 \x1
1\xab\x1d\x01\x80\xc0\x04\x08\x00\x00\x00\x00\x00\x05\x00\x000\x03\x1d\
x97\xcbh\x80\x012Y\x1a\xd9\x9a\x00\x00\x82\x80D\x00\x02\x00\x01\x00\x0\
x08(\x10\x08\x00\x06\x02\x00\x021\xca\xe6\x96\xa7\xa0\x9dR\x001\x00P\xx
x00\x00\x00\x00\x00\x00\x00\x00\x08\x01\x8fR\xe9m\x80\x00*$\x12\xc5\x00
\x00\x01\x10\x00\x02\x88\x85@\x00\x90\x00\x10\x00\x8ac6[6\xe5a\xb7K\xaa
\xd1\x00\x08\xcb\x8b\xe5x\xdb\xfag\xddk\xa1+\xb1\x08@\x08\x00\x00\x05\x
00\x00\x00\x00\x00\x00c\xb2\xe2\xd8\x00\x00\x81\x1a\xb2\x00\x00\x00
...
00\x00\x00\x00\x00\x00c\xb2\xe2\xd8\x00\x00\x81\x1a\xb2\x00\x00\x00
```

Hashed Image

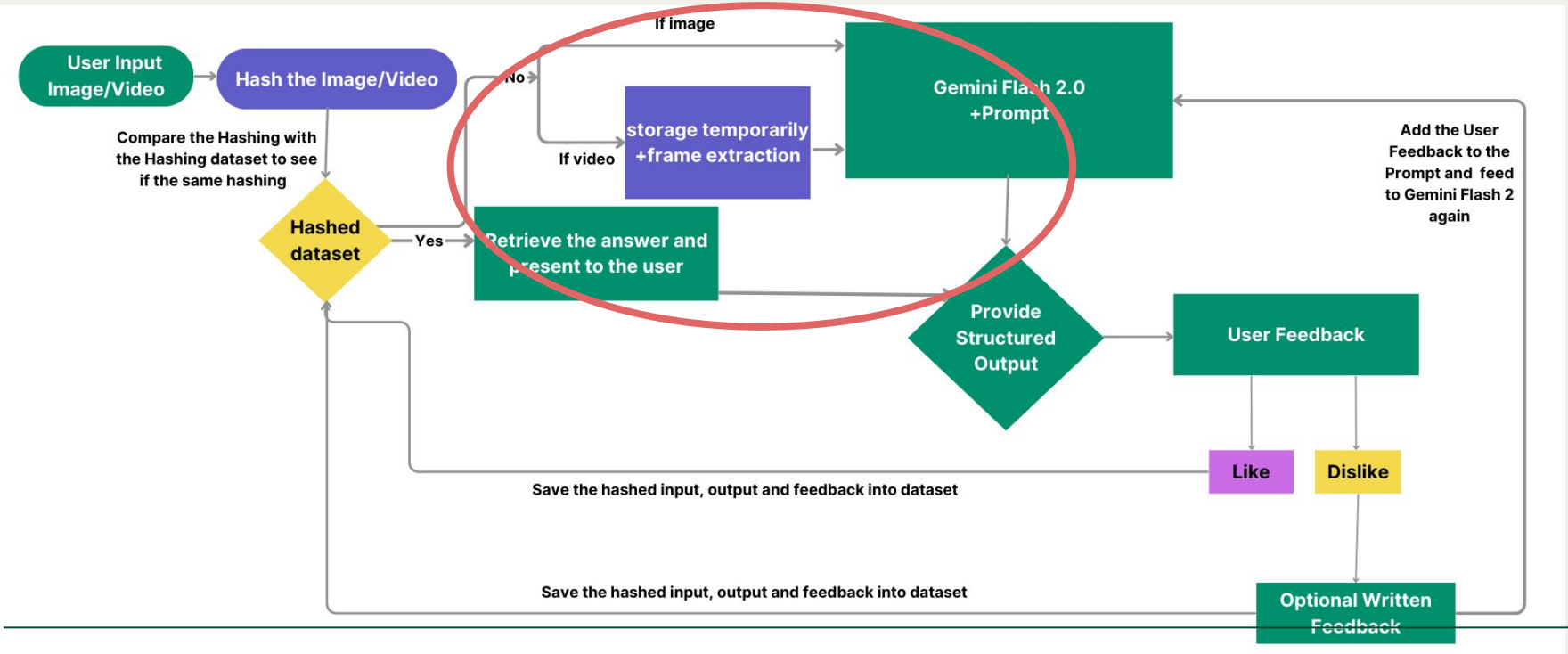
```
adc9a6bf707caaf59683d5bebb7ea45122d7beda3d13c5720a134715de270bf7
```

Hashed file

SHA-256

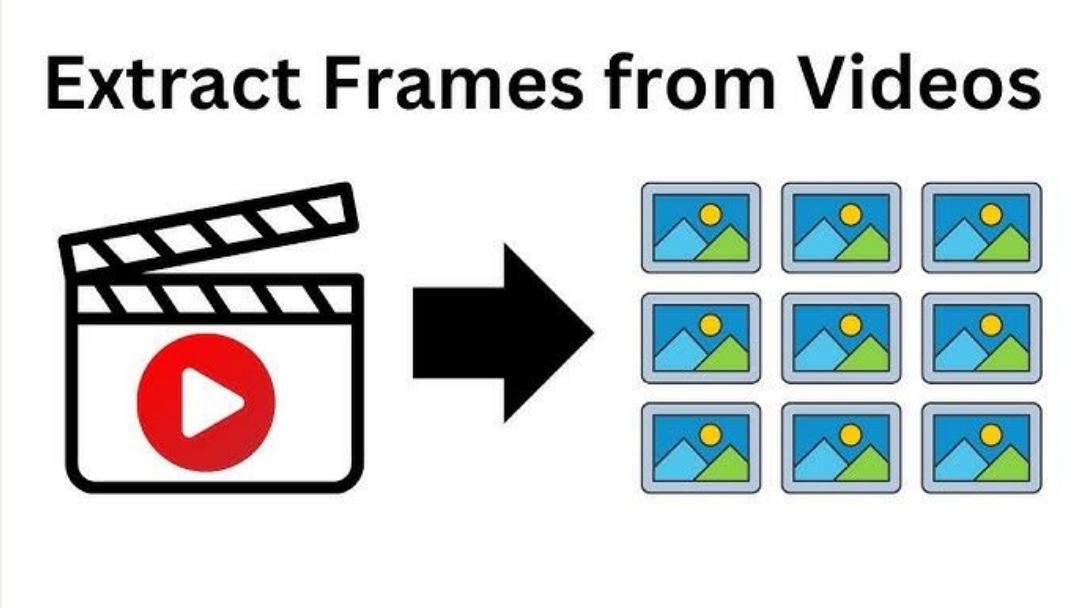
Hash function

Model Pipeline

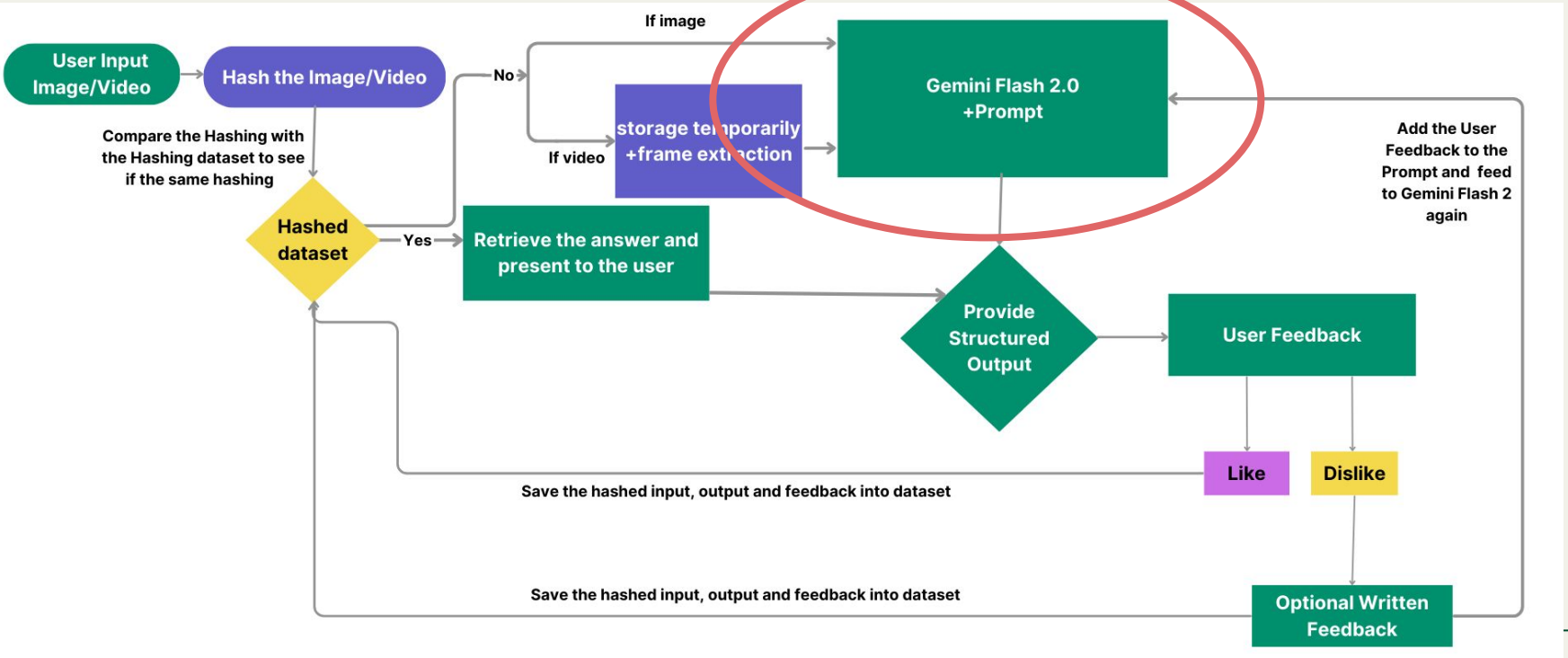


Video Frame Extraction

- **OpenCV for frame extraction**
- **10 frames/images**
- **Caption each image**
- **Combine the 10 captions into 1 output**



Model Pipeline

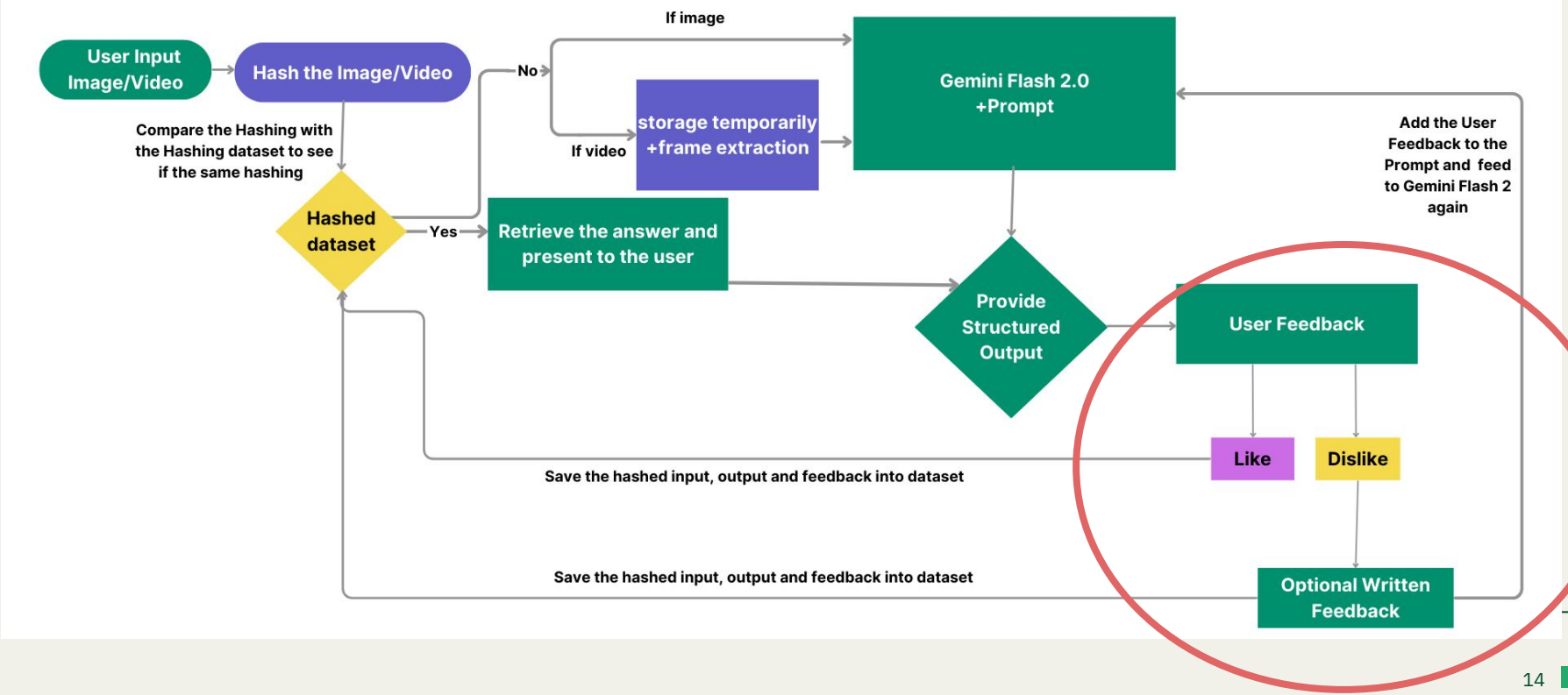


Prompt Engineering

```
| Variable             | Conditions
|----------------------|-----------------------------------------------------------------------------------------------
| Weather Visibility   | Clear, Foggy, Rainy, Snowy, Hazy, Obscured, Sunny, Glare (weather-related)
| Lighting Conditions  | Daytime, Nighttime, Dawn/Dusk (lighting-related)
| Traffic Flow         | Congestion, Moderate, Free Flow, Stopped
| Setting              | Urban, Suburban, Rural
| Road Surface         | Dry, Wet, Damp, Puddles, Flooded, Icy, Snow-Covered
| Road Infrastructure  | Uneven Surface, Good Lane Markings, Faded Lane Markings, Construction Zone
| Obstacles            | None, Debris, Pedestrians, Animals, Construction, Other Vehicles
| Intersections        | Signalized, Uncontrolled, Blind Spot, No Intersection
| Risk Level (1-5)     | [Number]
| Risk Factors         | [Brief explanation of factors leading to the risk level]
| Recommended Actions  | [Choose all that apply, in order of priority: Reduce Speed, Increase Following Distance, Change

  **Risk Assessment Scale:**

  - **1 - Negligible Risk:** No hazards; proceed normally.
  - **2 - Low Risk:** Minor issues; proceed with caution.
  - **3 - Moderate Risk:** Hazards need moderate adjustments.
  - **4 - High Risk:** Major obstacles; adjust significantly or reroute.
  - **5 - Critical Risk:** Imminent danger; stop or evade immediately.

---

## Important Considerations:
Format: Use bold headings and a compact table for easy scanning.
Safety First: If a variable cannot be reasonably determined from the image, choose the option that would be the most cons
Never assume beyond the image. Focus on what is immediately visible.
The table values MUST be derived from the paragraph generated in Part 1. Don't analyze the image twice.
In Part 2, adhere strictly to the provided option lists. Do not invent new categories or values.
"""
```

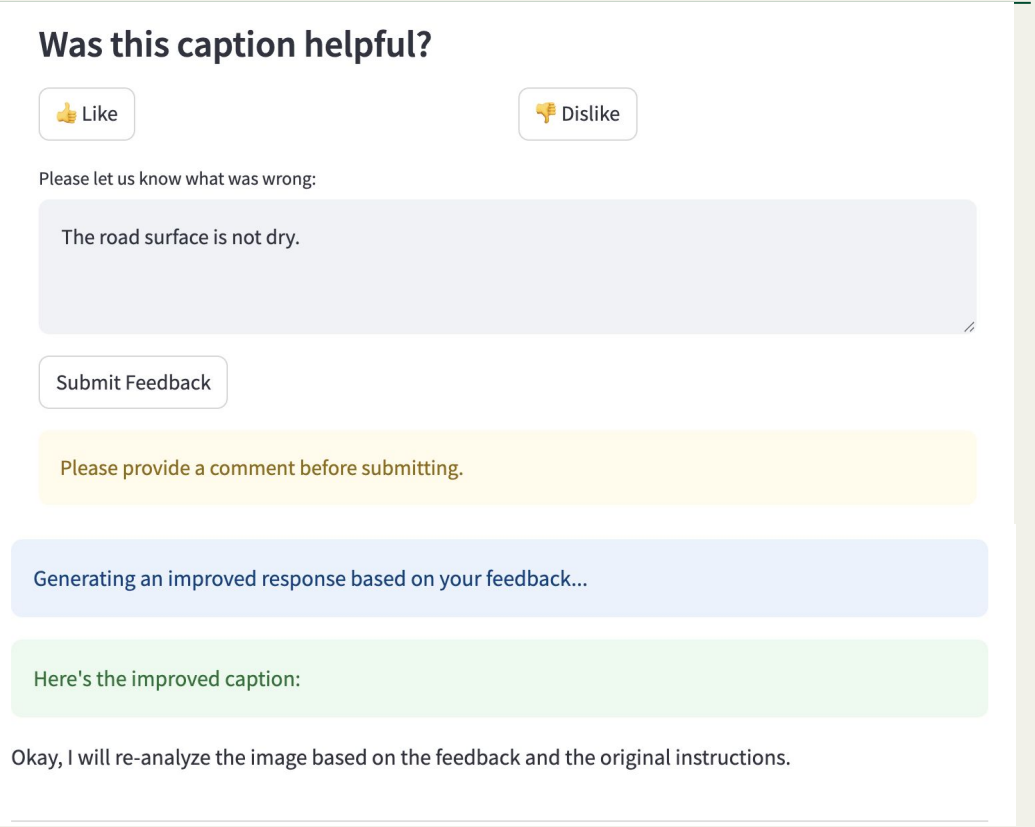
- Paragraph description
- Structured output table
- Risk level assessment
  - 3 levels → 5 levels
- csv saved output

Model Pipeline



User Feedback + Refined Output

- **Incorporate user inputs/suggestions**
  - Like
  - Dislike → comment → incorporate feedback → 2nd output generation



04 Final Product

Live Demo

05 Limitation

- **The model is not tested on a large dataset**
- **AI is not accountable for its suggestions and the usage of it is high risk (especially about people’s LIFE)**
- **video feature**
- **Relies on users to improve the output of the model**

06 Further Direction

- **Integrate multi-model for reflector & evaluator**
- **Larger Dataset and Fine tuning a model**
- **More user testing sample size**
- **Explore faster video processing model**