

UAV Flight Log Generation and High-Level Mission Planning Using Vision-Language Models

Author: Tsung-Yen, Yu

Advisor: Dr. Wen-Hung Liao

content

- Abstract & Background
- Methods
- Result
- Future

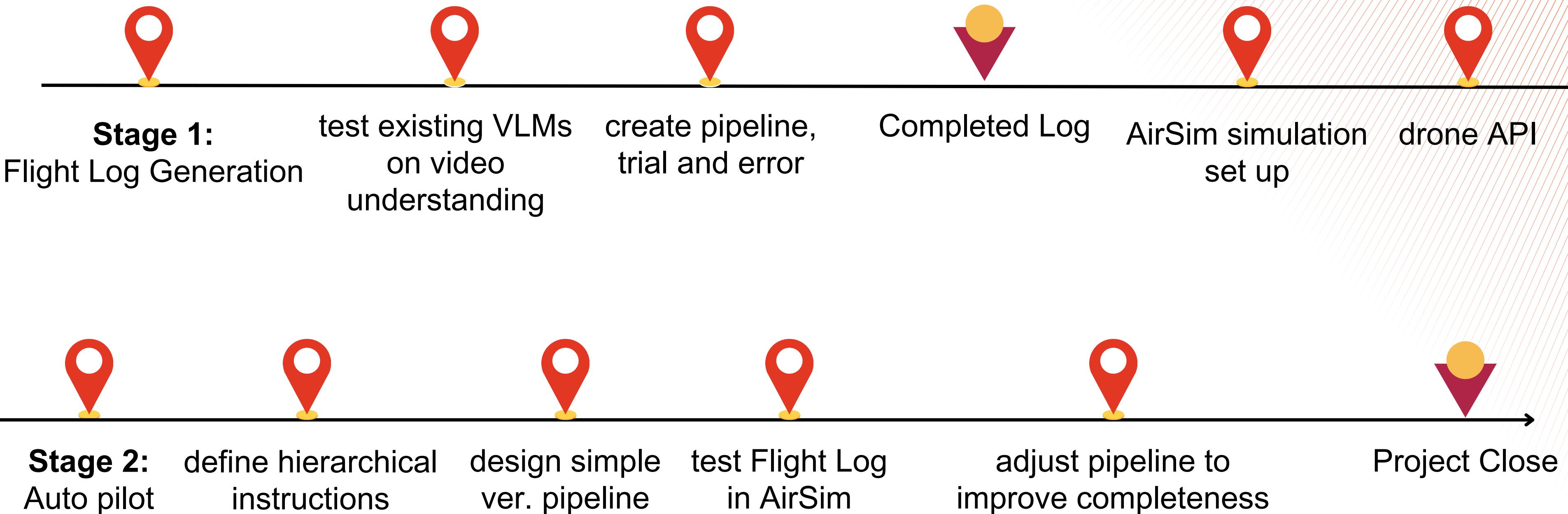
Abstract & Background

- The rapid rise of AI in computer vision is accelerating the future of autonomous systems
- For drones, this opens the door to move beyond simple GPS routes to true situational awareness
- Simultaneously, advances in NLP let us "talk" to AI in new ways

The Goal: To merge these fields, enabling a human to effortlessly collaborate with a drone

1. Integrate Vision-Language Models (VLMs) with drone control via multi-layer pipeline
2. Use a logbook-like approach to document the flight details

Methods - Overview



Stage 1: Flight Log Generation



Stage 1:
Flight Log Generation

test existing VLMs
on video
understanding

create pipeline,
trial and error

Methods - VLM Selection

To test existing VLMs, I first tested their ability to summarize a few consecutive frames.



- LLaVA (13b):
 - Run on local machine, fastest and cost-free
 - NLP backbone is too weak to reliably solve complex vision tasks



- ChatGpt:
 - Requires API key access
 - Richest contextual descriptions at a moderate speed



- Claude:
 - Requires API key access
 - Accurate contextual descriptions at a high speed
 - Accepts the longest context window



- Gemini:
 - Requires API key access
 - Accurate contextual descriptions at a moderate speed
 - Provides video understanding capabilities

← Act as the “eyes”

← Act as the “brain”

← Act as the
“second agent”



Methods - VLM Selection

Stage 1:
Flight Log Generation

test existing VLMs
on video
understanding

create pipeline,
trial and error

Base on the vision ability of the VLMs, decide the possible Flight Log Formation:

1. Flight Identification

- Date
- Start Time
- End Time
- Total Duration

2. Flight Purpose

- Purpose of Flight
- Type of Operation

3. Environment Observation

- Location Name/Description
- GPS Coordinates (Takeoff)
- GPS Coordinates (Landing)
- Weather Conditions:
 - Wind Speed, Wind Direction, Visibility, Temperature, Cloud Cover, Precipitation

4. Flight Parameters

- Maximum Altitude
- Maximum Distance
- Flight Pattern: Hover Linear
 Orbit Waypoint Free Flight
- Key Waypoints/Locations
- Flight Path Summary

5. Camera & Recording Settings

- Video Resolution
- Frame Rate
- Recording Format

6. Safety Consideration

- Obstacles Present
- People in Area
- Emergency Landing Sites

7. Incidents

- Any Issues Encountered
- Wildlife Interactions
- Signal Loss Events
- Weather Changes
- Equipment Malfunctions

8. Notes & Lessons Learned

- Flight Performance
- Footage Quality
- Areas for Improvement
- Future Considerations



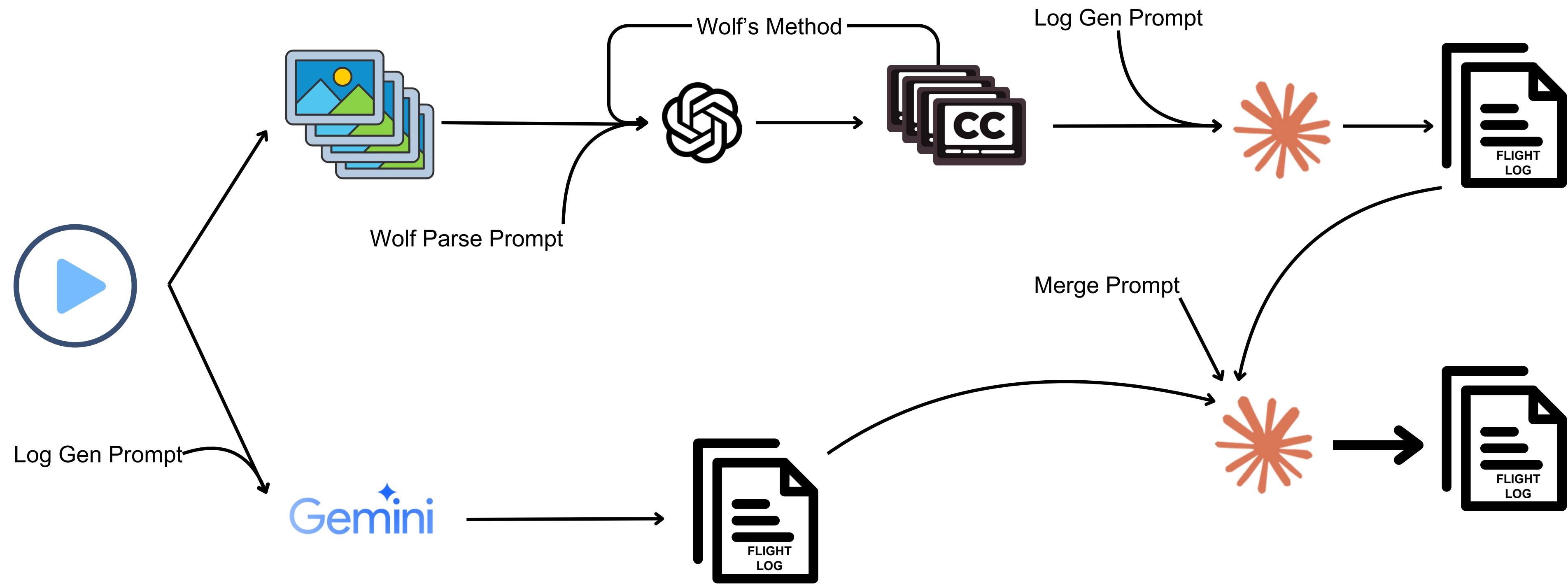
Methods - Pipeline Design

test existing VLMs
on video
understanding

create pipeline,
trial and error

Completed Log

Now, we merge the 2 pipelines to complement each other:



Interlude: AirSim Simulator



Completed Log

AirSim simulation
set up

drone API

Methods - Simulation

Why use simulation?

1. Reduce experimental costs
2. Shorten the development cycle
3. Avoid real world network connection overhead

Why AirSim?

1. Built on top of Unreal Engine, suitable for vision tasks
2. Integrated with ROS2, capable of easily adapting to real-world applications when hardware requirements are met



Completed Log

AirSim simulation
set up

drone API

Methods - Drone Basic

There are multiple ways to control the drone in AirSim, for simplicity, we first allow only:

1. move forward a distance
2. rotate an angle
3. move vertically a height

By utilizing these 3 methods, the drone can reach everywhere

Stage 2: Autonomous Flight

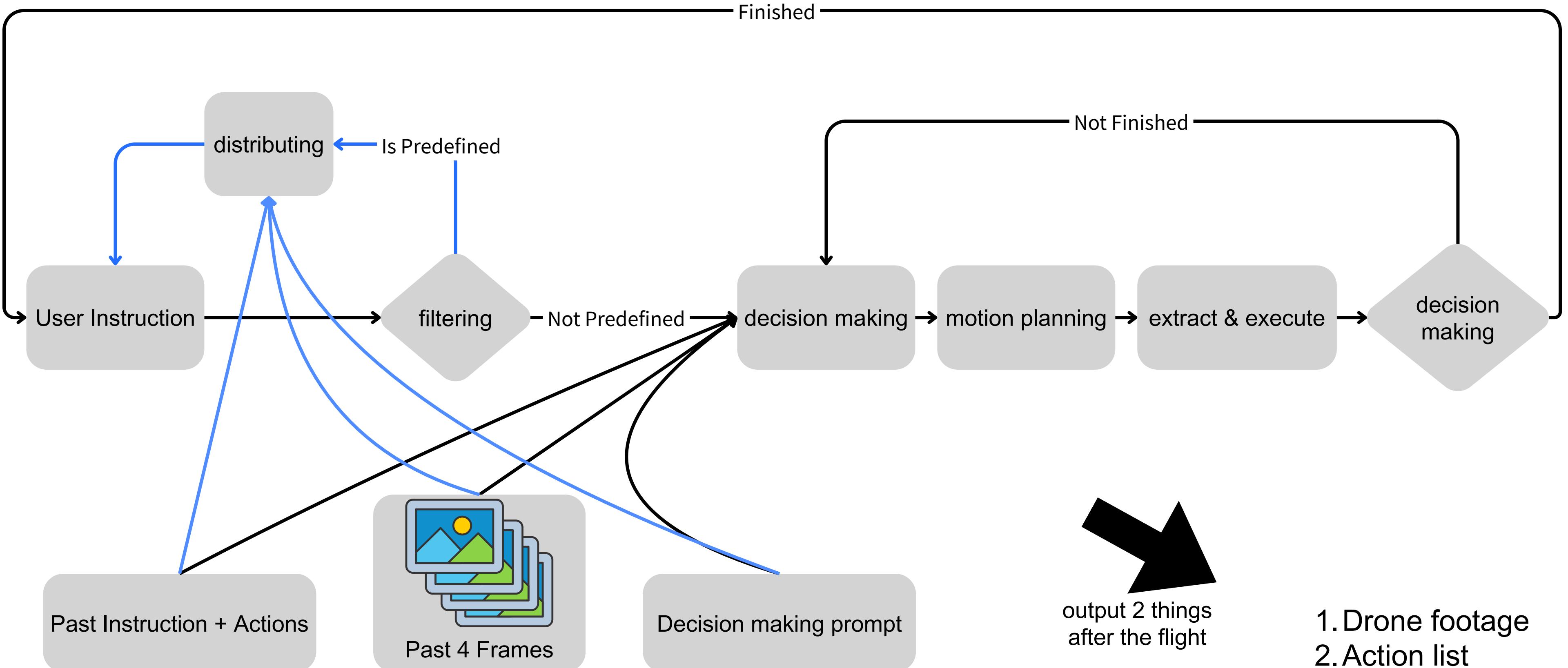


Methods - Define Levels

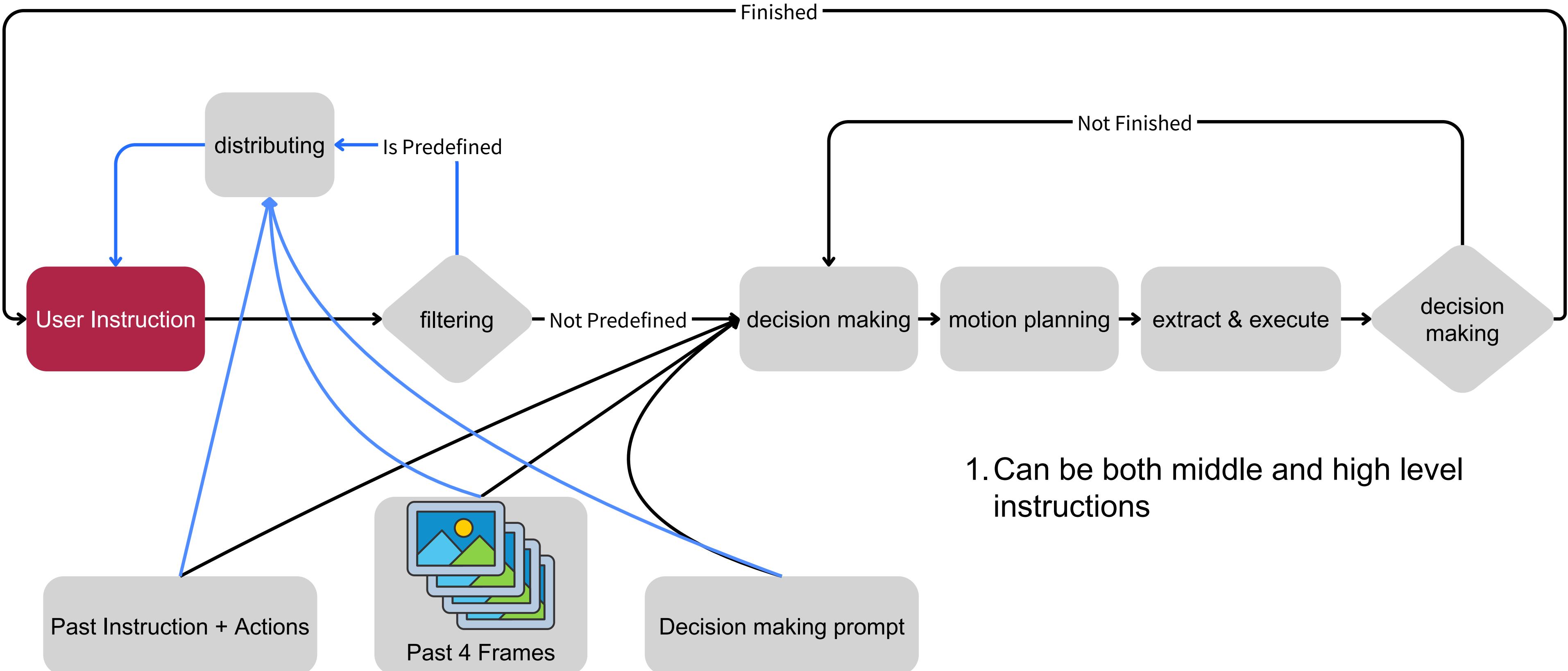
Stage 2: define hierarchical
Auto pilot instructions design simple
ver. pipeline

	Name	Description	Example
High Level	(Vision) Task	Ambiguous, high-level goals without specific "how-to" instructions.	"Find me an old temple", "Fly until you see a roundabout"
High Level	Predefined Mission	Complicated but commonly used. Hard for VLMs to handle	"Return Flight", "Areal Scan"
Middle Level	Navigation	Gives a specific idea of "how" to fly	"Fly ahead for 100m and turn left", "fly in a square that has a length of 100m"
Low Level	Action	The most specific commands (APIs) that directly control the system	move_forward(x), rotate(x)

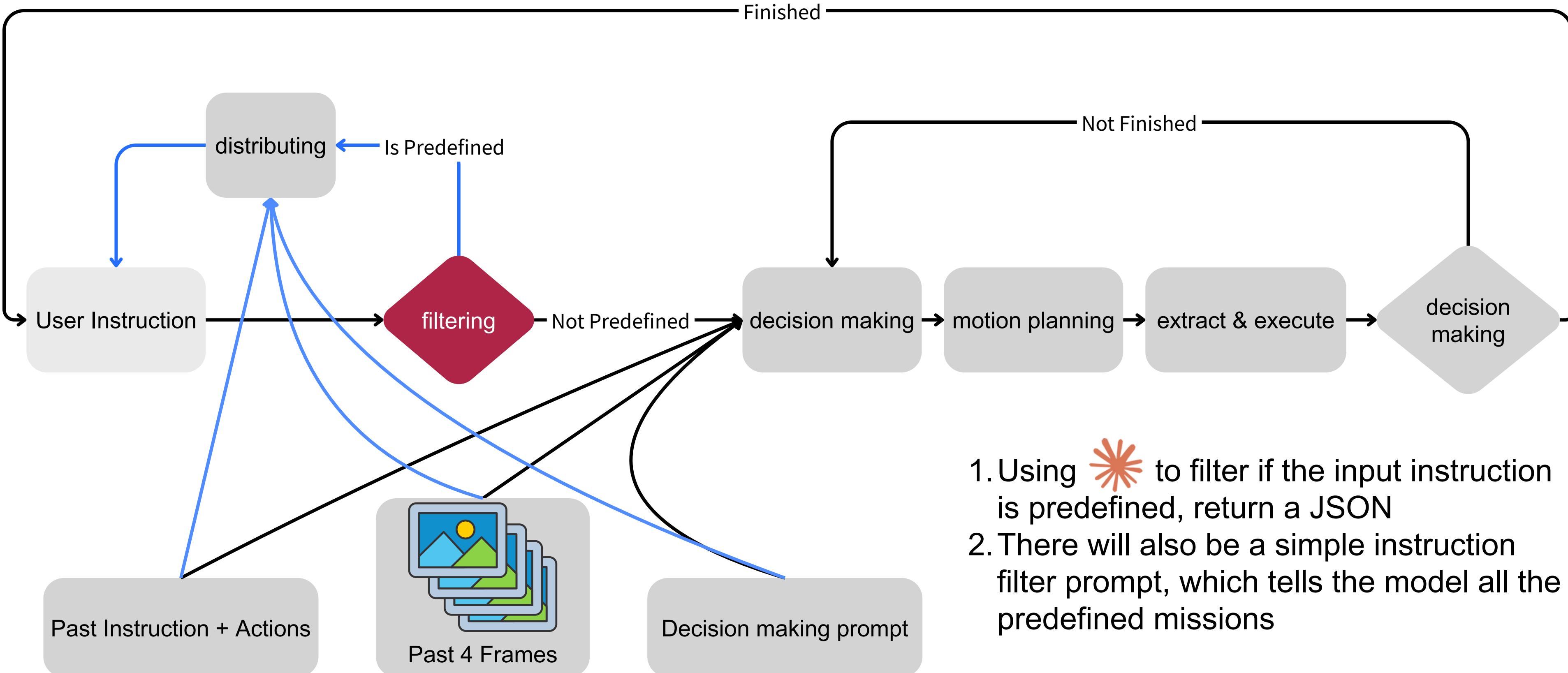
Methods - Pipeline Design



Methods - Pipeline Design



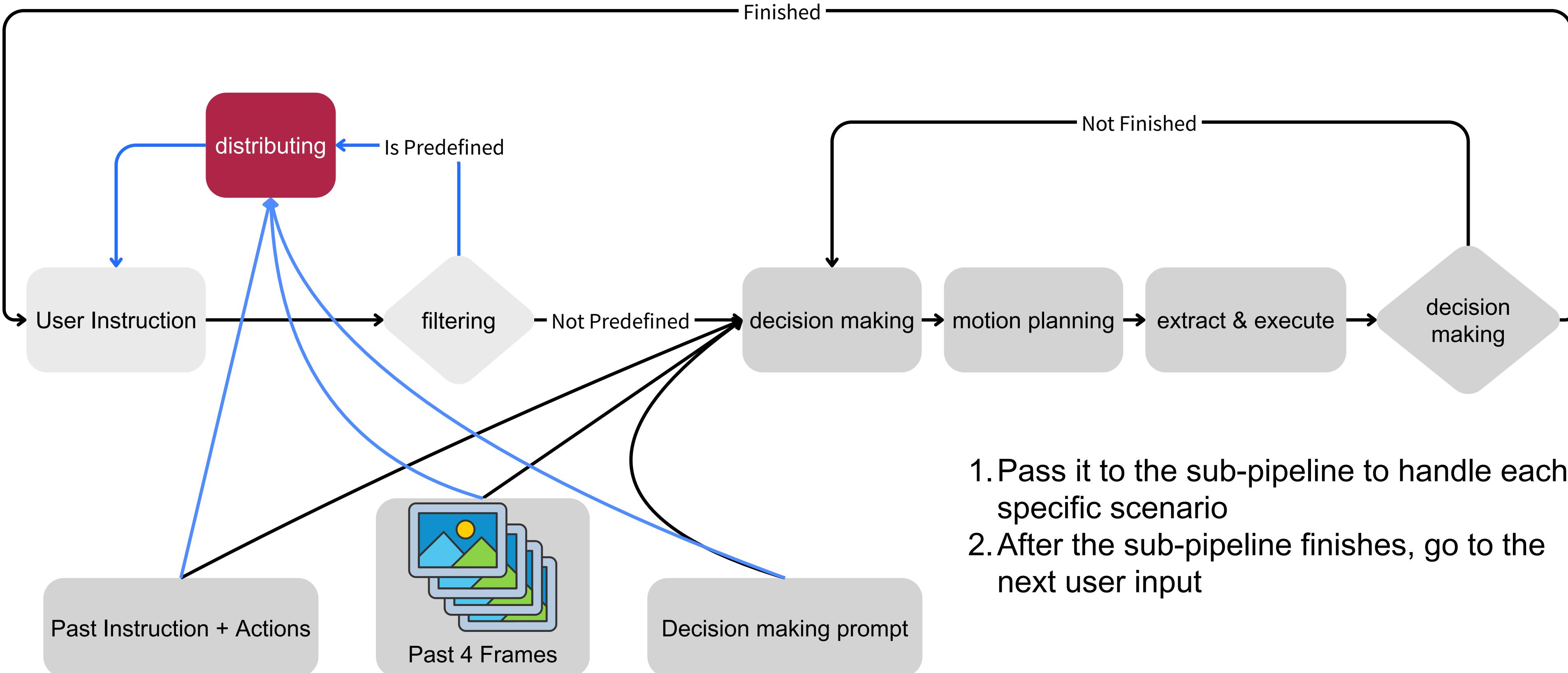
Methods - Pipeline Design



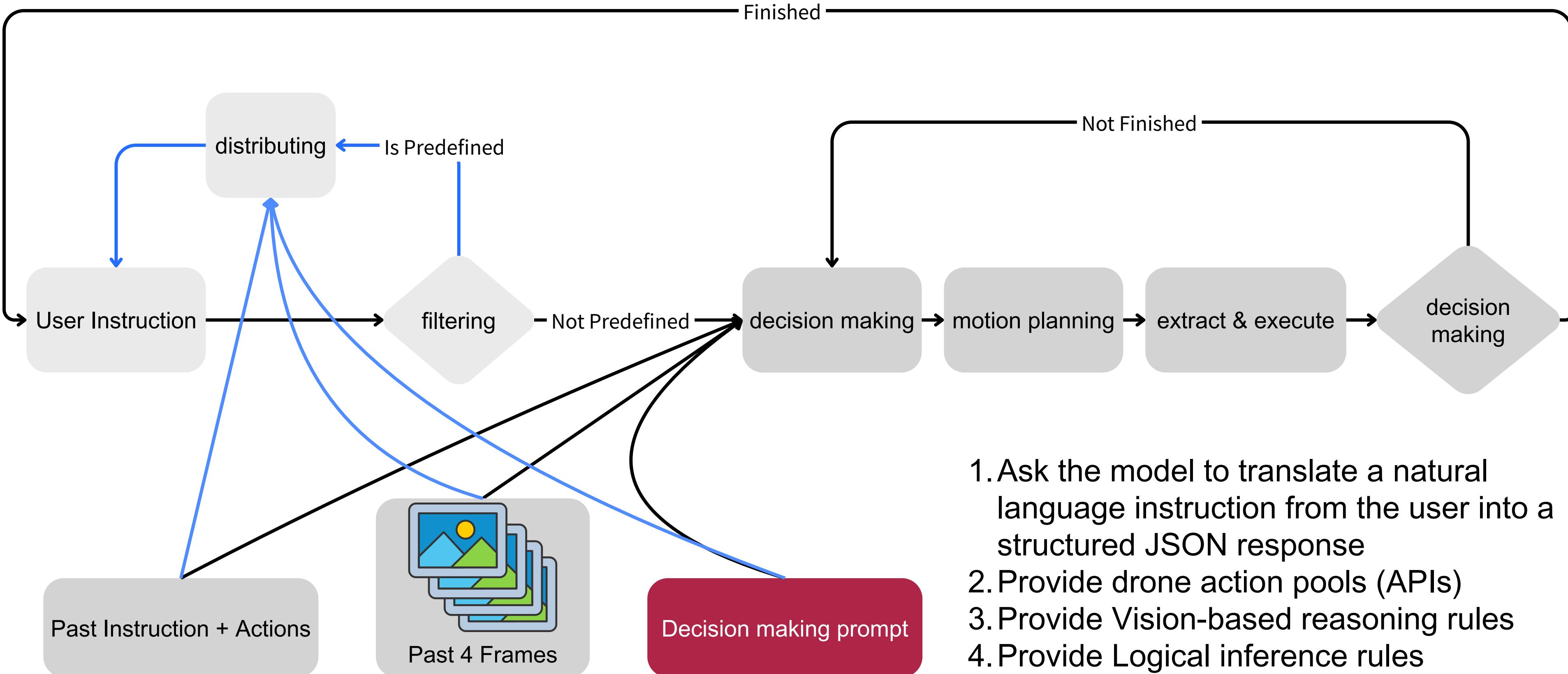
Methods - Pipeline Design



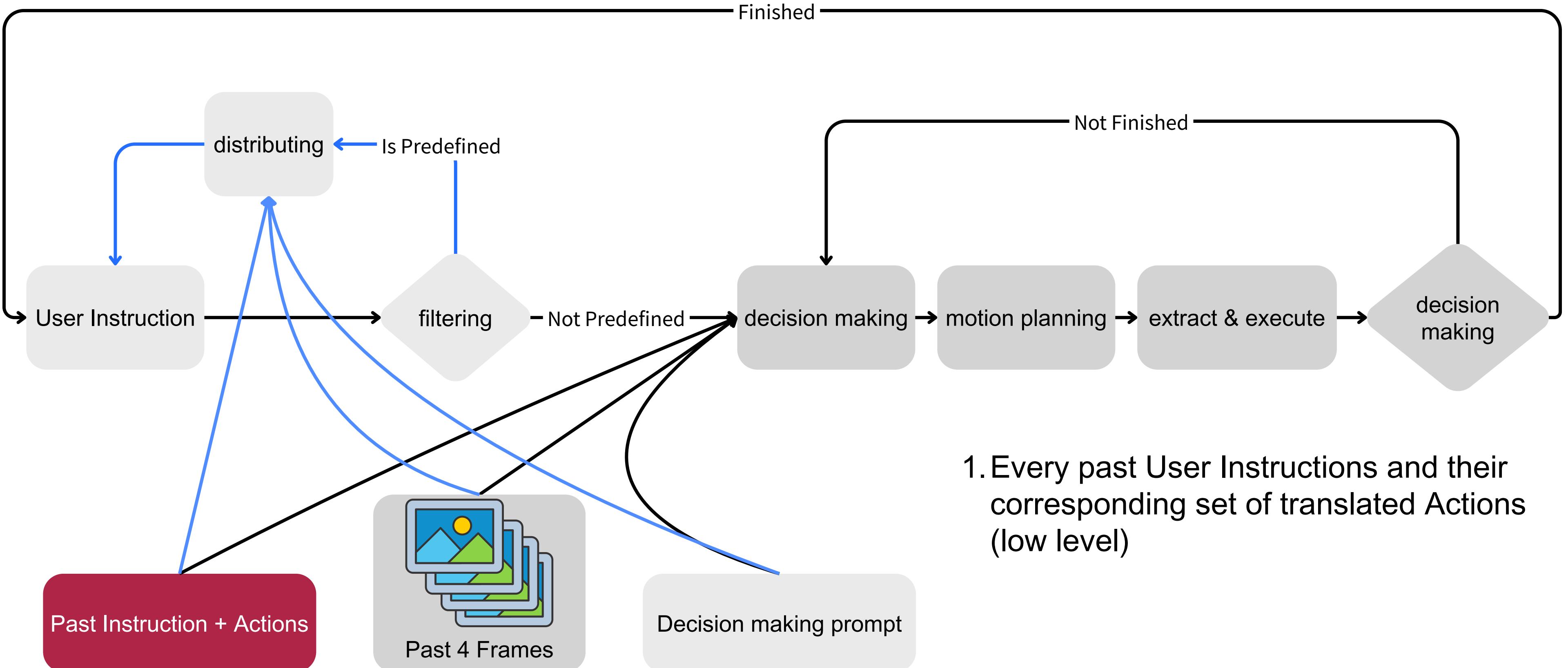
define hierarchical
instructions design simple
ver. pipeline test Flight Log
in AirSim



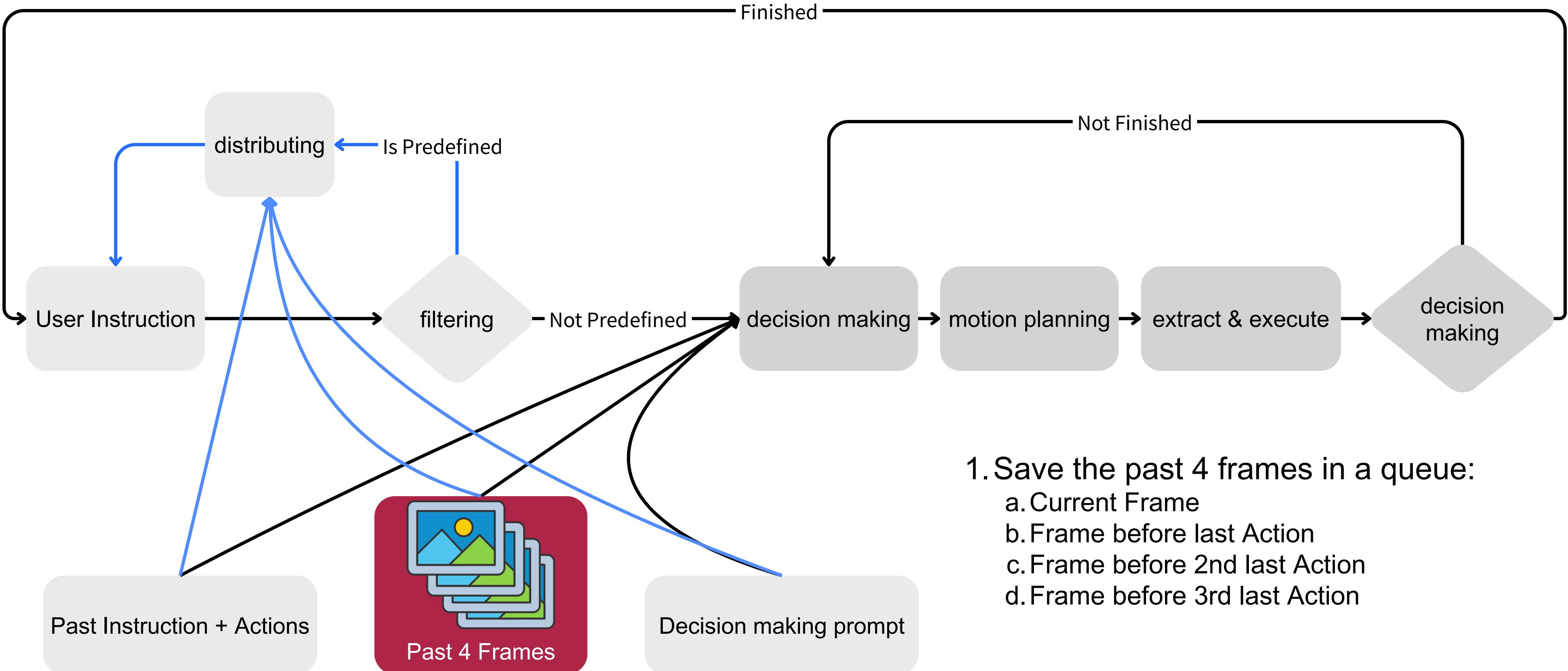
Methods - Pipeline Design



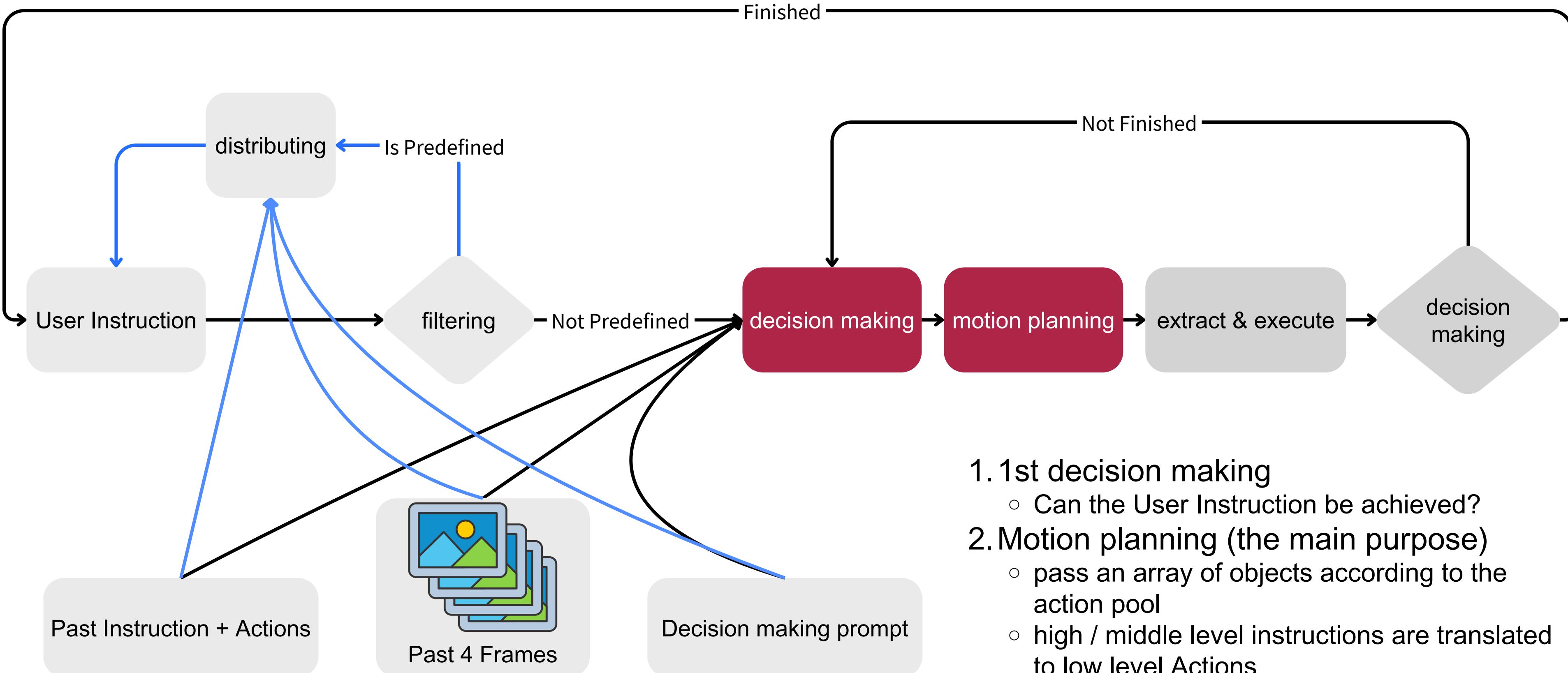
Methods - Pipeline Design



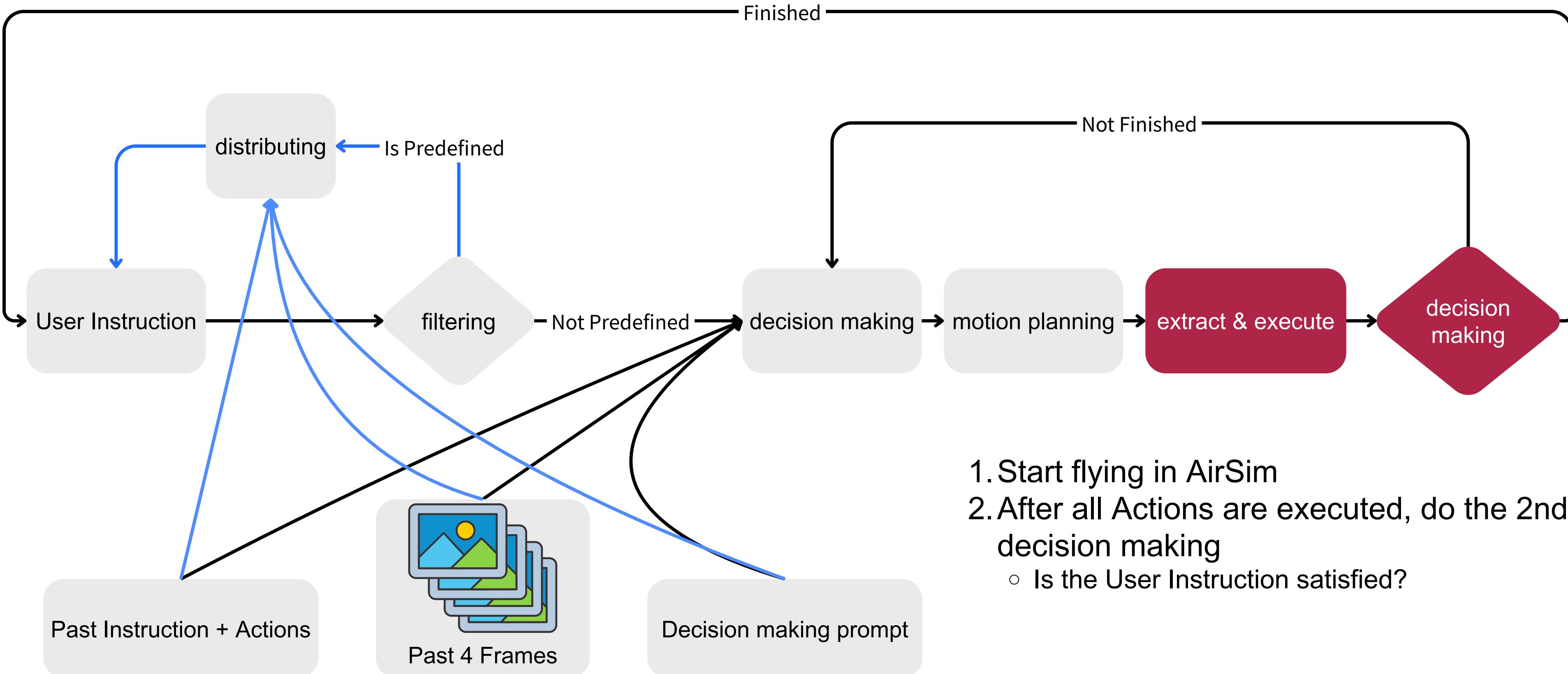
Methods - Pipeline Design

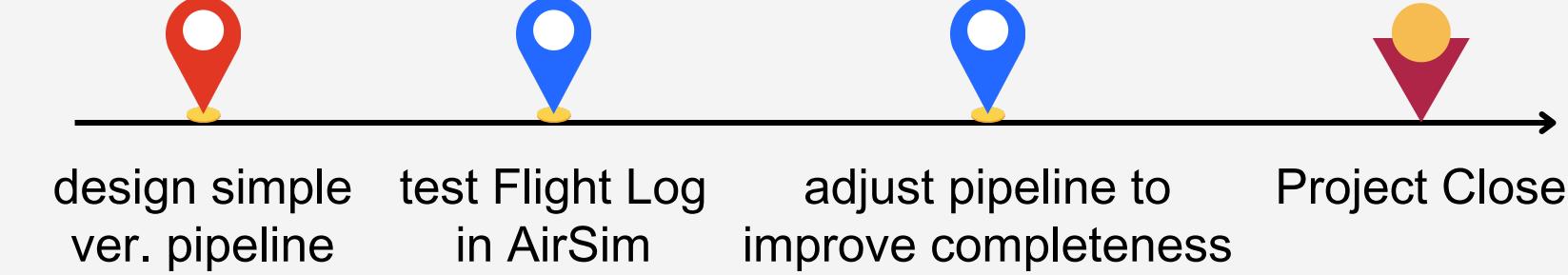


Methods - Pipeline Design



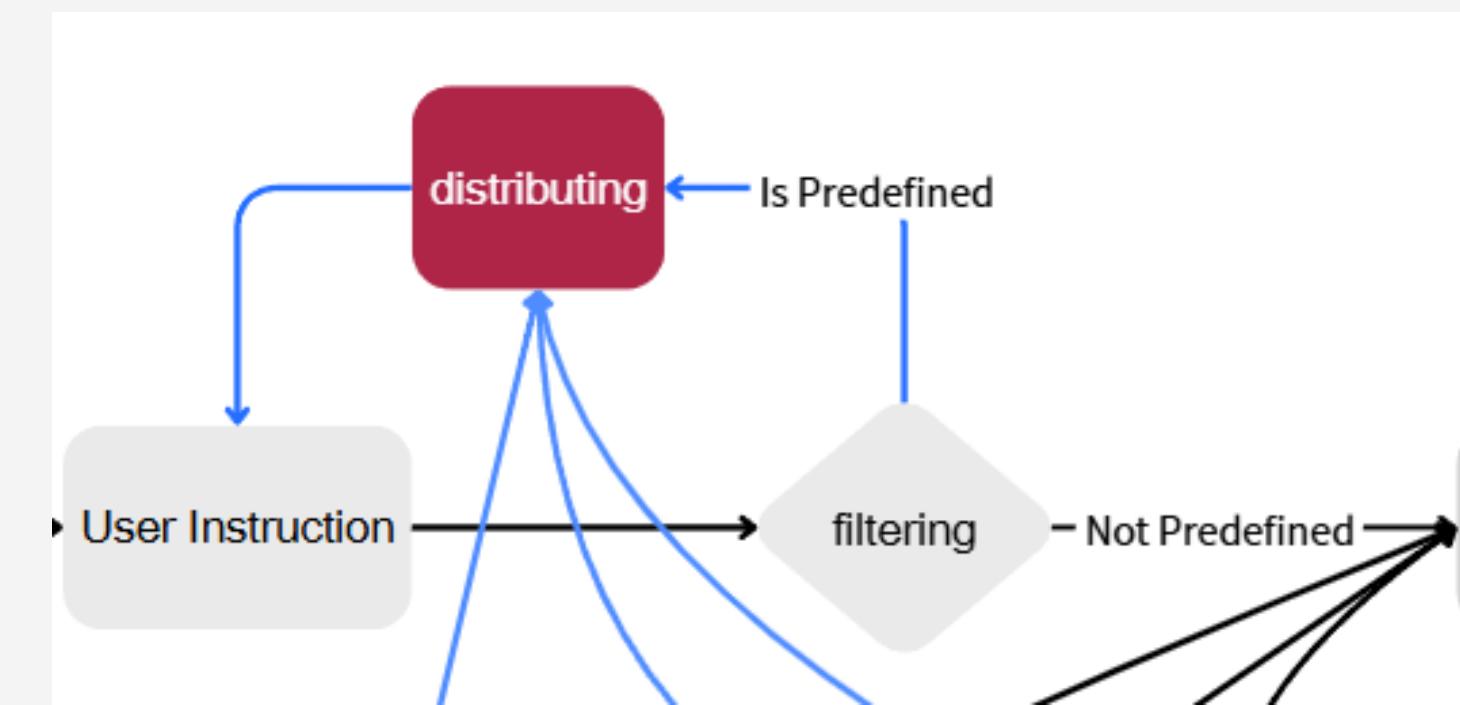
Methods - Pipeline Design

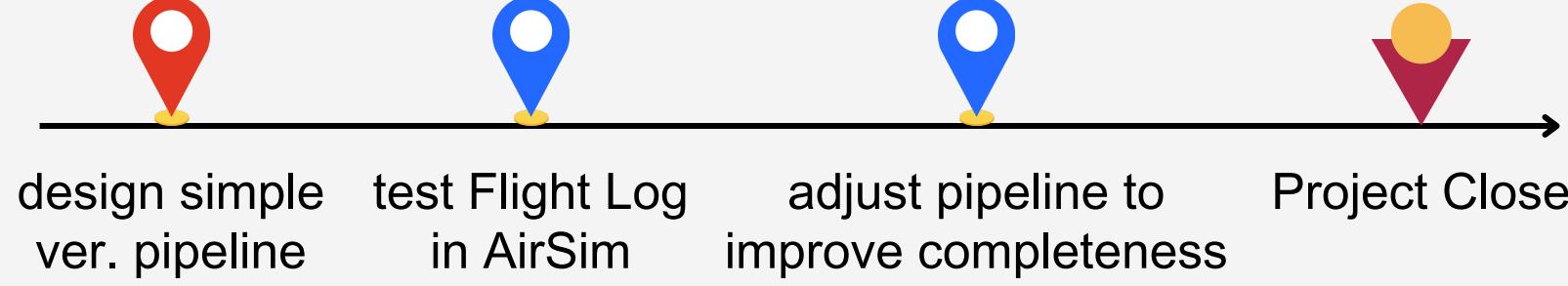




Methods - Adjustment

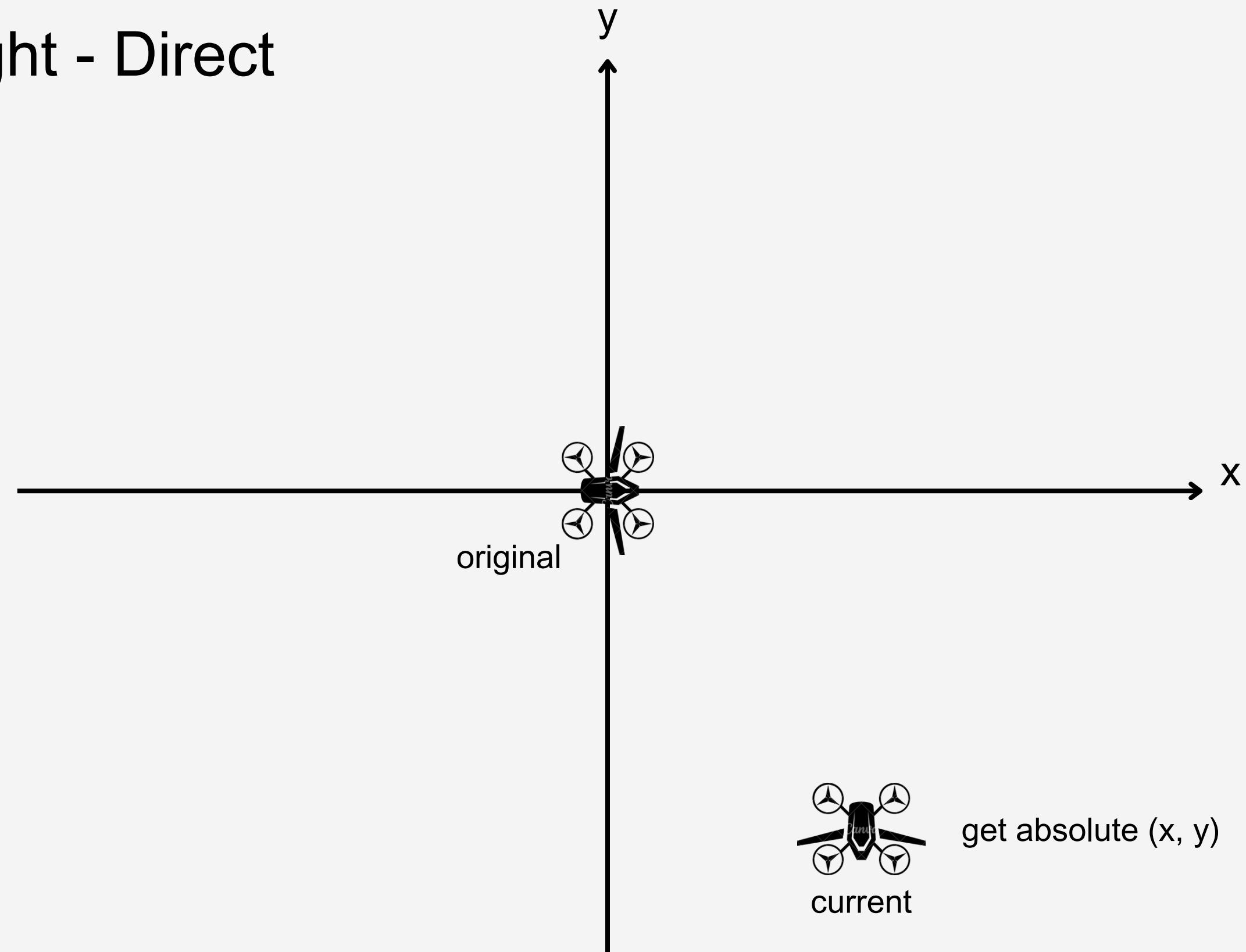
1. Try out what Flight Logs will look like in the AirSim simulation environment
2. Start customizing sub-pipelines for predefined missions

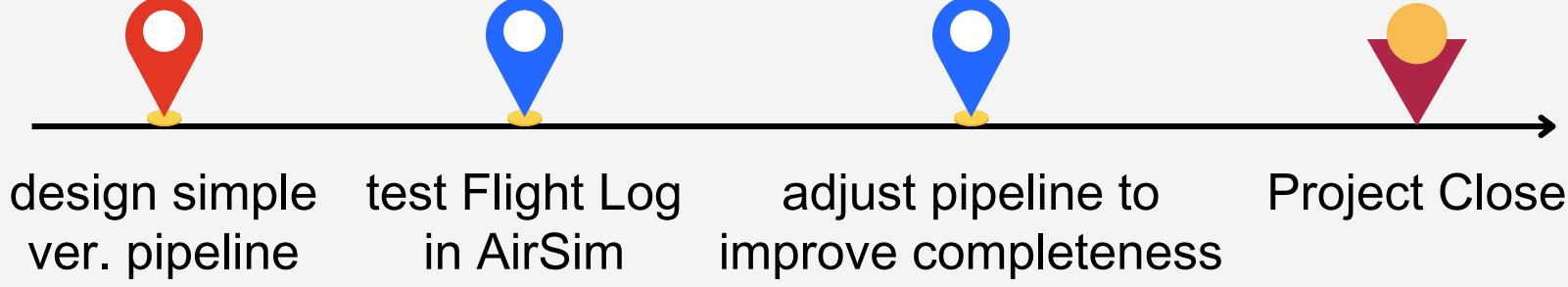




Methods - return flight

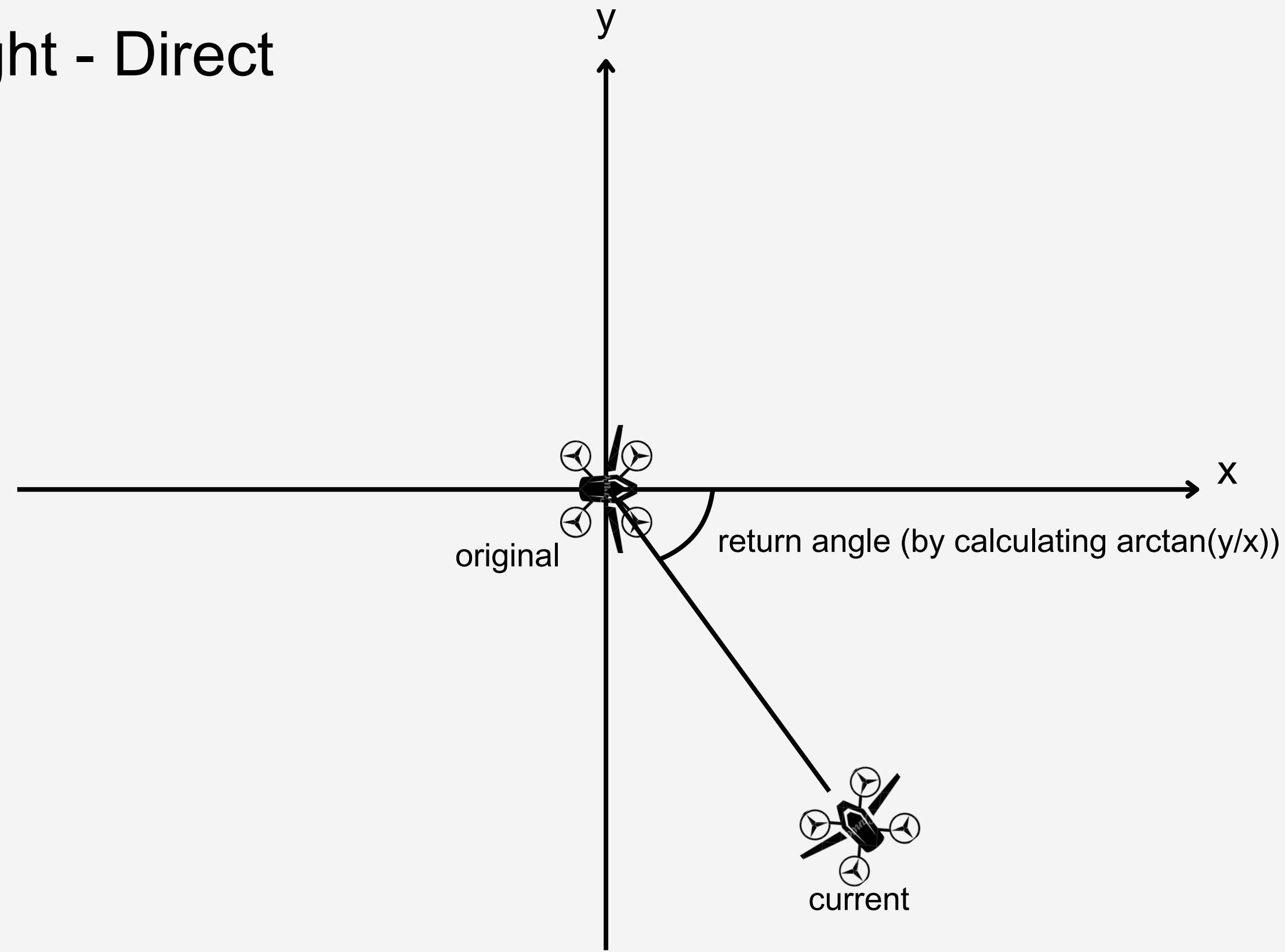
Return Flight - Direct

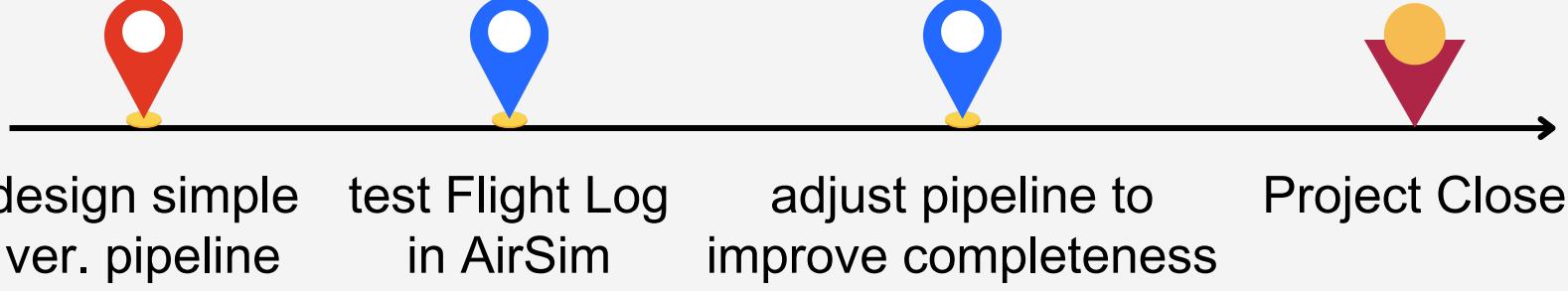




Methods - return flight

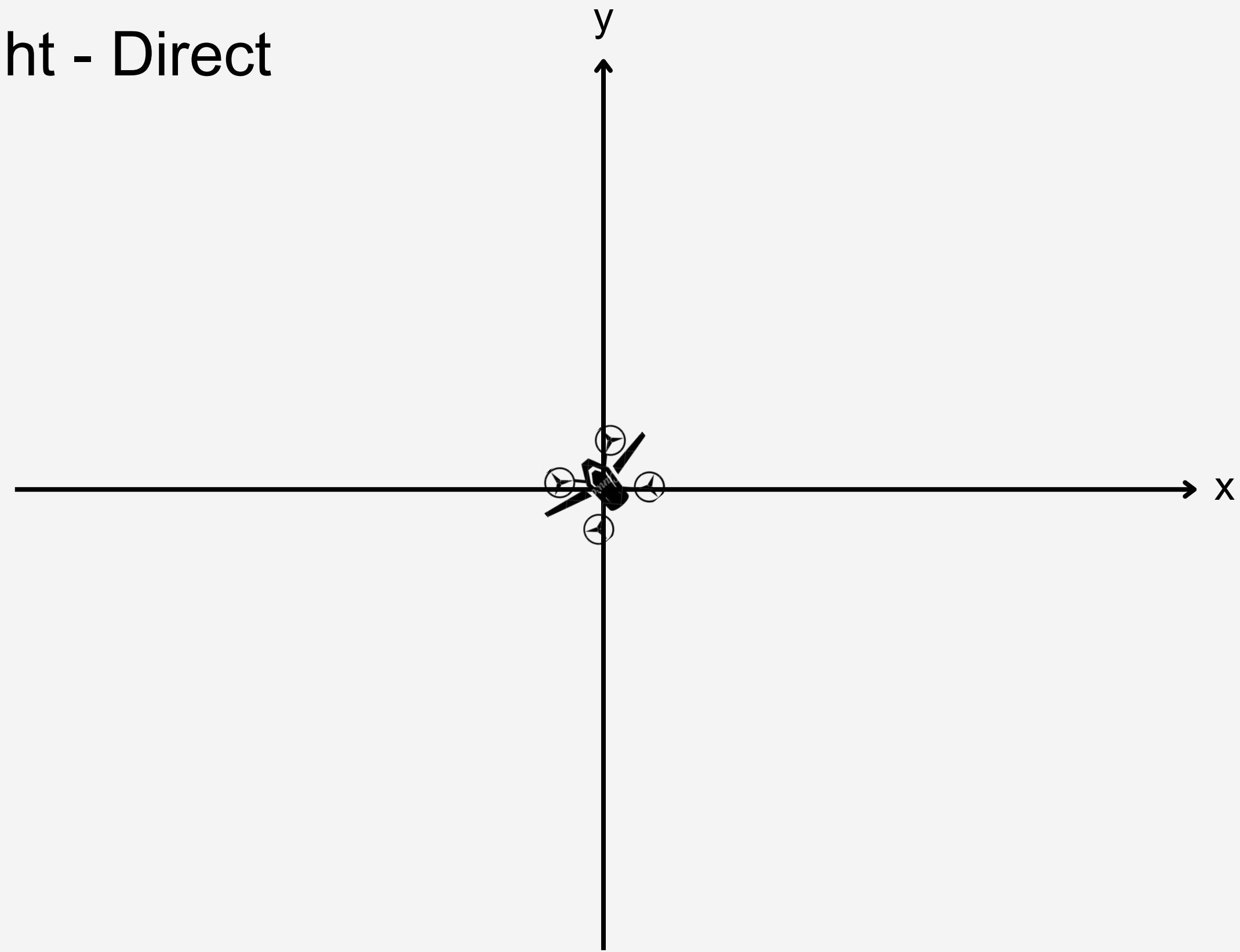
Return Flight - Direct

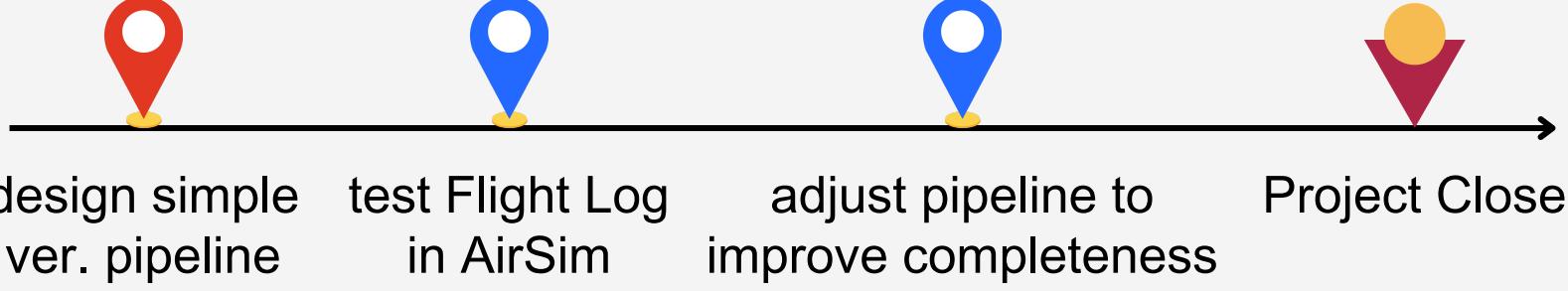




Methods - return flight

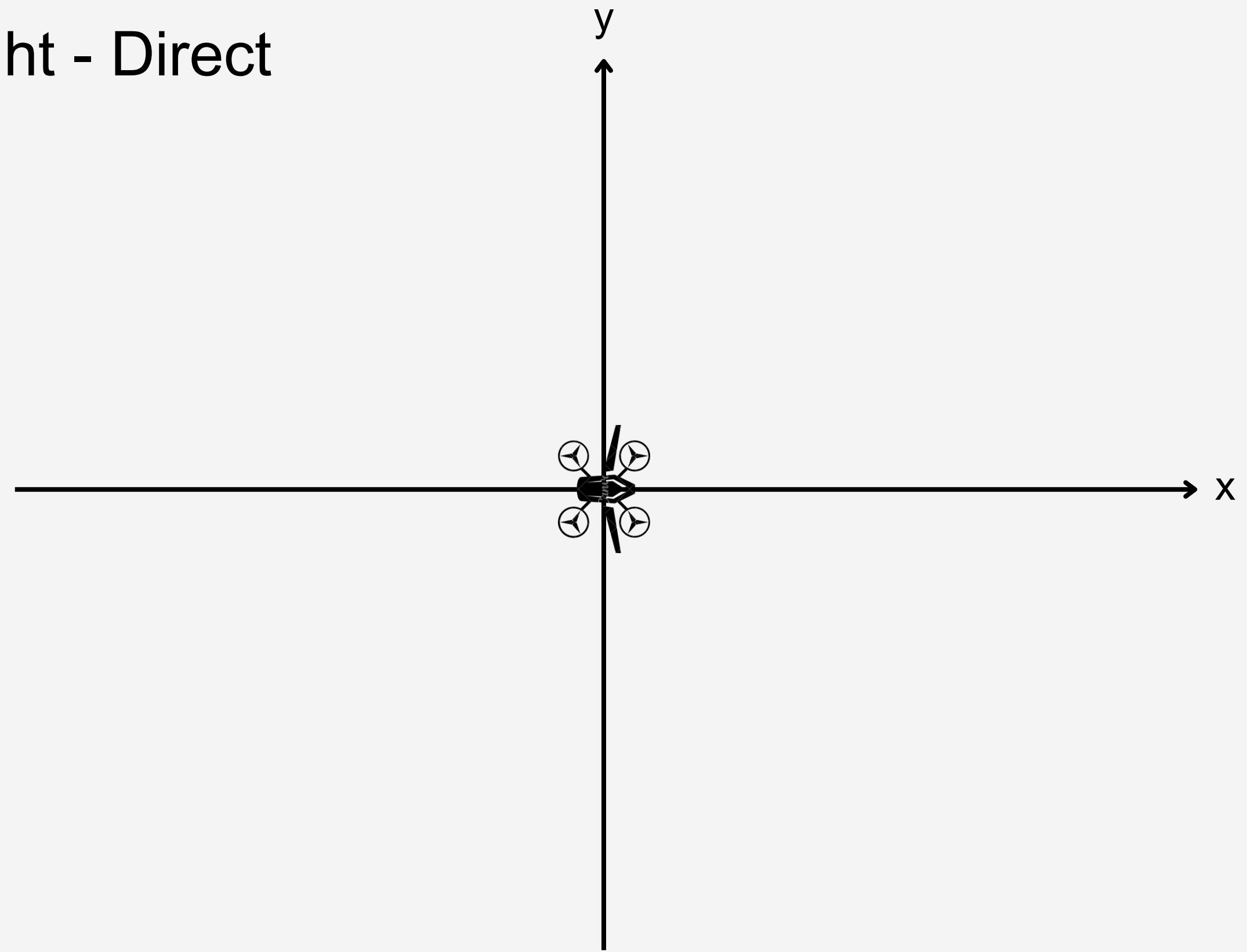
Return Flight - Direct



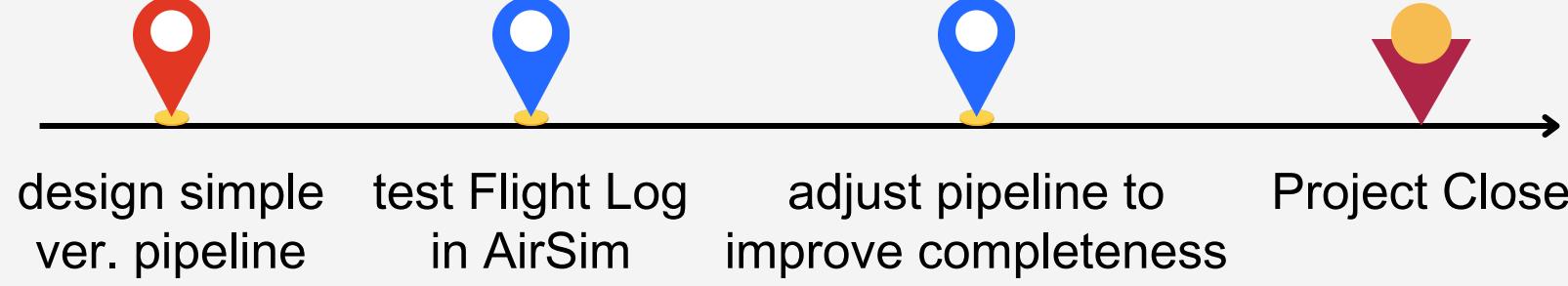


Methods - return flight

Return Flight - Direct

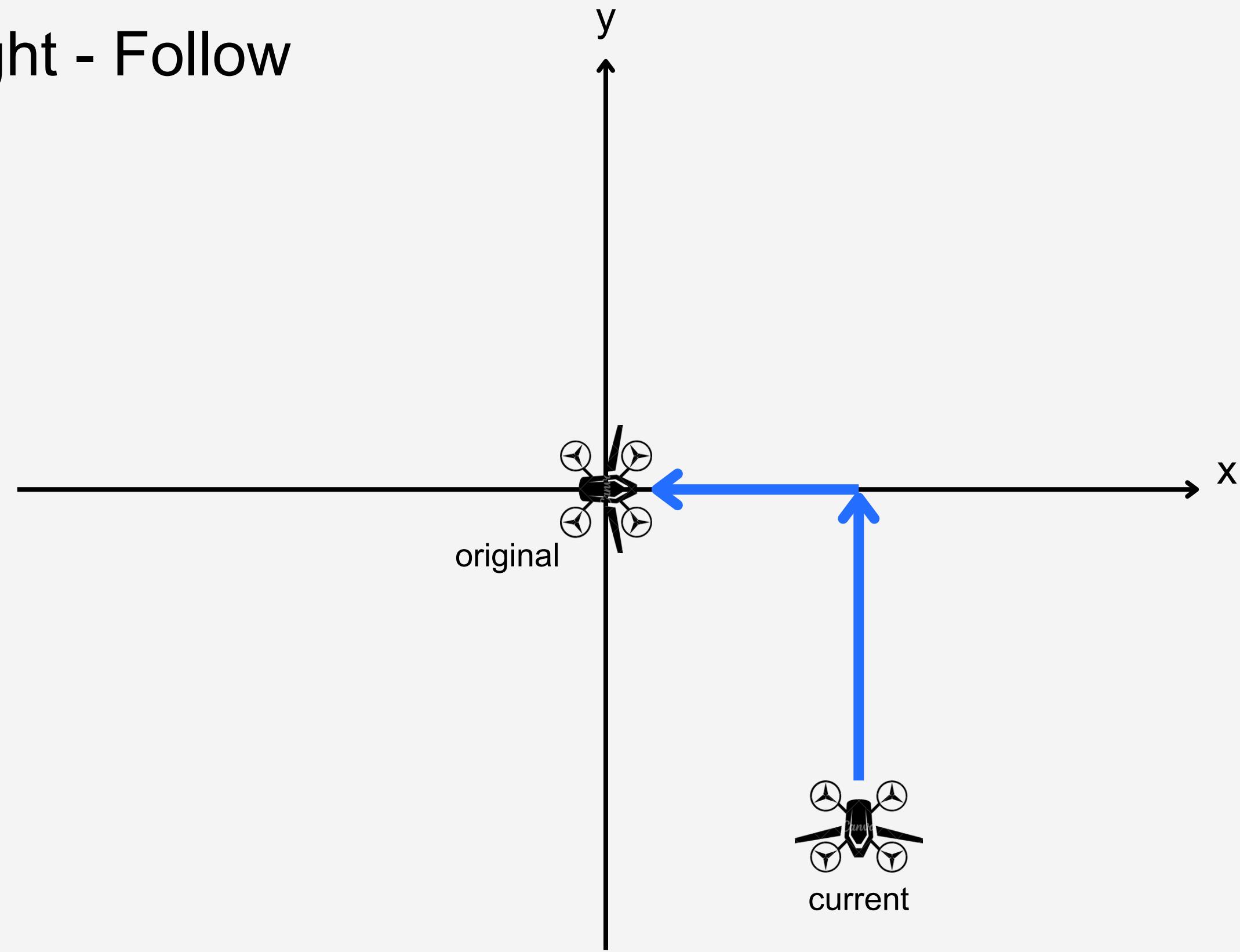


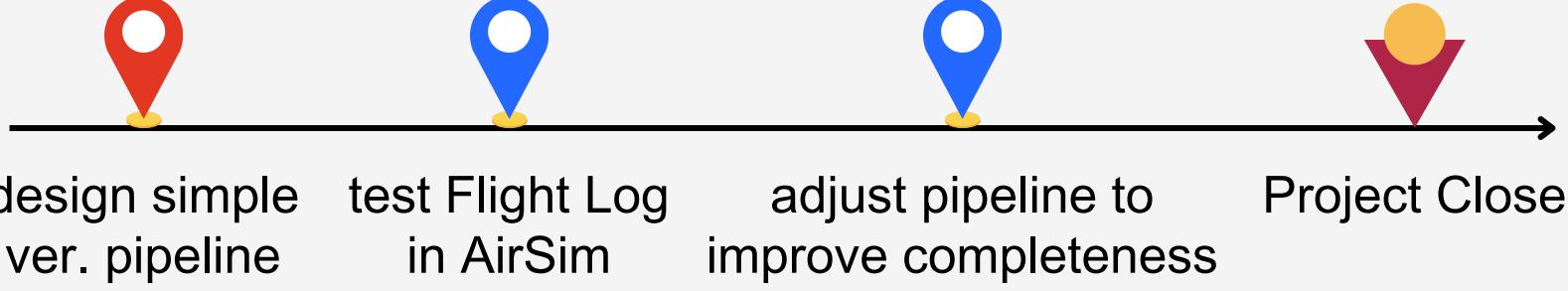
Done!



Methods - return flight

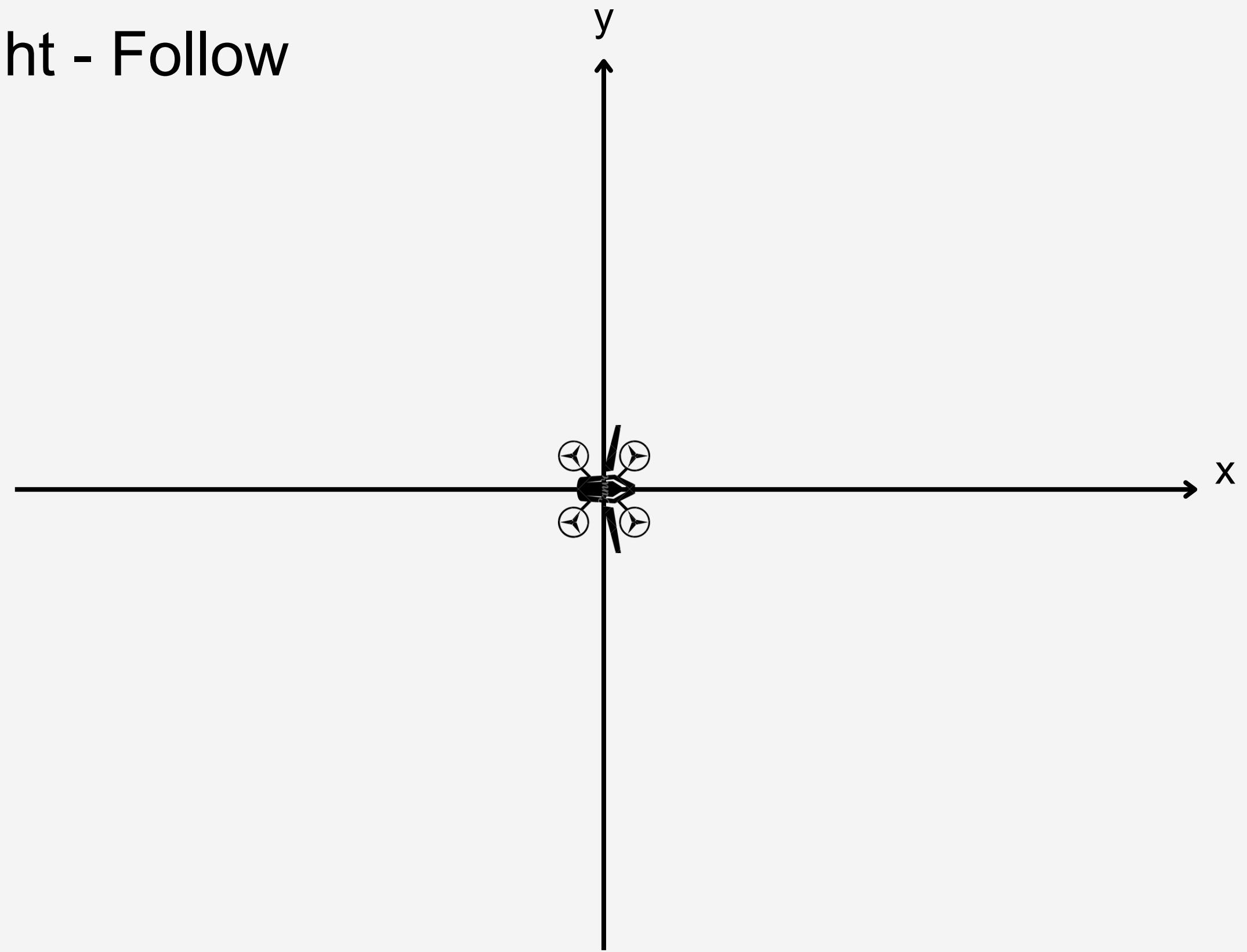
Return Flight - Follow





Methods - return flight

Return Flight - Follow



Done!



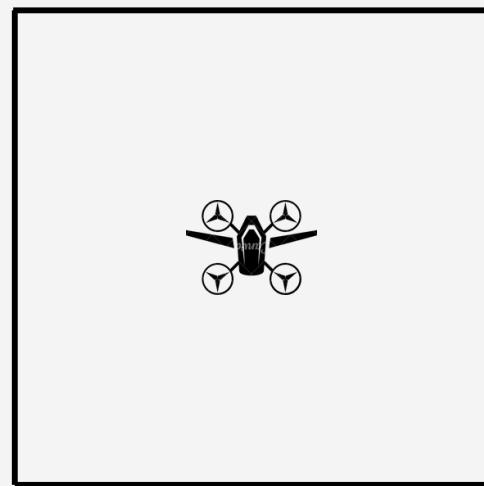
design simple
ver. pipeline test Flight Log
in AirSim adjust pipeline to
improve completeness Project Close

Methods - areal scan

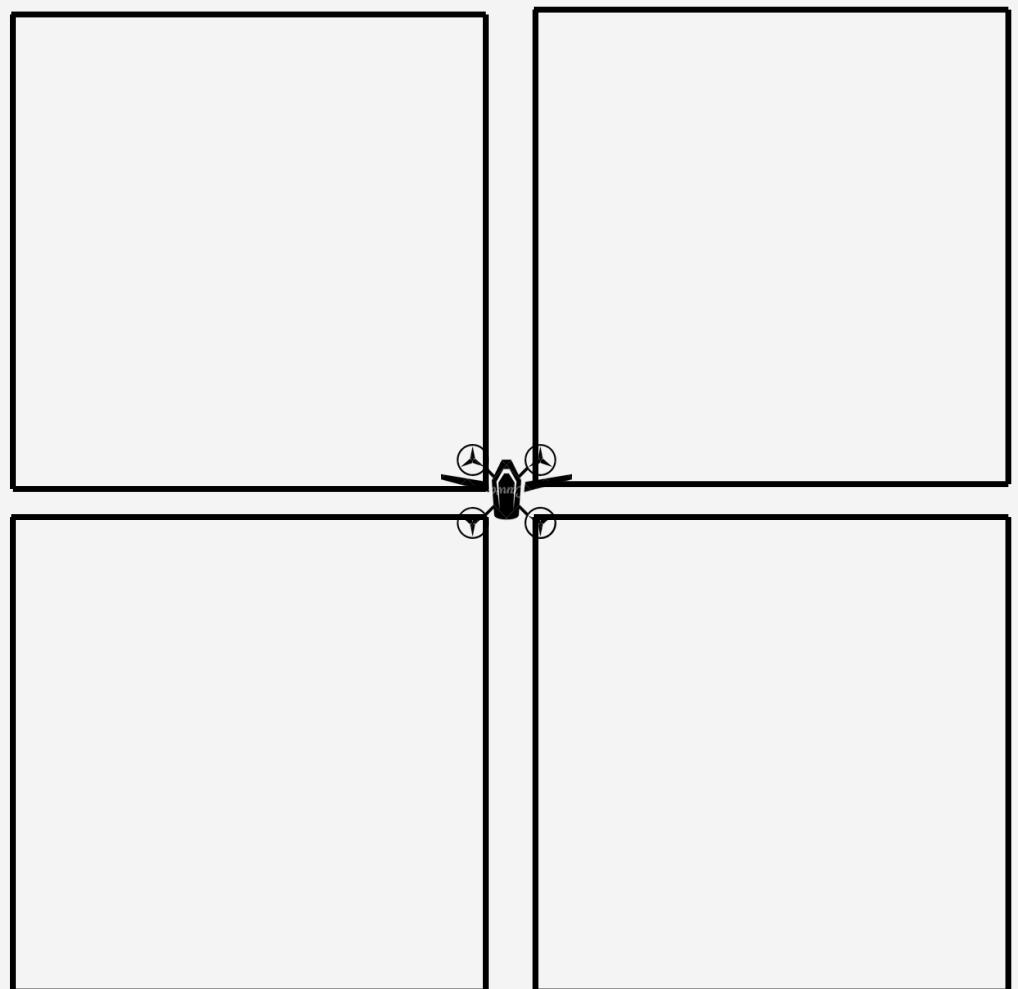
Areal Scan

9 types of scanning method

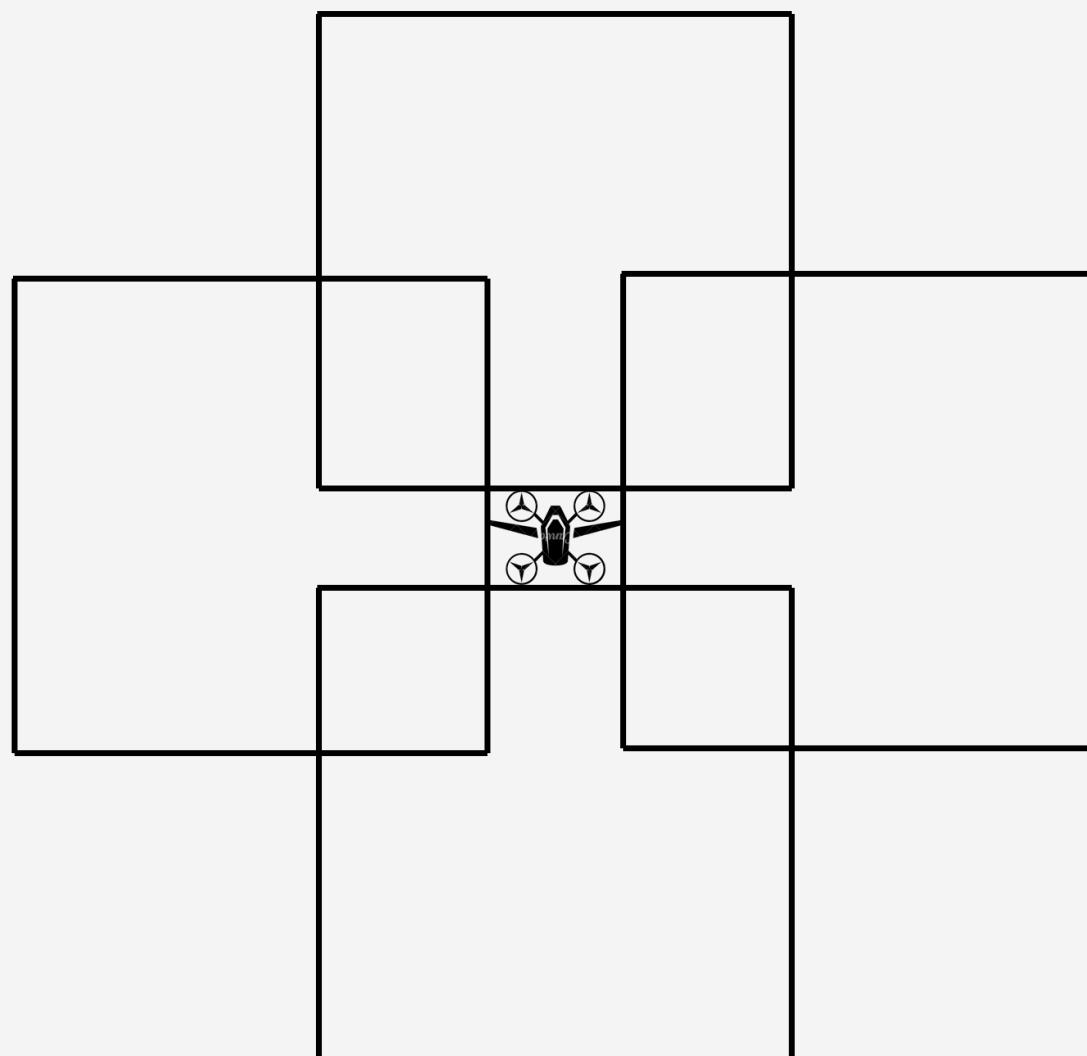
What we need:
nothing

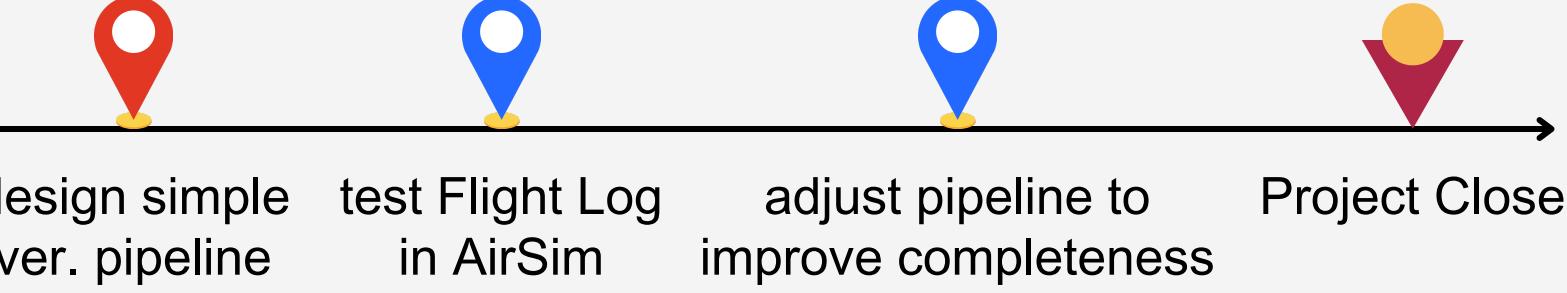


Spiral



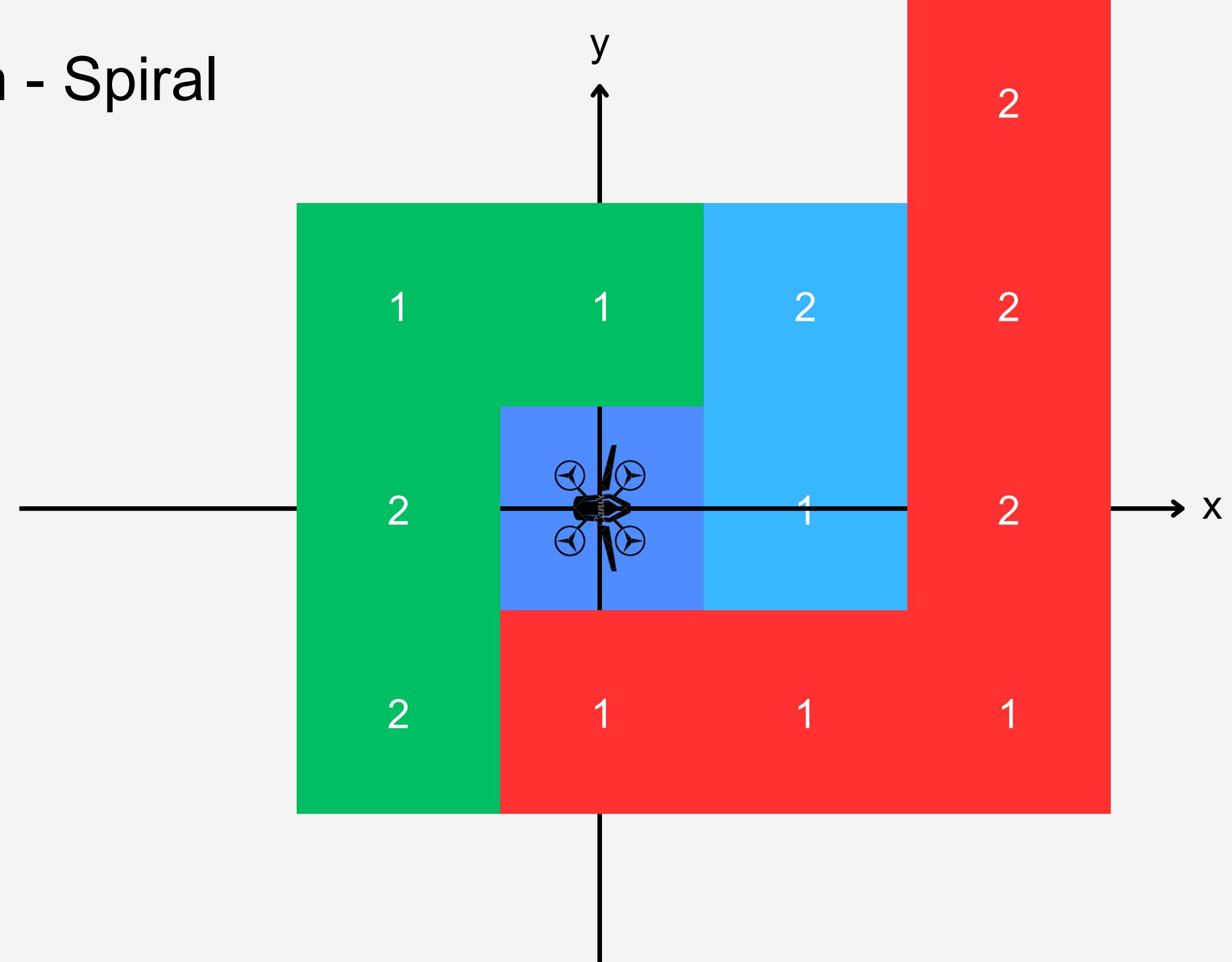
Grid

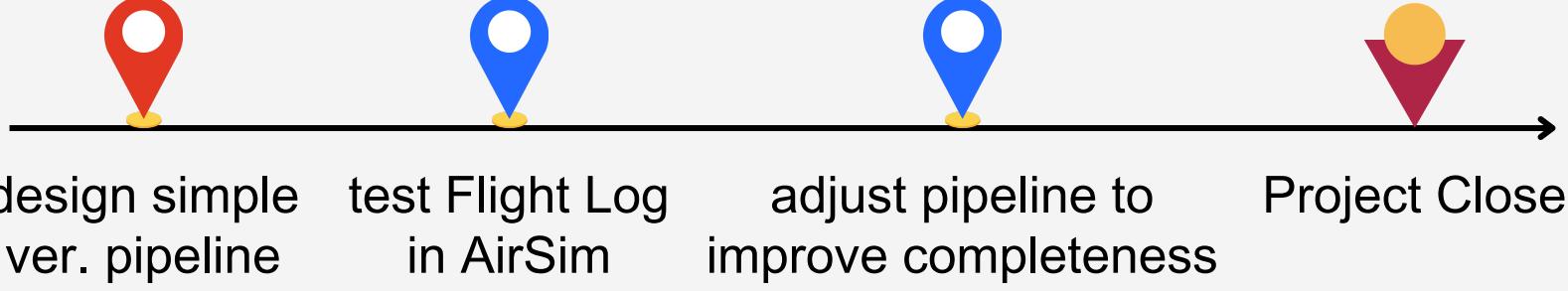




Methods - areal scan

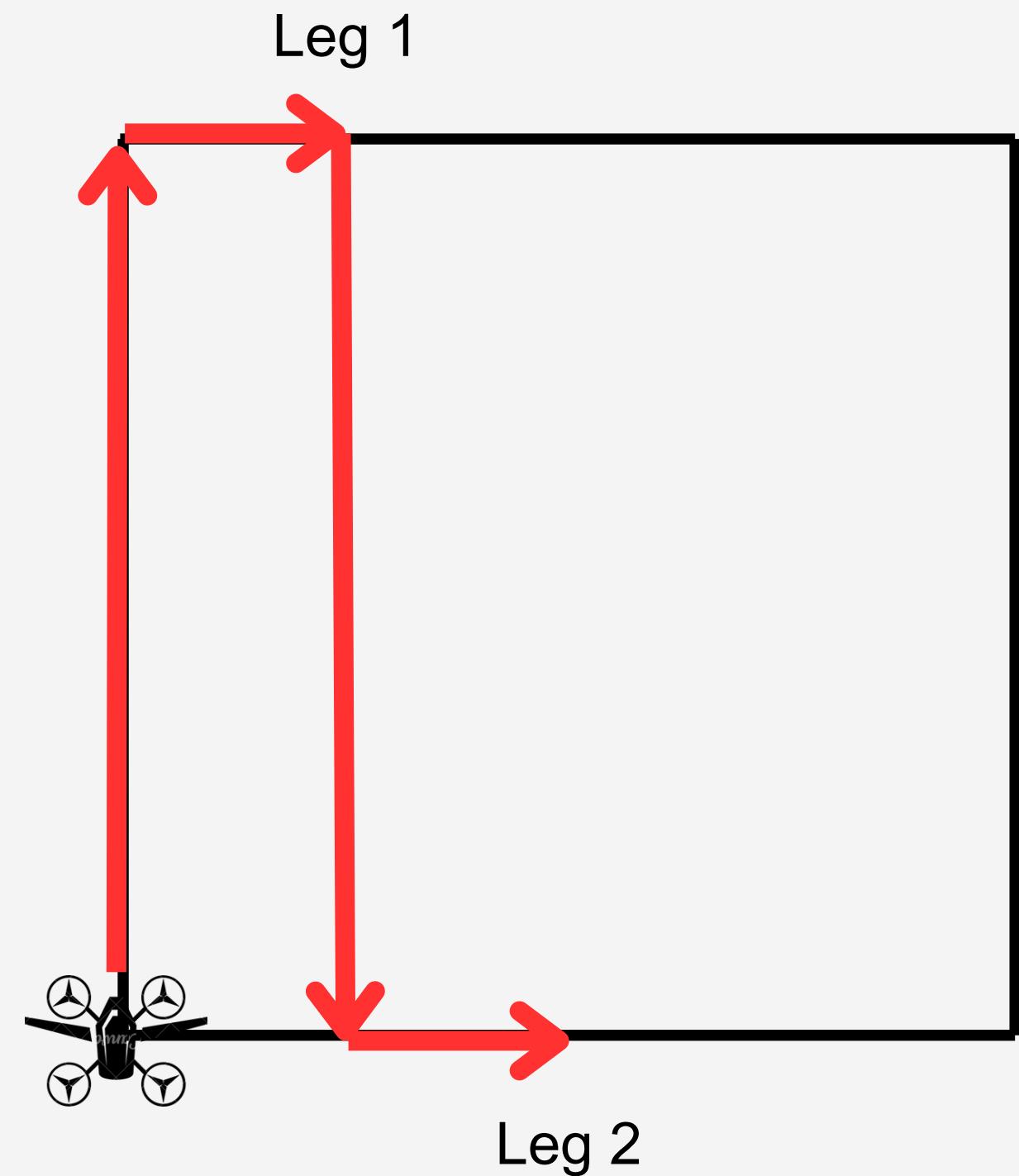
Areal Scan - Spiral

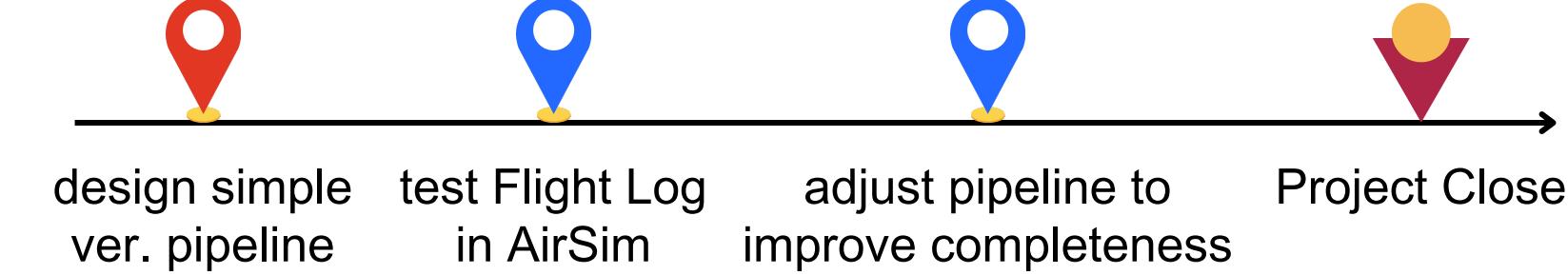




Methods - areal scan

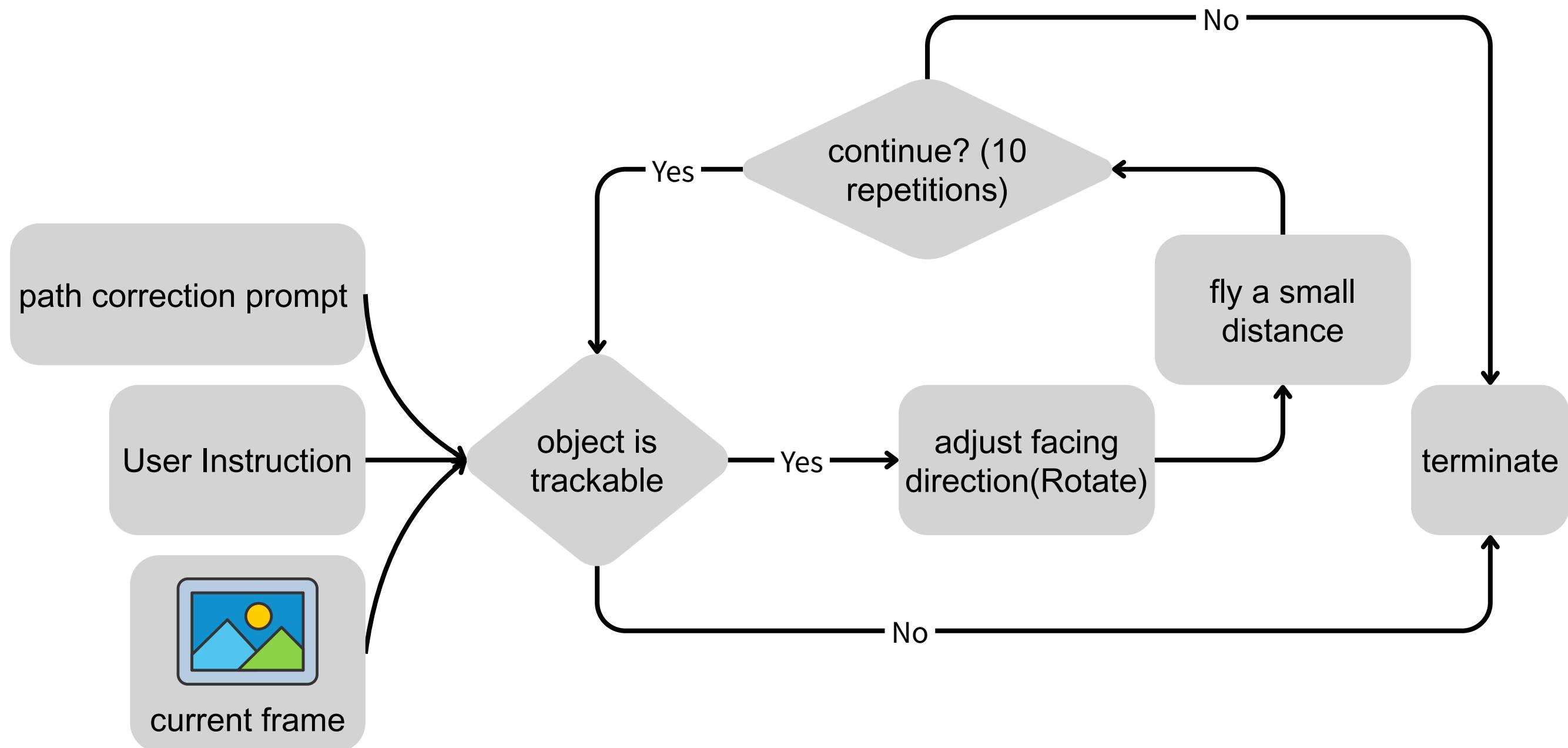
Areal Scan - Grid

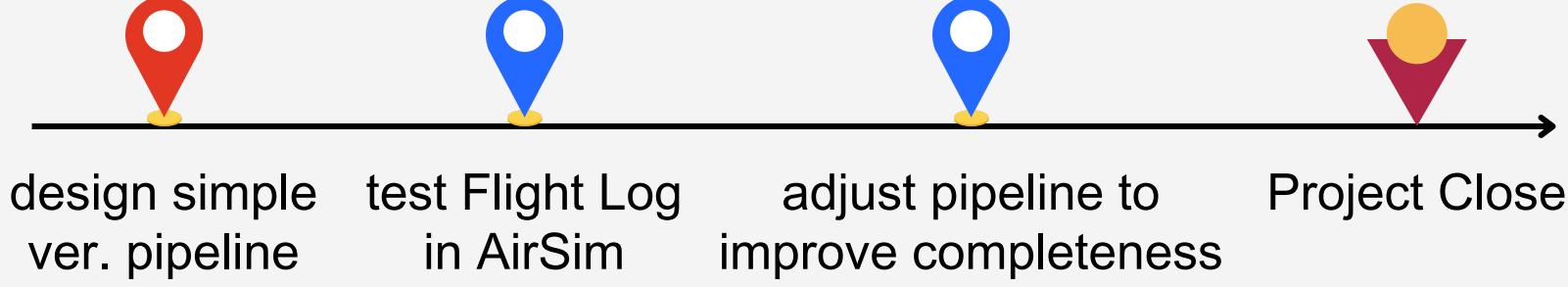




Methods - follow path

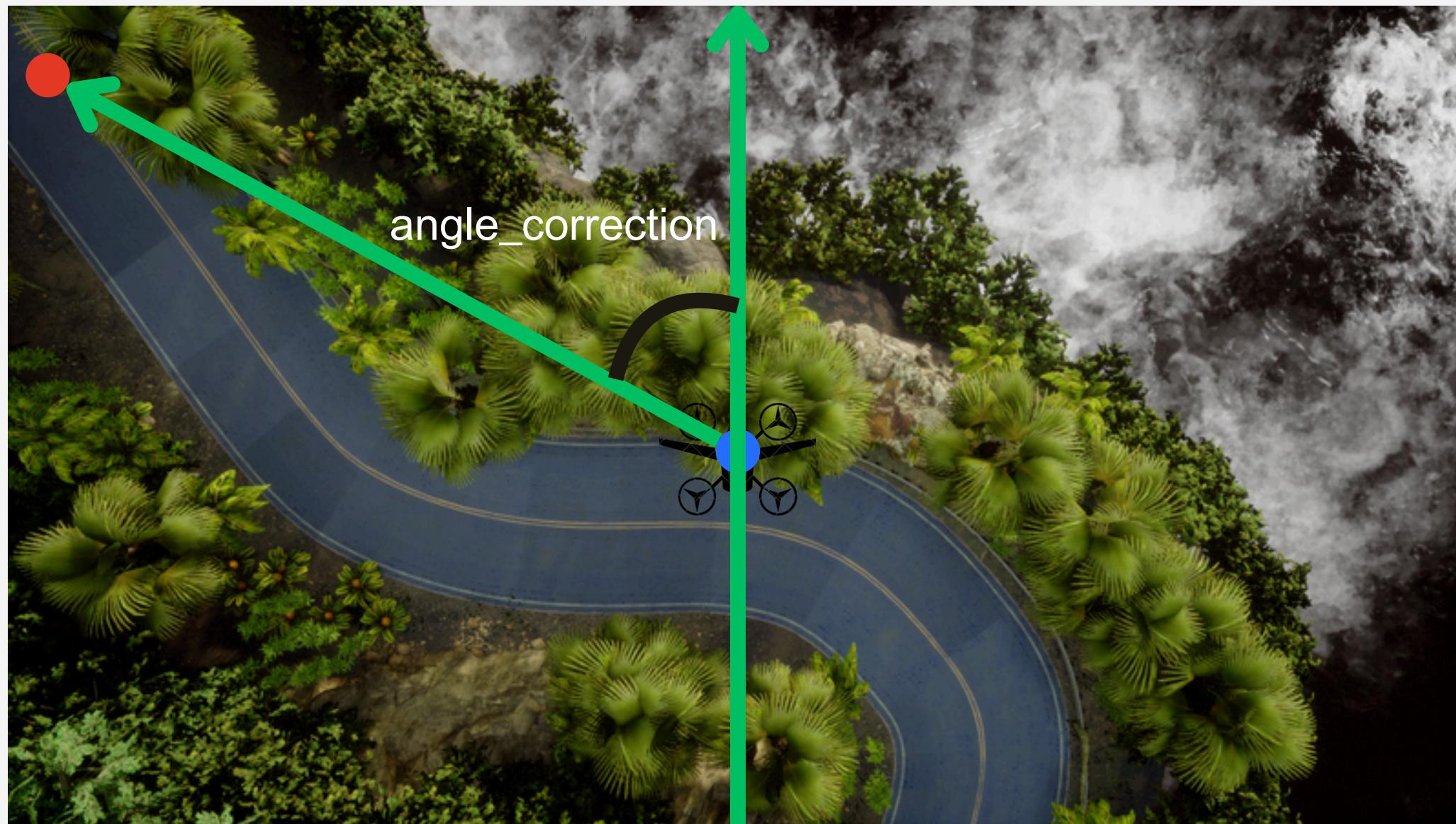
Follow Path (Static Object)





Methods - follow path

Follow Path (Static Object)

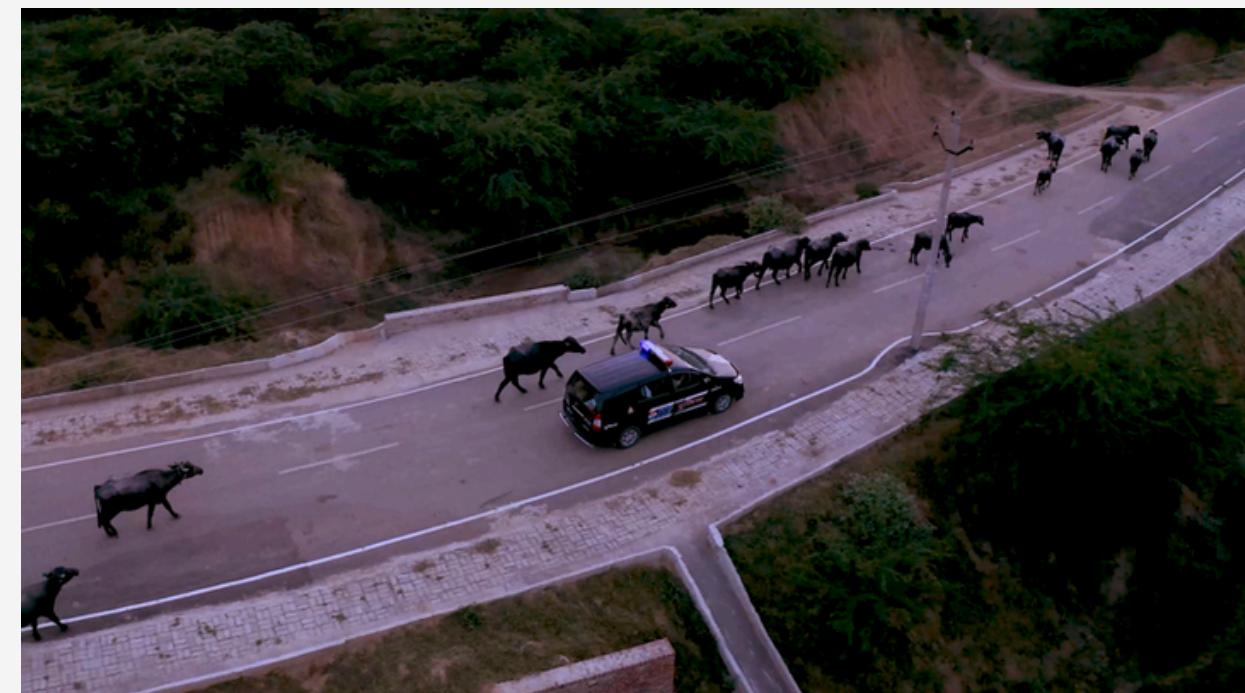


Result

The following result will be demonstrated in pictures. For a better experience, visit the [GitHub](#) page to check out the videos.

Result - Flight Log Generation

Following a police car



Result - Flight Log

Following a police car

DRONE FLIGHT LOG

FLIGHT IDENTIFICATION

Date: 2025-08-20
Start Time: 10:00
End Time: 10:05
Total Duration: 4 min 41 sec

FLIGHT PURPOSE & OPERATIONS

Purpose of Flight: observe and document police activity during a road check

Type of Operation: Aerial Surveillance Flight

LOCATION & ENVIRONMENT

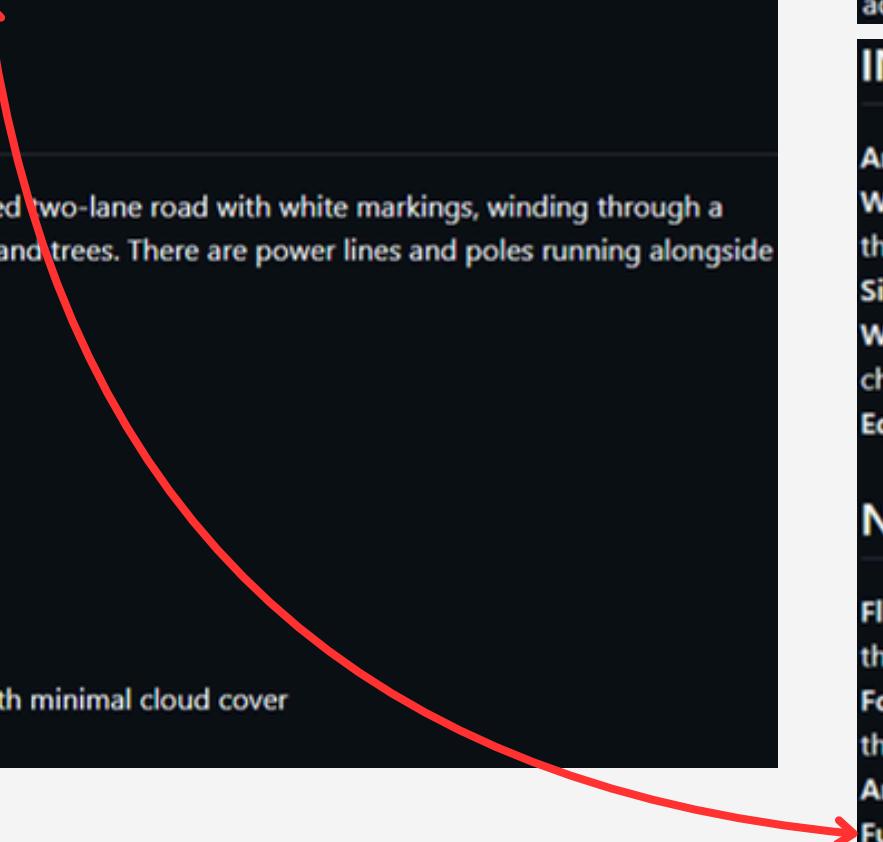
Location Name/Description: America, country road. The road is a narrow, paved two-lane road with white markings, winding through a landscape of rolling, eroded hills covered in dense, low-lying green shrubbery and trees. There are power lines and poles running alongside parts of the road. Some cultivated fields are visible further along the path.

GPS Coordinates (Takeoff): police station

GPS Coordinates (Landing): same as takeoff place

Weather Conditions:

- Wind Speed: Low (vegetation shows minimal movement)
- Wind Direction: Not mentioned
- Visibility: Clear conditions (bright lighting throughout)
- Temperature: Not mentioned
- Cloud Cover: Overcast or hazy sky, no distinct clouds visible / Clear day with minimal cloud cover
- Precipitation: None



FLIGHT PARAMETERS

Maximum Altitude: Approximately 50-100 meters (estimated, based on perspective relative to trees and hills)

Maximum Distance: Not mentioned

Flight Pattern: Free Flight

Key Waypoints/Locations: Initial scene with a tractor, police car, and people; rural road through lush landscape, bridge/culvert structure/elevated sections, cultivated fields, areas with cattle/herd of buffalo on the road

Flight Path Summary: The drone begins at a lower altitude, observing a stationary tractor and police car with people gathered around. It then ascends slightly and follows the police car as it drives along a winding rural road through dense greenery, elevated sections, and agricultural areas. The drone maintains a relatively consistent altitude and follows the road through the vegetated, hilly terrain, encountering a motorcycle and later livestock.

CAMERA & RECORDING SETTINGS

Video Resolution: 3840x2160

Frame Rate: 29.97 fps / 29.97002997002997 fps

Recording Format: mp4

SAFETY CONSIDERATIONS

Obstacles Present: Trees Power Lines Other: Rolling, eroded hills/ravines, steep embankments, utility poles, other vehicles (tractor, motorcycle), and livestock

People in Area: Small Group (police officers, person standing by vehicle, initially near the tractor/car, then individuals on motorcycle and walking)

Emergency Landing Sites: The paved road itself, open cultivated fields visible throughout flight path, or relatively flat, open patches of ground adjacent to the road

INCIDENTS & OBSERVATIONS

Any Issues Encountered: None mentioned/None visible

Wildlife Interactions: A herd of cattle/buffalo was encountered on the road, scattered across pavement interacting with police vehicle, causing the police car to slow down

Signal Loss Events: None mentioned/None visible

Weather Changes: Lighting transitions from early morning/late afternoon golden hour to dusk with purple hues / No significant weather changes observed during the flight

Equipment Malfunctions: None mentioned/None visible

NOTES & LESSONS LEARNED

Flight Performance: The flight appears stable and smooth, with good control over altitude and movement. Steady tracking of police vehicle throughout rural terrain

Footage Quality: High quality 4K footage with consistent framing and smooth movement. The footage is clear, well-focused, and well-exposed throughout the duration

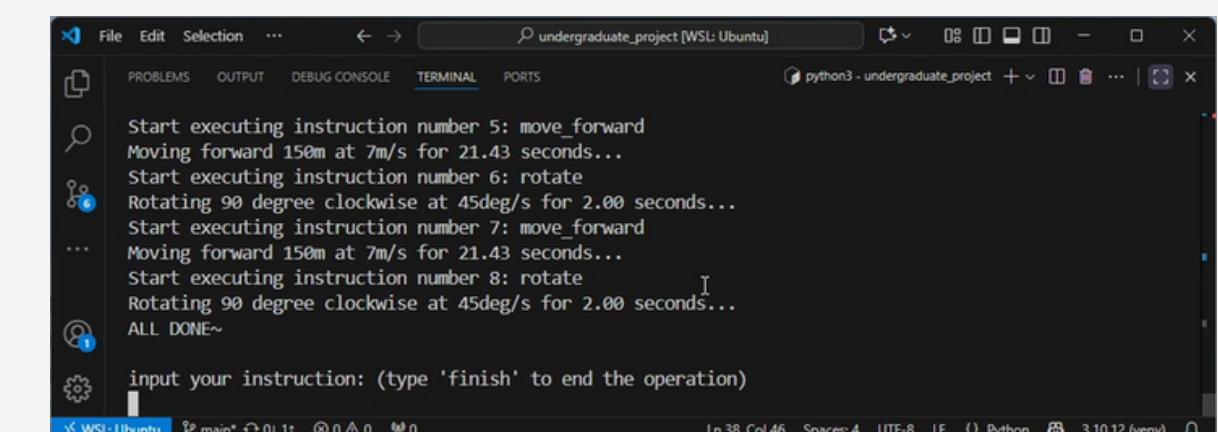
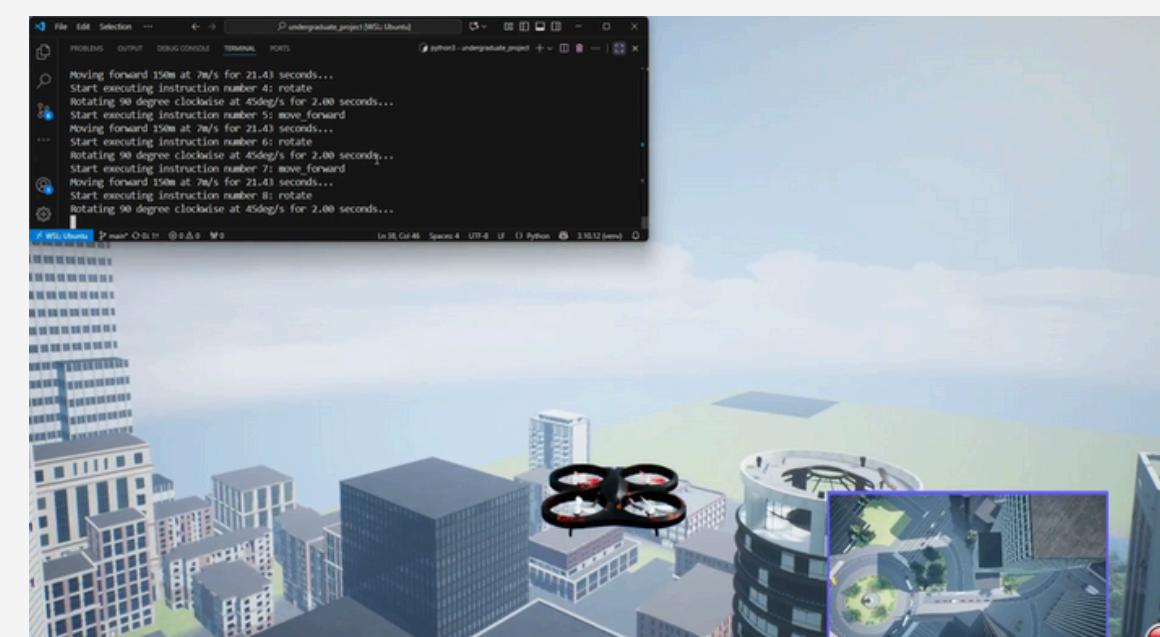
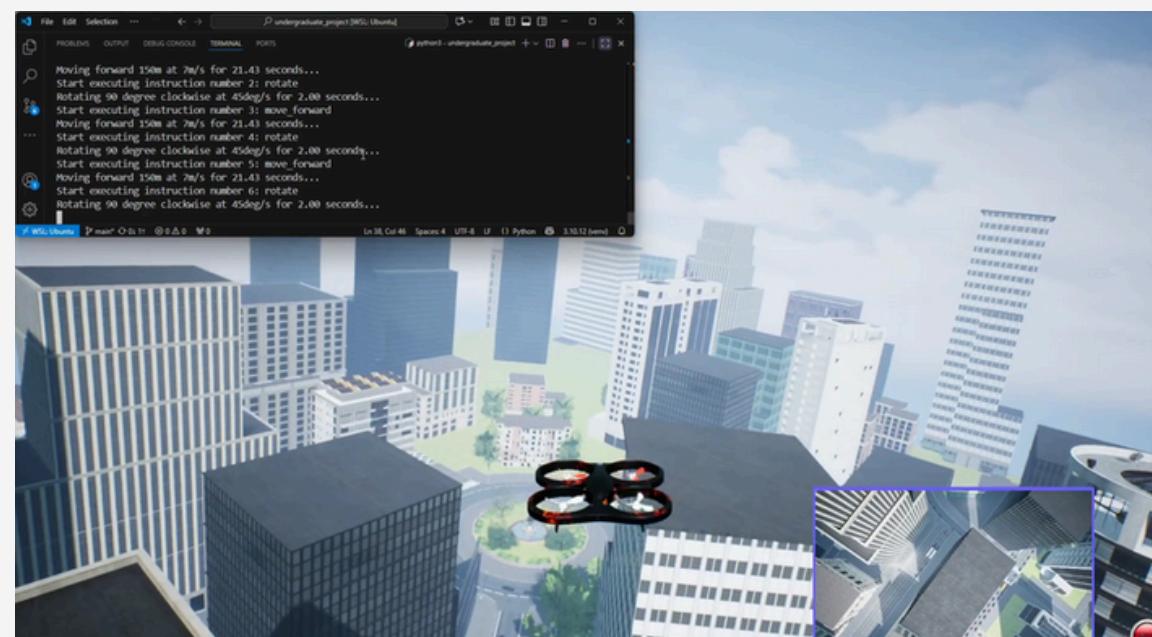
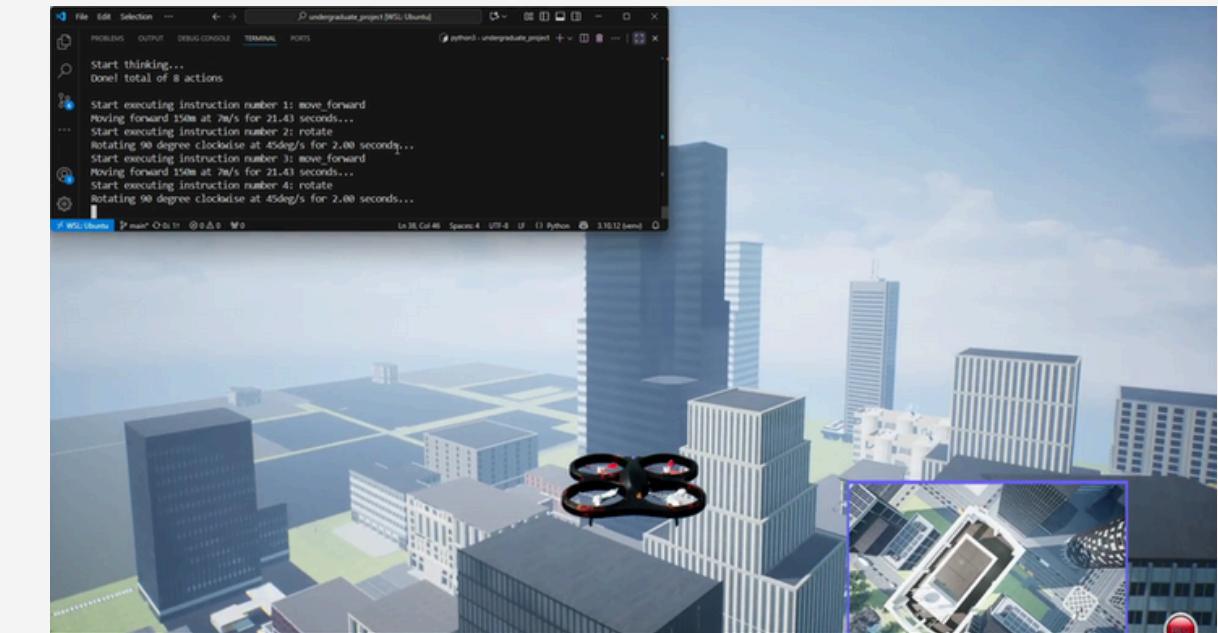
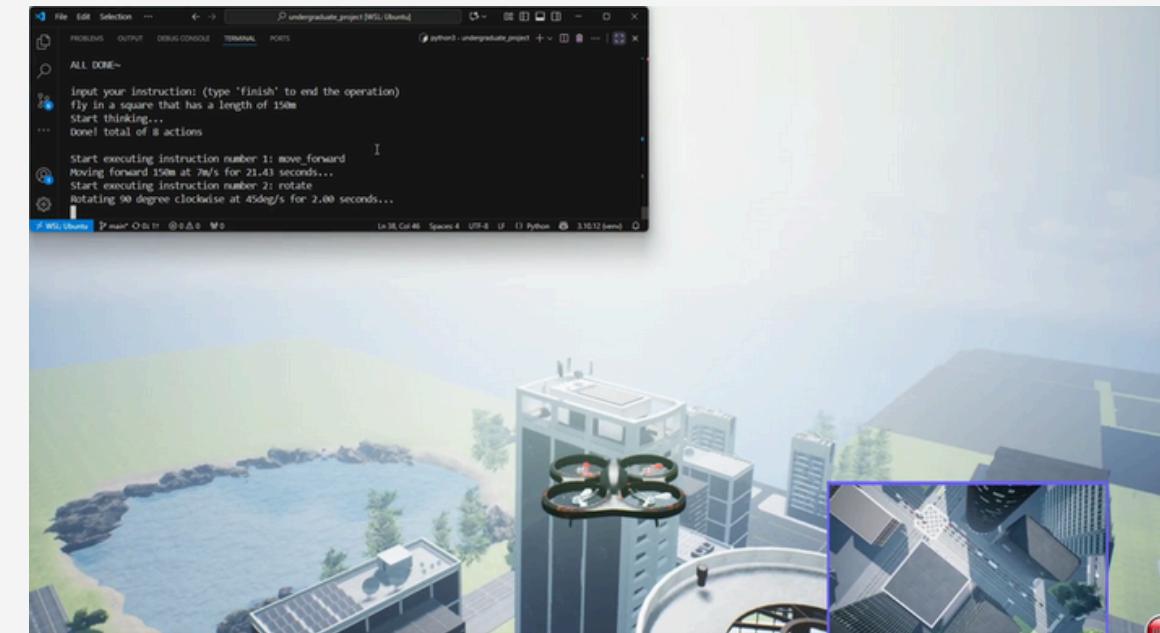
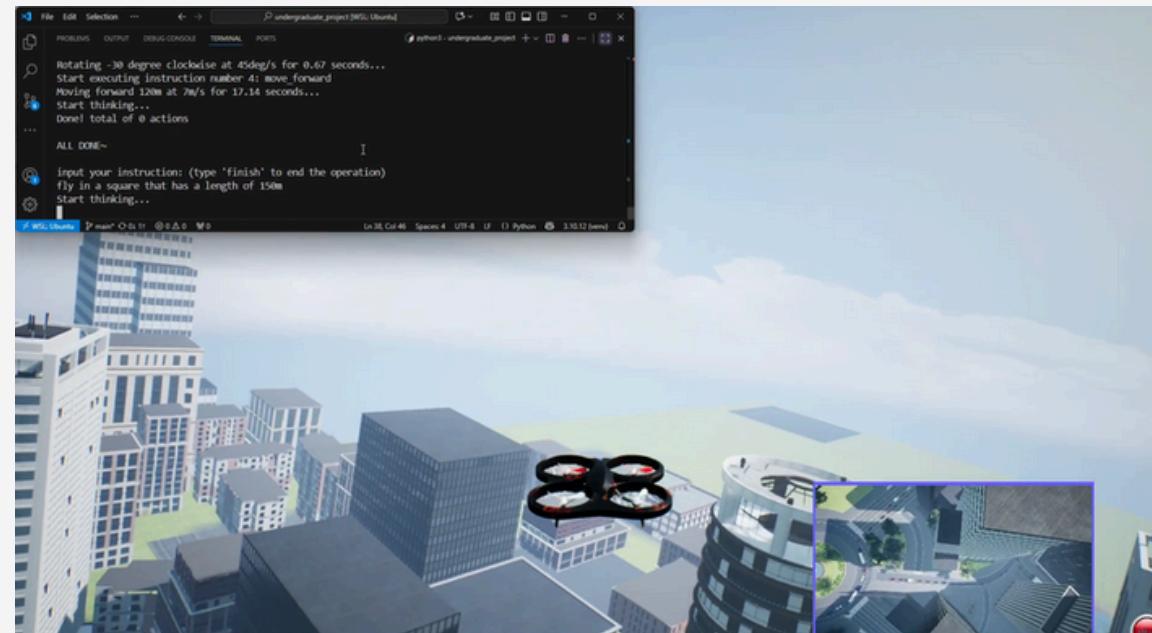
Areas for Improvement: Not mentioned

Future Considerations: For police activity documentation flights, establish multiple predetermined observation altitudes to capture both wide area context and detailed vehicle interactions. Consider coordination protocols when wildlife (cattle) may interfere with traffic stops. Plan for extended flight times during rural patrol documentation as distances between incidents may be greater than urban operations.

did mention cultivated field, but didn't see the lawnmower

Result - Autonomous Flight

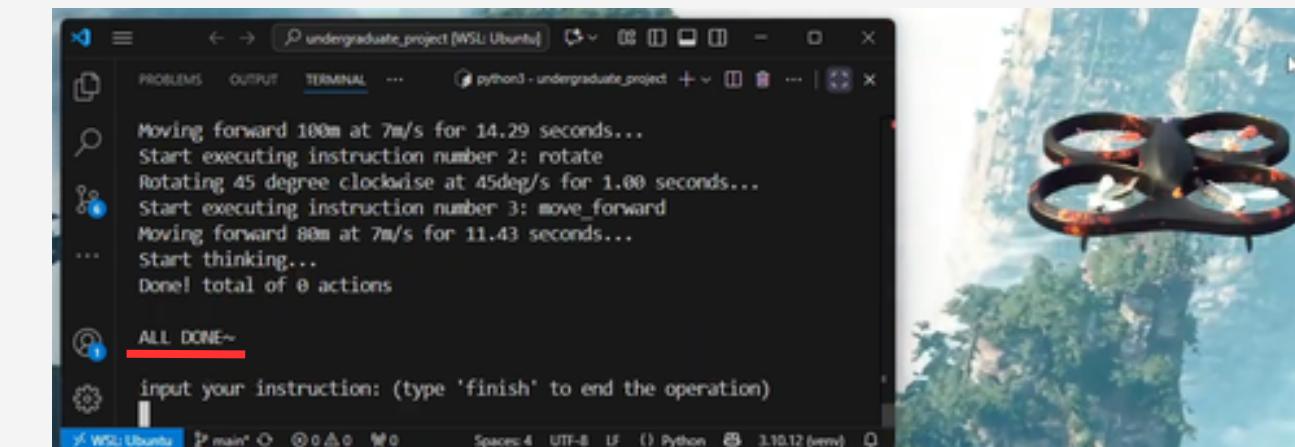
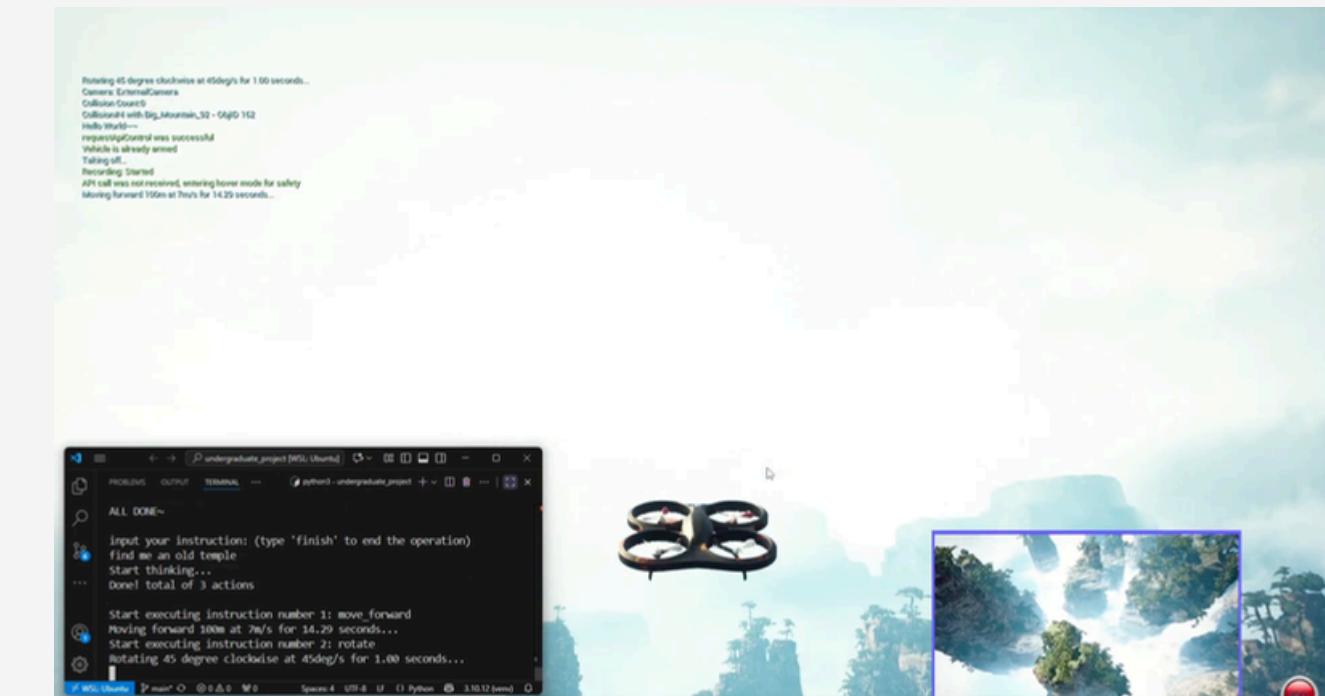
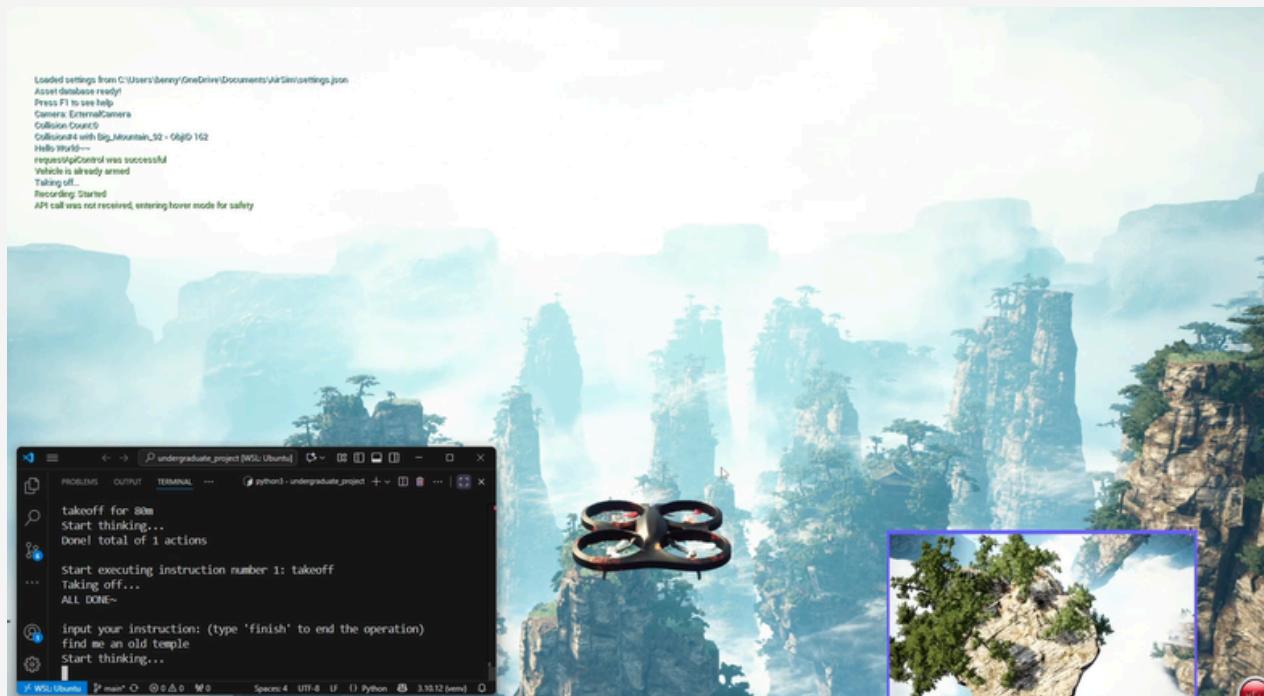
Navigation: “fly in a square that has a length of 150m”



We can see it flew back to where it started

Result - Autonomous Flight

(Vision) Task: “find me an old temple”

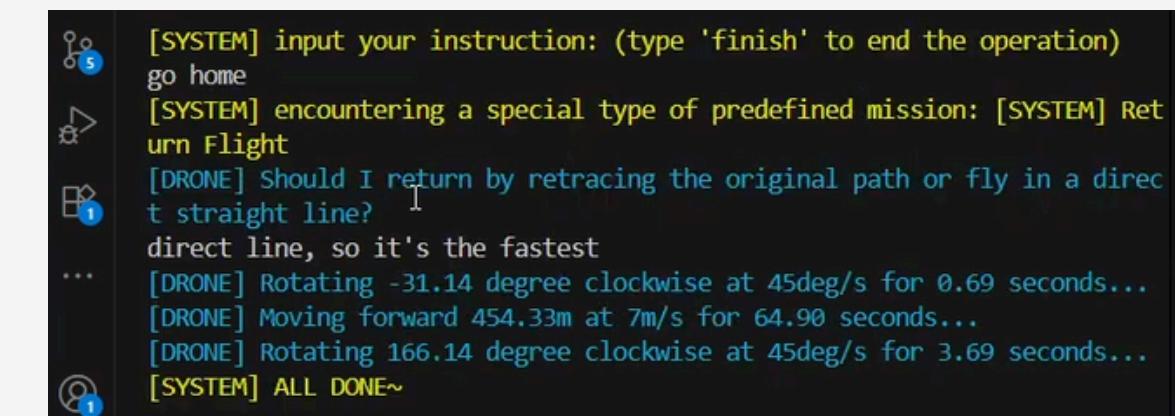
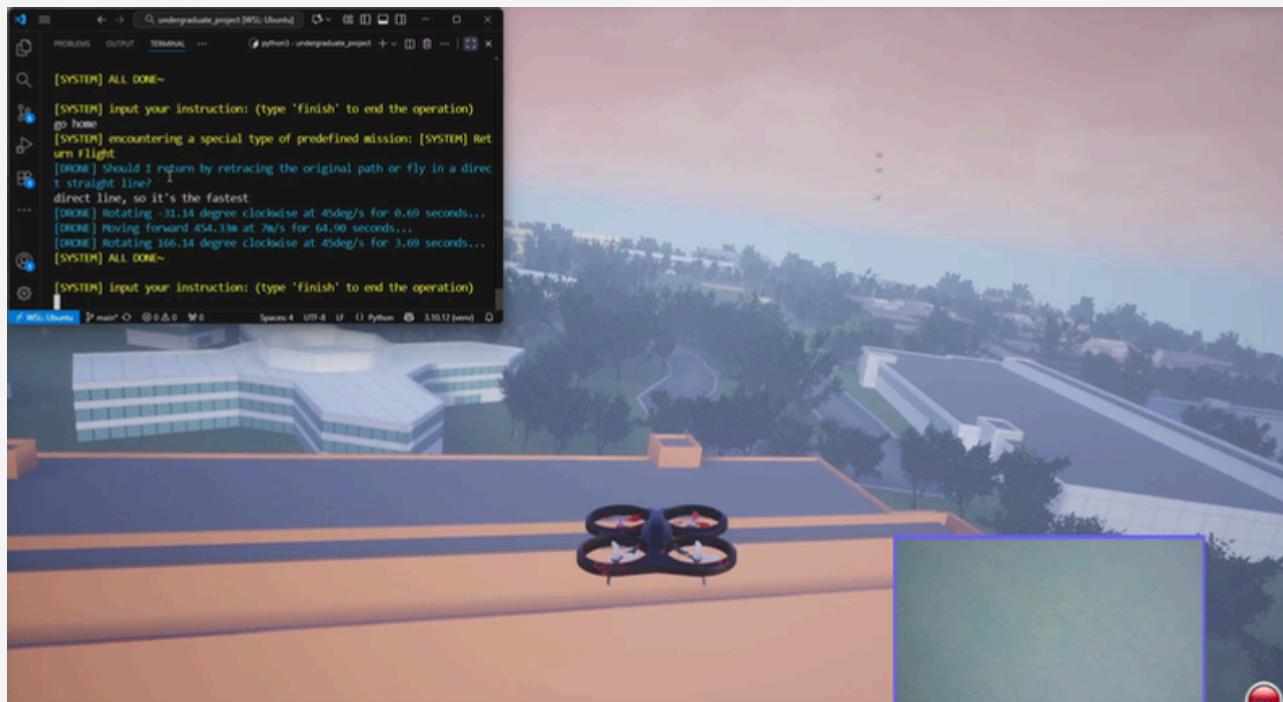
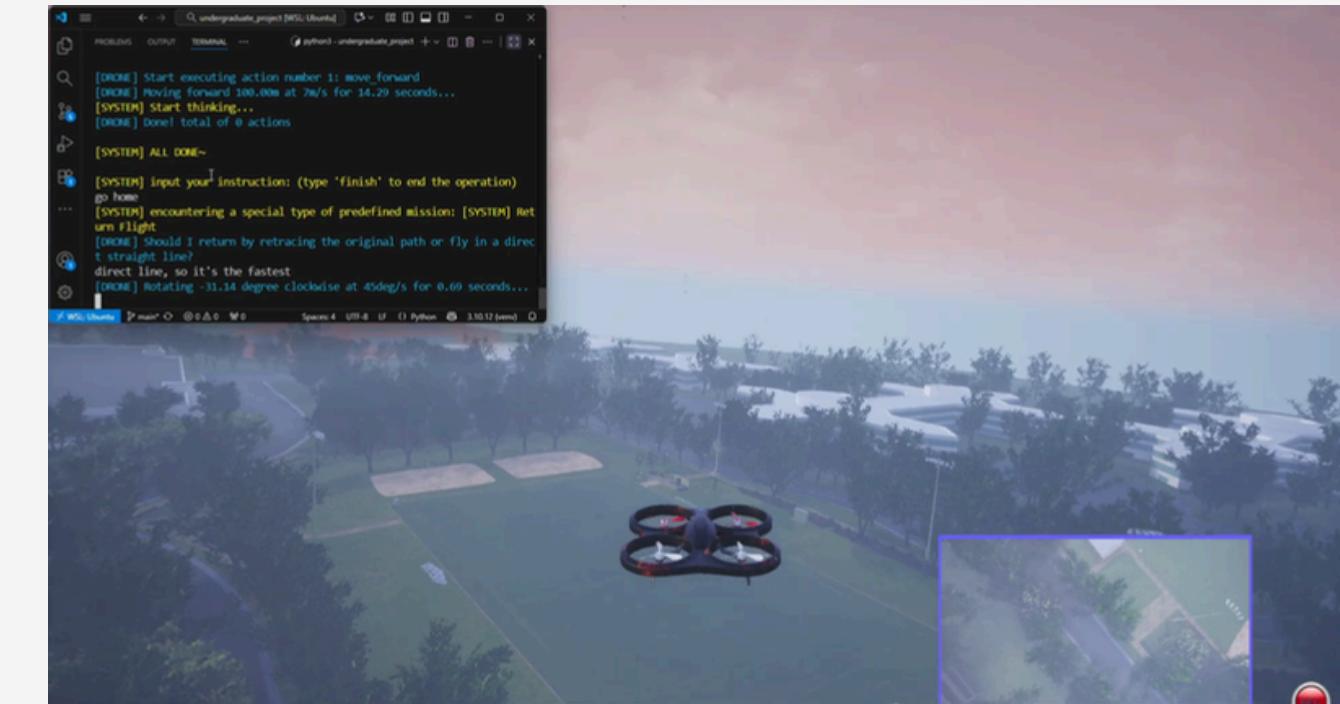
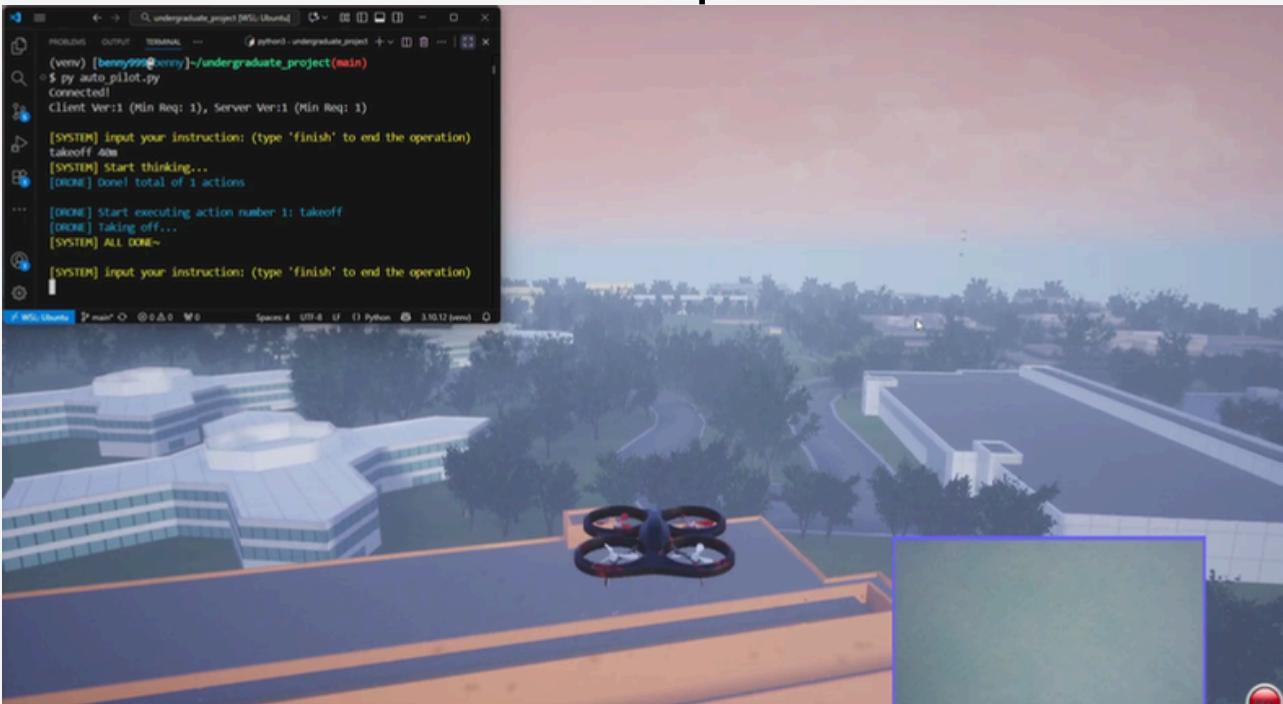


We can see it found a temple and stopped

Result - Autonomous Flight

Predefined Mission: “Return Flight - direct” (“go home”)

Takeoff place



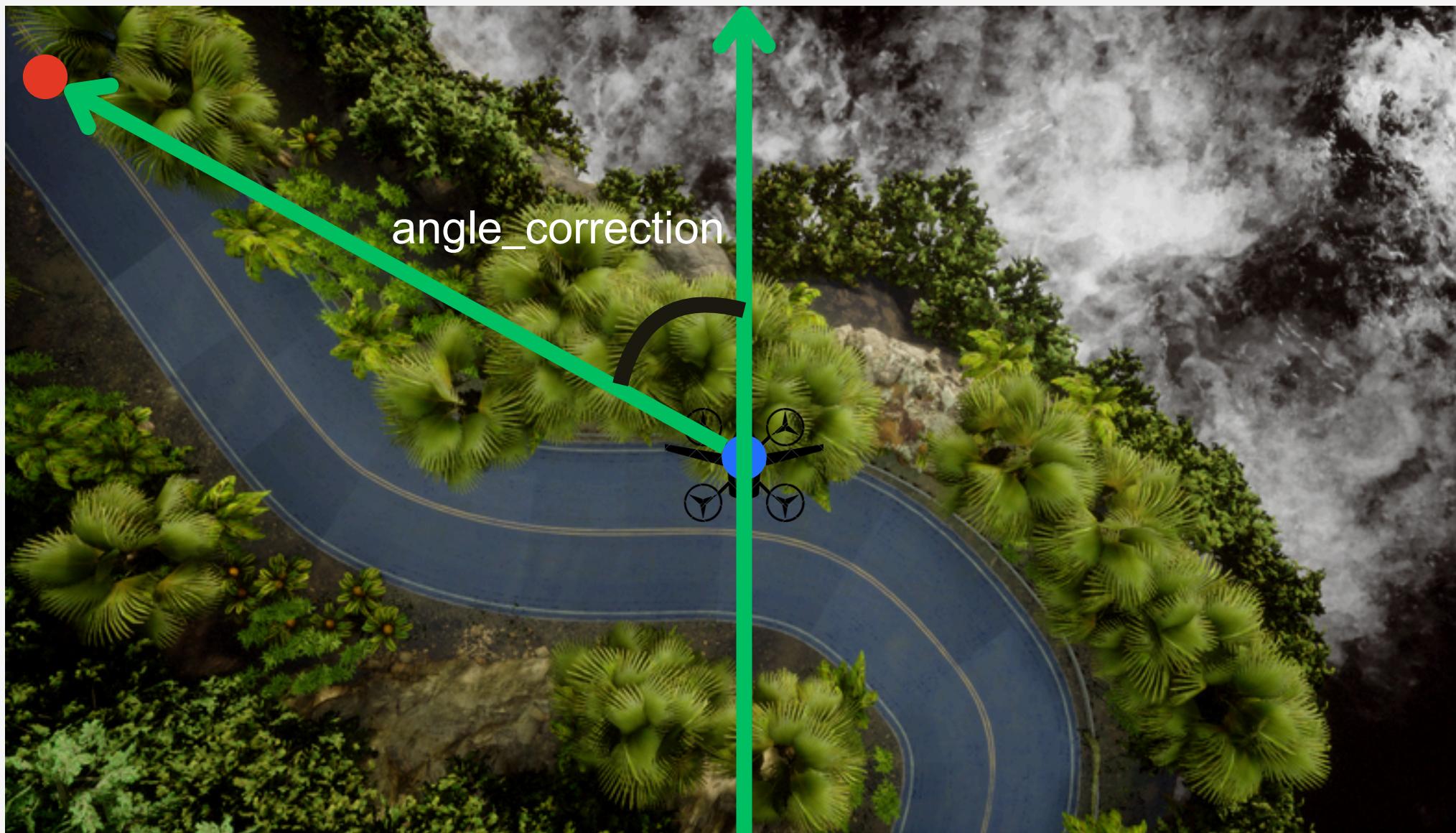
We can see it flew back to the takeoff place

Future

Future

Problem: VLM's vision ability

It's challenging to successfully implement the predefined mission: Follow Path.
The main issue is that VLM's vision abilities are still lacking.



End