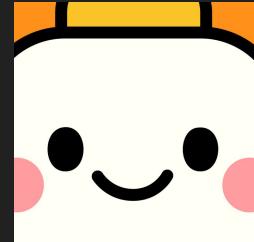


Autonomous Video Hunter: AI Agents for Real-Time OSINT

Kevin Dela Rosa

Recon Village @ DEFCON 33





Kevin Dela Rosa

kdr@cloudglue.dev



@kdrwins



Co-founder & CTO @
Cloudglue

Building agentic systems
that see & understand
video

- previously -



amazon

NIST



Carnegie Mellon University
Language
Technologies
Institute

SPAWAR



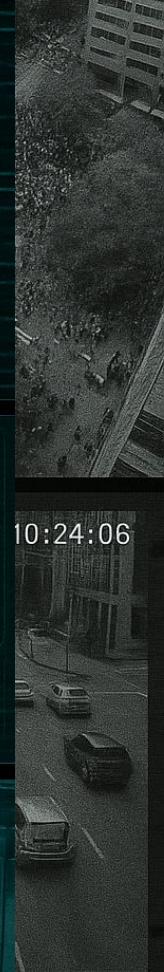
10:11

LIVE

The New OSINT Battlefield: Multimodal



LIVE



@usern
Lorem ip:

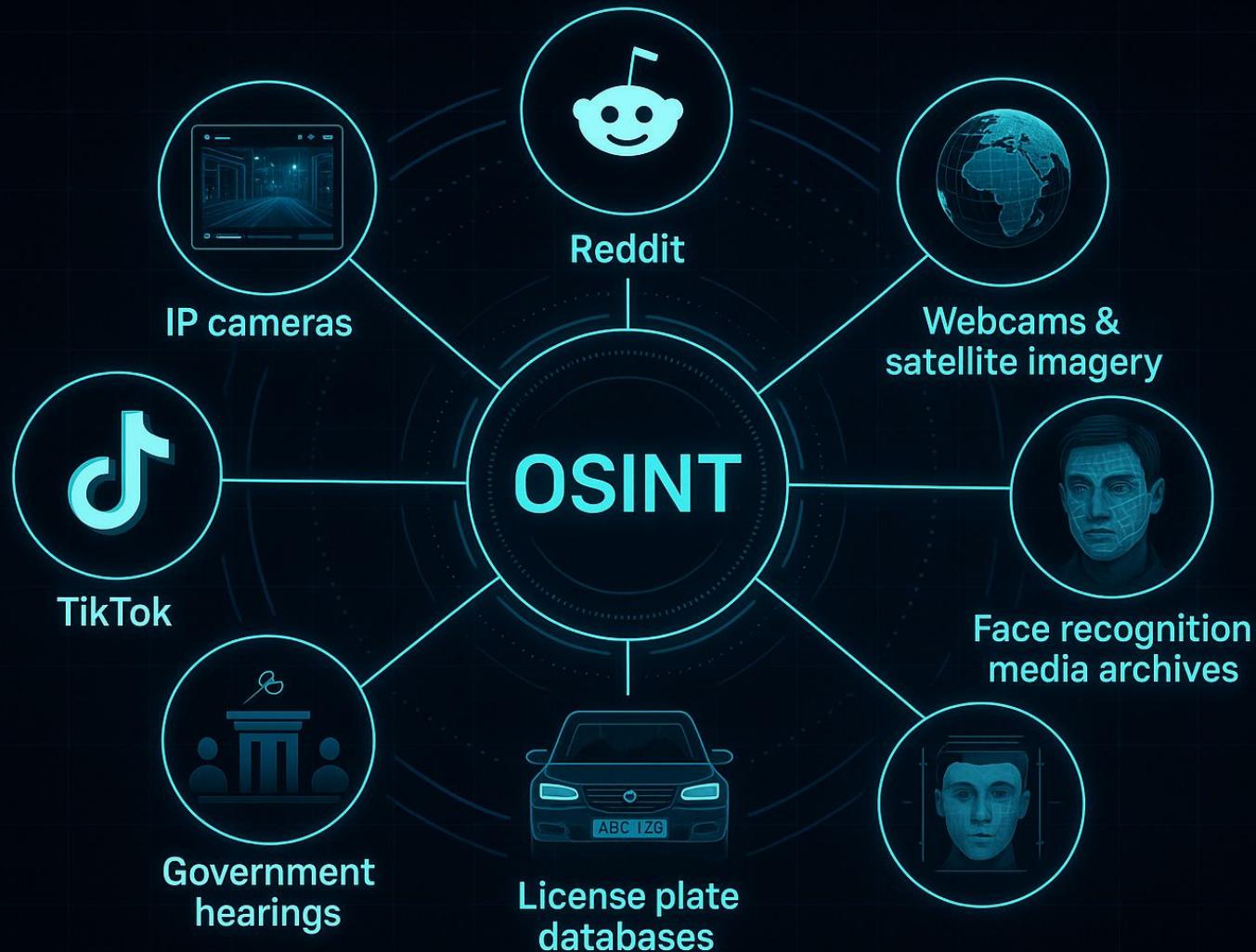


CAMERA 3



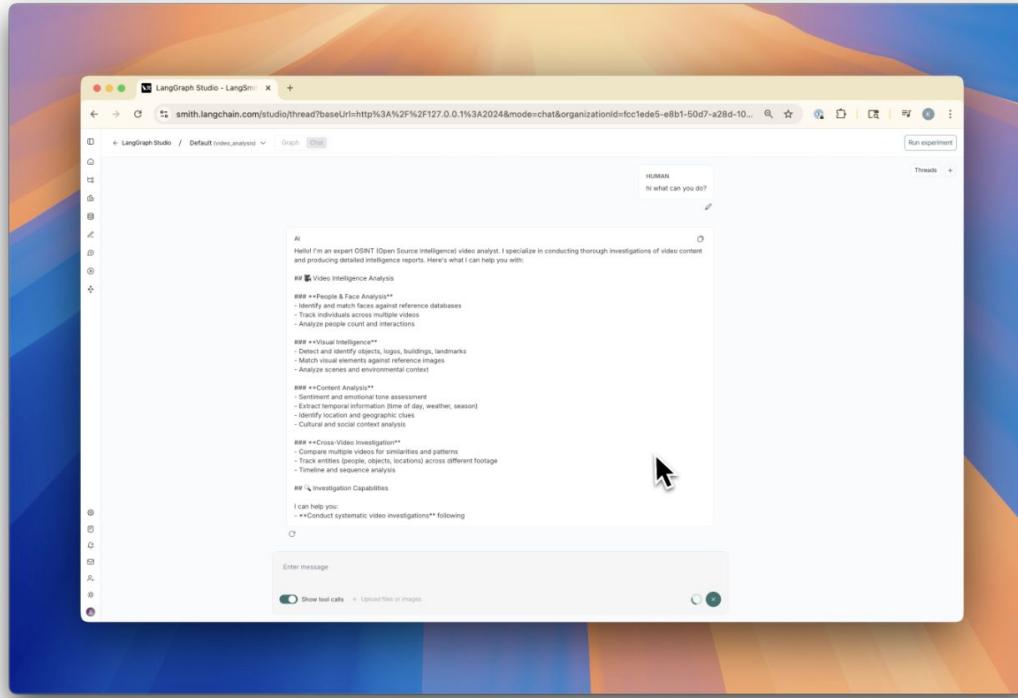
CAMERA 04

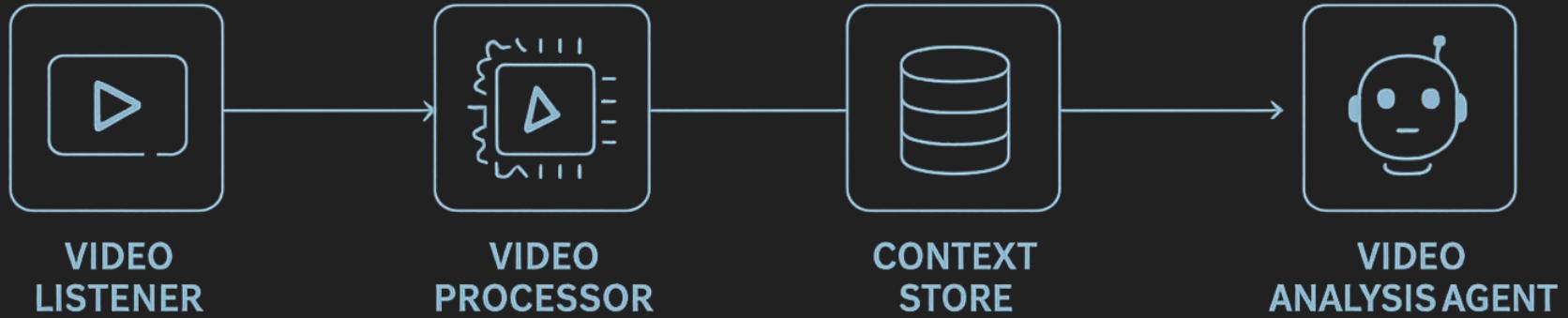




Autonomous Video Hunter

A proof-of-concept (POC) OSINT (Open Source Intelligence) video analysis system built with LangGraph. This implementation demonstrates how deep research agents can be used for automated video analysis, including face detection, object recognition, image matching, sentiment analysis, and cross-video correlation for intelligence gathering use cases.





Autonomous Video Hunter: Architecture



TARGET

OERLIKON

ZURICH



TARGET



ZURICH



TARGET

ZURICH





TARGET

ZURICH





TARGET

ZURICH





TARGET



ZURICH





TARGET



ZURICH





LOCATE D



ZURICH



Gumloop



Scrape popular tiktoks

listen in on the conversation on tiktok about a particular topic

kdr@aviaryhq.com
Personal Account

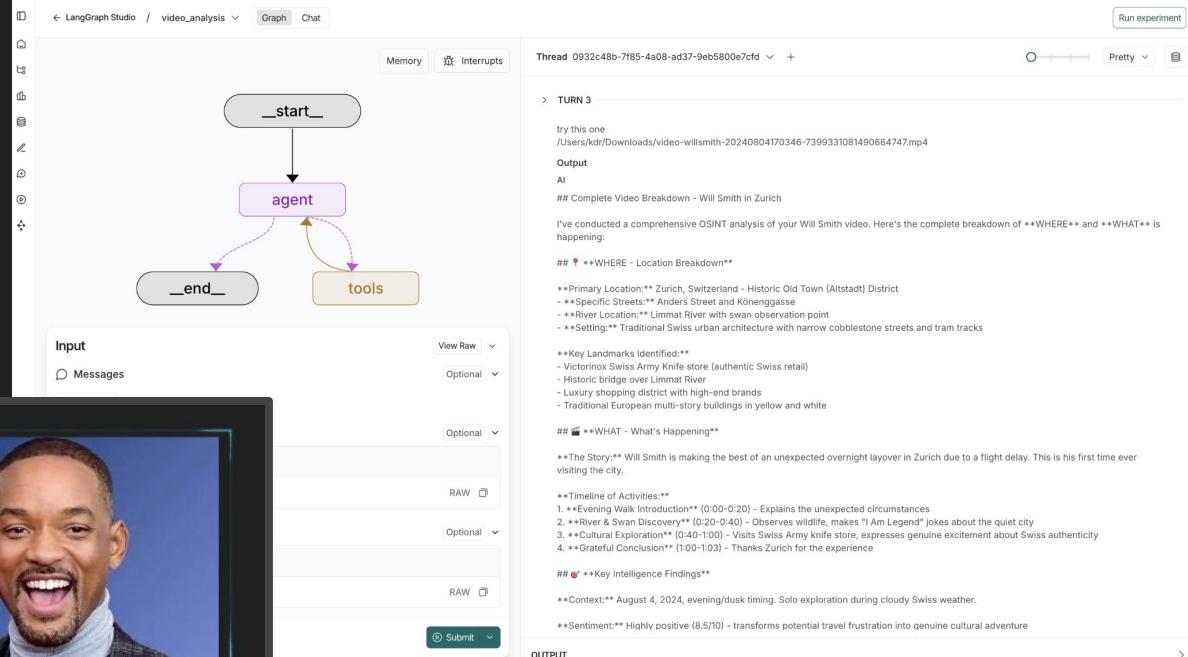
Hashtag

zurich

email

kdr@aviaryhq.com

Submit and run flow



Demo: Find Will

Social Media Intelligence Video Listener



Mon Aug 4 1:13AM

Gumloop | Scrape popular tiktoks | scrape tiktok | Gumloop | Tiktok videos zurich - kdr@aviaryhq.com | +

gumloop.com/interface/Scrape-Popular-Tiktoks-x8GVrRV5rB7rEv9aziMc9U

Gumloop View Workflow

SOCIAL MEDIA INTELLIGENCE LISTENER



Scrape popular tiktoks

listen in on the conversation on tiktok about a particular topic

kdr@aviaryhq.com Personal Account

Hashtag

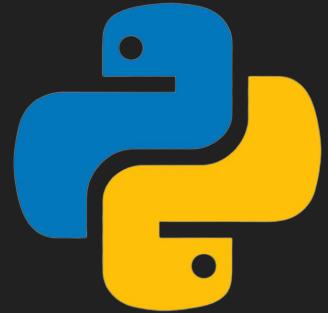
Enter text... *Required

email

Enter text... *Required

Submit and run flow

Video Processor + Context Store



video_processing_sample_code.py

...

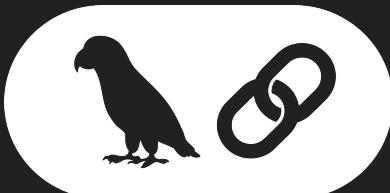
```
media > video_processing_sample_code.py > ...
1 import os
2 from video_processor import VideoProcessor
3 from video_helper import VideoHelper
4
5 vh = VideoHelper(media_dir="media")
6 vp = VideoProcessor(api_key=os.getenv("CLOUDGLUE_API_KEY"))
7
8 urls = [url.strip() for url in open('media_urls.txt', 'r').readlines()]
9 video_path = vh.download_from_url(urls[0])
10
11 result = vp.process(video_path,
12 save_to_db=True)
13
14
15
16
17
18
19
```

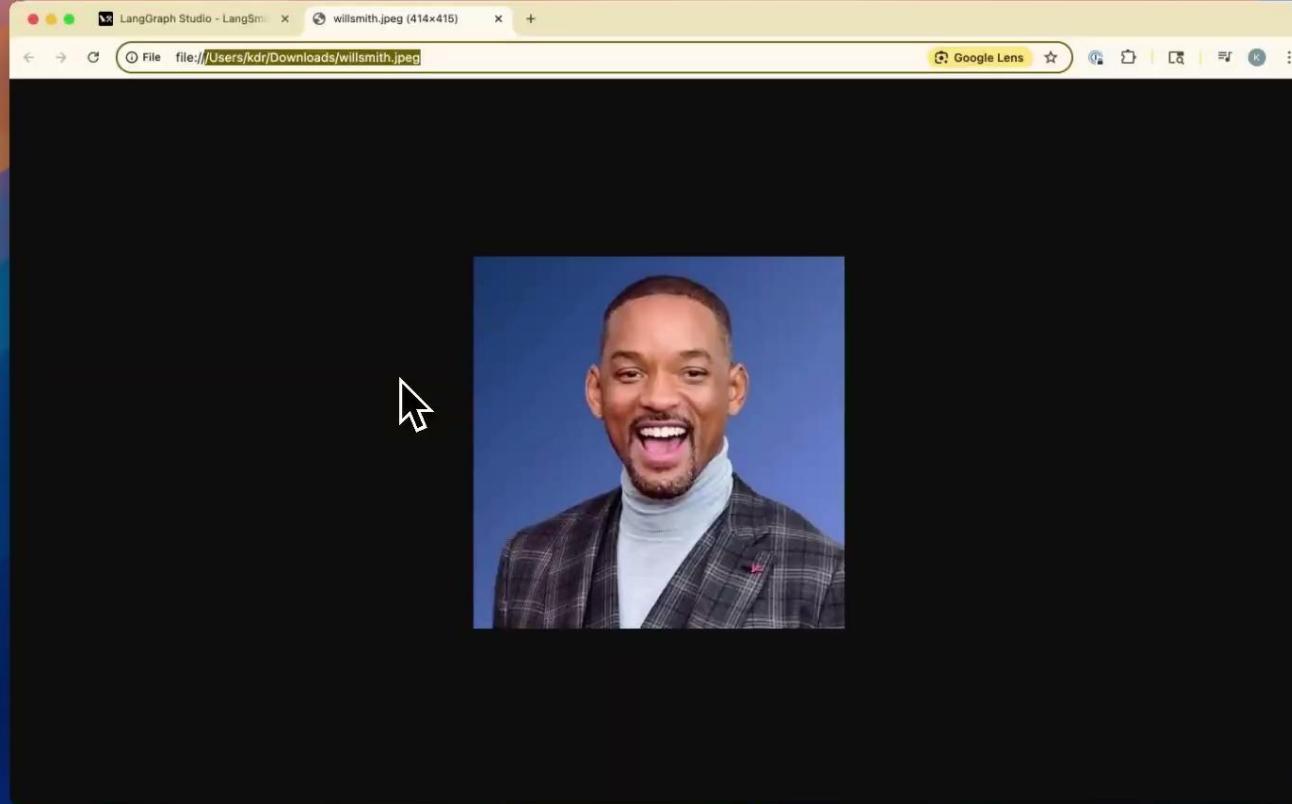
{} abbreviated.jsonl

media > {} abbreviated.jsonl

```
1 {
2     "video": "video-tomdaley-20210805.mp4",
3     "cloudglue_uri": "cloudglue://files/81dcd53c...",
4     "duration": 23.4,
5     "description": """# Video Document
6     ## Title
7     Tom Daley's Olympic Cardigan Reveal
8     ## Summary
9     Tom Daley presents his finished Olympic cardigan outdoors, pointing out details
10    like "Tokyo" in Japanese, the Union Jack, "GBR", and Olympic rings. He narrates
11    the details and invites opinions, with bright visuals and bold on-screen text.
12    ## Scenes
13    ### Scene [00:00 - 00:23]
14    **Visual:** Tom in white cardigan under blue sky, close to camera
15    **Speech:** "I finished the jumper... this says Tokyo... GBR... Union Jack..."
16    **Text on Screen:** "FINISHED MY | OLYMPIC CARDIGAN!"
17    ..."",
18    "frames": ...,
19    "has_logo": true,
20    "logos": ["Adidas", "Team GB", "Olympic Rings"],
21    "has_face": true,
22    "has_speech": true,
23    "has_text_on_screen": true,
24    "is_outdoors": true
}
```

Video Analysis Agent





Investigation Complete

I found one positive match for the person in your reference image across the 8 Zurich-related videos analyzed.

Key Findings:

POSITIVE IDENTIFICATION

- **Video:** [video-willsmith-20240804170346-7399331081490664747.mp4](#)
- **Frame Location:** /Users/kdr/dev/github/aviary-labs/agent-pocs/media/cbc7e344-a673-4304-8358-4d2913647d09.jpg
- **Cropped Face:** /Users/kdr/dev/github/aviary-labs/agent-pocs/faces/05187c99-ac84-4384-9ad2-1c9a21753759.jpg
- **Timestamp:** 21.013 seconds into the video
- **Confidence:** 0.357 (medium confidence)

Context of Zurich Visit:

Why he was in Zurich:

- Unplanned overnight stay due to flight disruption - his plane couldn't take off
- This was his first time ever visiting Zurich
- Made the best of travel delays by exploring the city

What he was doing:

- Solo evening walk through central Zurich streets
- Observing swans in the river
- Window shopping at Swiss Army knife store (Victorinox)
- Exploring luxury shopping areas (Anders Street, Könengasse)
- Creating content about his spontaneous city exploration
- Making playful "I Am Legend" references while looking for his dog

His emotional state:

- Very positive and delighted - called it "a treat for my life"
- Amazed by the city's quietness compared to his usual experiences
- Excited about experiencing authentic Swiss culture
- Grateful for the unexpected opportunity to explore

The person appears relaxed, walking alone without visible security, and genuinely enjoying this unplanned detour in Zurich during what was likely August 2024.

Investigation Complete

I found one positive match for the person in your reference image across the 8 Zurich-related videos analyzed.

Key Findings:

POSITIVE IDENTIFICATION

- **Video:** [video-willsmith-20240804170346-7399331081490664747.mp4](#)
- **Frame Location:** /Users/kdr/dev/github/aviary-labs/agent-pocs/media/cbc7e344-a673-4304-8358-4d2913647d09.jpg
- **Cropped Face:** /Users/kdr/dev/github/aviary-labs/agent-pocs/faces/05187c99-ac84-4384-9ad2-1c9a21753759.jpg
- **Timestamp:** 21.013 seconds into the video
- **Confidence:** 0.357 (medium confidence)

Context of Zurich Visit:

Why he was in Zurich:

- Unplanned overnight stay due to flight disruption - his plane couldn't take off
- This was his **first time ever** visiting Zurich
- Made the best of travel delays by exploring the city

What he was doing:

- Solo evening walk through central Zurich streets
- Observing swans in the river
- Window shopping at Swiss Army knife store (Victorinox)
- Exploring luxury shopping areas (Anders Street, Könengasse)
- Creating content about his spontaneous city exploration
- Making playful "I Am Legend" references while looking for his dog

His emotional state:

- Very positive and delighted - called it "a treat for my life"
- Amazed by the city's quietness compared to his usual experiences
- Excited about experiencing authentic Swiss culture
- Grateful for the unexpected opportunity to explore

The person appears relaxed, walking alone without visible security, and genuinely enjoying this unplanned detour in Zurich during what was likely August 2024.



Investigation Complete

I found one positive match for the person in your reference image across the 8 Zurich-related videos analyzed.

Key Findings:

POSITIVE IDENTIFICATION

- **Video:** [video-willsmith-20240804170346-7399331081490664747.mp4](#)
- **Frame Location:** /Users/kdr/dev/github/aviary-labs/agent-pocs/media/cbc7e344-a673-4304-8358-4d2913647d09.jpg
- **Cropped Face:** /Users/kdr/dev/github/aviary-labs/agent-pocs/faces/05187c99-ac84-4384-9ad2-1c9a21753759.jpg
- **Timestamp:** 21.013 seconds into the video
- **Confidence:** 0.357 (medium confidence)

Context of Zurich Visit:

Why he was in Zurich:

- Unplanned overnight stay due to flight disruption - his plane couldn't take off
- This was his **first time ever** visiting Zurich
- Made the best of travel delays by exploring the city

What he was doing:

- Solo evening walk through central Zurich streets
- Observing swans in the river
- Window shopping at Swiss Army knife store (Victorinox)
- Exploring luxury shopping areas (Anders Street, Königgasse)
- Creating content about his spontaneous city exploration
- Making playful "I Am Legend" references while looking for his dog

His emotional state:

- Very positive and delighted - called it "a treat for my life"
- Amazed by the city's quietness compared to his usual experiences
- Excited about experiencing authentic Swiss culture
- Grateful for the unexpected opportunity to explore

The person appears relaxed, walking alone without visible security, and genuinely enjoying this unplanned detour in Zurich during what was likely August 2024.



Investigation Complete

I found one positive match for the person in your reference image across the 8 Zurich-related videos analyzed.

Key Findings:

POSITIVE IDENTIFICATION

- **Video:** [video-willsmith-20240804170346-7399331081490664747.mp4](#)
- **Frame Location:** /Users/kdr/dev/github/aviary-labs/agent-pocs/media/cbc7e344-a673-4304-8358-4d2913647d09.jpg
- **Cropped Face:** /Users/kdr/dev/github/aviary-labs/agent-pocs/faces/05187c99-ac84-4384-9ad2-1c9a21753759.jpg
- **Timestamp:** 21.013 seconds into the video
- **Confidence:** 0.357 (medium confidence)

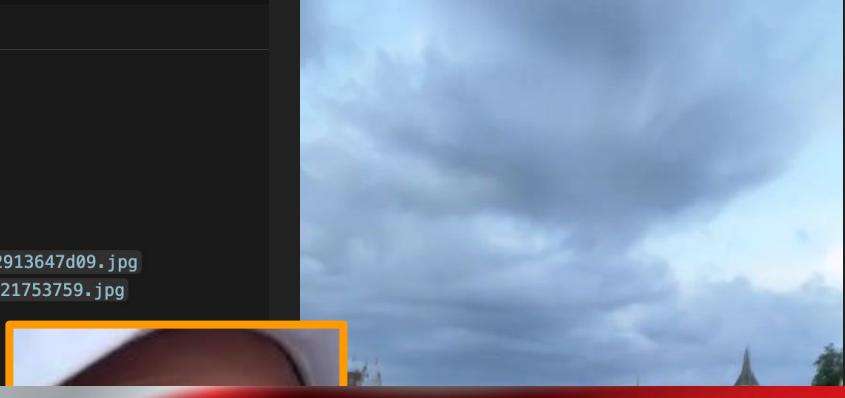
MISSION ACCOMPLISHED

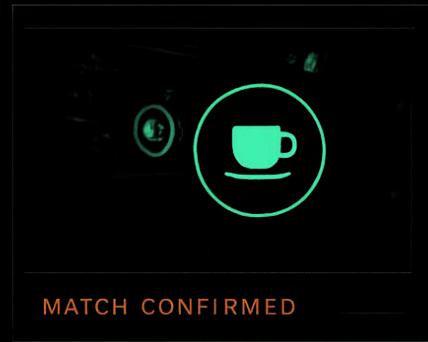
- Solo evening walk through central Zurich streets
- Observing swans in the river
- Window shopping at Swiss Army knife store (Victorinox)
- Exploring luxury shopping areas (Anders Street, Könengasse)
- Creating content about his spontaneous city exploration
- Making playful "I Am Legend" references while looking for his dog

His emotional state:

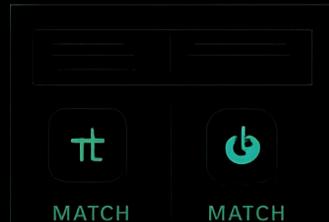
- Very positive and delighted - called it "a treat for my life"
- Amazed by the city's quietness compared to his usual experiences
- Excited about experiencing authentic Swiss culture
- Grateful for the unexpected opportunity to explore

The person appears relaxed, walking alone without visible security, and genuinely enjoying this unplanned detour in Zurich during what was likely August 2024.





MATCH
CONFIRMED



40.712° N
74.005° W

18-23:45
02:17-2024



TARGET

Demo: Logo -> Context Intelligence

Investigation Complete

Based on my comprehensive analysis of the logo `/Users/kdr/Downloads/sb-target.jpg`, here are the key findings:

Answer to Your Question:

The logo appears in 4 out of 15 videos in the database.

Videos Containing the Logo:

1. `video-therealdri-20250703184117` (HIGHEST CONFIDENCE)

- 2 strong matches: 61.1% and 58.4% confidence
- Context: Starbucks promotional content showing exterior signage

2. `video-blondebeac-20250726205821`

- 2 matches: 46.5% and 42.5% confidence
- Context: Customer testimonial featuring Starbucks cup

3. `video-iheartdest-20250704130649`

- 1 match: 44.4% confidence
- Context: Employee demonstration video behind Starbucks counter

Highest Confidence Frame Paths:

- **61.1% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/ce45b1a0-b9a8-41de-a1cf-1a1155dfe61a.jpg`
- **58.4% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/d02f6f77-bd16-417a-9e41-187832f6a262.jpg`
- **46.5% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/b199e3ee-5efc-42e9-b650-8213616ef9ef.jpg`

Aggregate Context Summary:

The logo appears in three distinct contexts:

1. **Exterior building signage** (highest confidence matches)
2. **Product packaging** on cups and containers
3. **Employee workspace equipment/signage**

All appearances are in legitimate Starbucks-related content - customer testimonials, employee demonstrations, and promotional material. The logo appears to be the official Starbucks brand mark, with all contexts aligning with authentic brand usage.

Investigation Complete

Based on my comprehensive analysis of the logo `/Users/kdr/Downloads/sb-target.jpg`, here are the key findings:

Answer to Your Question:

The logo appears in 4 out of 15 videos in the database.

Videos Containing the Logo:

1. `video-therealdri-20250703184117` (HIGHEST CONFIDENCE)

- 2 strong matches: 61.1% and 58.4% confidence
- Context: Starbucks promotional content showing exterior signage

2. `video-blondebeac-20250726205821`

- 2 matches: 46.5% and 42.5% confidence
- Context: Customer testimonial featuring Starbucks cup

3. `video-iheartdest-20250704130649`

- 1 match: 44.4% confidence
- Context: Employee demonstration video behind Starbucks counter

Highest Confidence Frame Paths:

• **61.1% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/ce45b1a0-b9a8-41de-a1cf-1a1155dfe61a.jpg`

• **58.4% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/d02f6f77-bd16-417a-9e41-187832f6a262.jpg`

• **46.5% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/b199e3ee-5efc-42e9-b650-8213616ef9ef.jpg`

Aggregate Context Summary:

The logo appears in three distinct contexts:

1. **Exterior building signage** (highest confidence matches)
2. **Product packaging** on cups and containers
3. **Employee workspace equipment/signage**

All appearances are in legitimate Starbucks-related content - customer testimonials, employee demonstrations, and promotional material. The logo appears to be the official Starbucks brand mark, with all contexts aligning with authentic brand usage.



Investigation Complete

Based on my comprehensive analysis of the logo `/Users/kdr/Downloads/sb-target.jpg`, here are the key findings:

Answer to Your Question:

The logo appears in 4 out of 15 videos in the database.

Videos Containing the Logo:

1. `video-therealdri-20250703184117` (HIGHEST CONFIDENCE)
 - 2 strong matches: 61.1% and 58.4% confidence
 - Context: Starbucks promotional content showing exterior signage
2. `video-blondebeac-20250726205821`



MISSION ACCOMPLISHED

`1a1155dfe61a.jpg`

- **58.4% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/d02f6f77-bd16-417a-9e41-187832f6a262.jpg`
- **46.5% confidence:** `/Users/kdr/dev/github/aviary-labs/agent-pocs/media/b199e3ee-5efc-42e9-b650-8213616ef9ef.jpg`

Aggregate Context Summary:

The logo appears in three distinct contexts:

1. **Exterior building signage** (highest confidence matches)
2. **Product packaging** on cups and containers
3. **Employee workspace equipment/signage**

All appearances are in legitimate Starbucks-related content - customer testimonials, employee demonstrations, and promotional material. The logo appears to be the official Starbucks brand mark, with all contexts aligning with authentic brand usage.



LOCATE
D

Under the hood

AUTONOMOUS VIDEO HUNTER

VIDEO
PROCESSOR

FACE
MATCHER

VIDEO
ANALYSIS
AGENT

IMAGE
MATCHER

ZERO-SHOT
OBJECT
DETECTOR

General Video Understanding

A dense multimodal description of the video, with short generated title, summary, and a play-by-play of each scene including:

- Visual descriptions
- Automatic Speech Transcription
- On Screen Text (OCR)

Used to populate the initial general video context provided to the agents for a given video

Summary

This video documents a Maryland Public Service Commission hearing held on May 22, 2024, regarding the application for a certificate of public convenience and necessity for the Wild Turkey community solar proj....

Scene [00:00 - 00:20]

Video Conference Call

A video conference call between multiple participants.

Visual Content:

- [00:00 - 00:20]

- The video begins with a black screen displaying the name "Doug".
- The scene transitions to a man identified as Law Judge.
- He is wearing a suit jacket

...

Speech:

- [00:02 - 00:03] Alright.

- [00:03 - 00:05] Good morning, everybody.

- [00:05 - 00:06] Everybody hear me okay

...

On-screen Text:

- [00:00 - 00:20] "Doug" | "Law Judge" | "Andy Flavin"



Targeted Extraction

A targeted structured extraction from each video keying into common points of interest for user queries as well as agent reasoning, including:

- presence of logos
- presence of faces
- presence of speech
- is setting outdoors
- presence of text

```
1 ✓ extract_prompt = """  
2 Extract whether the video contains:  
3     - a visually identifiable human face  
4     - spoken narration  
5     - is outdoors  
6     - has text on screen  
7     - has logos  
8     - visible logo name (provide if known)  
9 """  
10  
11 ✓ extract_schema = {  
12     "has_face": False,  
13     "has_speech": True,  
14     "is_outdoors": False,  
15     "has_text_on_screen": False,  
16     "logos": ["name of logo if known"],  
17     "has_logo": True,  
18 }  
 {  
     "has_logo": true,  
     "logos": ["Louis Vuitton", "S.RIDE"],  
     "has_face": true,  
     "has_speech": true,  
     "is_outdoors": true,  
     "has_text_on_screen": true  
 }
```

Face Matcher

Compares reference images of face(s) against provided video frames to find matches

Deep learning models used

- Retina Face (face bounding box detector)
- VGG Face (face understanding model)

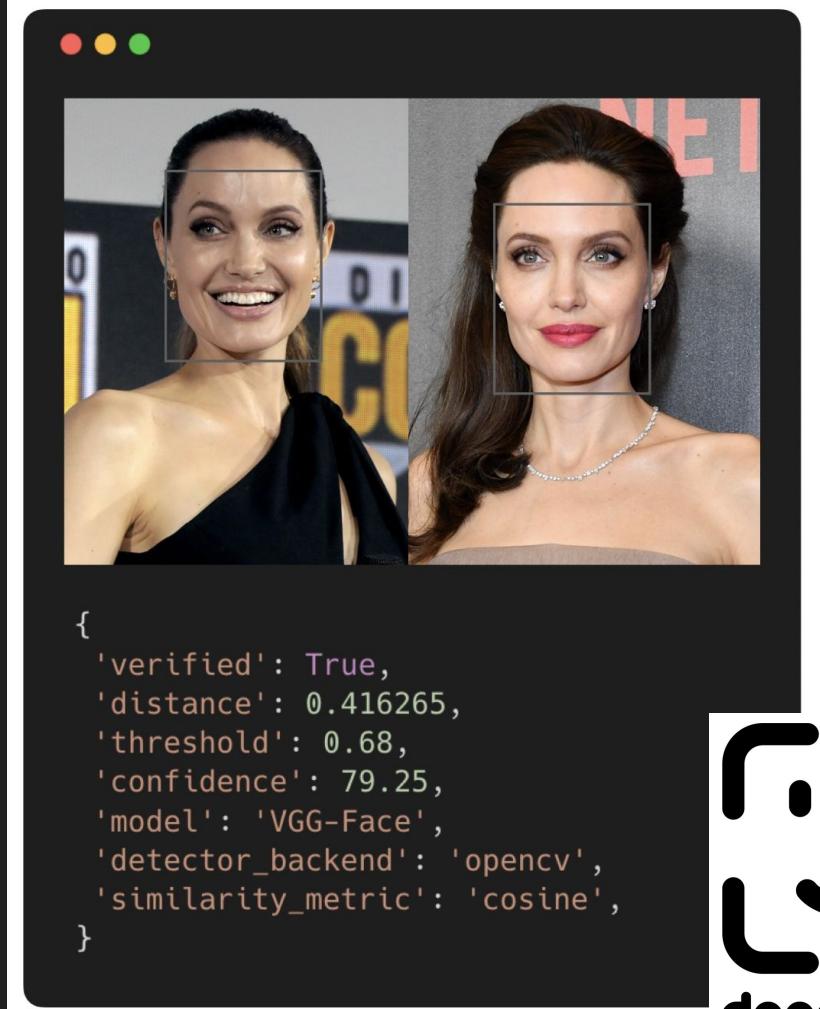


Image Matcher

(near) exact matches to reference images, helps spot:

- Logos
- Posters
- Buildings

Powered by OpenCV using

- SIFT/ORB based keypoint detectors
- RANSAC based matching algorithm



Zero Shot Object Detector

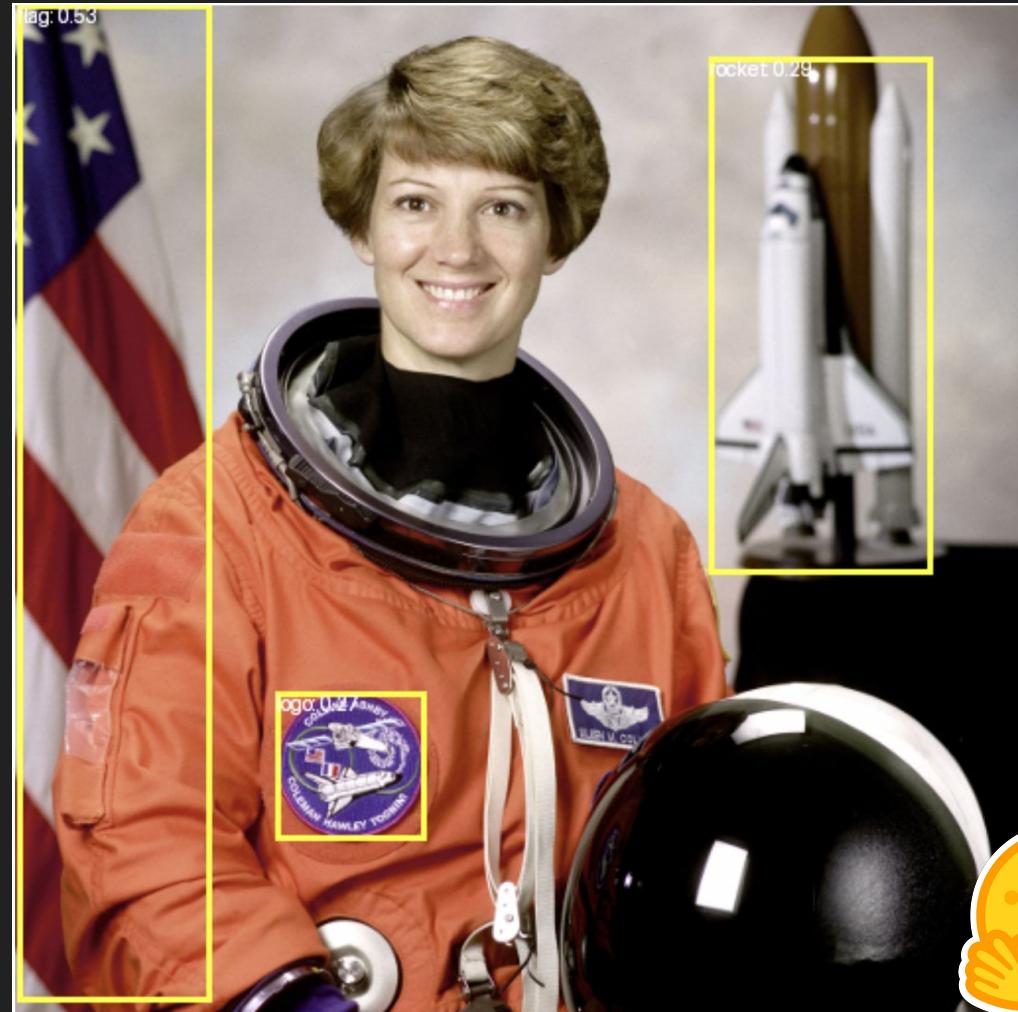
Helps localize arbitrary objects (vehicles, rockets, logos, backpacks, etc...), finding specific concepts on the fly

Uses hugging face transformer library to run inference of Owl-V2 model

[google/owlv2-base-patch16-ensemble](https://github.com/google/owlv2-base-patch16-ensemble)

Ref: Scaling Open-Vocabulary Object Detection

<https://arxiv.org/abs/2306.09683>



Visual Analysis (Deep) Agent

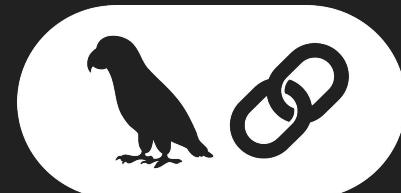
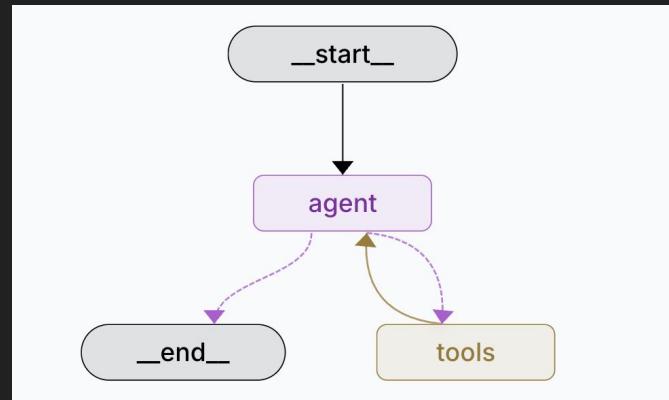
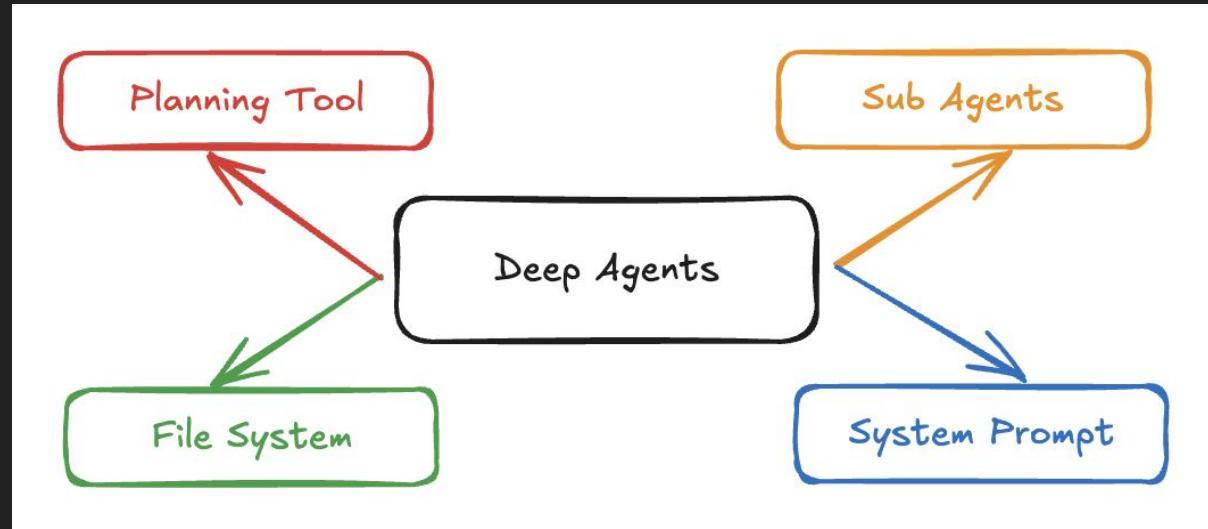
The main agent is a ReAct based deep research agent featuring:

- Planning
- Sub Agents
- Access to file system
- Detailed system prompt

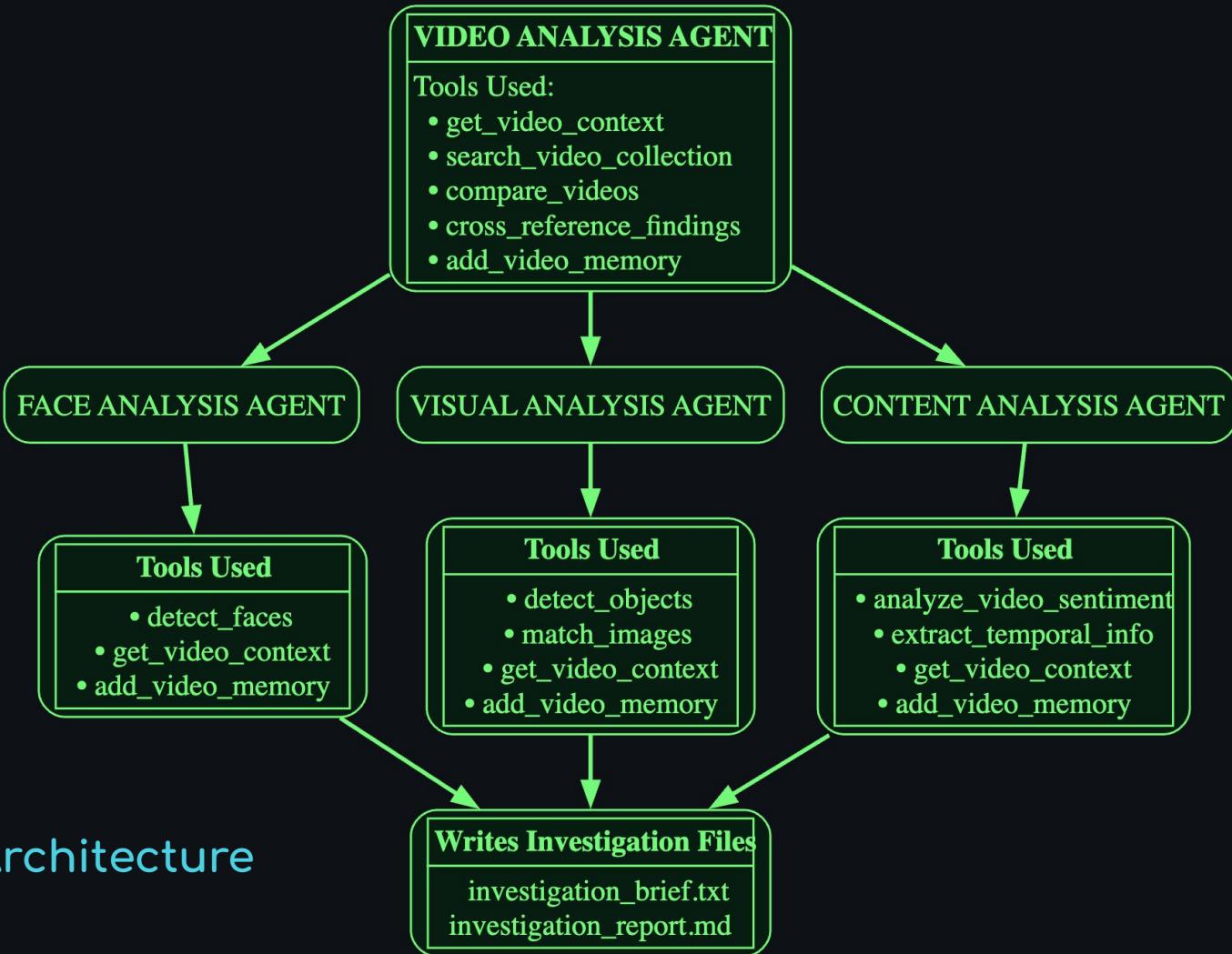
Our example agent has 3 sub agents:

- Facial Analysis Agent
- Visual Analysis Agent
- Content Analysis Agent

Interacts with a RAM based “Video Agent Memory” for storing context across conversation / cached video analysis output



Agent Architecture



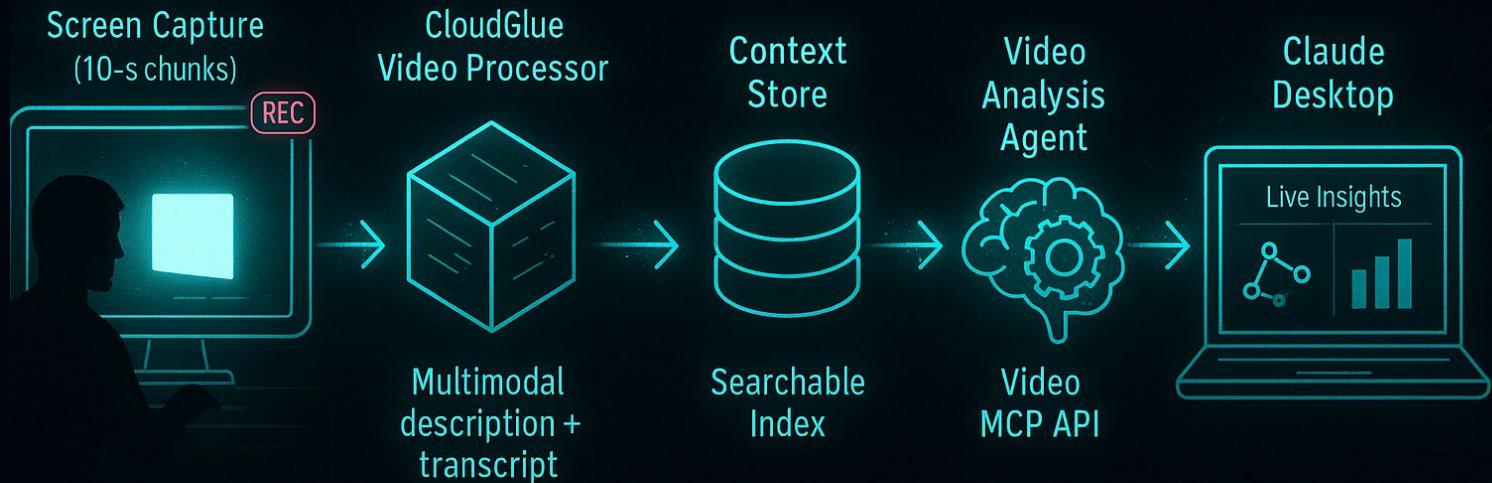
Applications /
Deployment Ideas



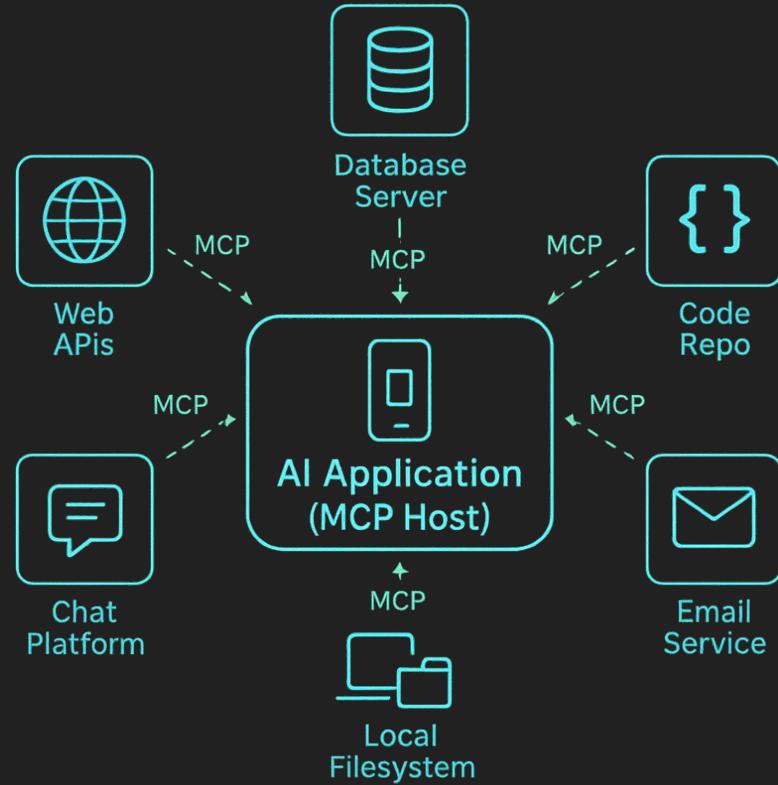
REAL-TIME SUMMARIZATION

Example Real-Time
Ad-hoc
Summarization

REAL-TIME SCREEN-TO-INSIGHT WORKFLOW



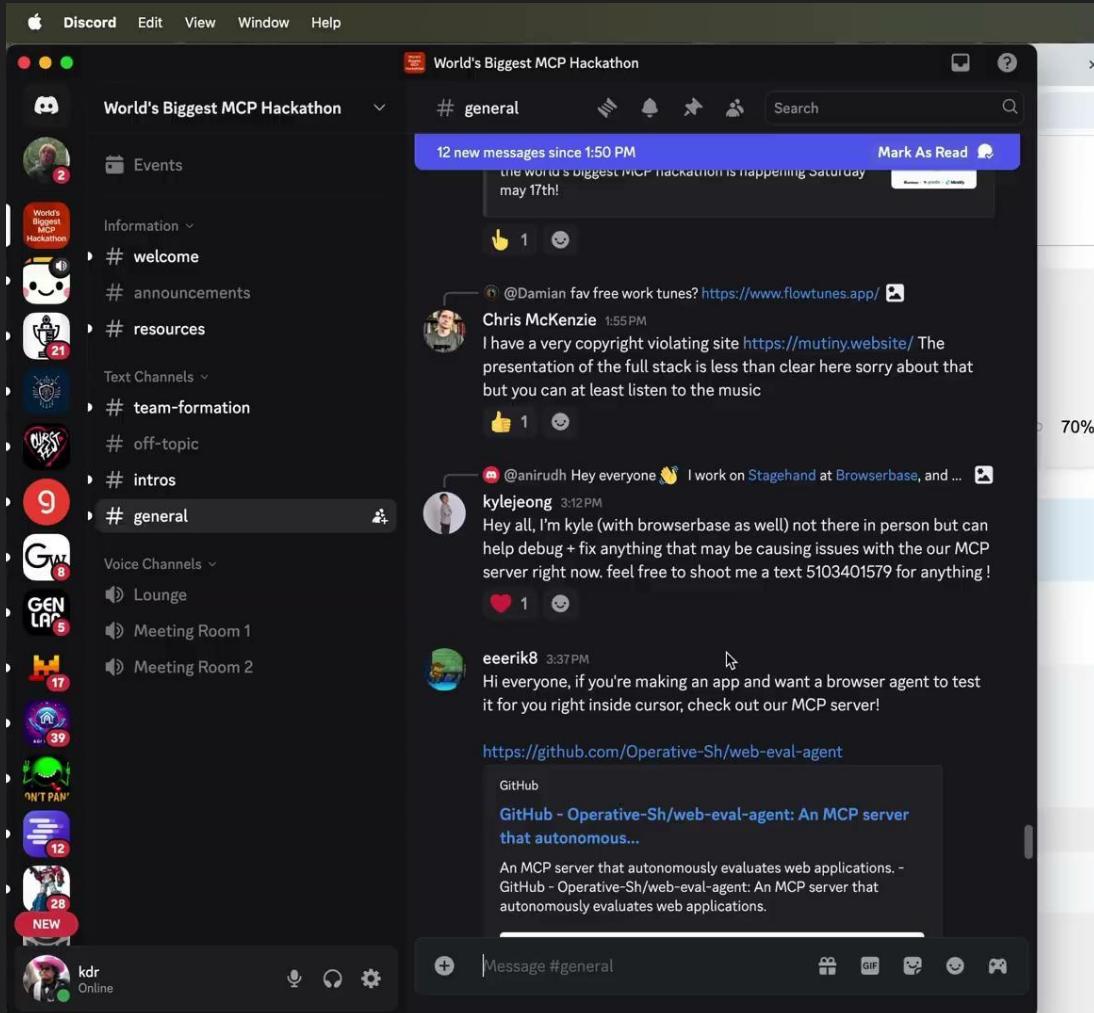
WHAT IS MCP?

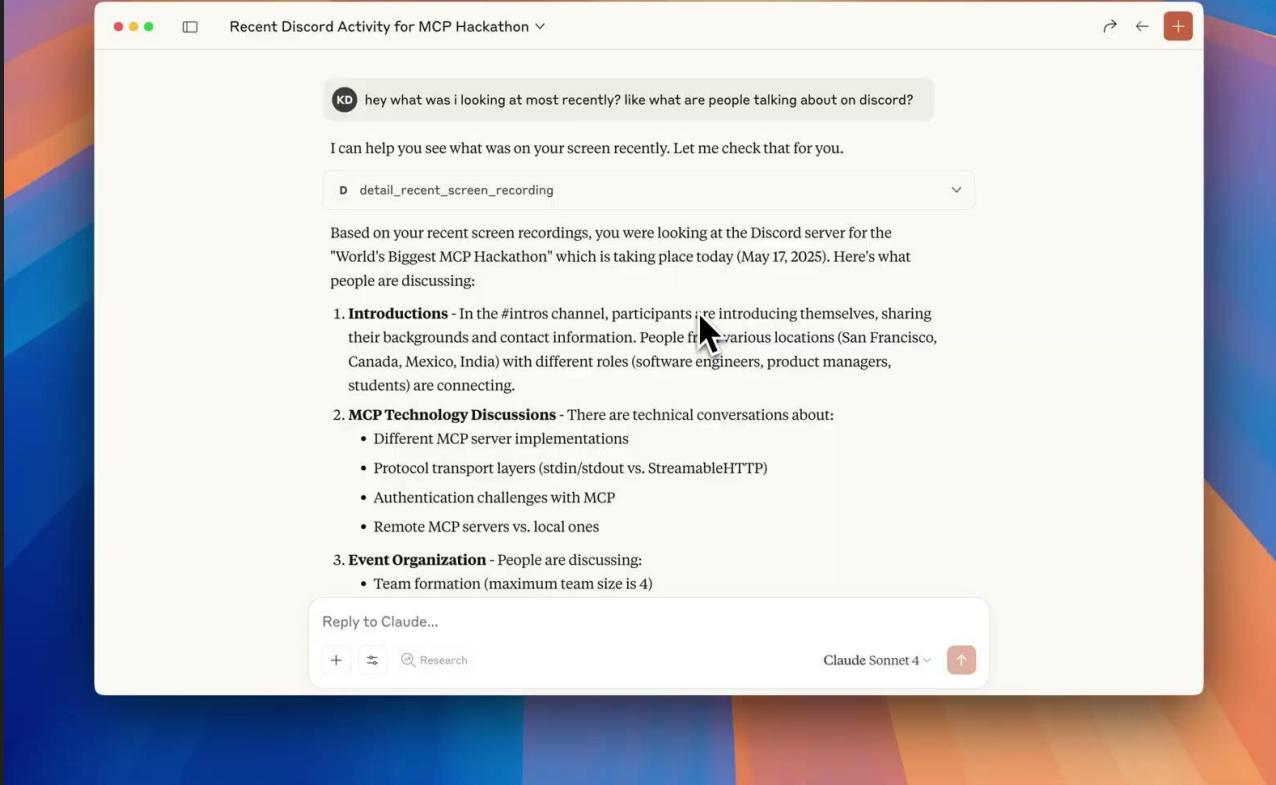


Setup

- Ad-hoc screen recorder setup on target computer
- Screen recordings uploaded to video processor at regular intervals (10 seconds) for indexing
- An MCP server deployed to surface insights using your favorite AI chat bot app (demo with claude)

Example: scrolling through discord channels during a hackathon





Cloudglue VideoMCP x Claude Desktop

APPLICATION PATTERNS

PROTEST MONITORING



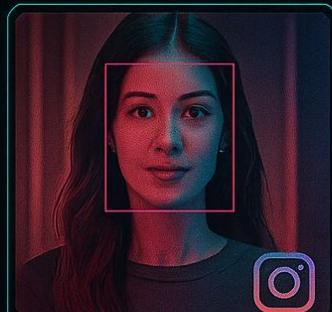
Detect flags & crowd mood

TRAFFIC DISPUTES



Extract plate numbers

INFLUENCER TRACKING



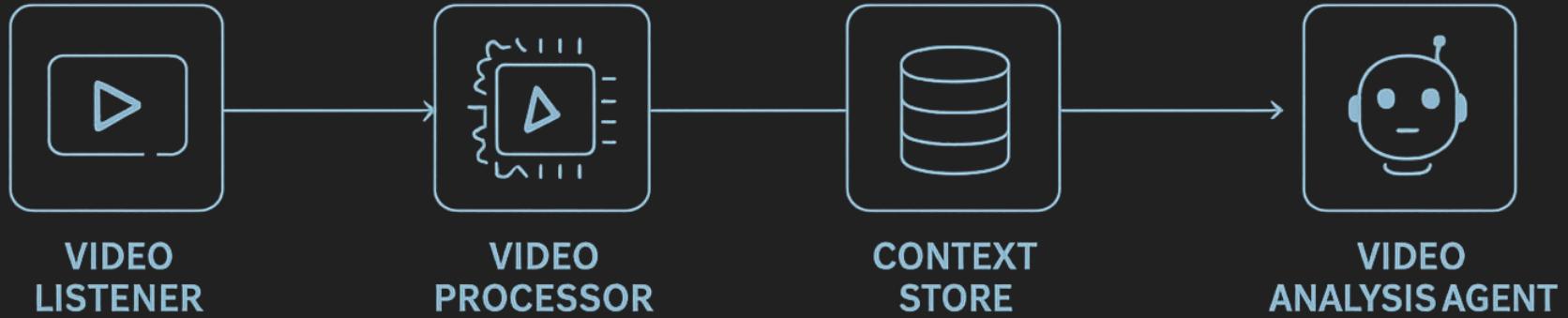
Track logo & person

GOVERNMENT SCRUTINY



Compare cross-platform

Recap



Autonomous Video Hunter: Architecture

CHALLENGES & ETHICS

PRECISION ↔ RECALL



Precision
↔ Recall

TRANSPARENCY ↔ SPEED



Transparency
↔ Speed

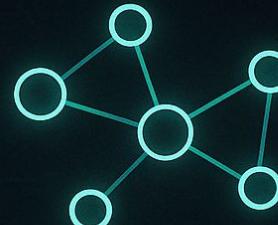
Cost ↔
Coverage



AUTONOMOUS VIDEO HUNTER

BEFORE – NOISE

KEY FINDINGS



AFTER – INTELLIGENCE

Limitations / Notes

- Proof of concept to demonstrate video tools can naturally play in deep research style agents
- Biased towards tools that can run locally / on CPU over SOTA



Kevin Dela Rosa

kdr@cloudglue.dev



@kdrwins



Thanks for listening!



Repos of interest

Main repo for this talk

- <https://github.com/kdr/autonomous-video-hunter>

Cloudglue VideoMCP

- <https://github.com/aviaryhq/cloudglue-mcp-server>

Video Screen Recorder / Copilot

- <https://github.com/kdr/video-screen-copilot>