Columbia University in the City of New York

Columbia University Data Science Institute
Smart Cities Center

Northwestern University

FlexiFly: Interfacing the Physical World with Foundation Models Empowered by Reconfigurable Drone Systems

SenSys 2025
Irvine, CA

Minghui (Scott) Zhao*Φ, Junxi Xia+Φ, Kaiyuan Hou*Φ, Yanchen Liu*, Stephen Xia+, Xiaofan (Fred) Jiang*

Columbia ICSL
Intelligent & Connected Systems Lab

*Columbia University
+Northwestern University

IMEC
Intelligent Mobile and Embedded Computing Lab

ΦCo-first authors

Columbia | Engineering
The Fu Foundation School of Engineering and Applied Science

Imagine AI that can truly help us in the physical world

But today's AI can’t adapt to dynamic physical tasks

Restricted by fixed sensors and static deployment locations

But today's AI can't adapt to dynamic physical tasks

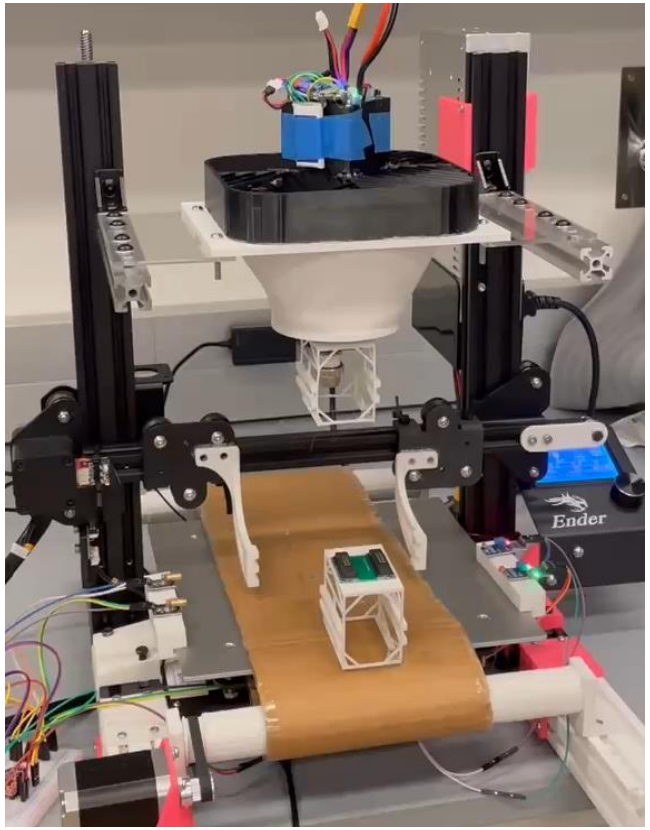
Restricted by fixed sensors and static deployment locations

Fixed Configuration

Coverage Gaps

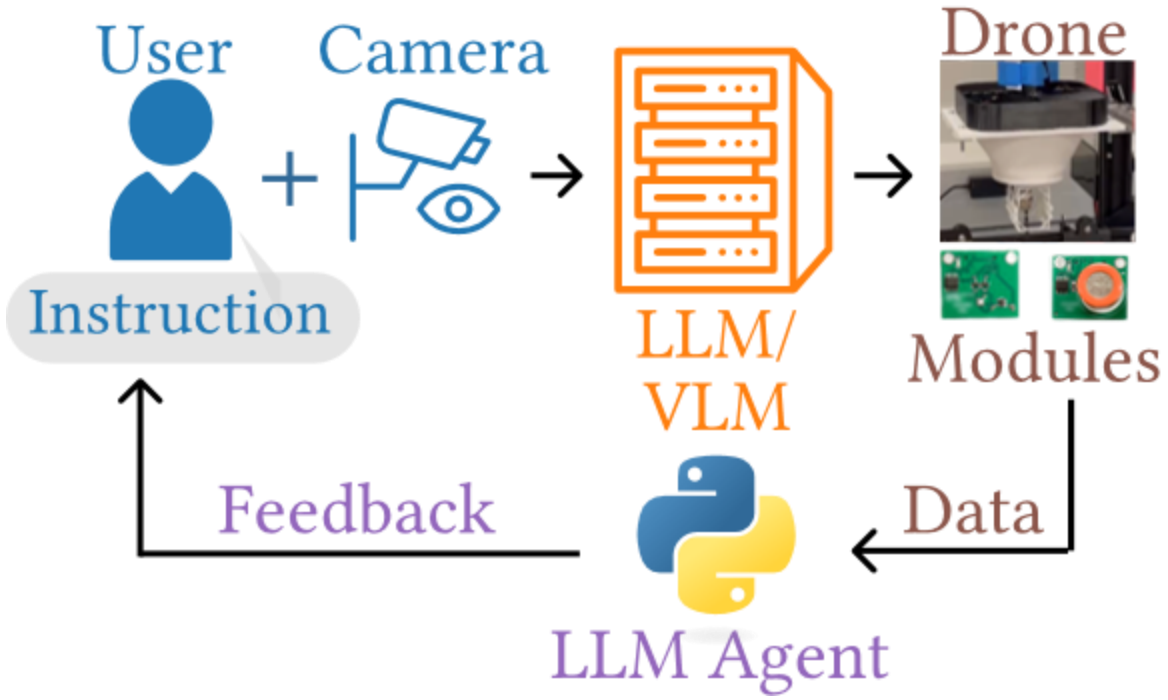
Physical Interactions

Creating AI That Can Sense, Move and Act in Our World



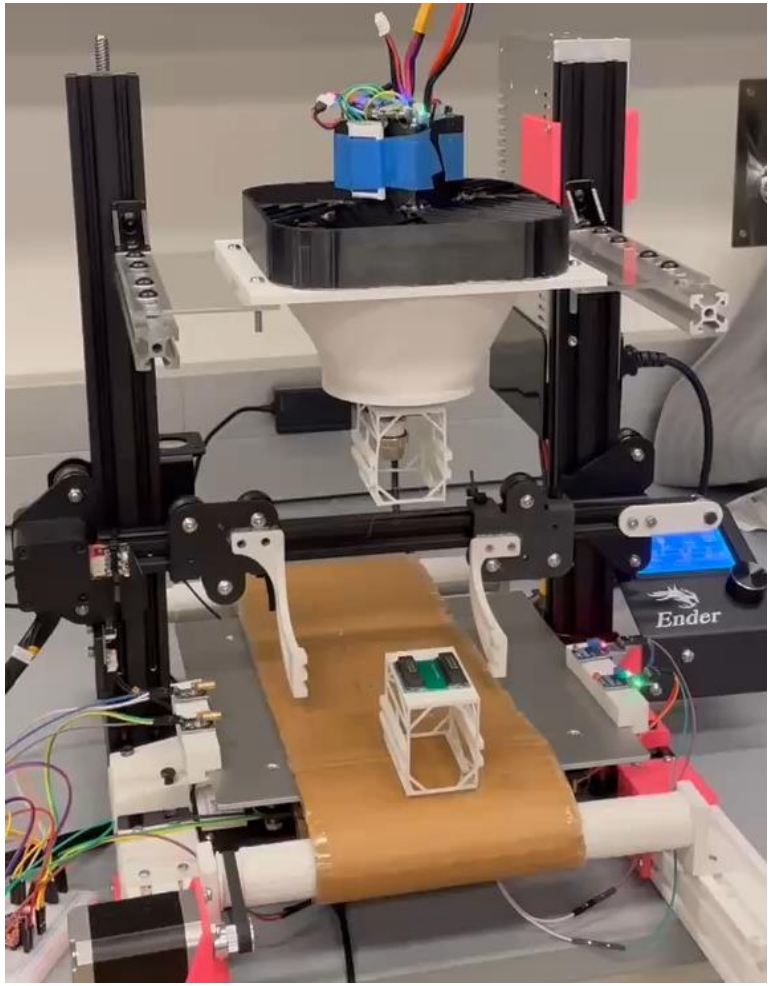
Reconfigurable Drone Platform

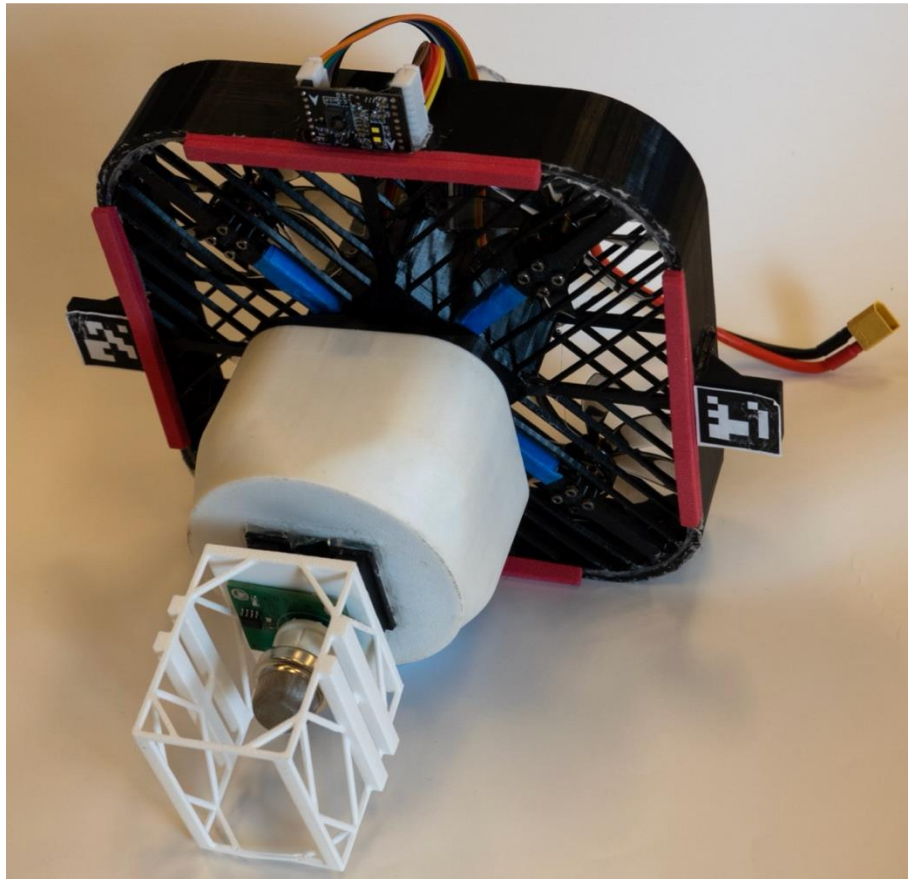
+



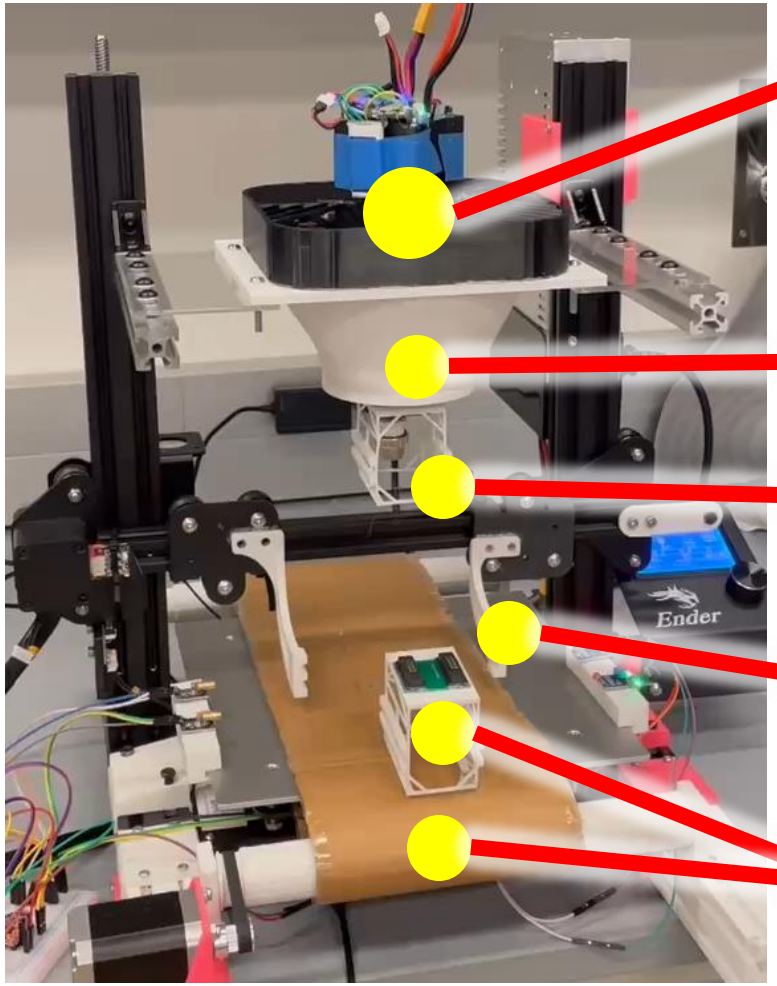
Foundation Model Pipeline

Building AI's Physical Presence

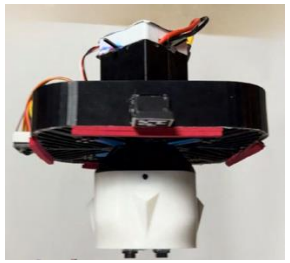




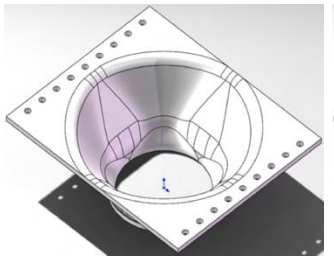
Building AI's Physical Presence

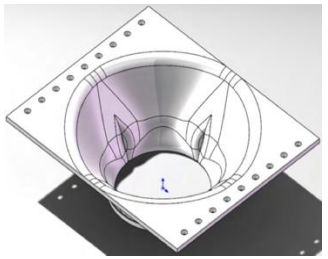


Drone
Carries modules

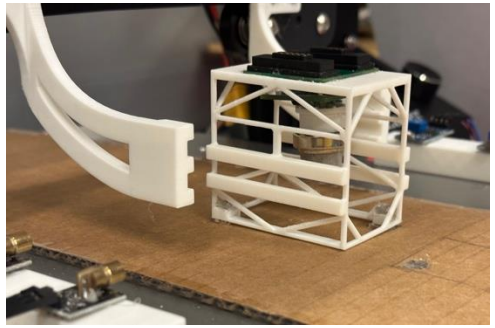


Landing Platform
Precise alignment





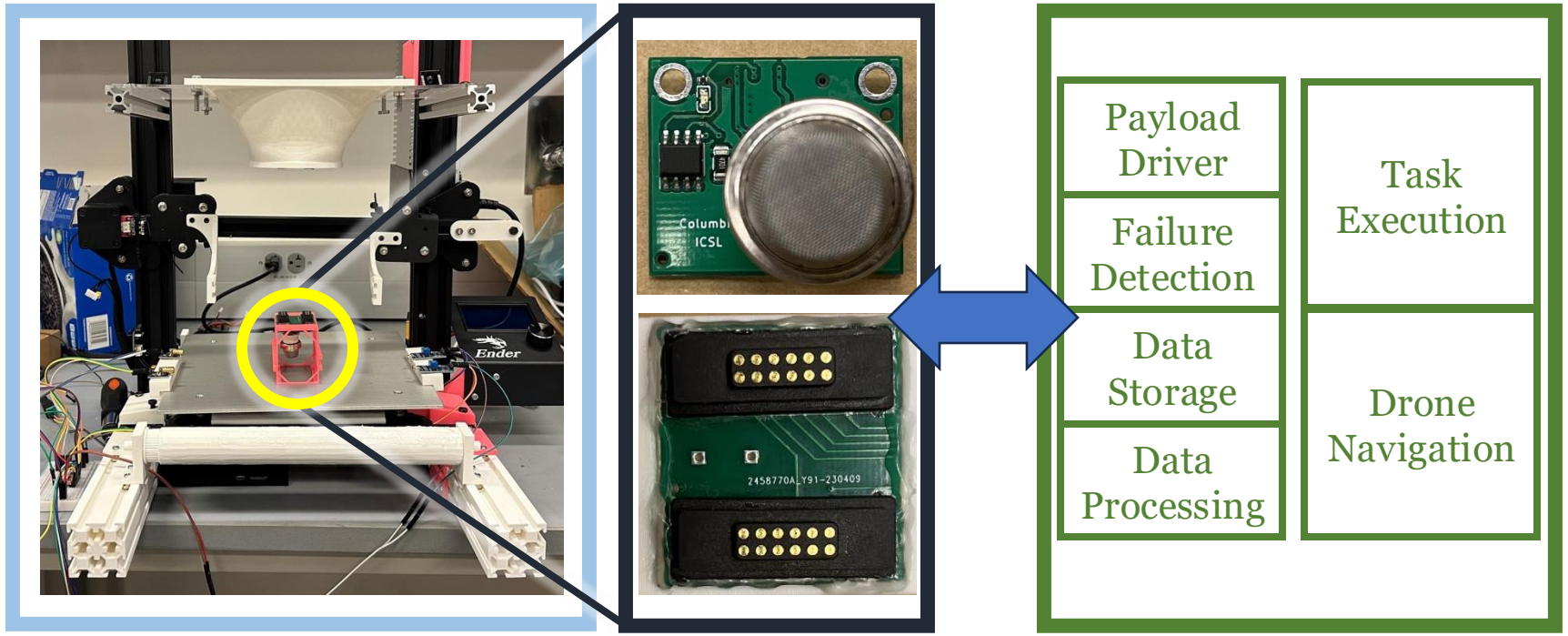
Swappable Module



Claw
Transports sensor modules

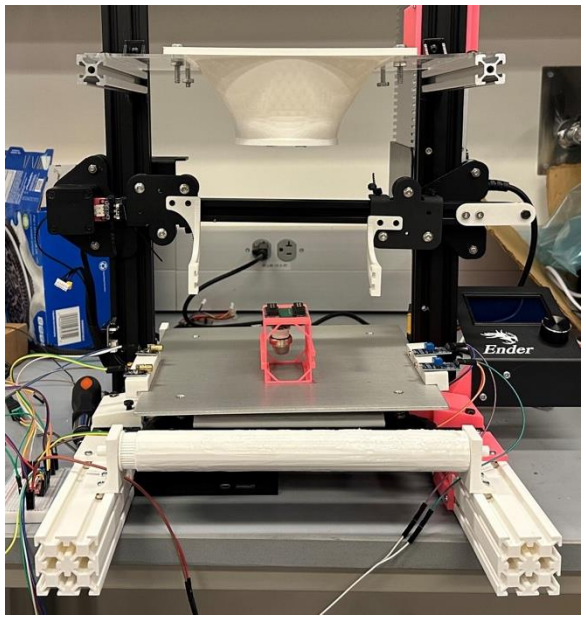
Module Repository

Reconfigurability Breaks Down to 3 Layers



Mechanical Layer

Physical Swapping

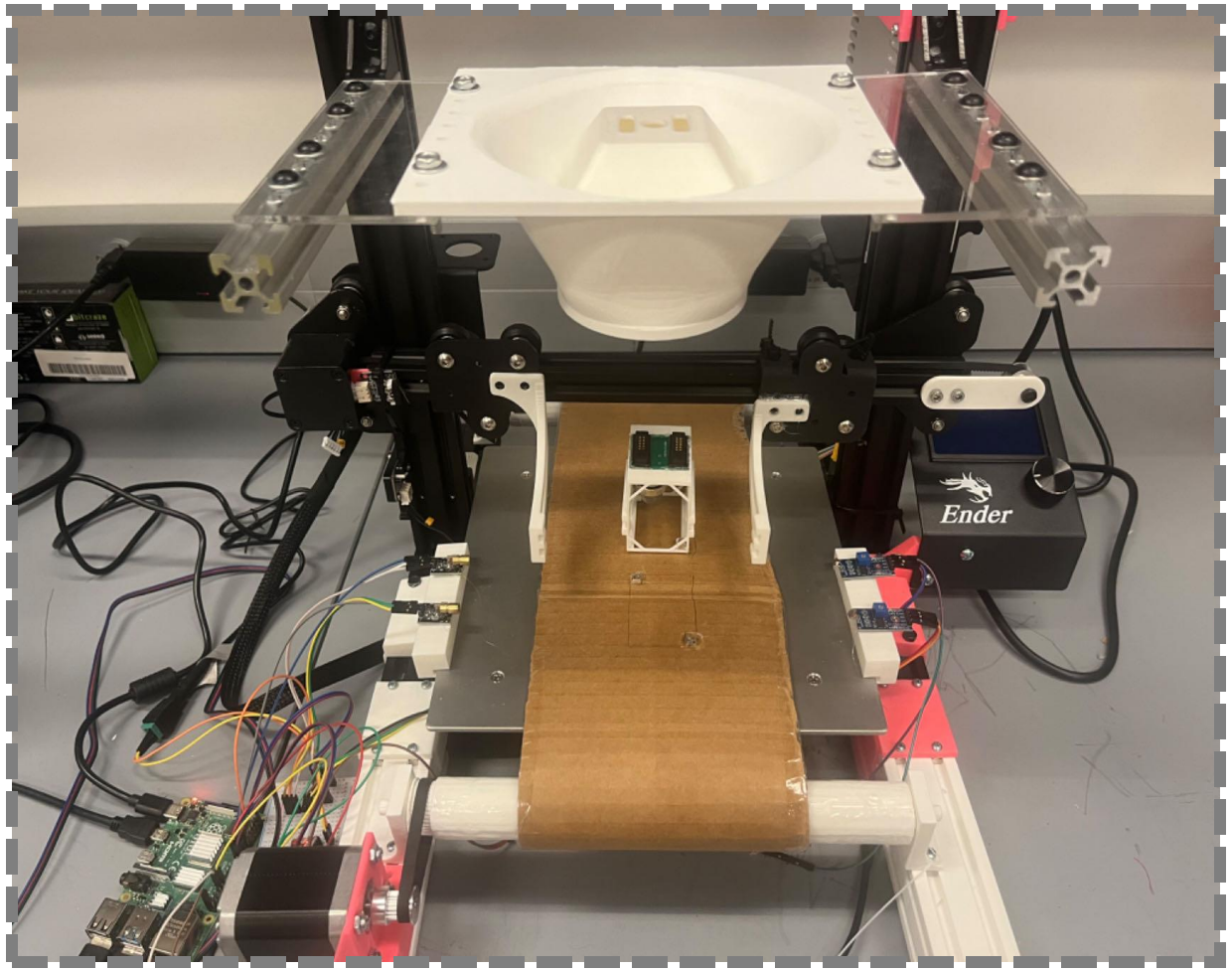


Mechanical Layer

Connection Layer

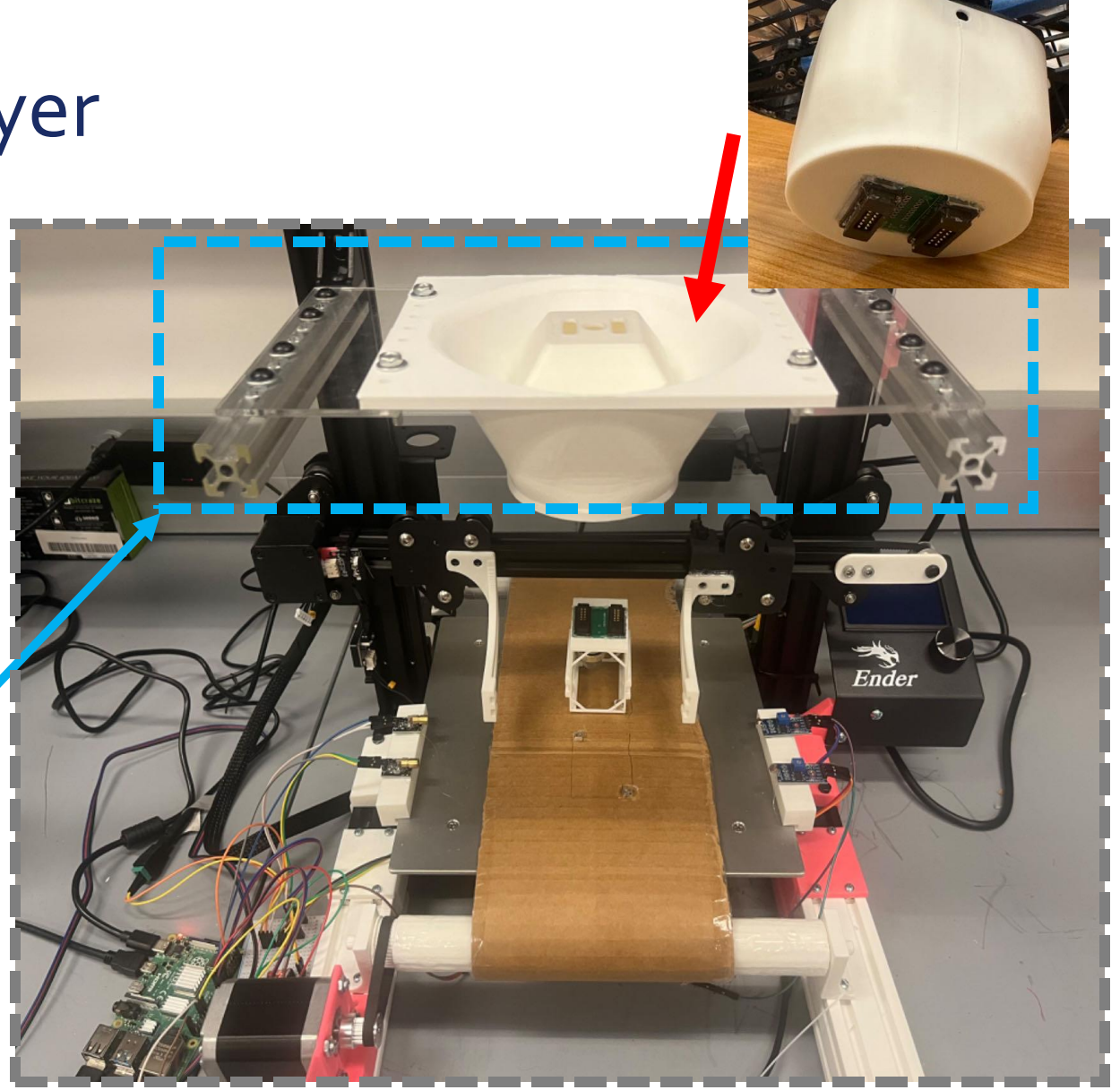
Software Layer

Mechanical Layer



Ender-3 Open-Source 3D Printer

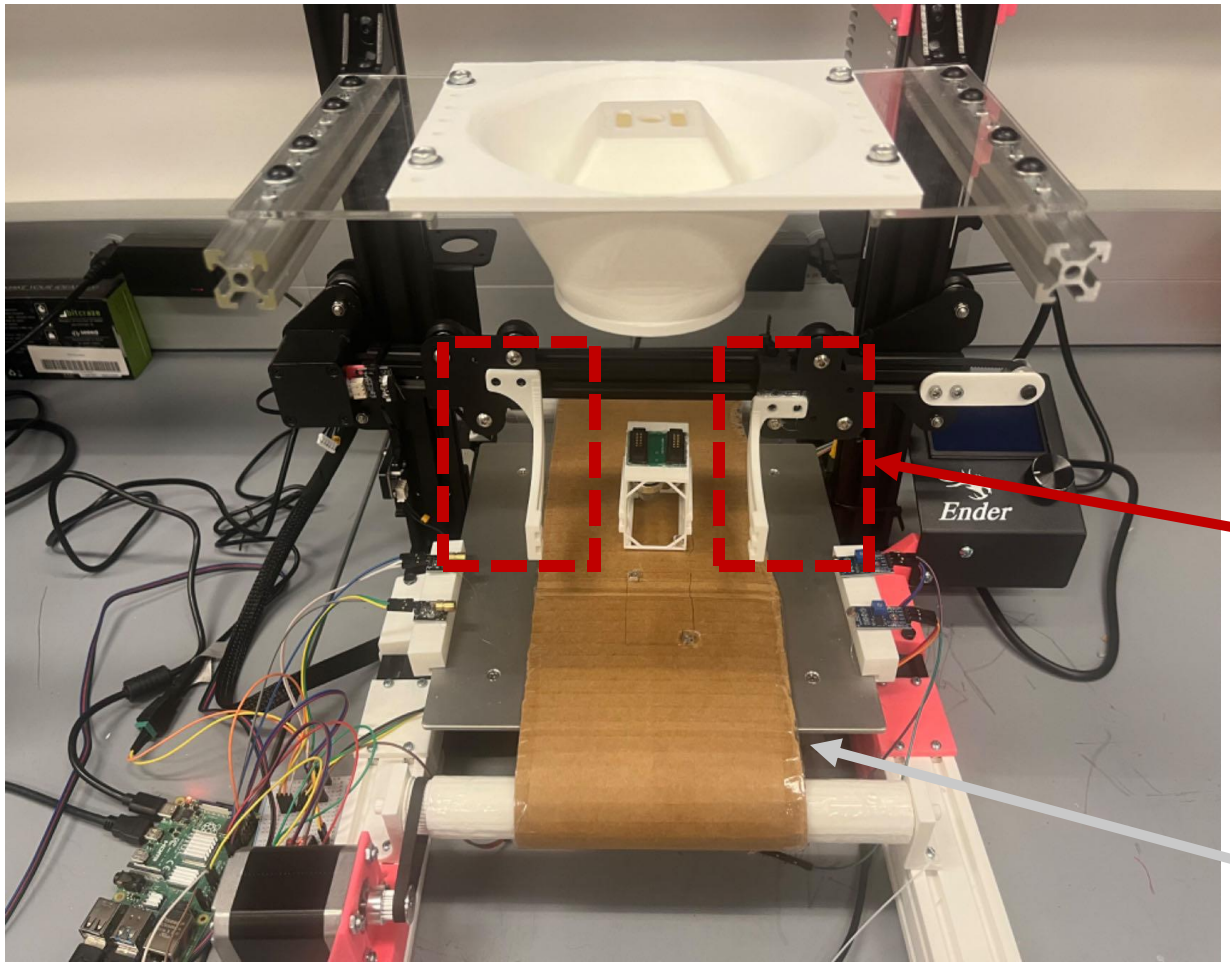
Mechanical Layer



Funnel-shaped Landing Station

Ender-3 Open-Source 3D Printer

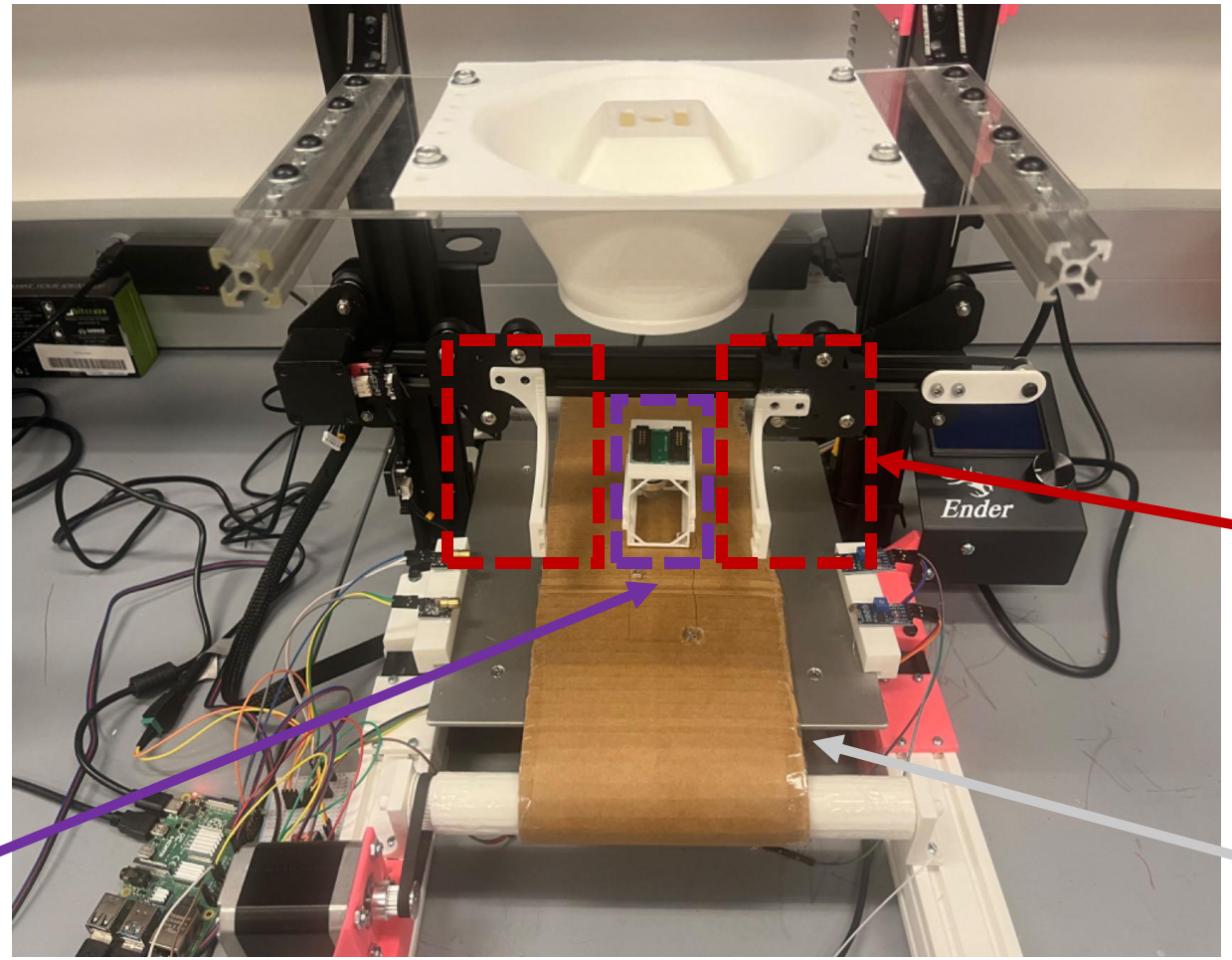
Mechanical Layer



Module Grippers

Conveyor Belt of Sensors

Mechanical Layer

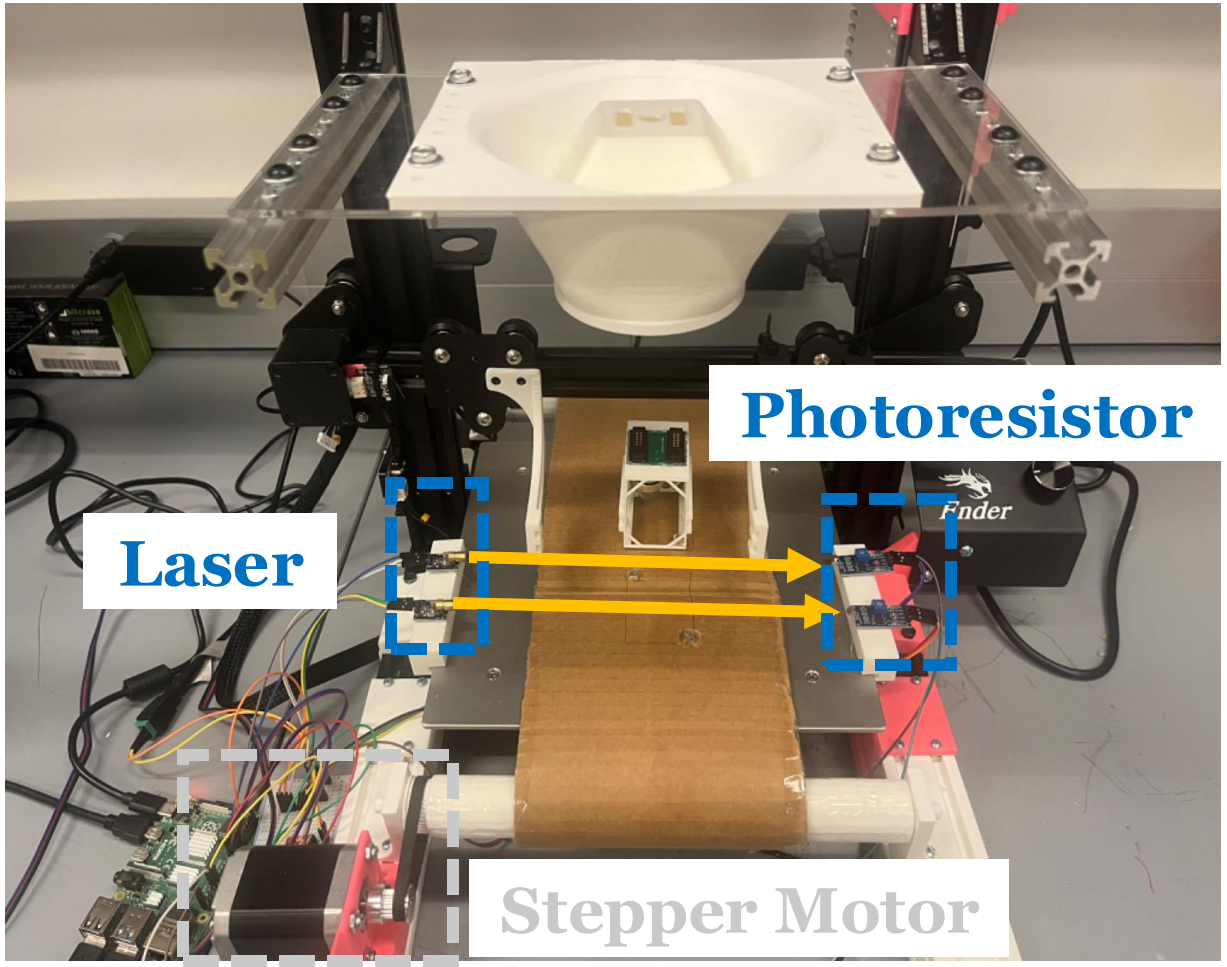


Module Grippers

Sensor Module

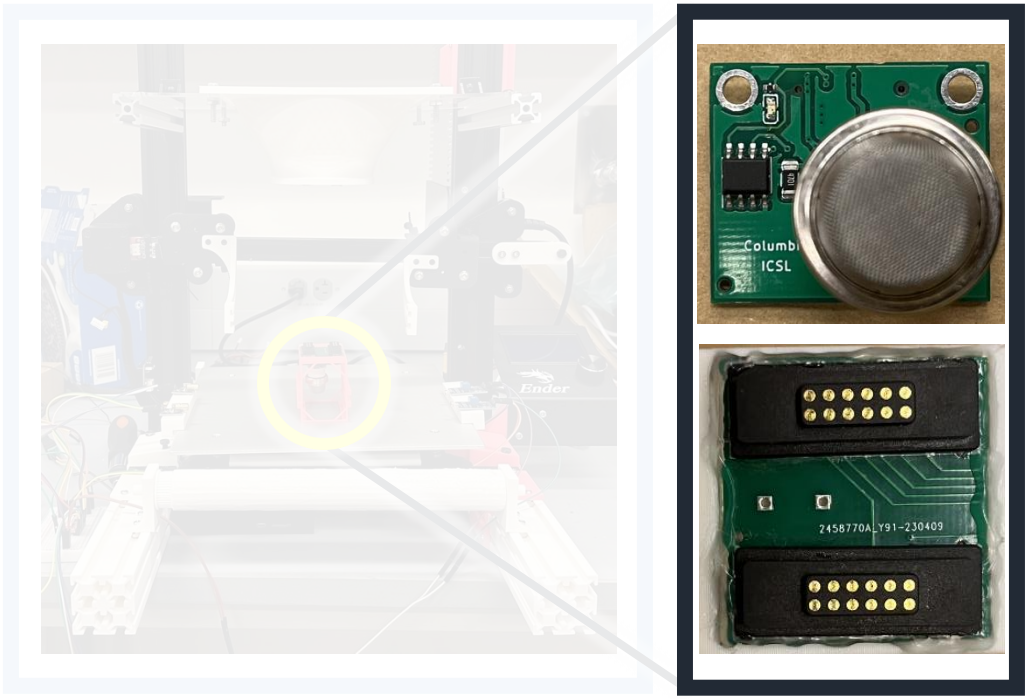
Conveyor Belt of Sensors

Mechanical Layer



Connection Layer

Maintains Connection



Connection Layer

Payload Driver

Failure Detection

Data Storage

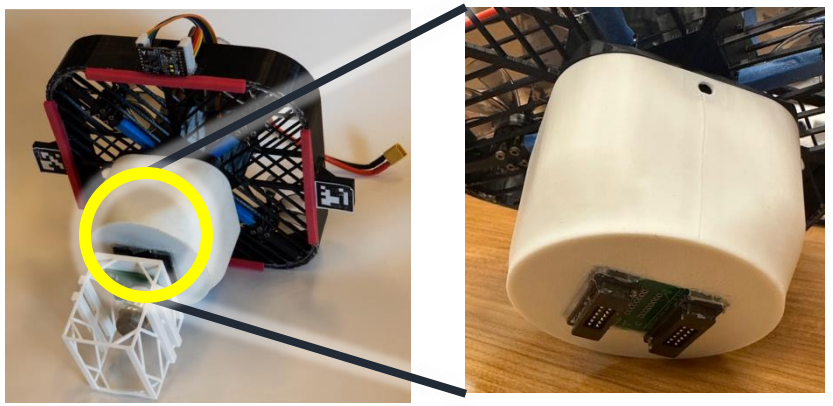
Data Processing

Task Execution

Drone Navigation

Physical Connection and Transportation

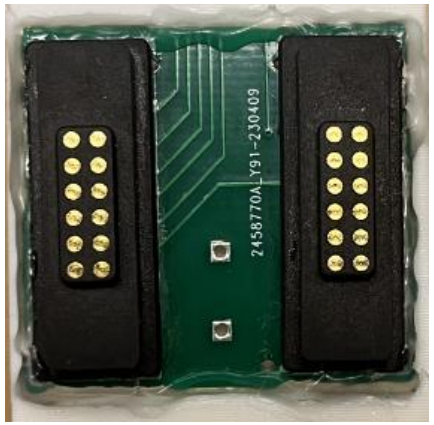
On the Drone



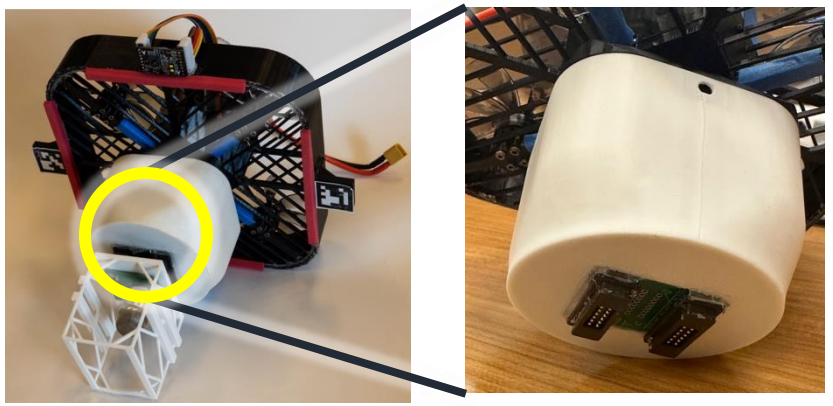
Drone

Physical Connection and Transportation

Sensor Module



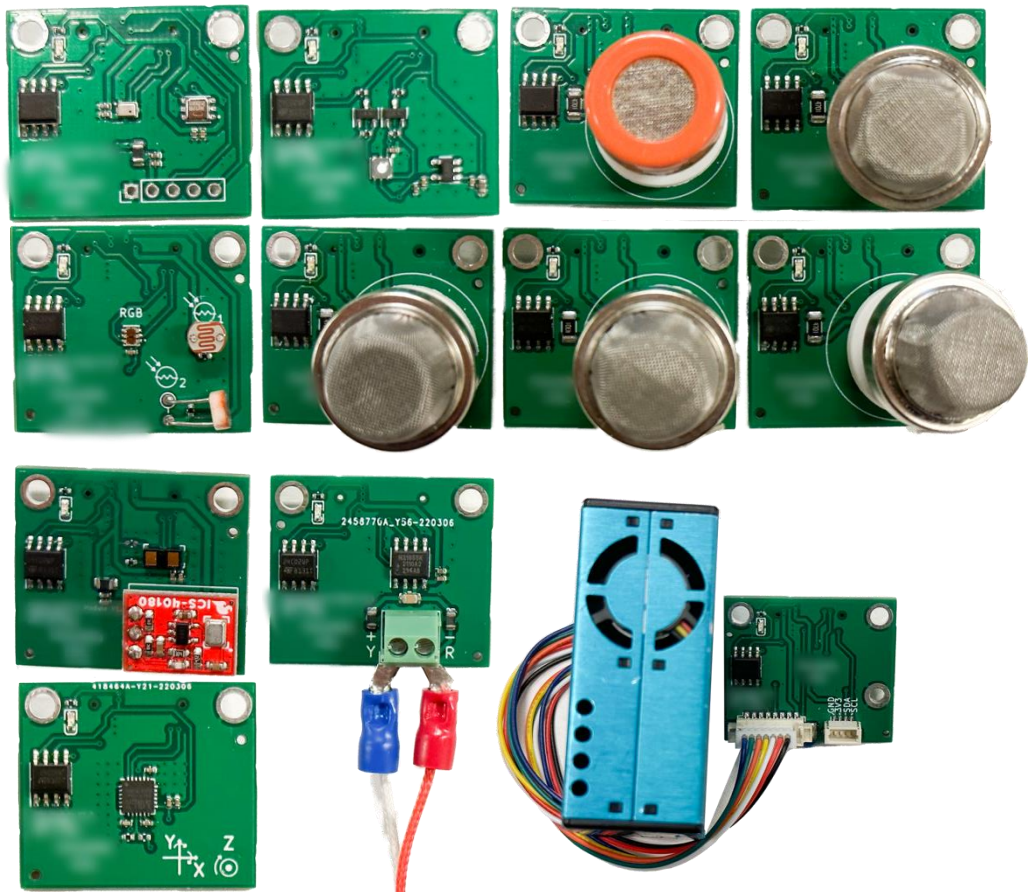
On the Drone



Physical Connection and Transportation

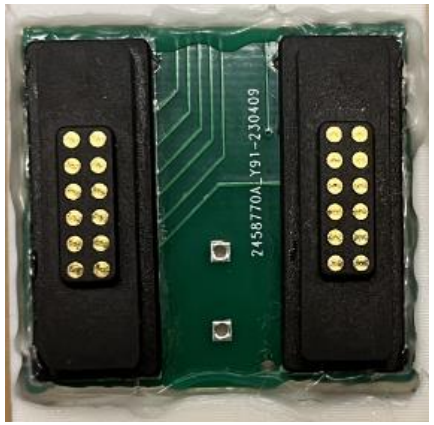
Sensor Module

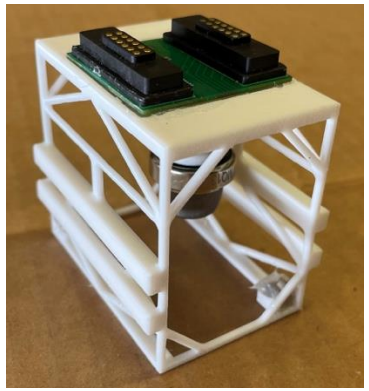
Designed Sensor Modules

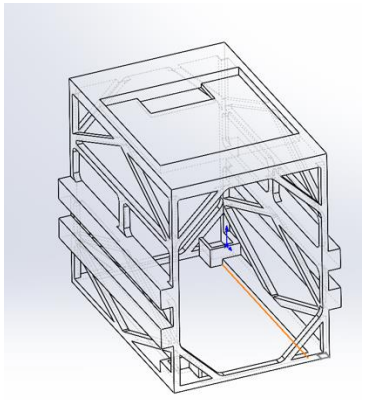


Physical Connection and Transportation

Sensor Module

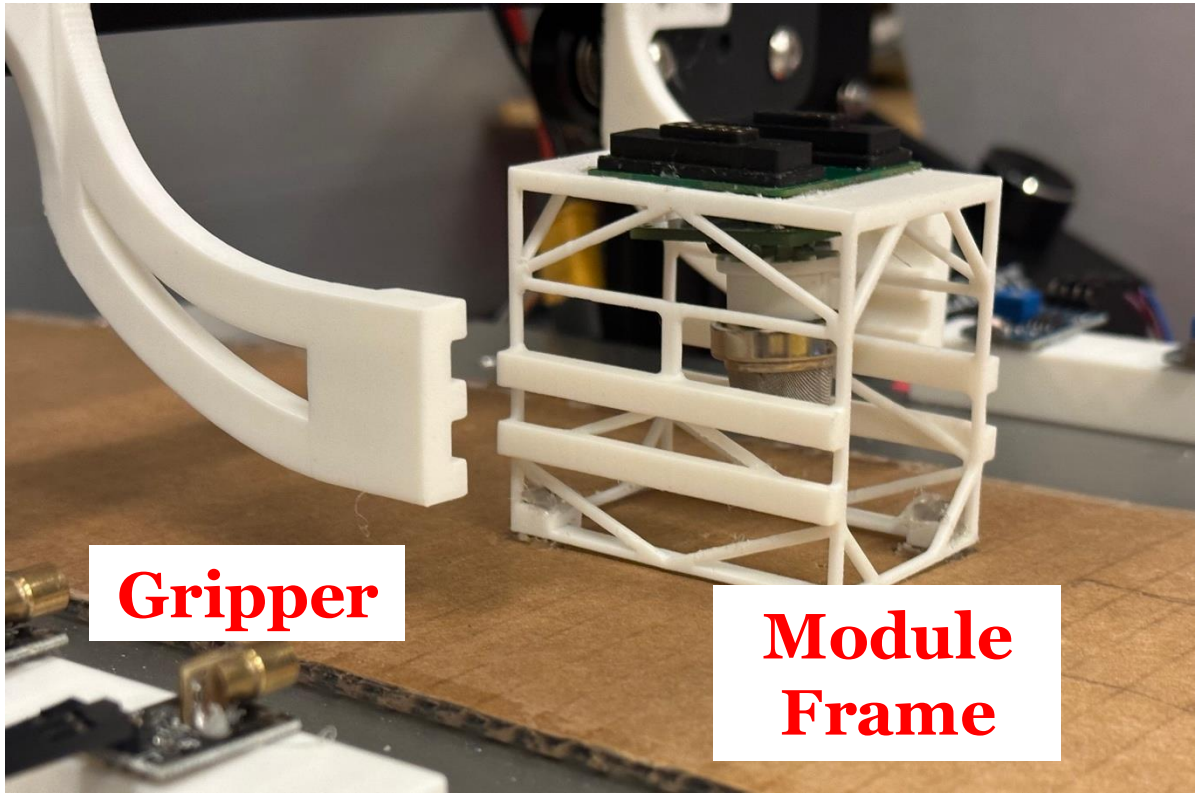






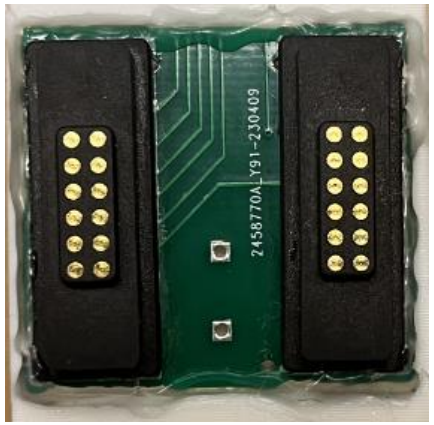
Module Frame

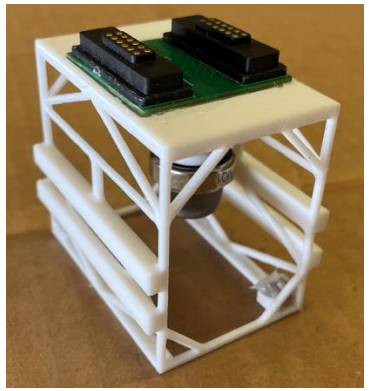
Gripper Mechanism

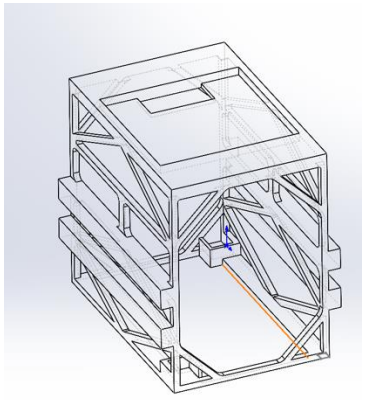


Physical Connection and Transportation

Sensor Module

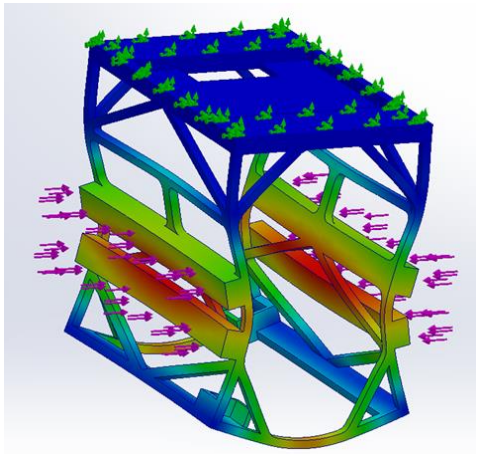




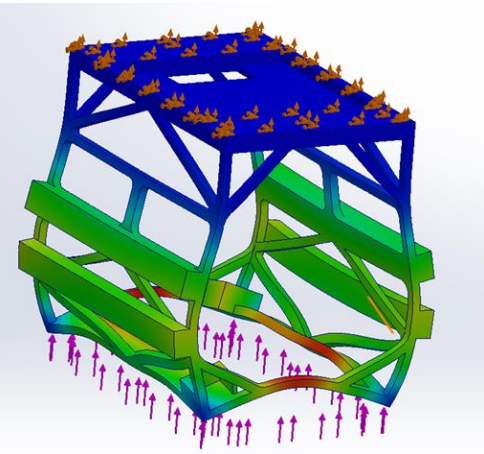


Module Frame

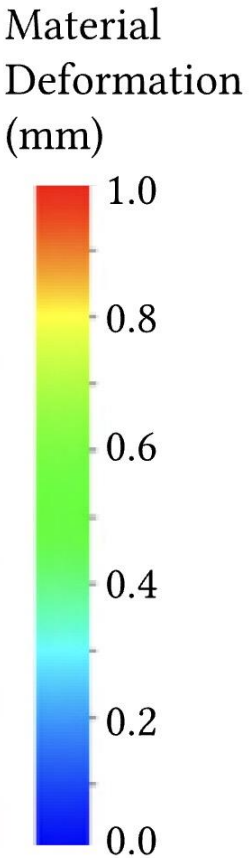
Force Simulation



Grabbing Module

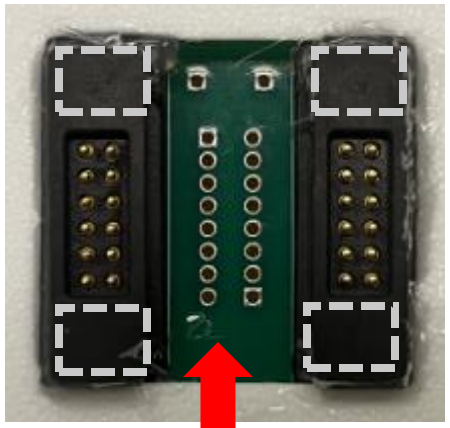


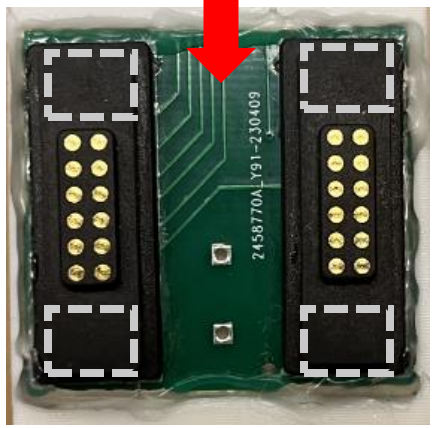
Placing Down Module



Electrical Connection

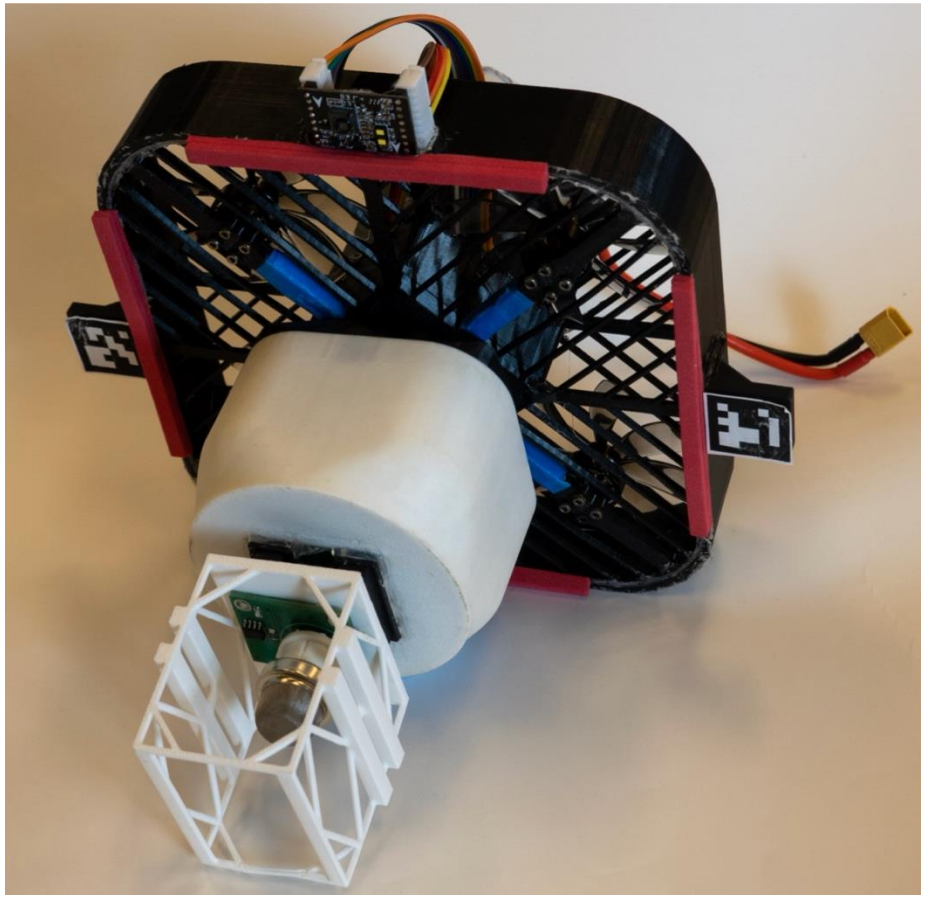
On the Drone





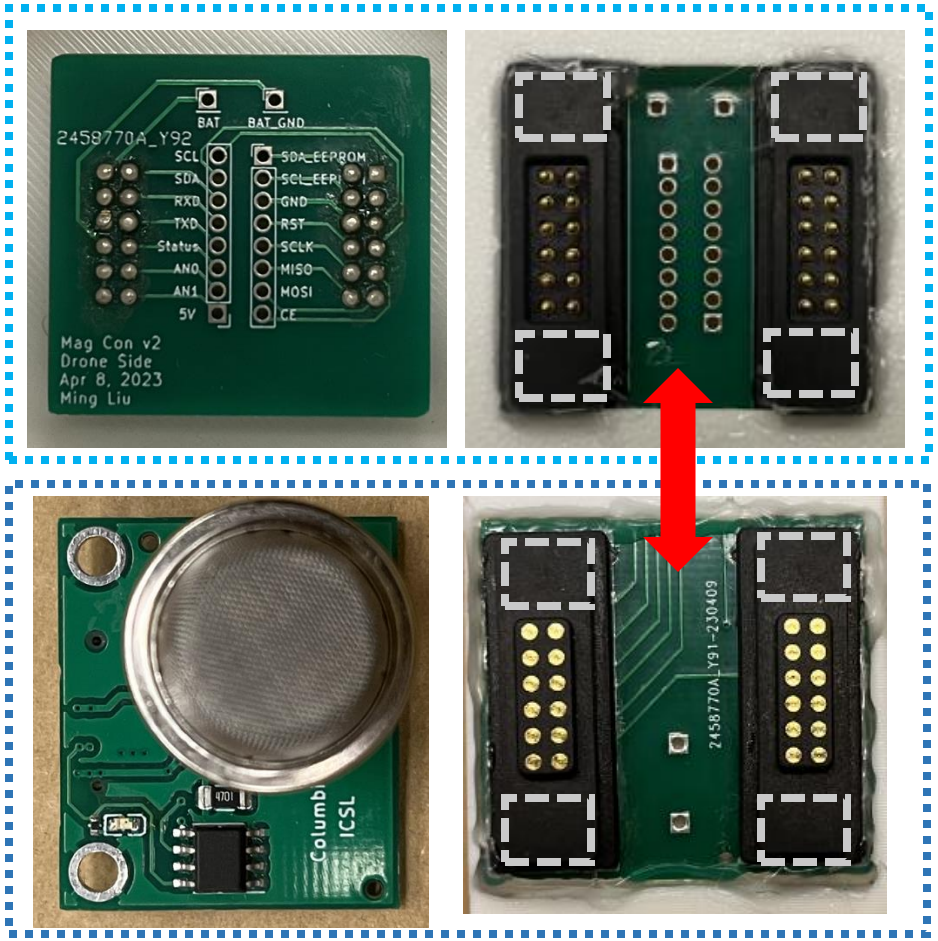
Sensor Module

Weak Magnets



Electrical Connection

On the Drone

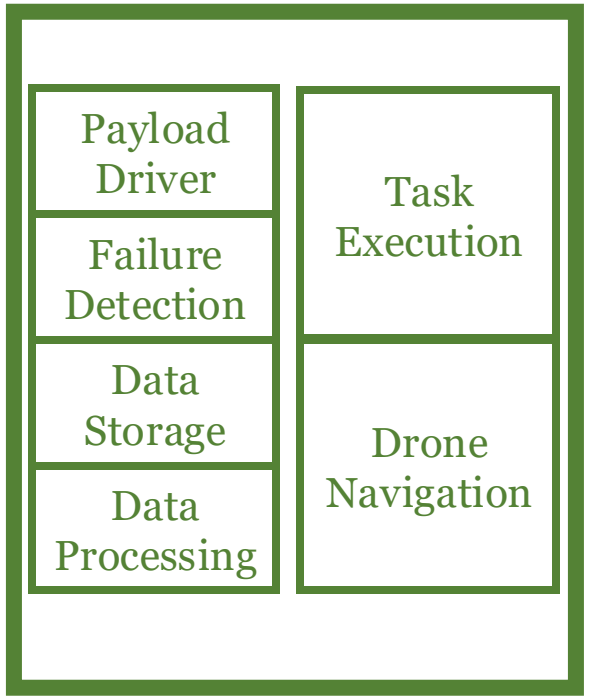


Unified Connector Interface

24 spring loaded pins

- ❖ I²C
- ❖ SPI
- ❖ Analog
- ❖ UART

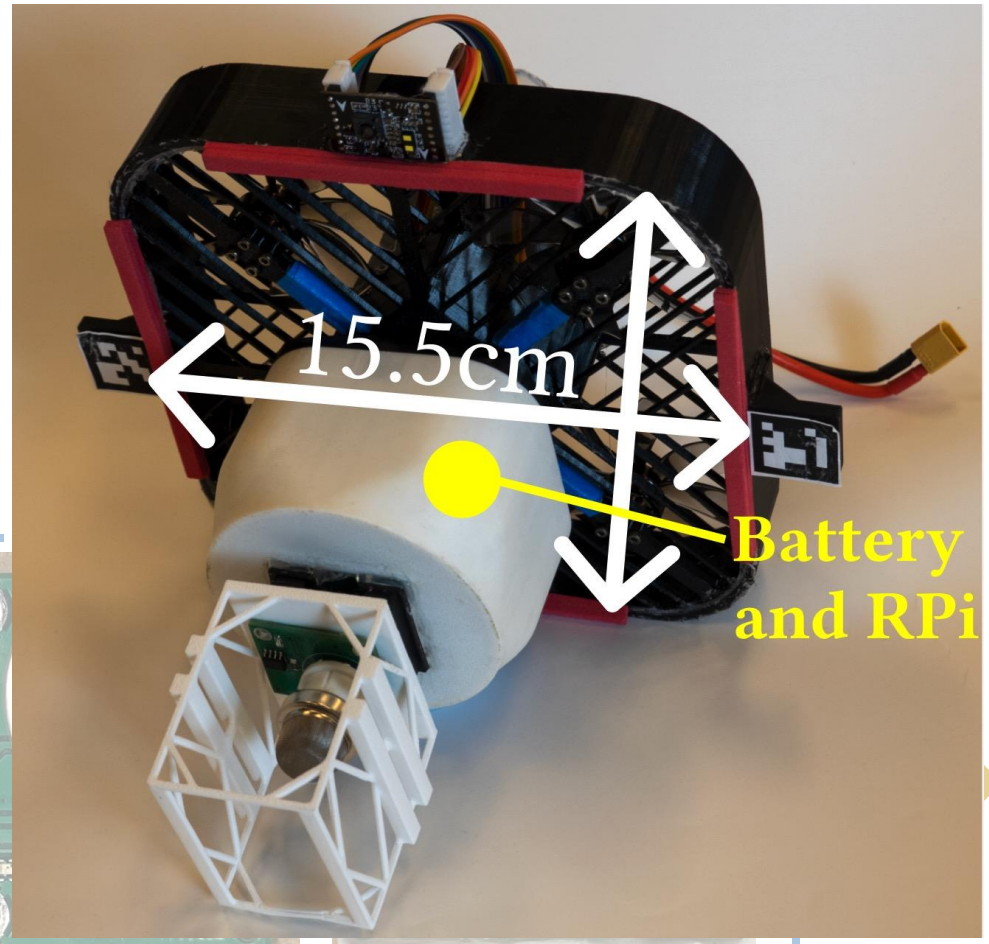
Software Layer



Software Layer

Software Layer – Sensor Data Access

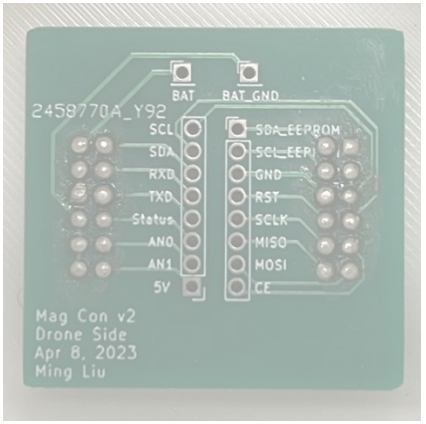
Sensor Driver	Failure Detection	Data Storage	Data Processing



Linux

Raspberry Pi Zero 2W Edge Computer

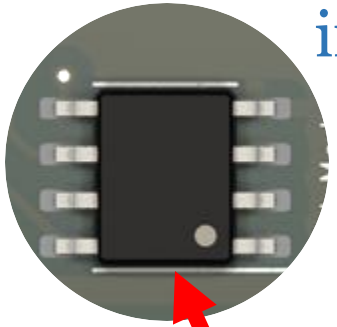
On the Drone



Software Layer – Sensor Data Access

Sensor Driver	Failure Detection	Data Storage	Data Processing

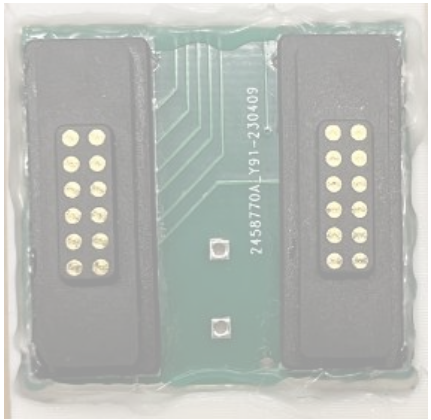
EEPROM with module's information



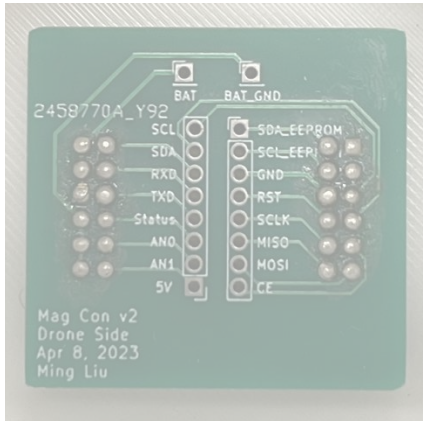
Linux

Raspberry Pi Zero 2W Edge Computer

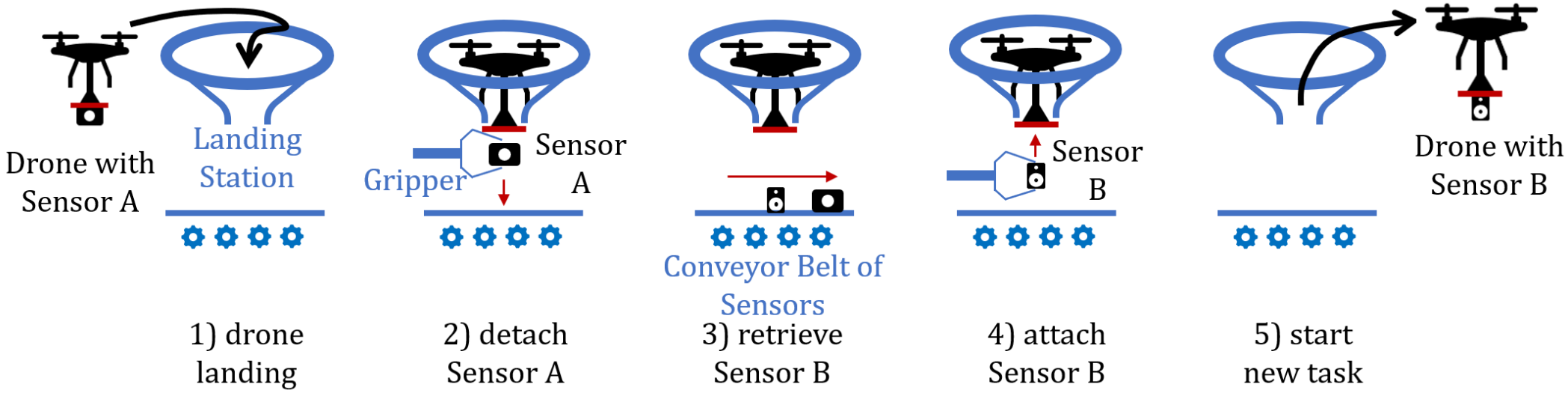
Sensor Module



On the Drone

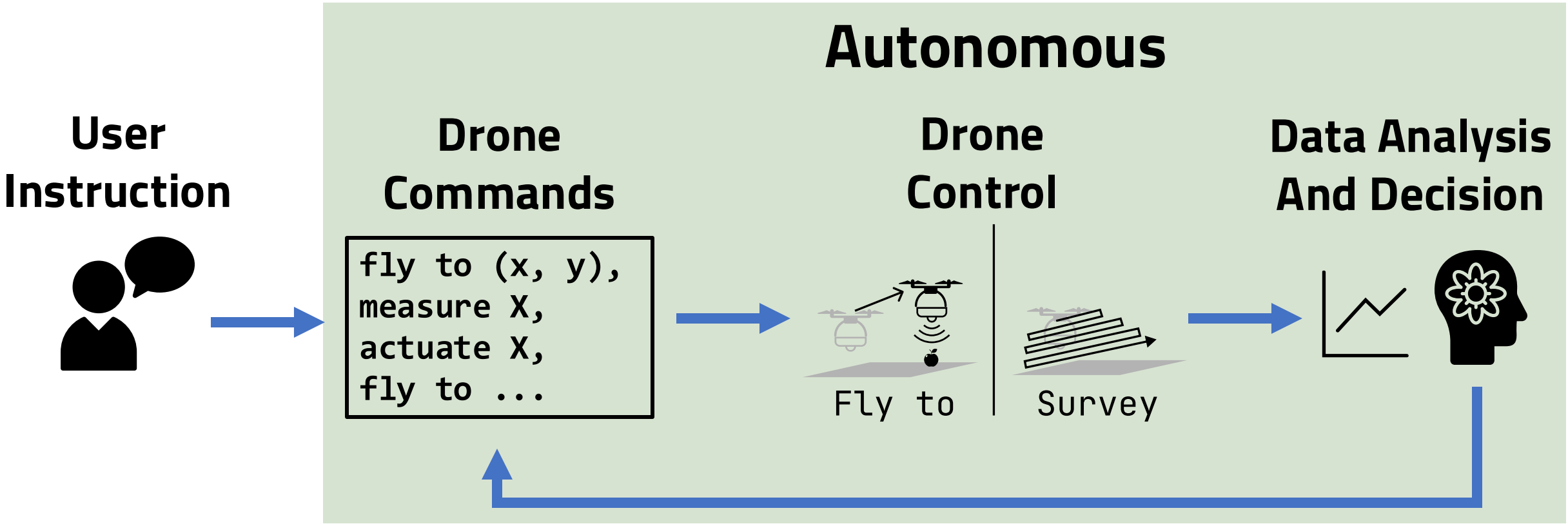


Sensor Swapping Process



Building the Intelligence

To create a ***fully autonomous*** drone system from simple user command

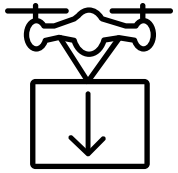


Users may ask…

Where is the warmest to sit?

Is the stove still on?

Monitor for grandpa falling



Bring snack to my pet

Task Comprehension

Example: **where** is the **warmest** to sit?

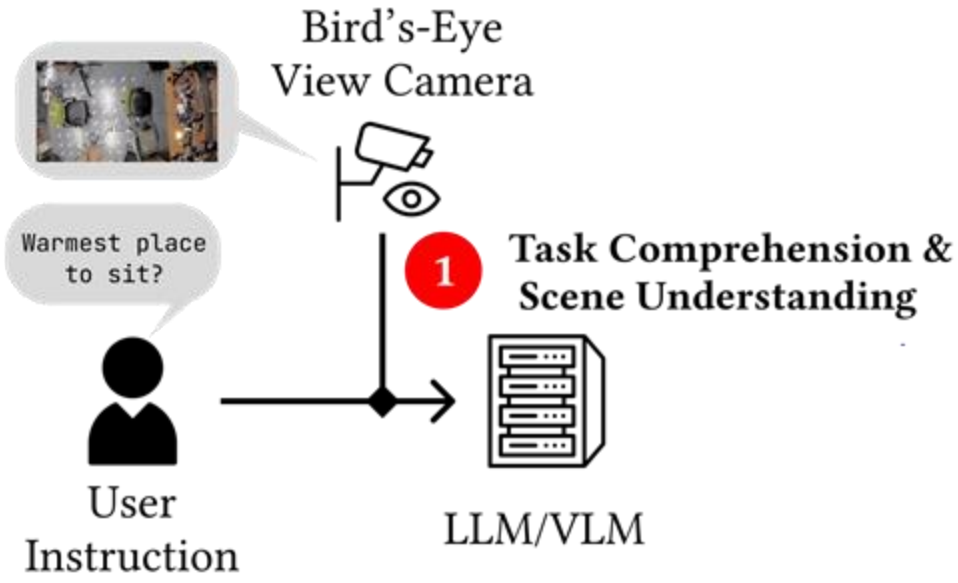
An LLM would understand what this sentence means, but it is **missing the actual knowledge of the environment**

Scene Understanding

Example: **where** is the **warmest** to sit?

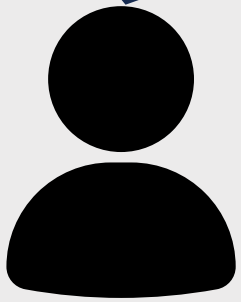
- Sensor input (in many cases, camera)
- Autonomous way to understand elements of the scene
- Vision language model (VLM) enables text + image interpretation

Task Comprehension + Scene Understanding

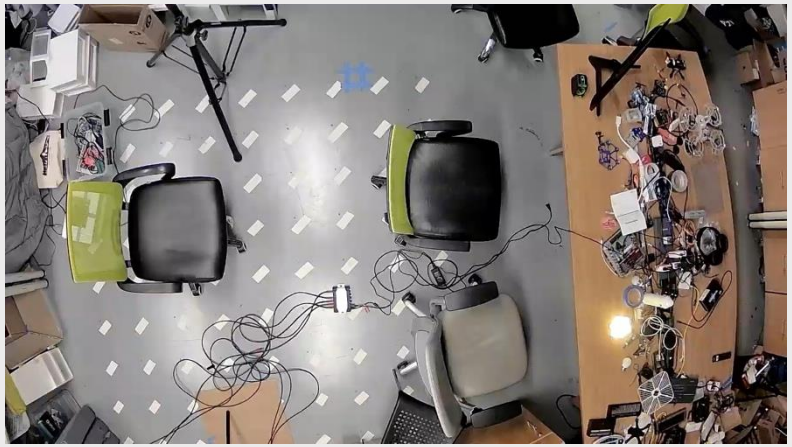


A Input

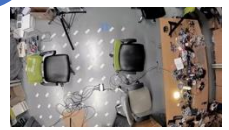
*Where is the **warmest** place to **sit**?*



Bird-eye Camera Snapshot



B Formulate Prompt

-  is a top-down view taken from the camera installed on the ceiling.
- I would like to know {***Where is the warmest place to sit?***}, and which type of sensor should I use to better assist this task? I have {***list of sensors***} available.
- Your response should only be:
  - Reason: (reason for determine the object and sensor)
  - Target: (the object of interest)
  - Payload: (sensor or payload to carry)"

VLM Output

"Chain-of-thought" for better reasoning

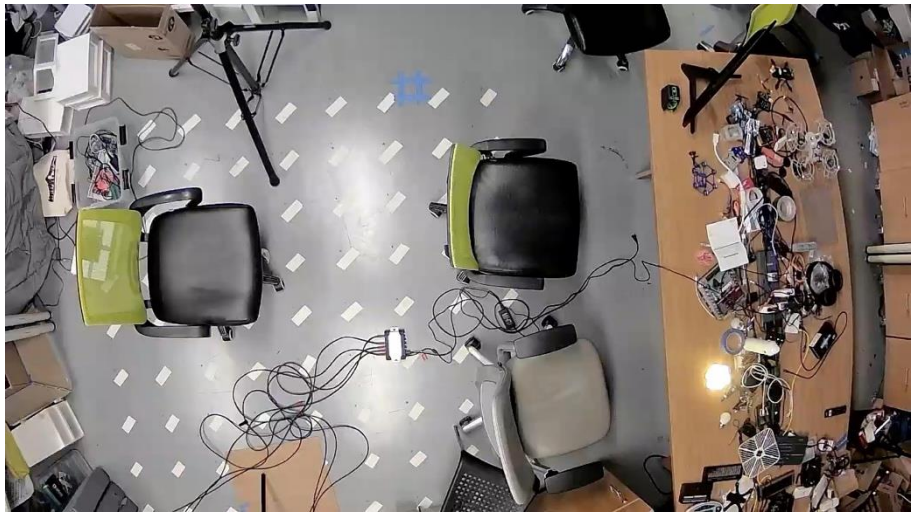
Reason: *The warmest place to sit can be determined by measuring the temperature at the various seating areas within the lab. These seating areas are influenced by heat generated from nearby electronic equipment, airflow from vents, and possibly direct sunlight or lighting. By focusing on these seating areas, we can obtain a clear picture of the temperature variations and identify the warmest spot.*

Target: Chairs

Payload: Temperature sensor

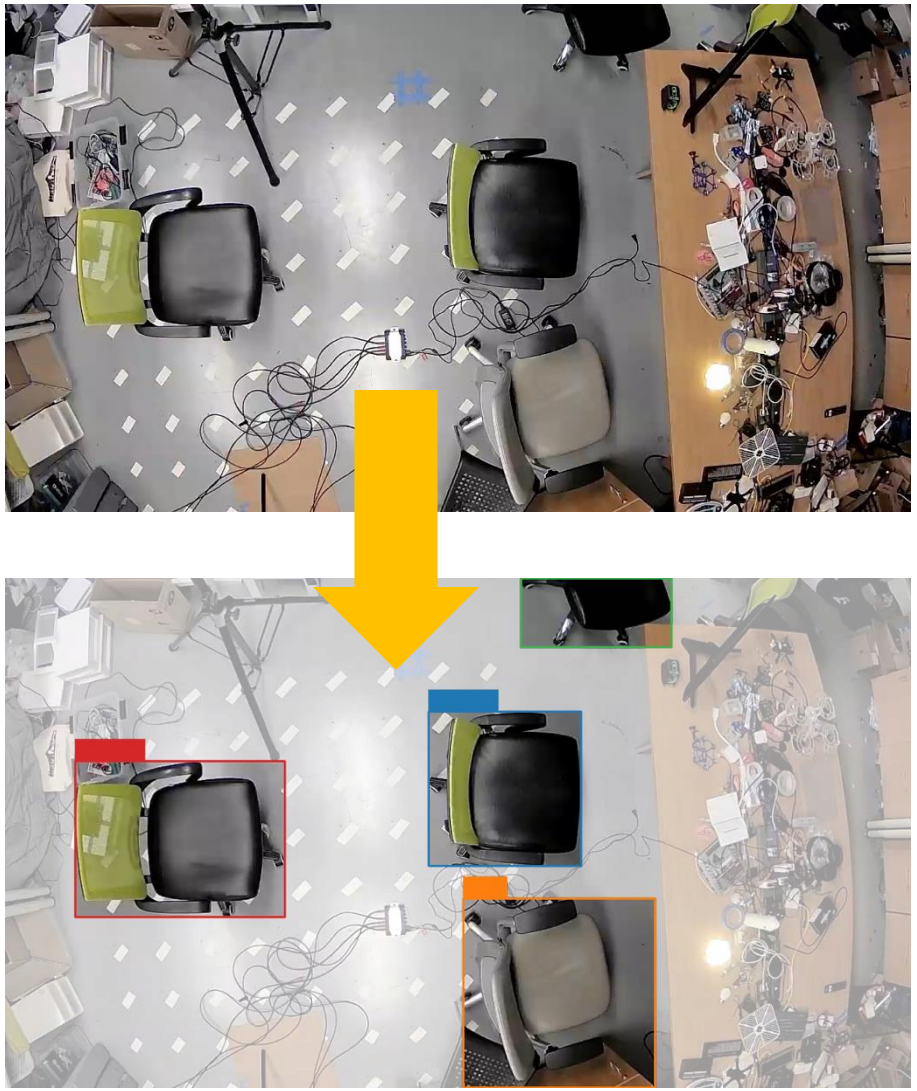
VLM Output

Target: Chairs
Payload: Temperature sensor



VLM Output

Target: Chairs
Payload: Temperature sensor

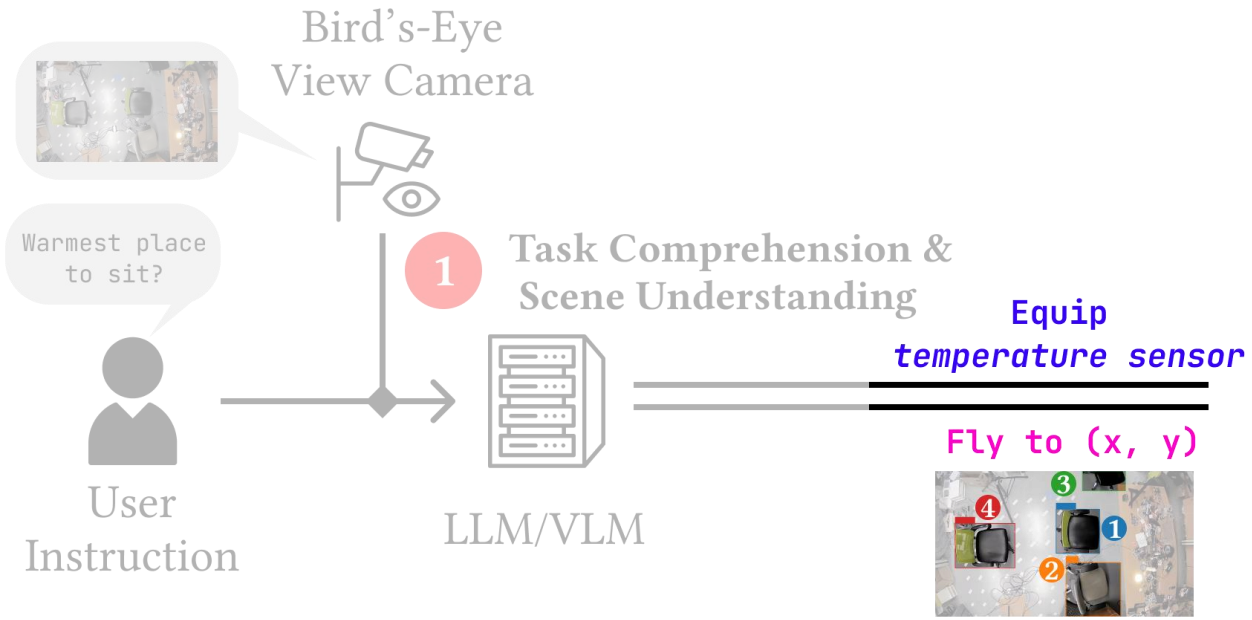


Open-set Object Detection Model

Input: Description of Object

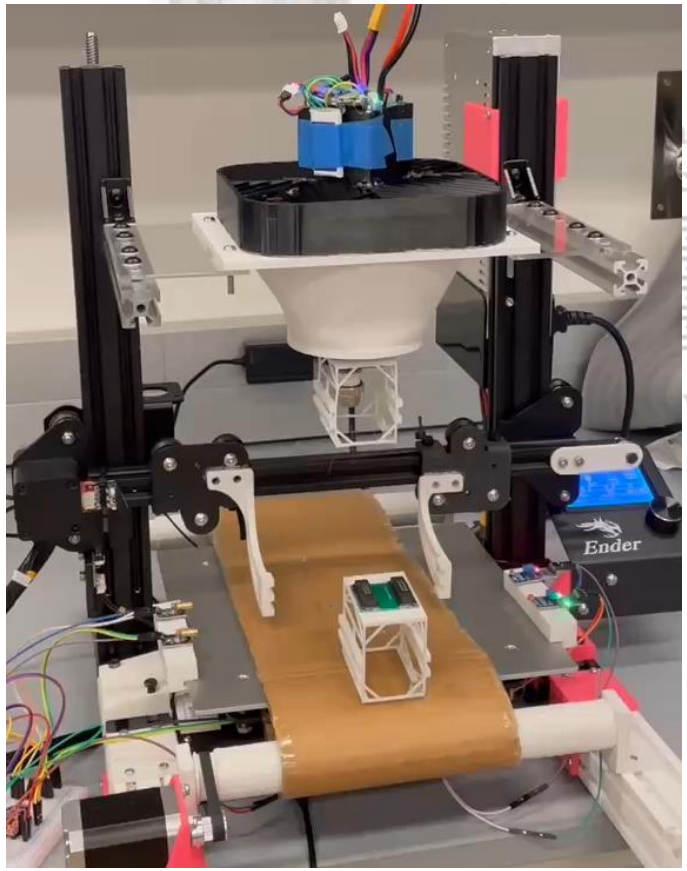
Output: Bounding Boxes of the Objects

Task Comprehension + Scene Understanding



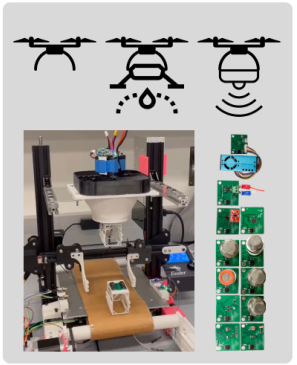
VLM Output

Automatic Payload Installation

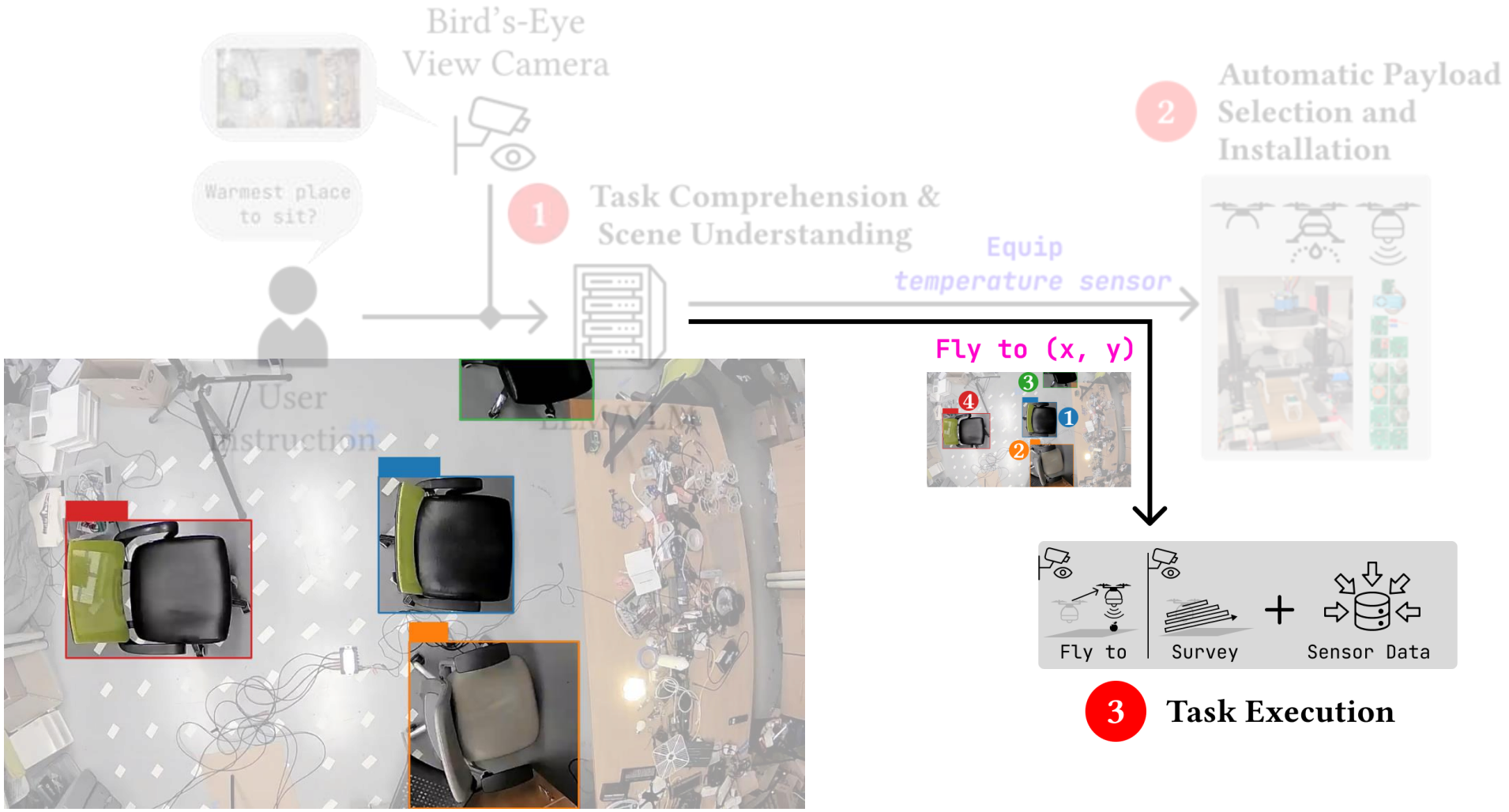


Equip
temperature sensor

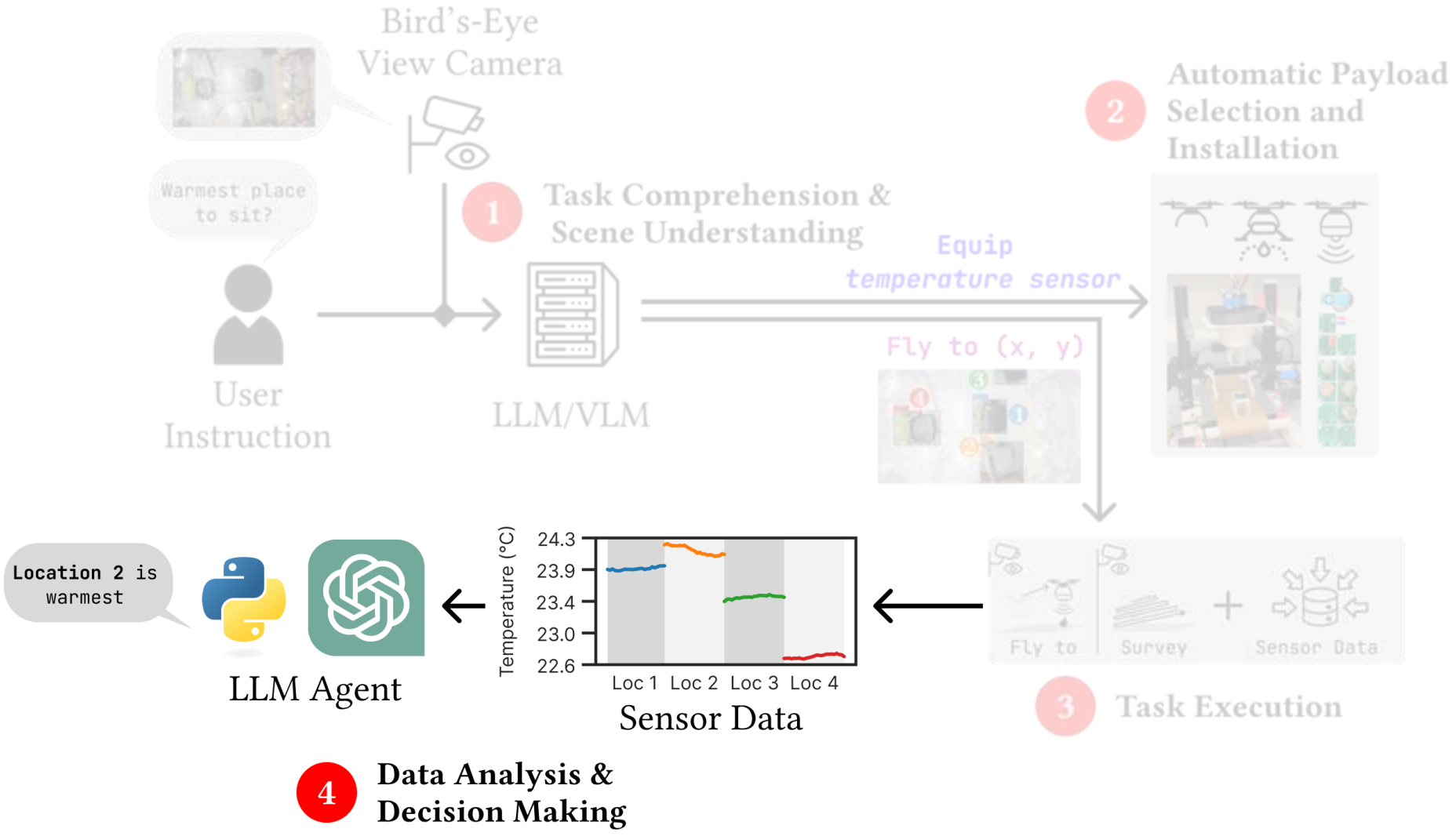
2 **Automatic Payload Selection and Installation**



Task Execution



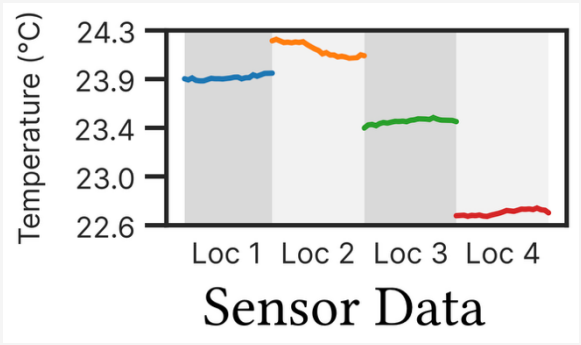
Data Analysis & Decision Making



Data Analysis & Decision Making

Input Data

1.csv
2.csv
3.csv
4.csv



Prompt

- This data is from {***temperature sensor***} measured by a drone tasked {***find warmest place to sit***}.
- The N.csv is taken while the drone is hovering above location X. The csv file contains …
- First perform EDA to understand the data, then perform necessary preprocessing.
- Only answer in the following format
  {"**answer**": the location ID, "**explanation_to_user**": "XX", "**analyze_process**": "XX"}
  In the explanation to user, try to add some numbers from your analysis.

Data Analysis & Decision Making

Analyzed

python Always show details Copy code

```python
import pandas as pd

# Load the data from the CSV files
files = {
    "0": "/mnt/data/0.csv",
    "1": "/mnt/data/1.csv",
    "2": "/mnt/data/2.csv",
    "3": "/mnt/data/3.csv"
}

dataframes = {key: pd.read_csv(files[key], header=None, names=["Time", "Temperature"]) for

# Display the first few rows of each dataframe to understand their structure
for key, df in dataframes.items():
    print(f"Data from file {key}.csv")
    print(df.head(), "\n")
    print("Basic Statistics:")
    print(df.describe(), "\n")
    print("-" * 50, "\n")
```

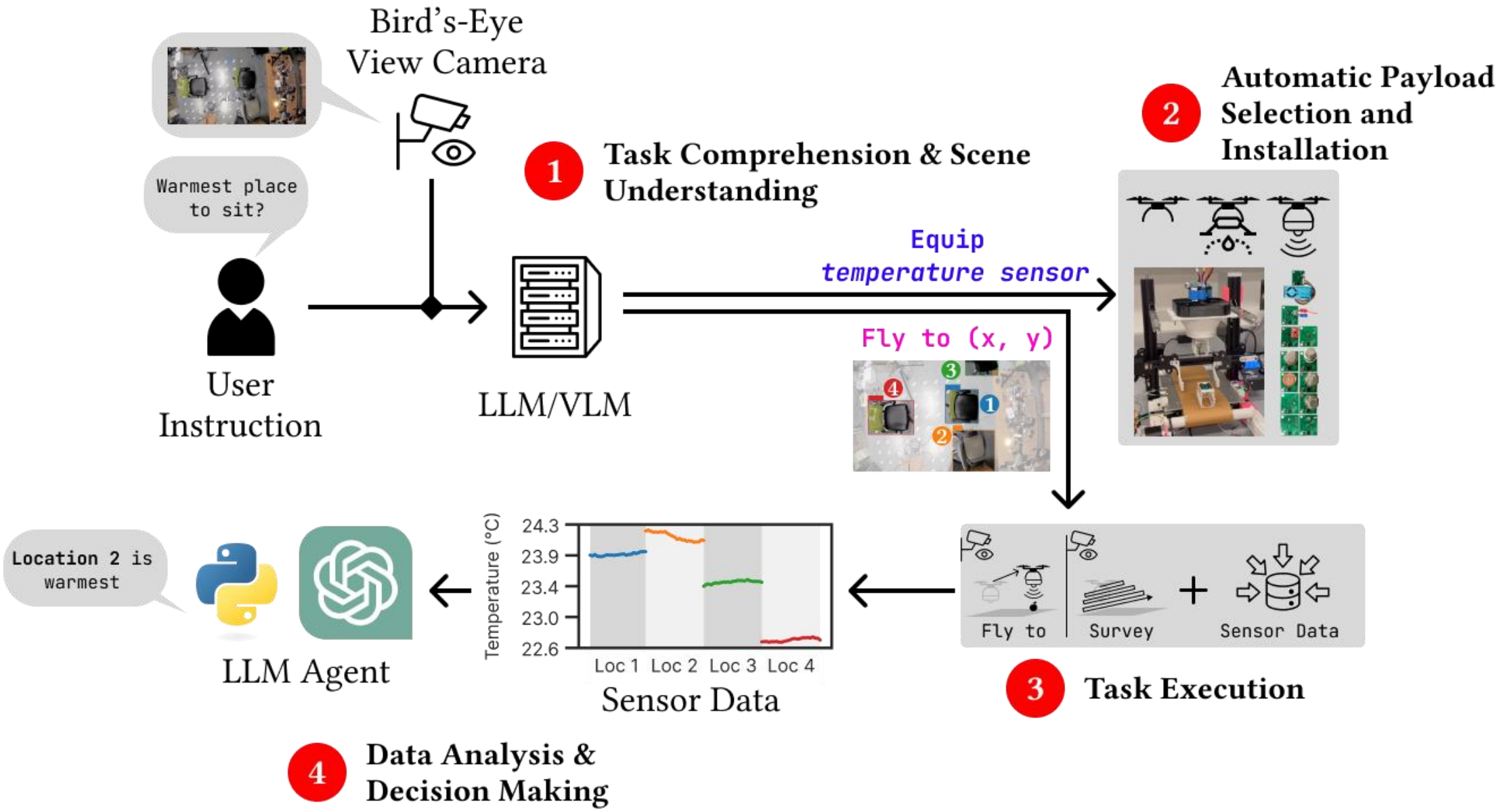
STDOUT/STDERR

```
Data from file 0.csv
           Time   Temperature
0  1.701395e+09  26.982403
1  1.701395e+09  27.047348
2  1.701395e+09  27.068615
3  1.701395e+09  26.999950
4  1.701395e+09  27.037144
```

Data Analysis & Decision Making

json Copy code

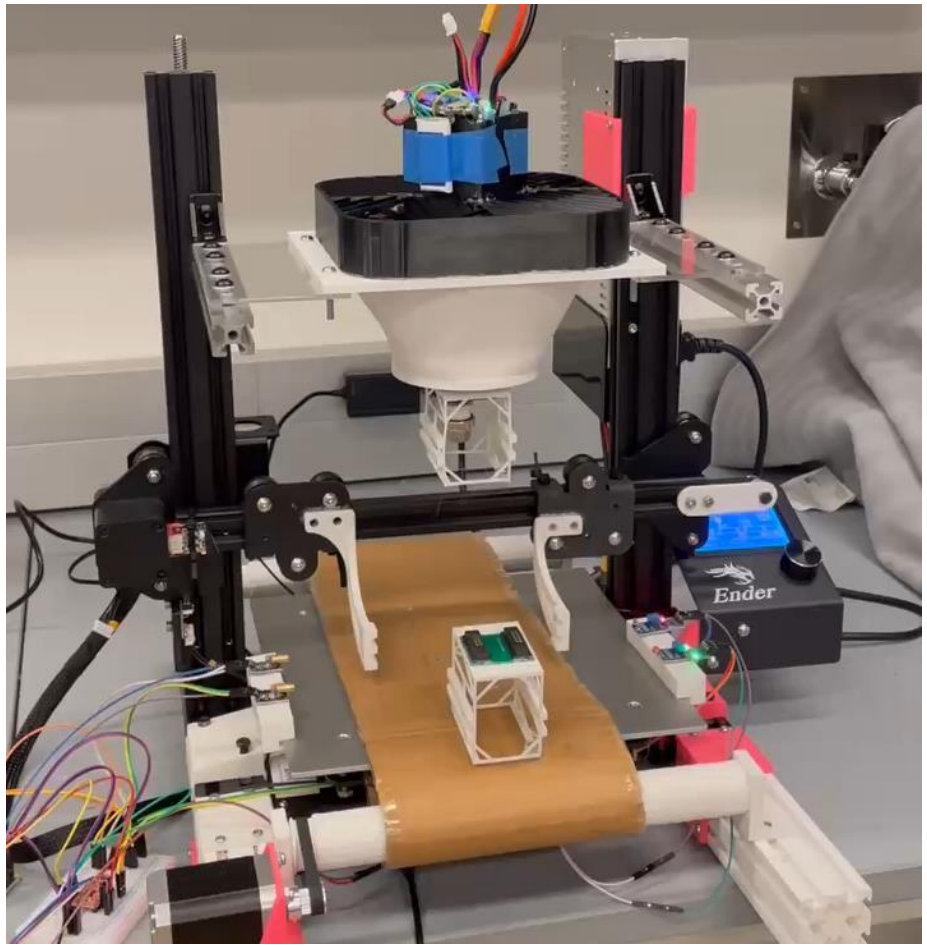
```
{
  "answer": "2",
  "explanation_to_user": "The data from the Temperature sensors
  across four locations (1, 2, 3, 4) shows that
  location 2 has the highest average Temperature
  at 24.2 degrees. Location 1 has an average
  Temperature of 23.9 degrees, location 3 has an
  average of 23.4 degrees, and location 4 has an
  average Temperature at 22.7.",
  "analyze_process": "Each CSV file was loaded and
  analyzed to calculate the average Temperature.
  The data was checked for basic statistics like mean,
  min, max, and standard deviation. The location with
  the highest average Temperature was identified
  as location 2."
}
```

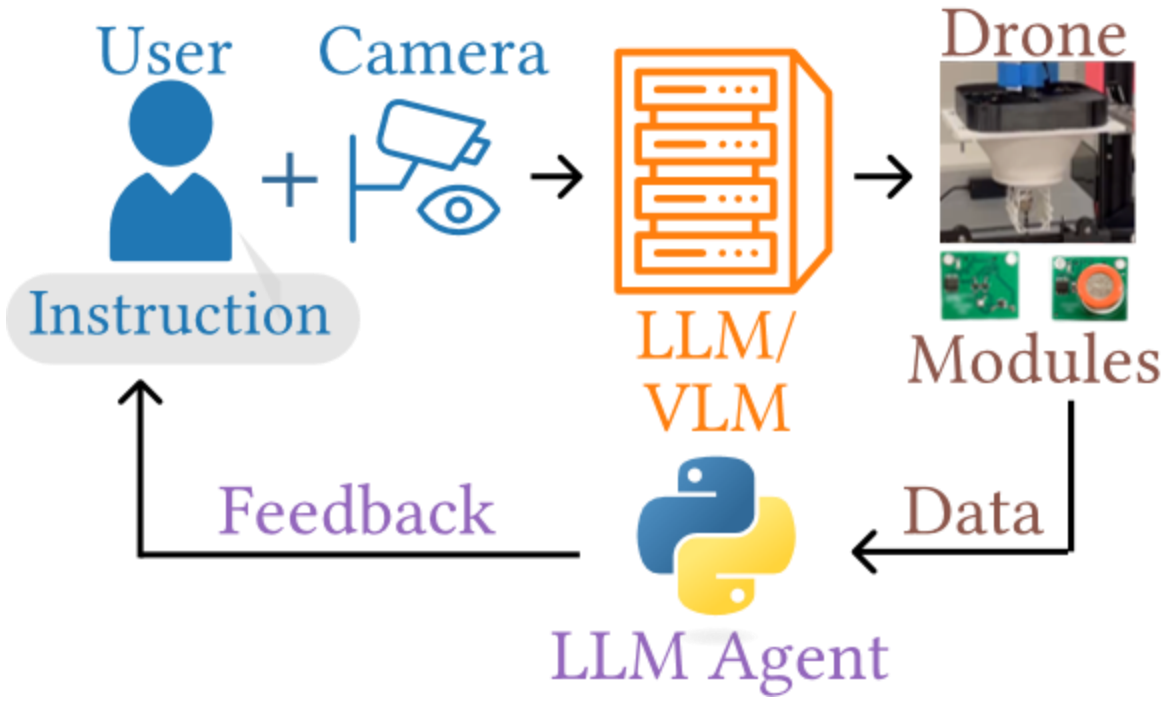

Bird's-Eye View Camera
Warmest place to sit?
User Instruction
1 Task Comprehension & Scene Understanding
LLM/VLM
Equip temperature sensor
Fly to (x, y)
2 Automatic Payload Selection and Installation
Fly to
Survey
Sensor Data
3 Task Execution
Temperature (°C)
24.3
23.9
23.4
23.0
22.6
Loc 1 Loc 2 Loc 3 Loc 4
Sensor Data
Location 2 is warmest
LLM Agent
4 Data Analysis & Decision Making

Evaluation

	Camera Only				Camera + **FlexiFly**				
Scenario	Precision	Recall	F-1	Accuracy	Sensor Used	Precision	Recall	F-1	Accuracy
Object / Location Identification									
Find Phone	68.85%	84.00%	75.68%	73.00%	Drone Cam	100.00%	84.00%	91.30%	92.00%
Find Key	78.05%	80.00%	79.01%	71.67%	Drone Cam	100.00%	80.00%	88.89%	86.67%
Sit - Temperature	23.47%	92.00%	37.40%	25.24%	Temperature	76.67%	92.00%	83.64%	91.26%
Sit - Humidity	25.81%	88.89%	40.00%	28.00%	Humidity	82.76%	88.89%	85.71%	92.00%
Sit - Light	20.62%	83.33%	33.06%	22.12%	Light Sensor	95.24%	83.33%	88.89%	95.19%
Average (ID)	43.36%	85.64%	53.03%	44.01%		90.93%	85.64%	87.69%	91.42%
State of Object / Location									
Faucet Open	80.00%	97.78%	88.00%	82.86%	Humidity	93.62%	97.78%	95.65%	94.29%
Stove Open	79.22%	87.14%	82.99%	77.27%	Temperature	96.83%	87.14%	91.73%	90.00%
Average (State)	79.61%	92.46%	85.50%	80.06%		95.22%	92.46%	93.69%	92.14%
Surveillance									
Food Burning	50.91%	70.00%	58.95%	51.25%	PM	93.33%	70.00%	80.00%	82.50%
Chemical Spill	25.40%	80.00%	38.55%	43.33%	Gas (Alcohol)	88.89%	80.00%	84.21%	93.33%
Average (Sur.)	38.15%	75.00%	48.75%	47.29%		91.11%	75.00%	82.11%	87.91%
Average (all)	46.54%	83.17%	55.70%	48.99%		91.93%	84.79%	87.78%	90.80%

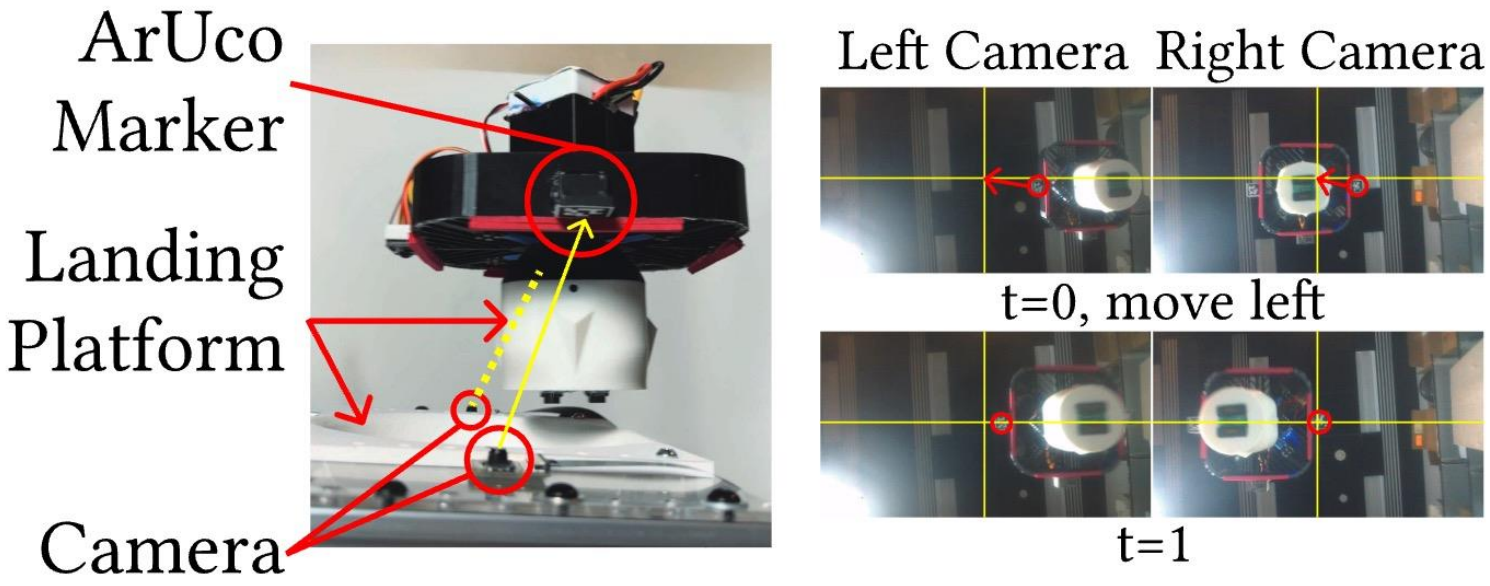
FlexiFly connects foundation models with the physical world using reconfigurable drone agents



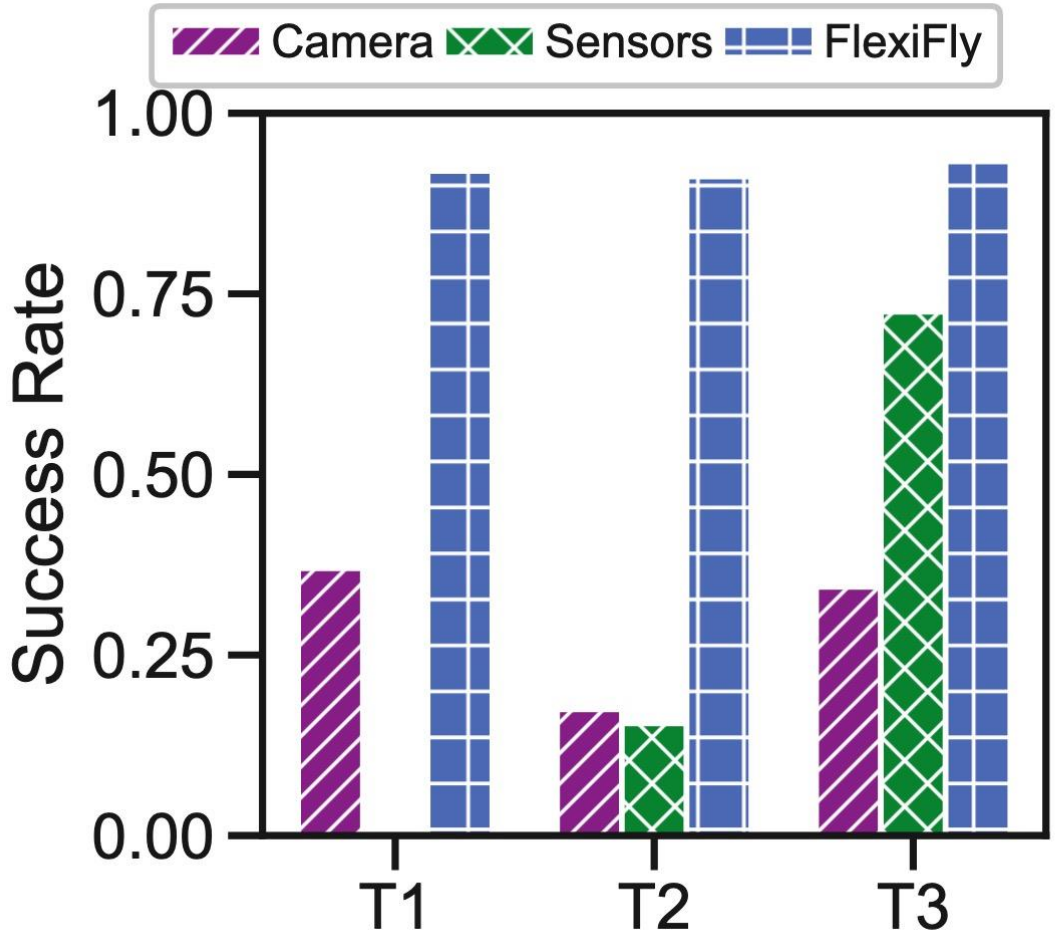


Backup Slides

Drone Landing

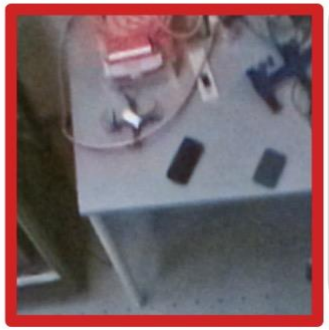


Motivation: Zoom In



Query: This is a top down view image. Where is my phone?
Model: *LLaVA v1.5 13b*
Config: Temperature=0.2 Top P=0.7

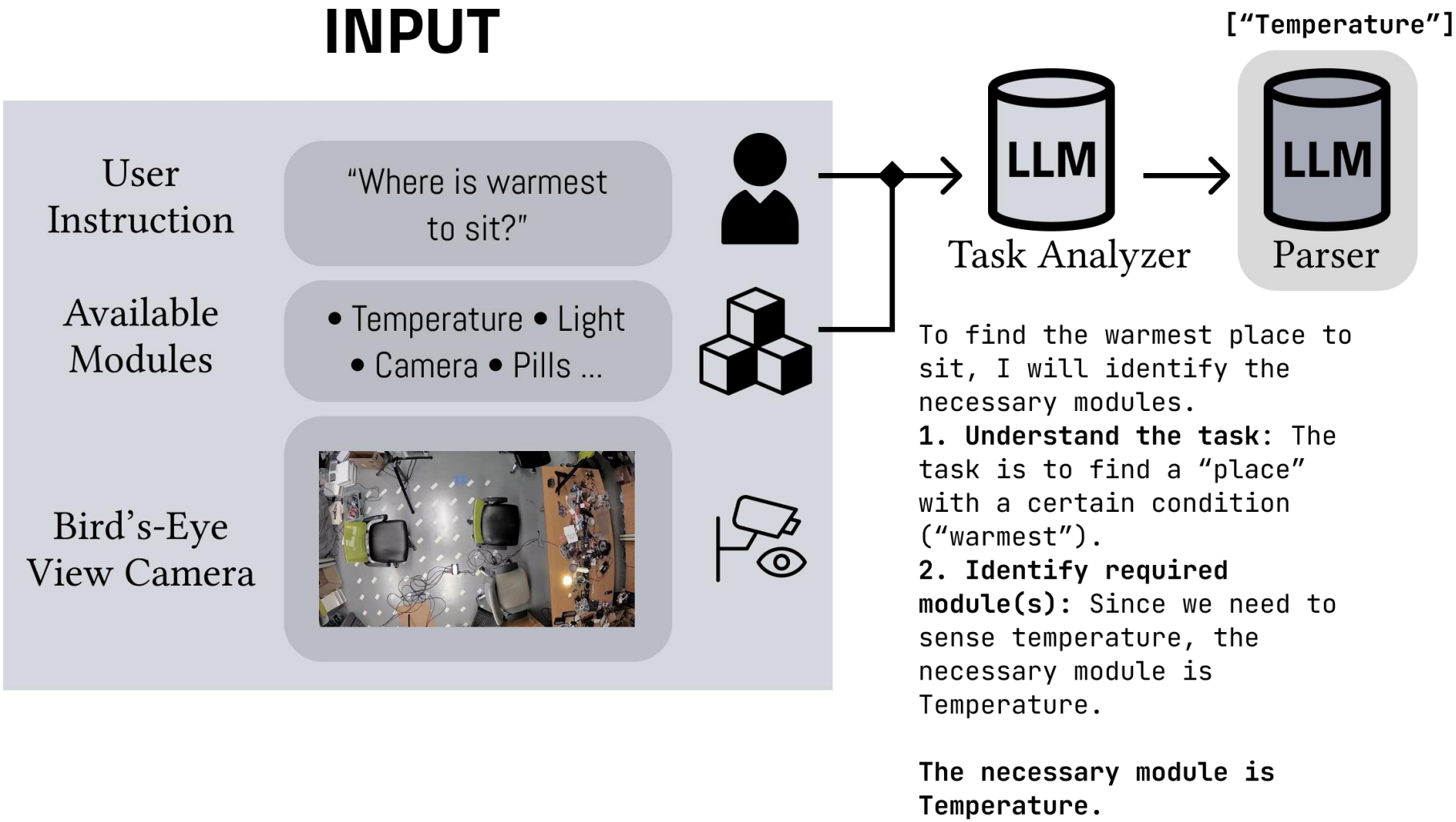
Manual Zoom In



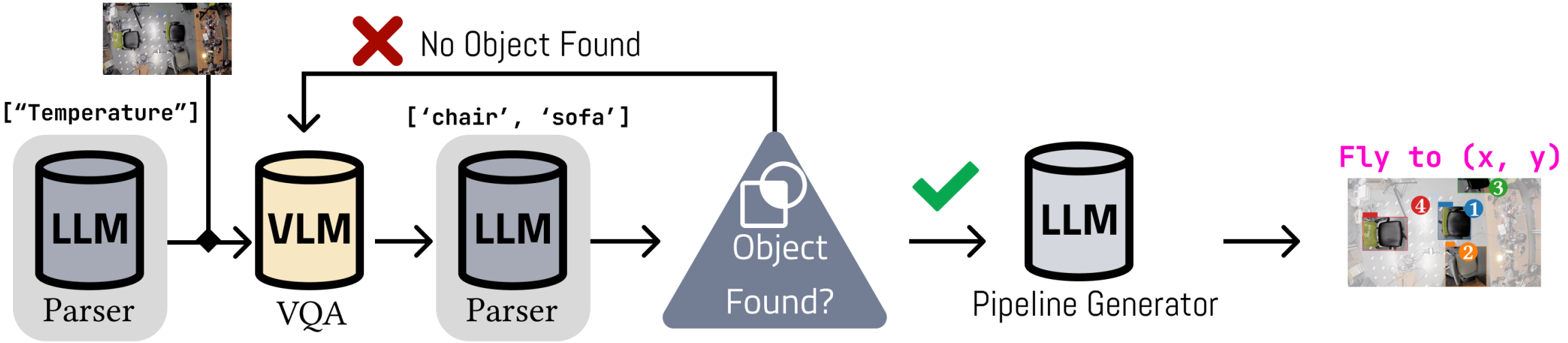
The image does not provide any information about the location of a phone.

The cell phone is placed on the table, likely for easy access and use while working on the electronics.

Example



Example

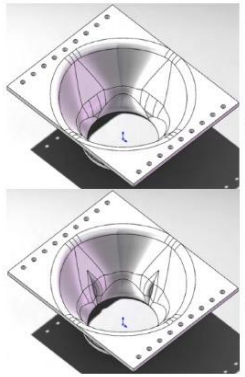



```
In the image provided, there are two
visible objects that are relevant for
sitting:
- Chair: There is a chair in the foreground
of the image. It appears to be empty and
could be considered as a warm place to sit
on.
- Sofa: To the right, there is a sofa
against the wall. It seems to be unoccupied
and could also serve as a comfortable spot
to sit on.
These two objects are suitable for sitting
in this context and would likely provide
warmth due to their upholstery materials.
```

```
(1, 'equip', "module='temperature'", None, None)
(2, 'flyto', "target='chair 1'", 'chair 1', (555, 667))
(3, 'measure', "module='temperature'", None, None)
(4, 'flyto', "target='chair 2'", 'chair 2', (393, 409))
(5, 'measure', "module='temperature'", None, None)
(6, 'flyto', "target='sofa 1'", 'sofa 1', (393, 409))
(7, 'measure', "module='temperature'", None, None)
(8, 'land', None, None, None)
```

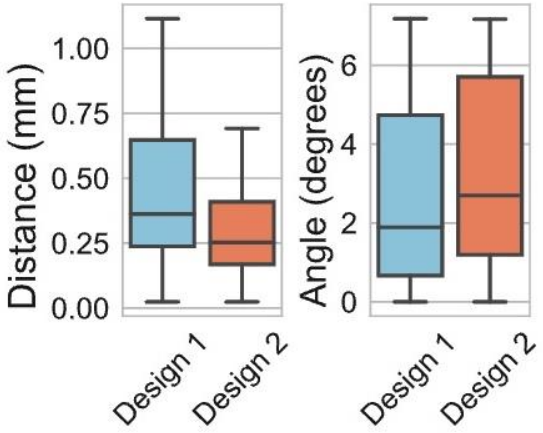
Evaluations: Reconfigurable Drone

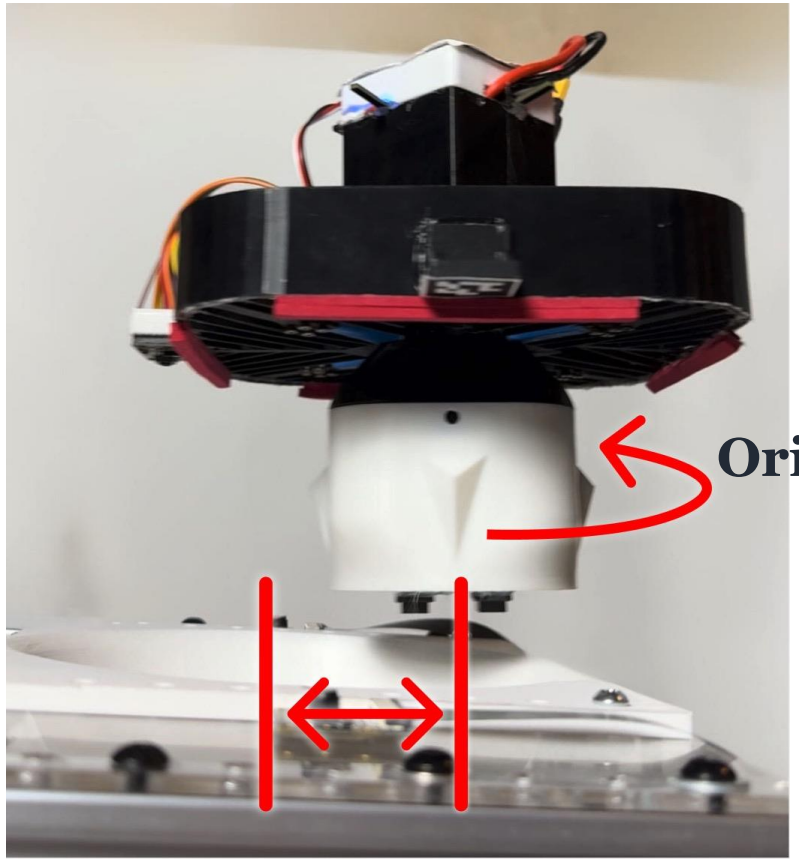
Swap Success Rate



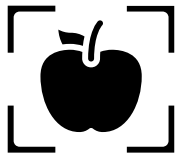
Landing Platform	Maximum Tolerable Orient. Err	Maximum Tolerable Offset	Swap Success Rate
Type 1	22.5°	30mm	93%
Type 2	27.5°	45mm	94%

Landing Accuracy





Classifying User Instructions



Object / Location Identification

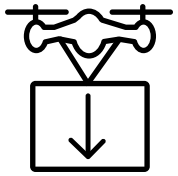
- Where is warmest to sit?
- Where is my phone?

State of something

- Is the stove still on?
- Did I turn off the faucet?

Surveillance

- Monitor my chemical experiment for spills
- Watch out my cooking food in the wok



Aerial Actuation

- Bring snack to my pet
- Put some rat poison next to the cabinet