AI for Scientists: Perception, Reasoning, & Discovery

Jennifer J. Sun

6/5/2025

A Bernese mountain dog giving a talk at the AI for animal science conference at ETH Zurich.

Veo 3

A Bernese mountain dog giving a talk at the AI for animal science conference at ETH Zurich.

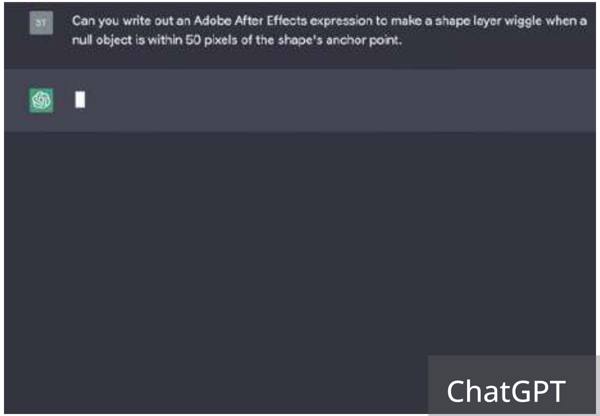
Veo 3

SAM 2

A Bernese mountain dog giving a talk at the AI for animal science conference at ETH Zurich.

Veo 3

SAM 2



ChatGPT


Raw Data
Insight

Raw Data → Insight

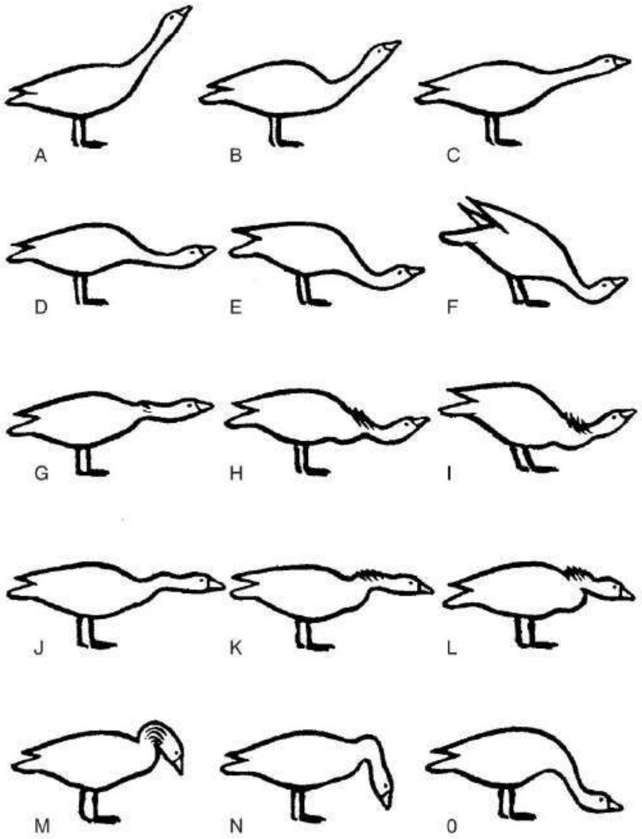
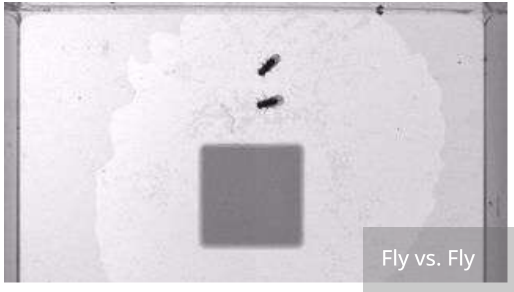
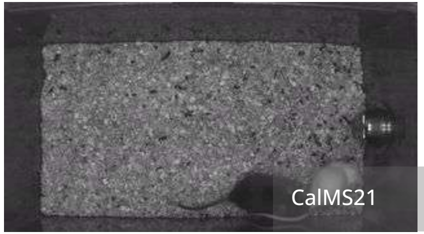


Figure 4

Konrad Lorenz,
On Aggression
~1963, p.97

Raw Data → Insight





Which animal where/when?

How many animals?

What behavior?

Are they healthy?

How does X affect Y?

Why does X affect Y?

Raw Data → Insight

How to best use AI to extract insight from raw data?

Which animal where/when?

How many animals?

What behavior?

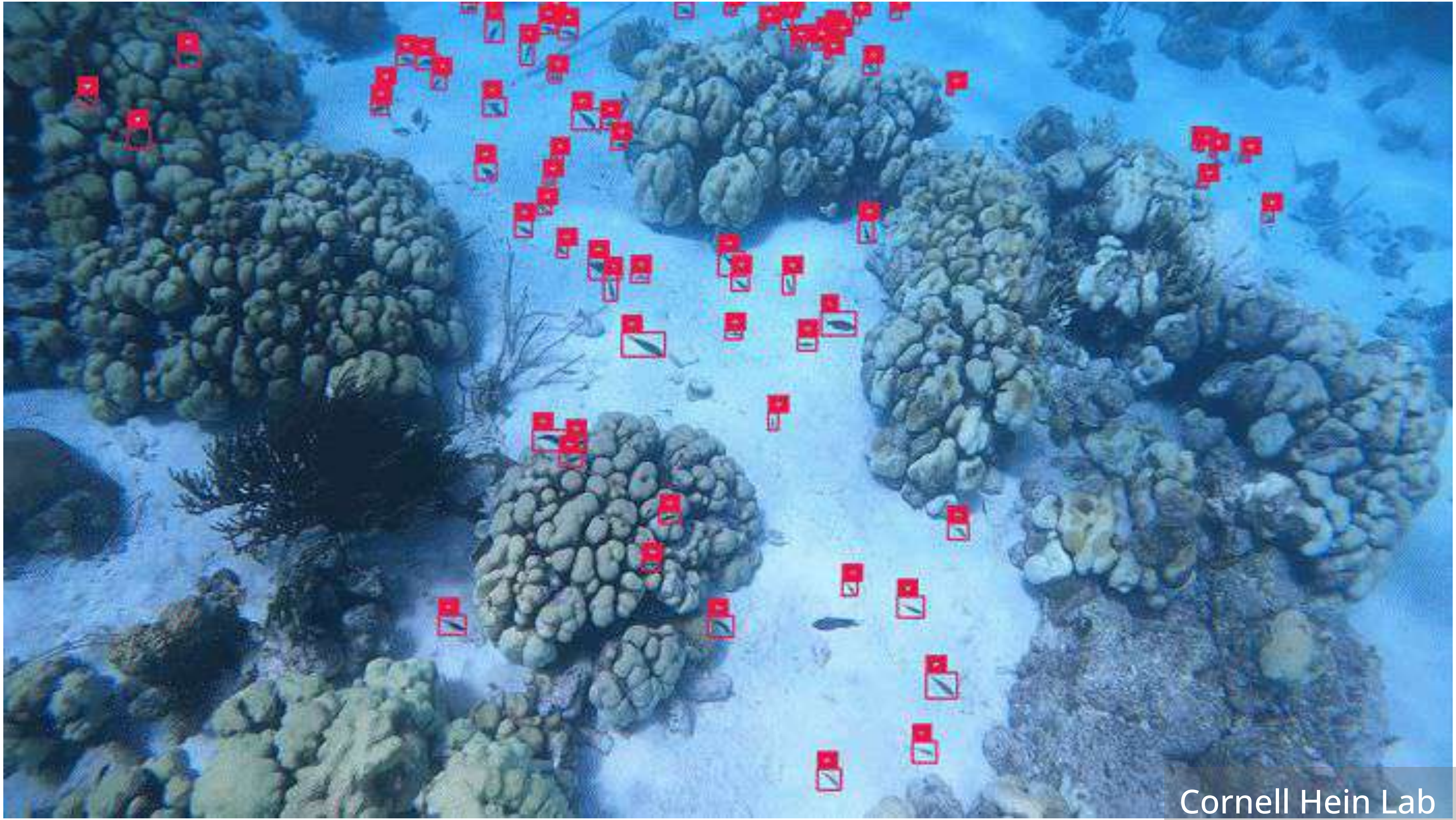
Are they healthy?

How does X affect Y?

Why does X affect Y?

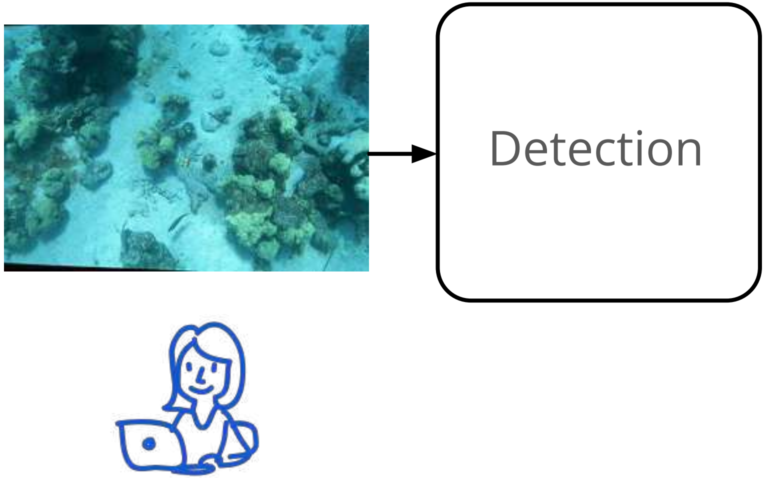
Can AI automatically track these fish?

Cornell Hein Lab



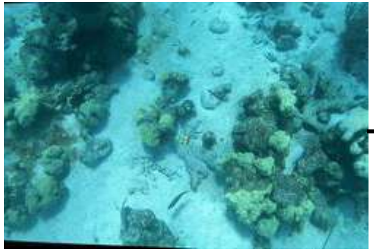
Cornell Hein Lab

Current Analysis Pipelines

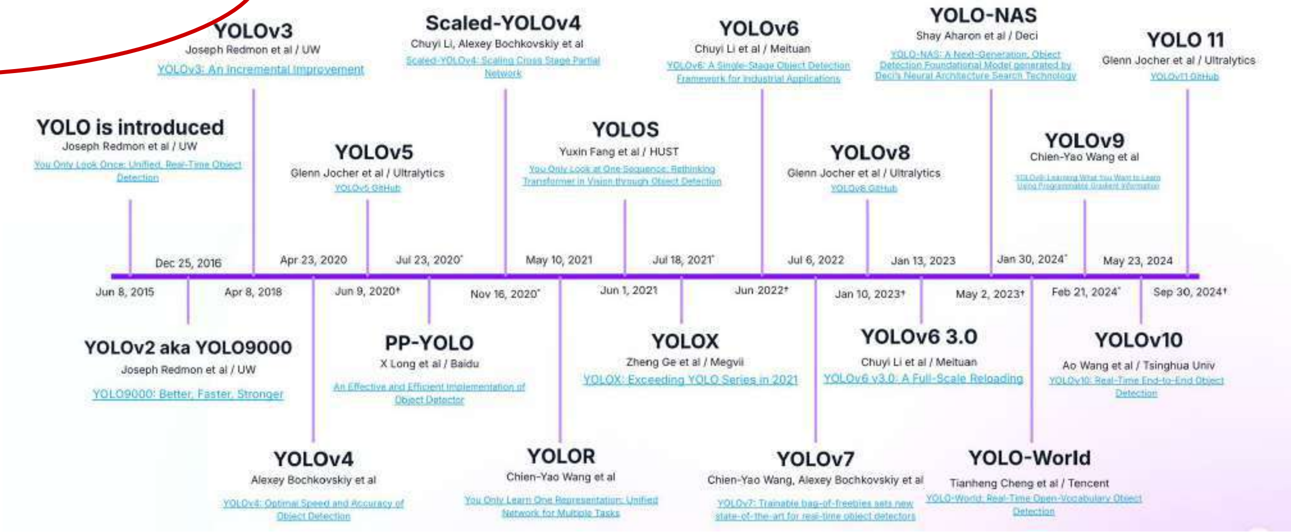


Abby Grassick

Current Analysis Pipelines

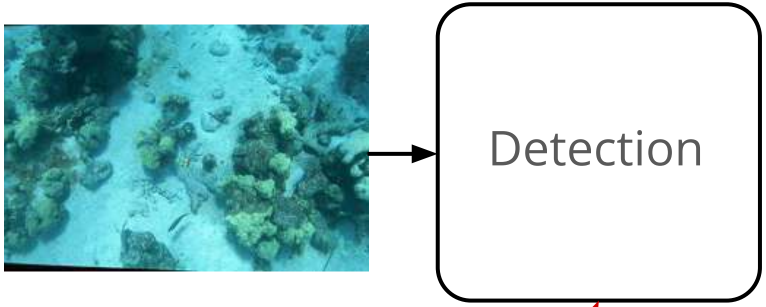


Detection

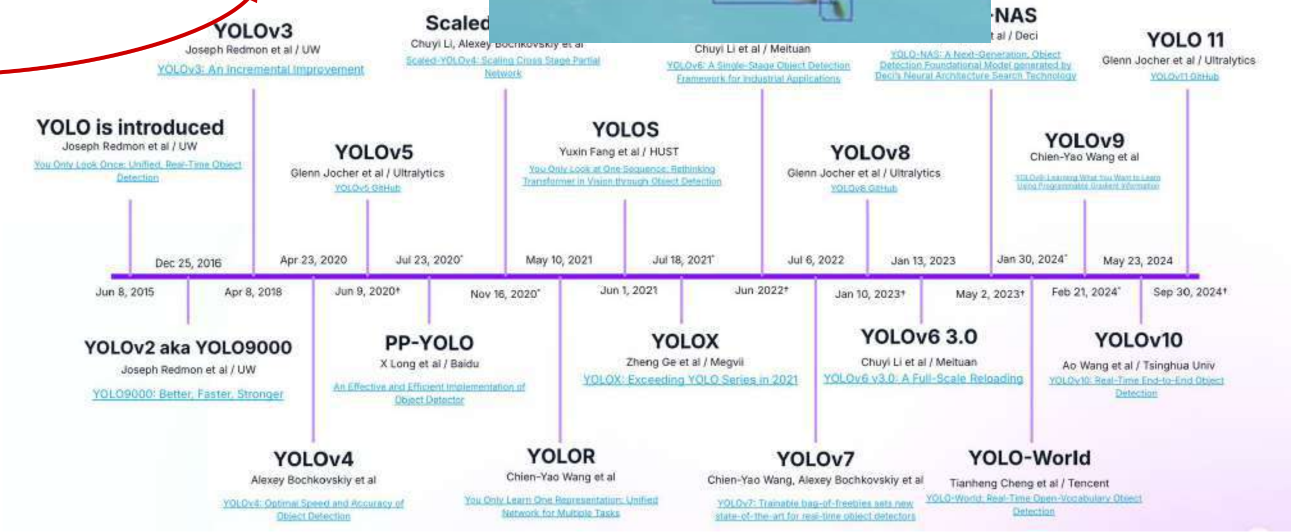


Current Analysis Pipelines

Challenge 1: Annotation bottleneck

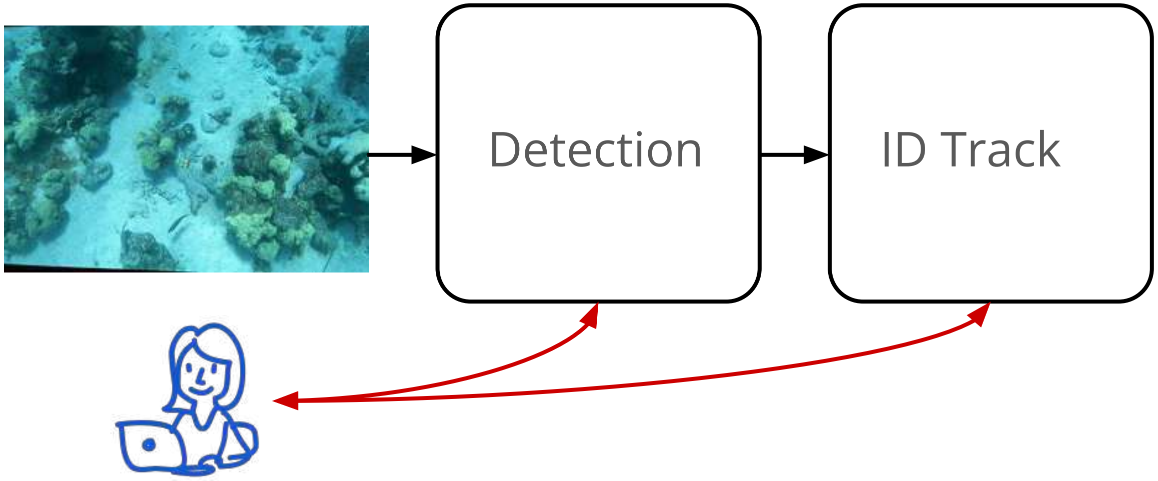


1000s of manual annotations



Challenge 1: Annotation bottleneck

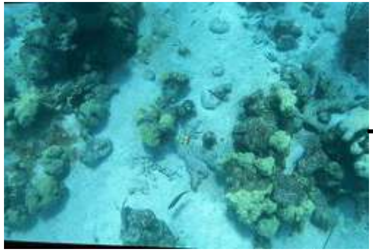
Current Analysis Pipelines



~50 “tracking” papers in vision conferences **last year**

Challenge 1: Annotation bottleneck

Current Analysis Pipelines



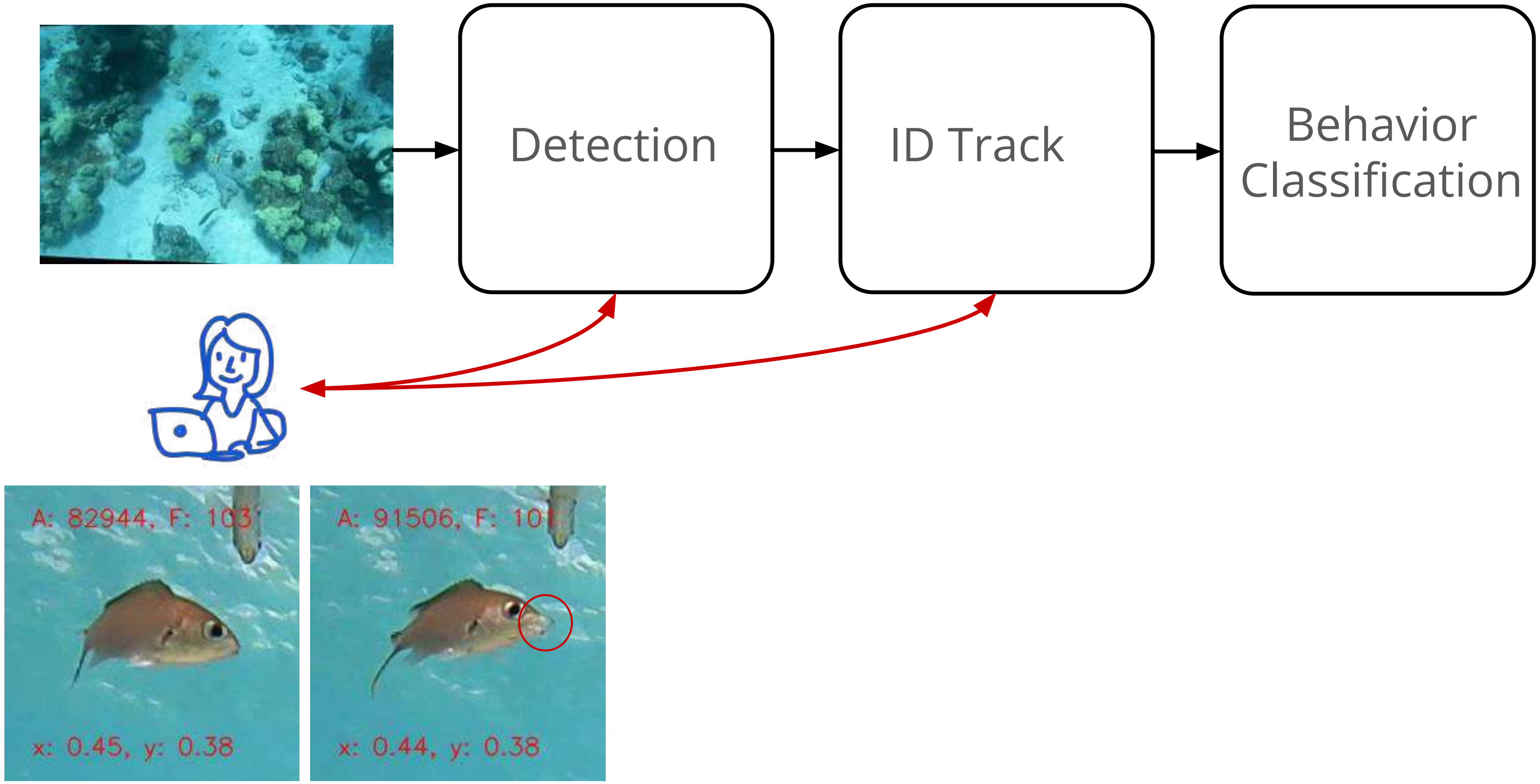
Detection

ID Track

1000s of manual corrections

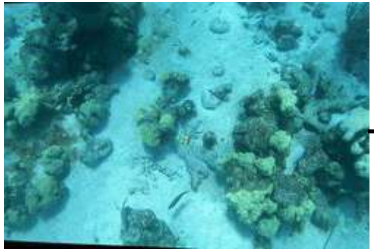
~50 “tracking” papers in vision conferences **last year**

Current Analysis Pipelines

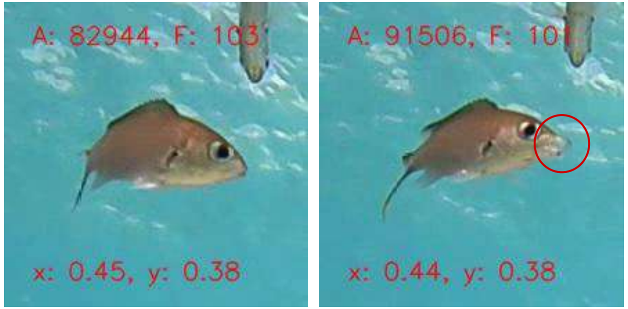


Current Analysis Pipelines

Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback

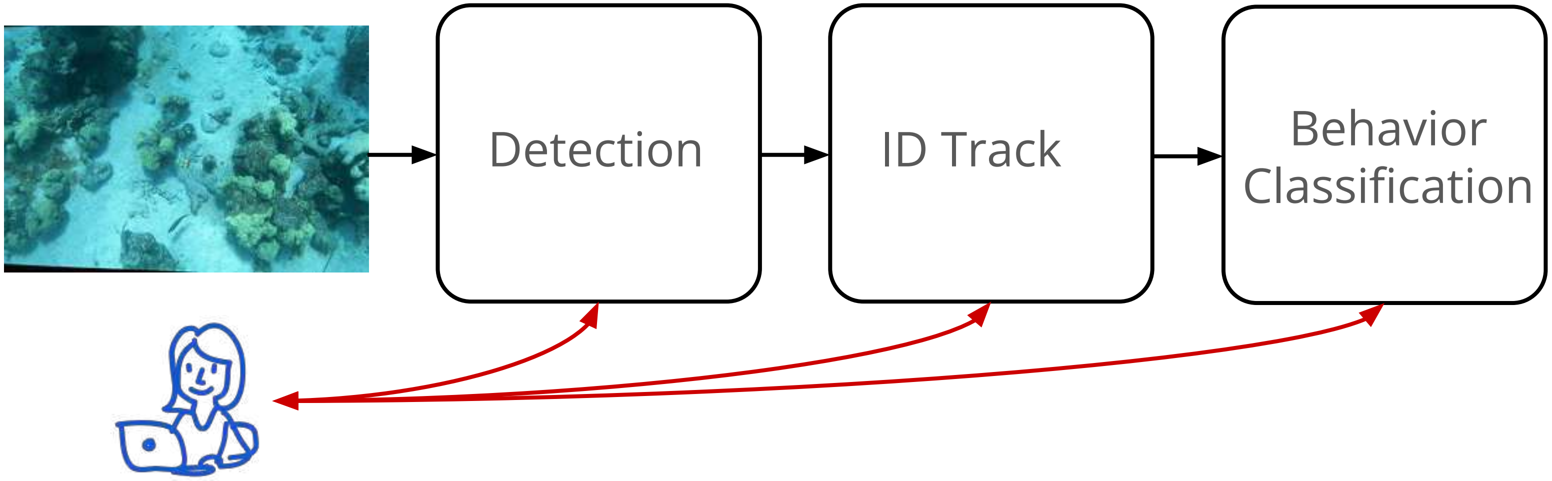


Detection → ID Track → Behavior Classification



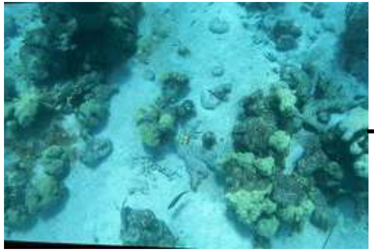
Current Analysis Pipelines

Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback



Current Analysis Pipelines

Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback



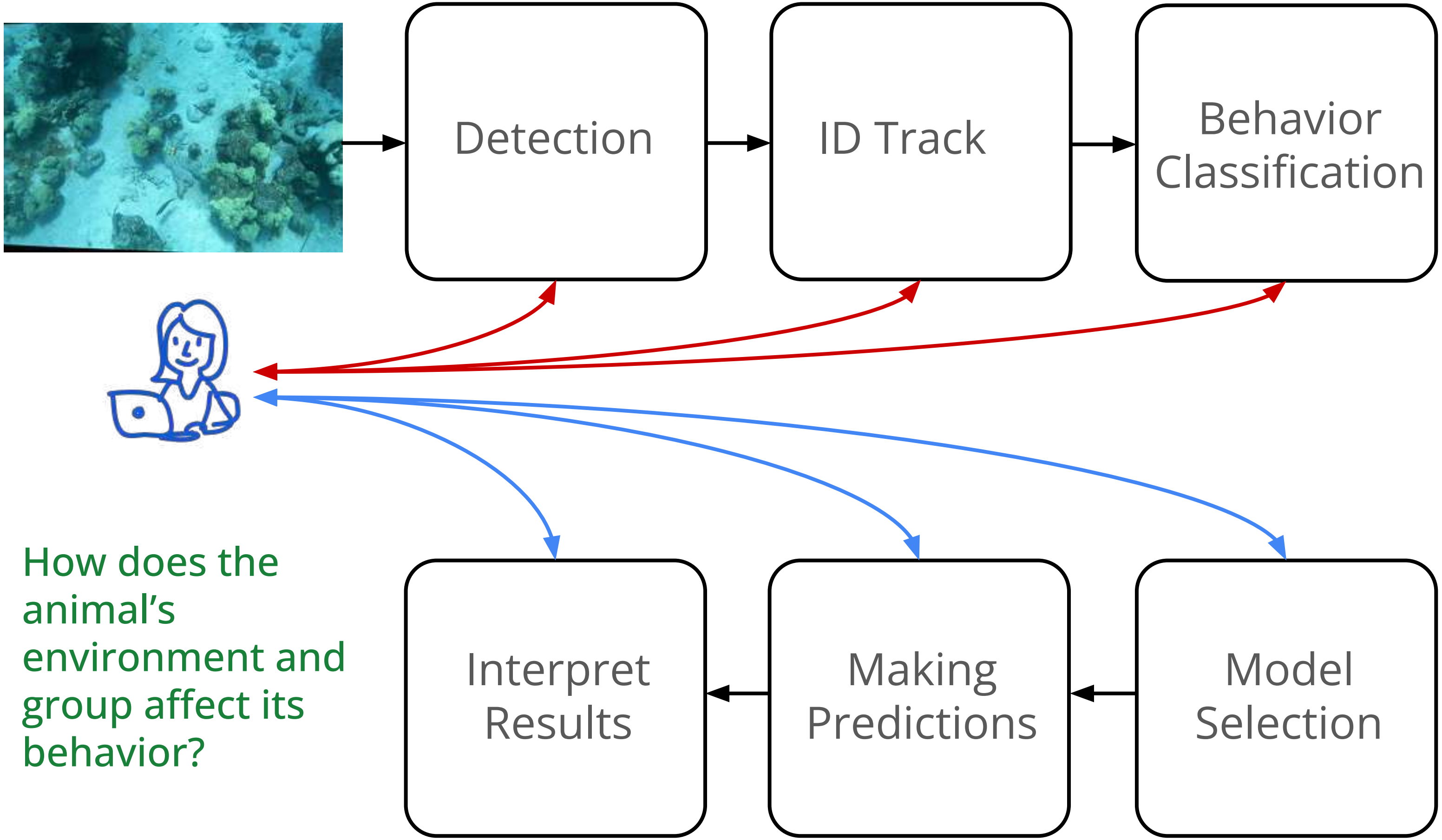
Detection → ID Track → Behavior Classification

- Complexity
- Resolution
- Data / theory
- Stochastic / deterministic
- ...

Model Selection

Current Analysis Pipelines

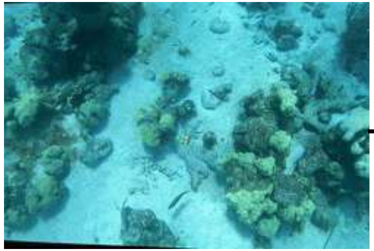
Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback
Challenge 3: Interpretability



How does the animal's environment and group affect its behavior?

Current Analysis Pipelines

Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback
Challenge 3: Interpretability



Perception

- Pose Estimation
- 3D
- Vision representations
- …

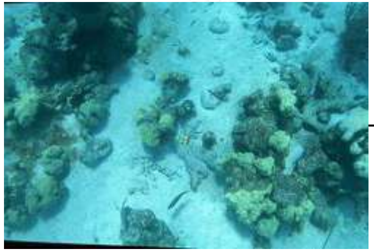
Discovery

How does the animal's environment and group affect its behavior?

Reasoning

- Forecasting
- Equation learning
- Multimodal
- …

Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback
Challenge 3: Interpretability



Perception → Reasoning → Discovery

Which animal where/when?

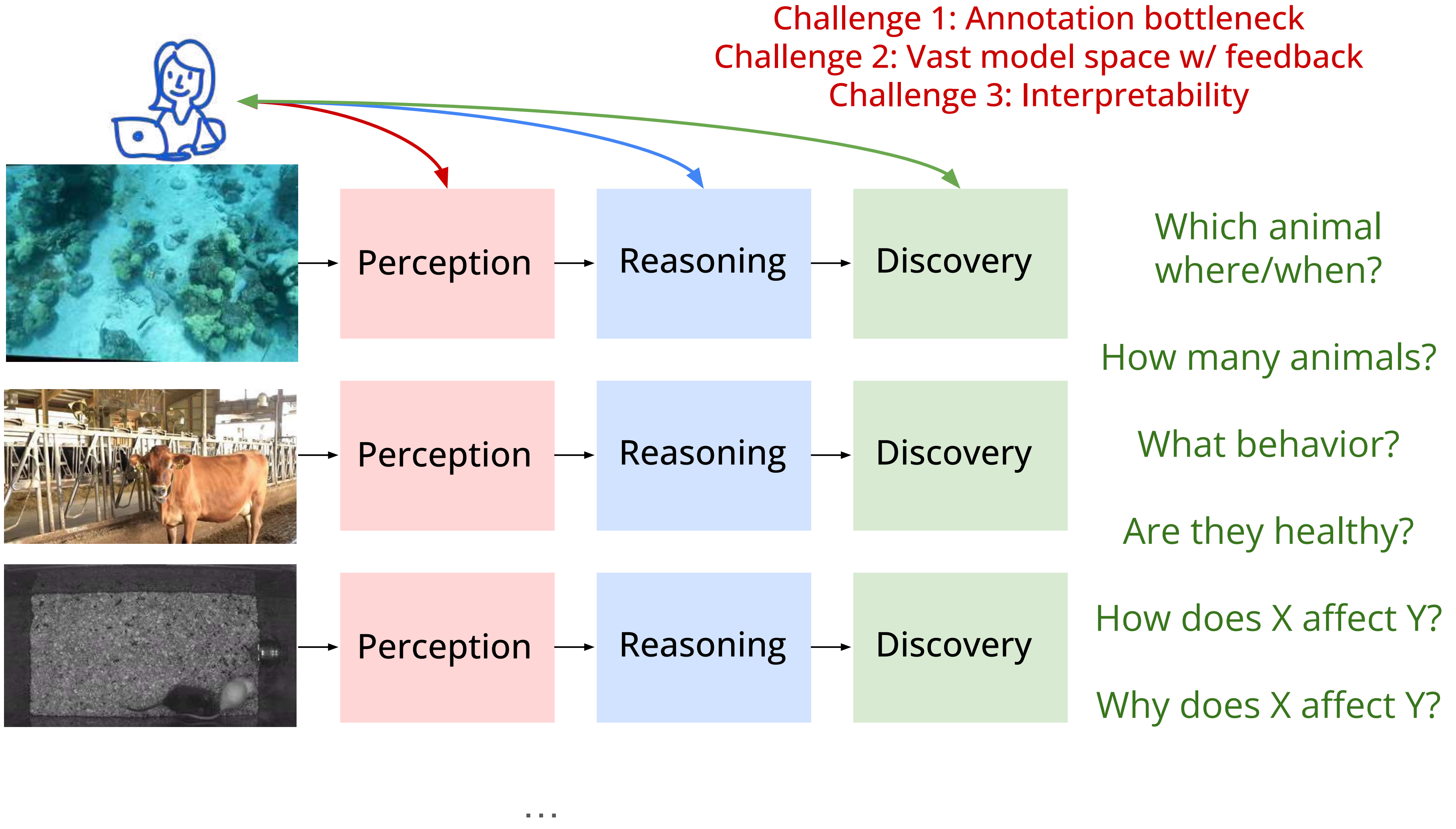
How many animals?

What behavior?

Are they healthy?

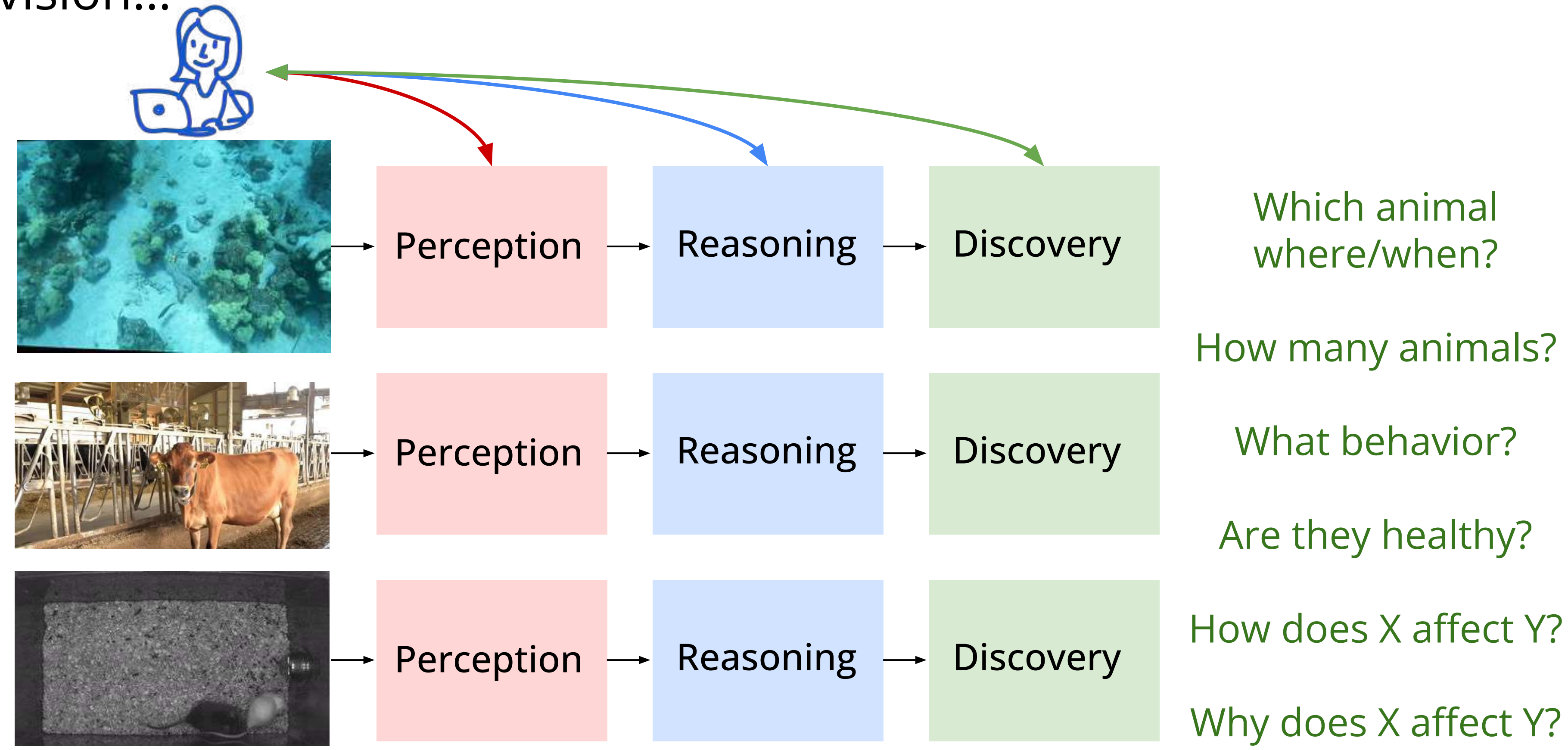
How does X affect Y?

Why does X affect Y?

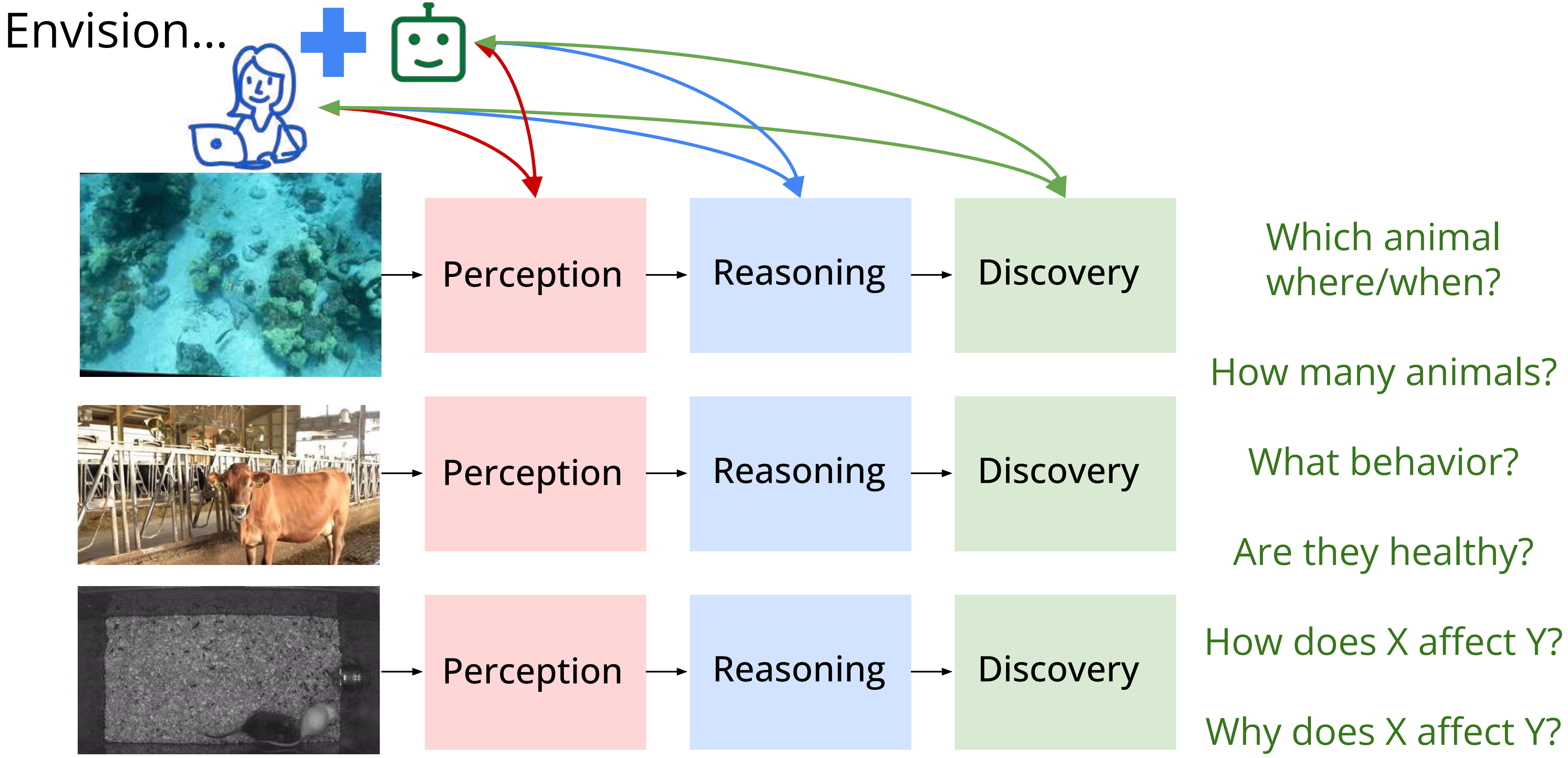


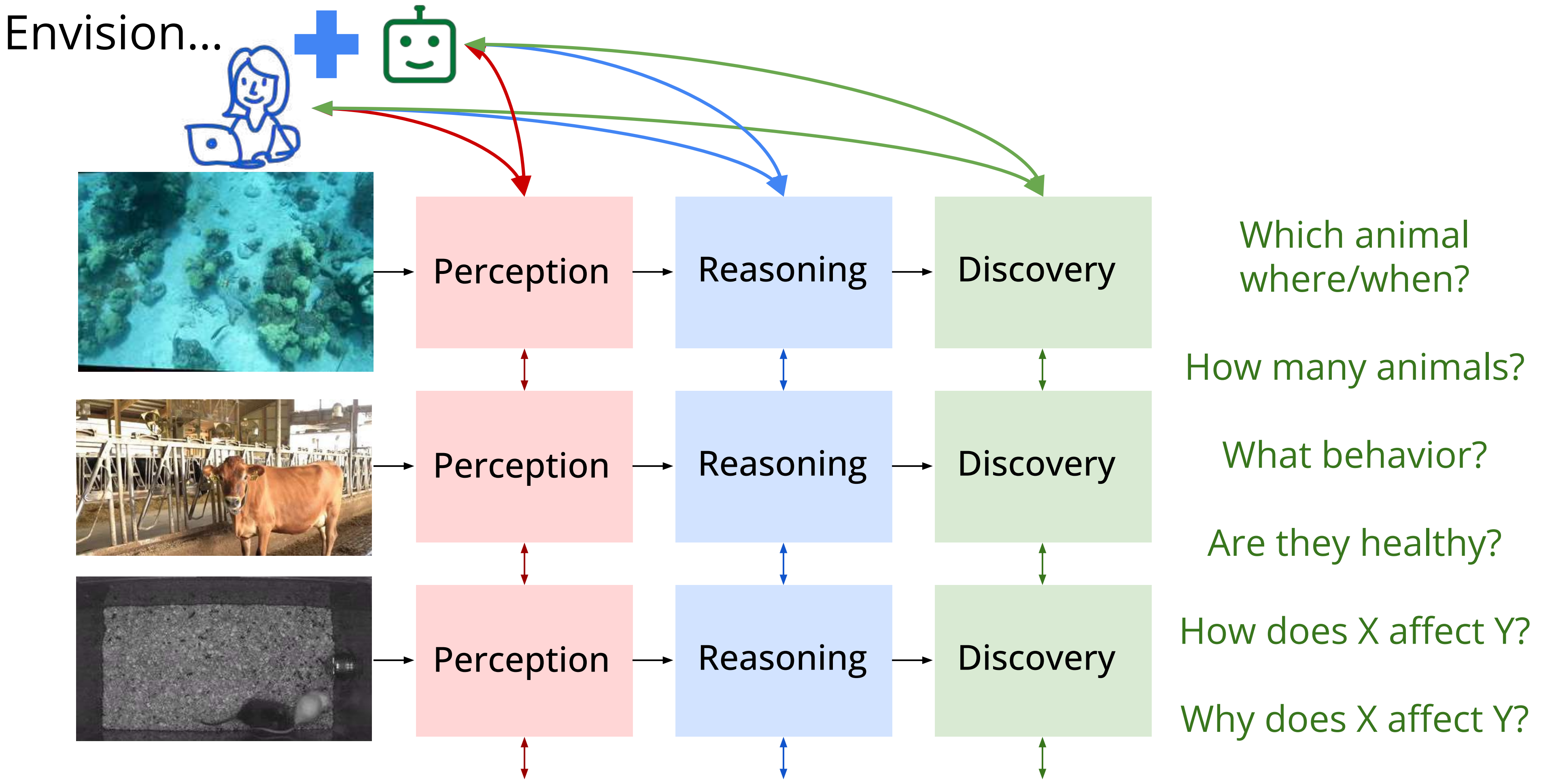
Challenge 1: Annotation bottleneck
Challenge 2: Vast model space w/ feedback
Challenge 3: Interpretability
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?
...

Envision...

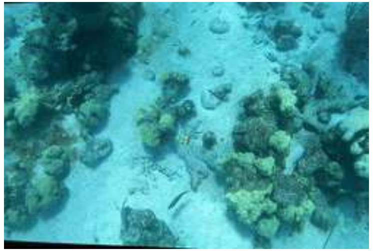


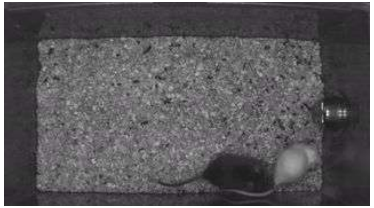
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?


Envision...
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?


Envision...
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?

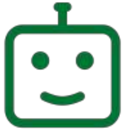
Our Approach





Scientists

AI Systems



Efficient & impactful collaborations between scientists & AI systems

Which animal where/when?

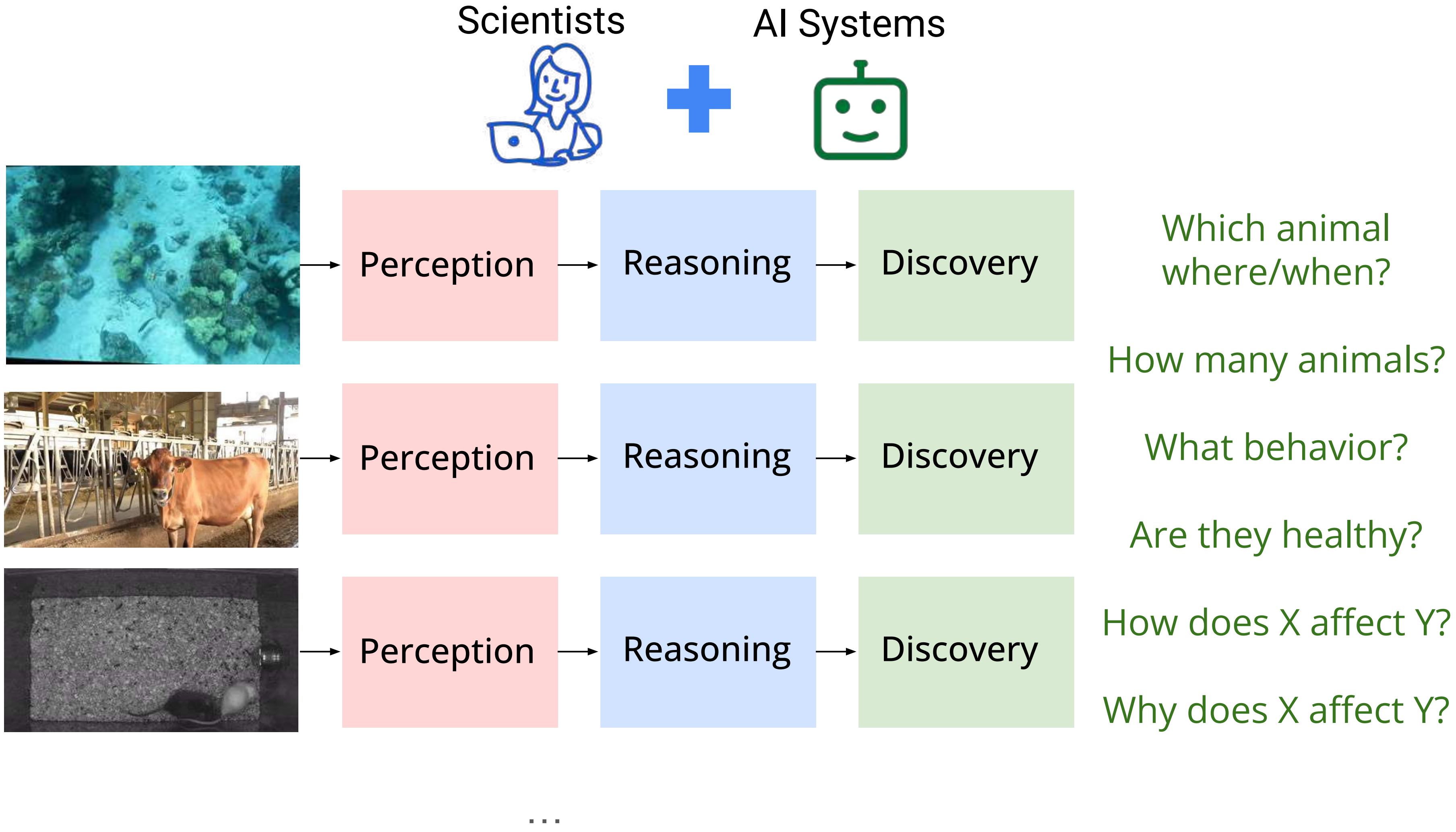
How many animals?

What behavior?

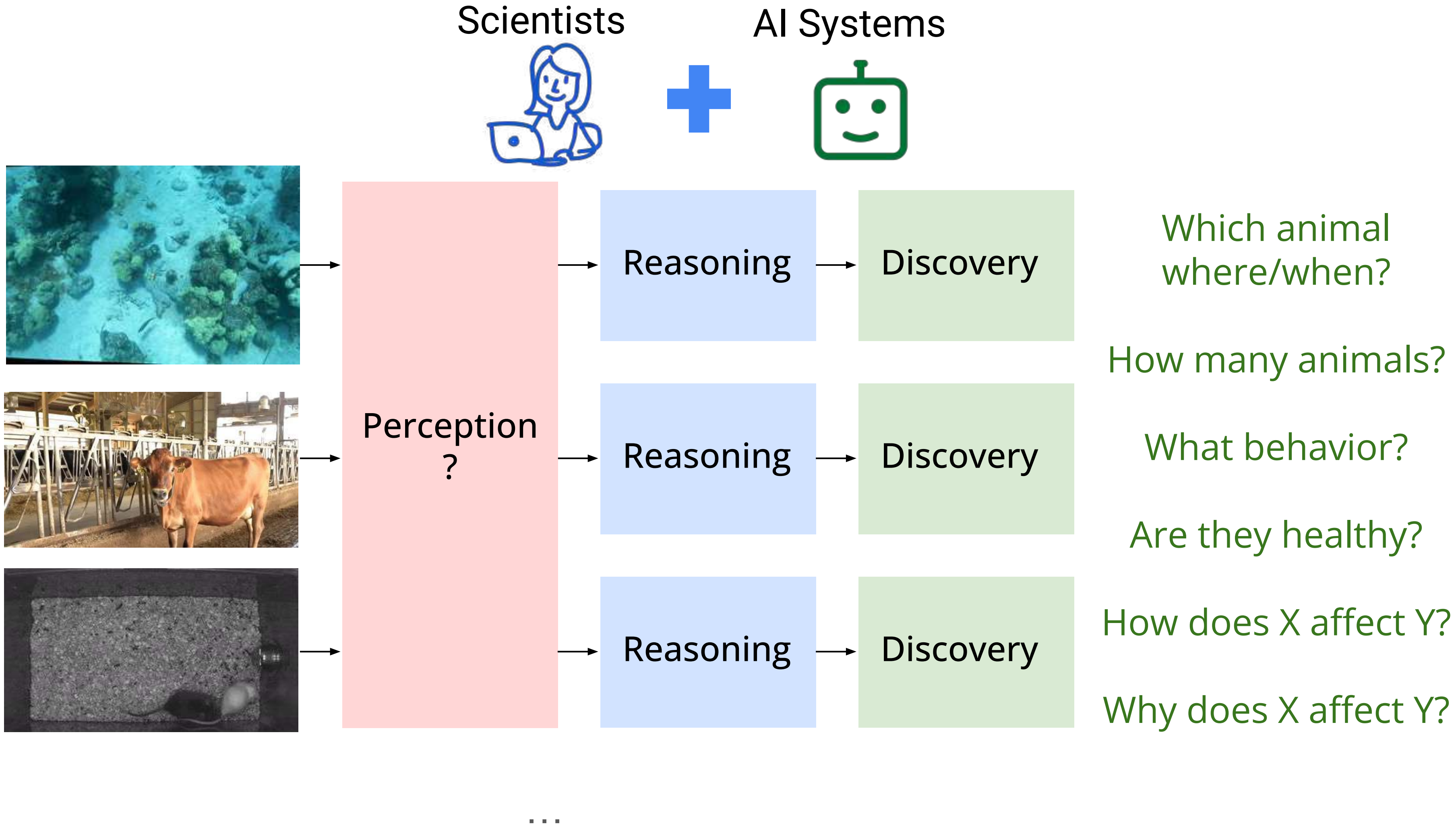
Are they healthy?

How does X affect Y?

Why does X affect Y?



Scientists
AI Systems
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?
...

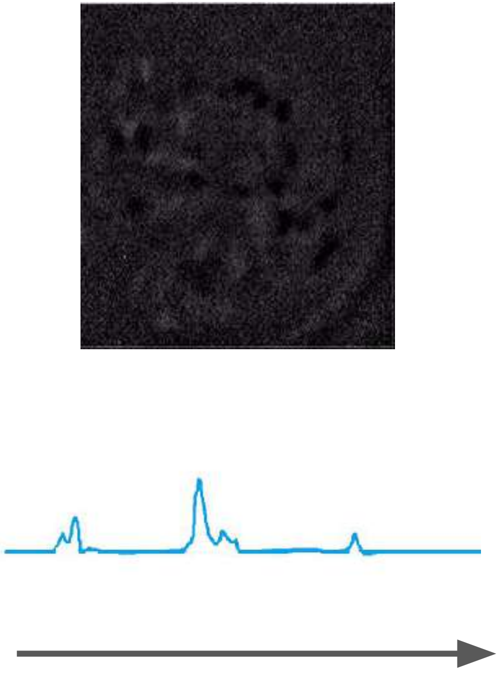

Scientists
AI Systems
Perception
?
Reasoning
Discovery
Reasoning
Discovery
Reasoning
Discovery
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?
...

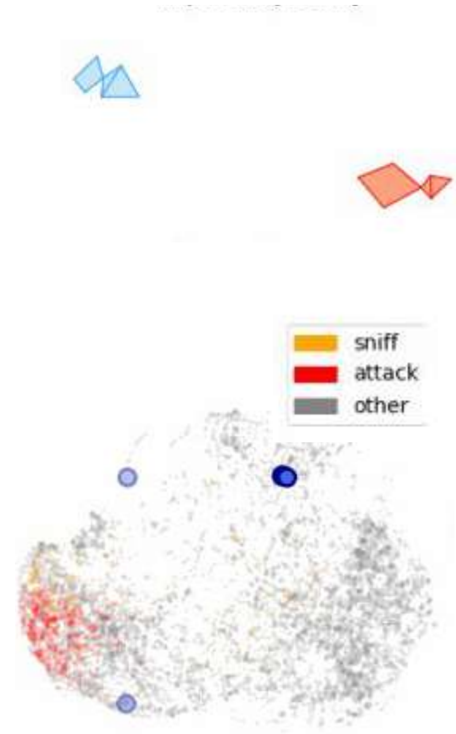
Perception

- Why is it important to extract symbolically interpretable representations?

Data has meaningful structure

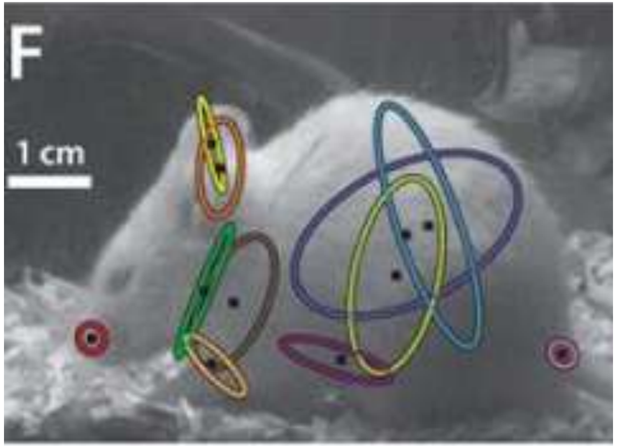
Human3.6M





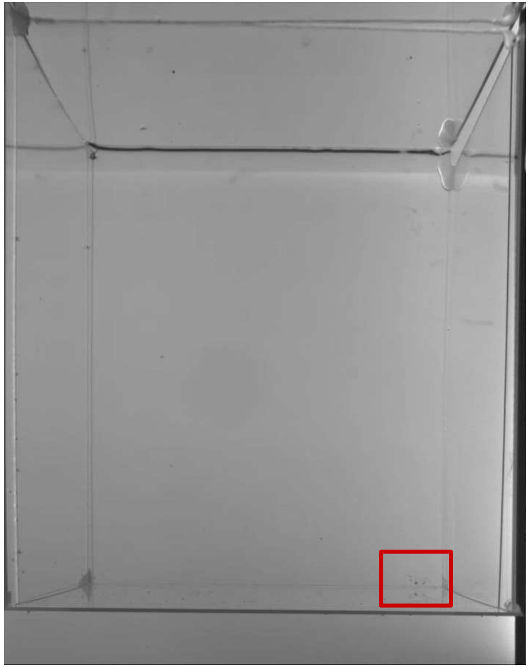
Challenges of extracting structure

Annotation Cost

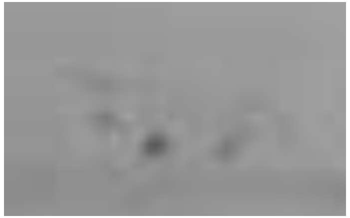


Segalin et al., 2021

Ambiguity & Variability



Brady Weissbourd at MIT

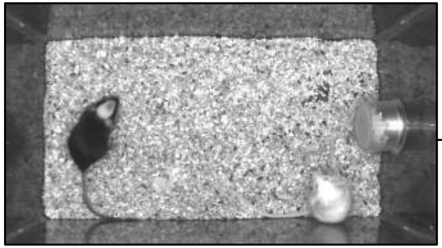


Low SNR

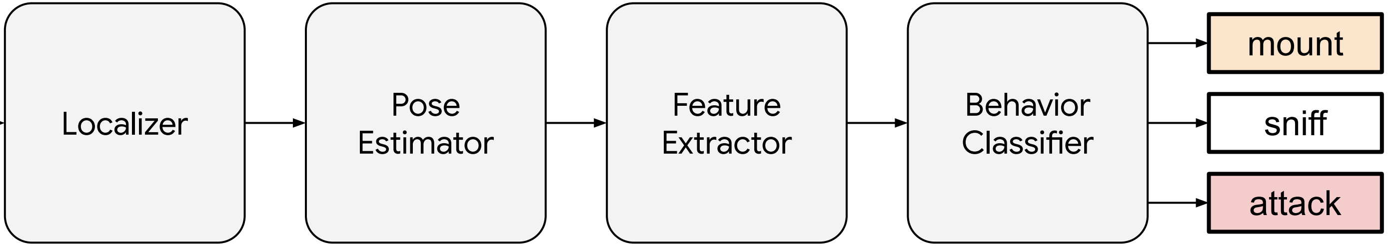
Perception

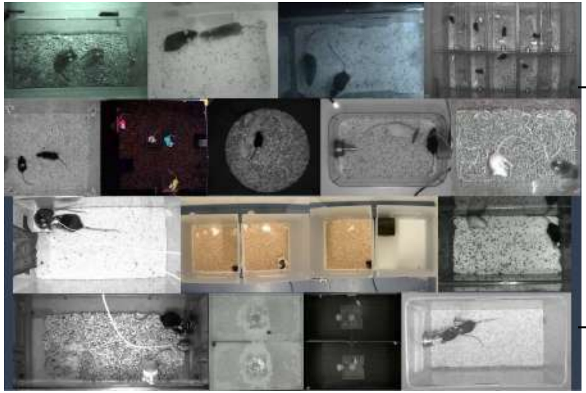
- Why is it important to extract symbolically interpretable representations?
- Can we have a general-purpose foundation model for learning representations?

Task-Specific Approach

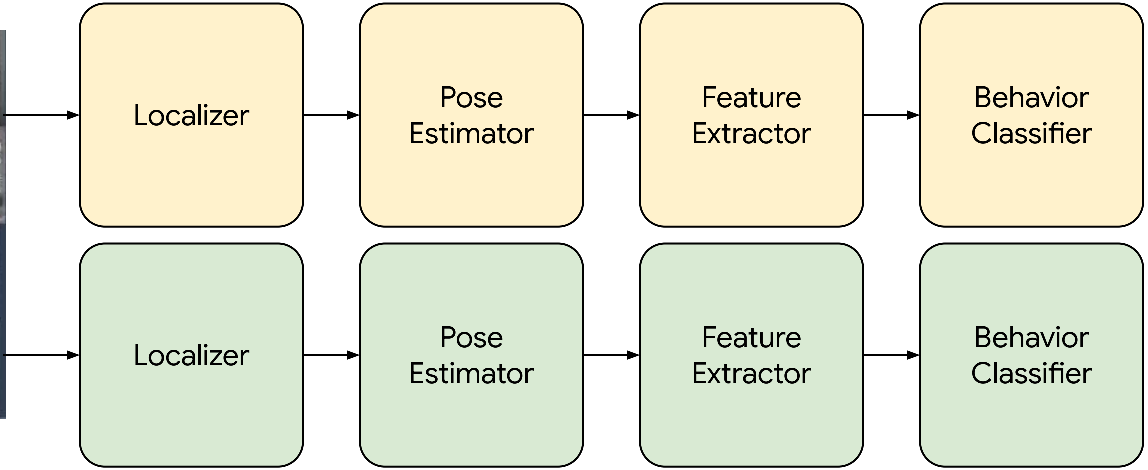


CalMS21 dataset

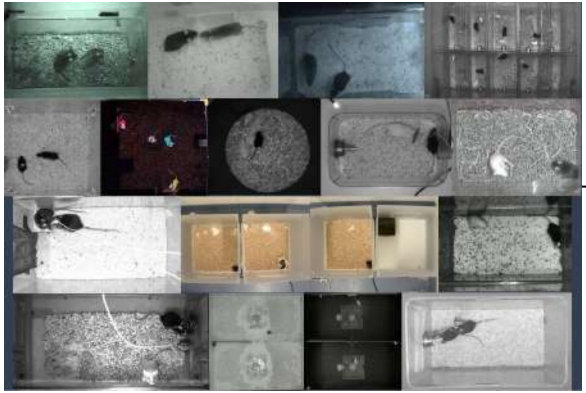




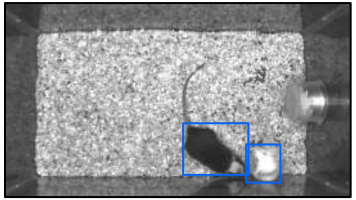
MABe2025 (upcoming dataset)



Foundation Model Approach



Foundation Model System



Localization

<sniff> <walk>

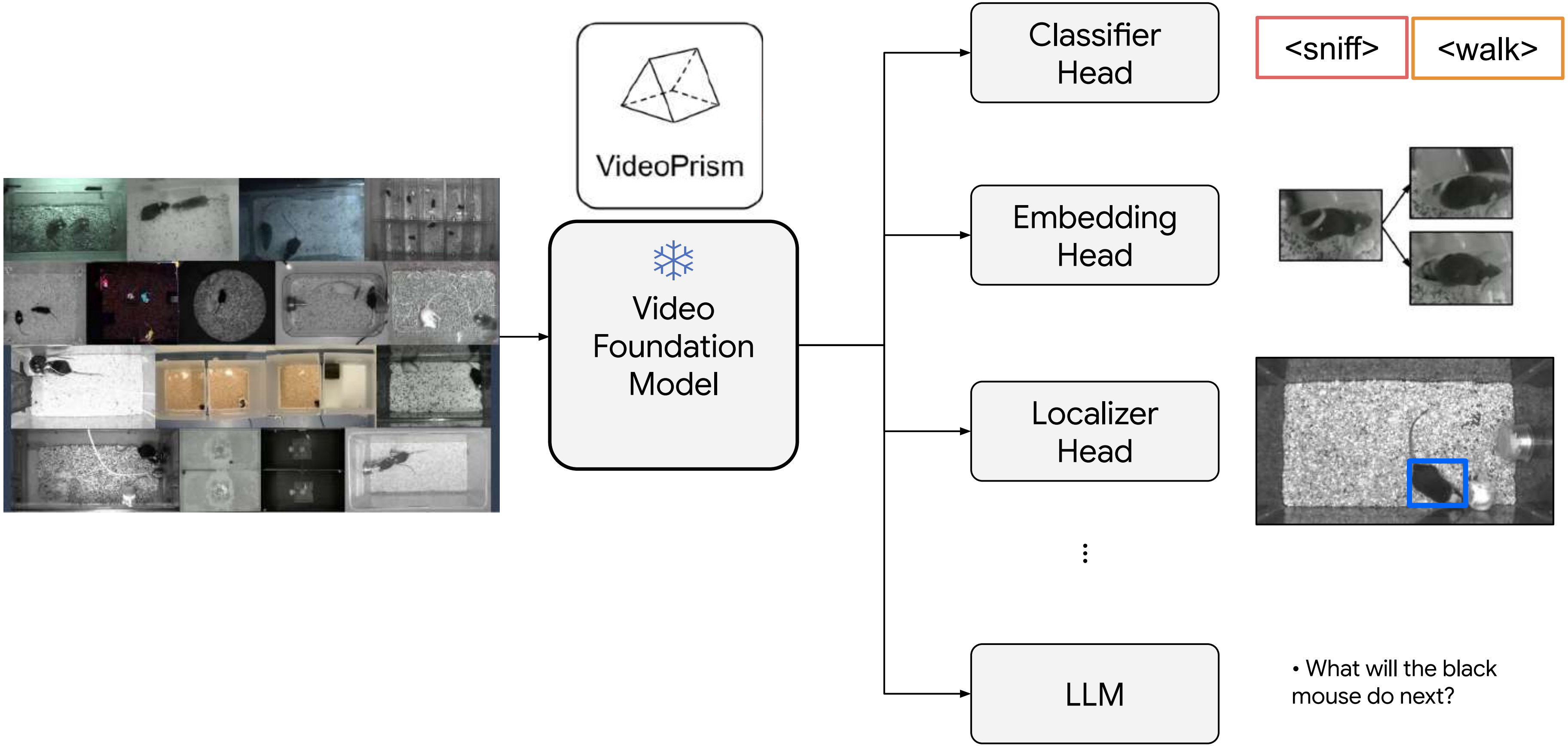
Classification

- What is the black mouse's sensory environment?
- What will the black mouse do next?

……

Scientific Video Analysis

Foundation Model Approach

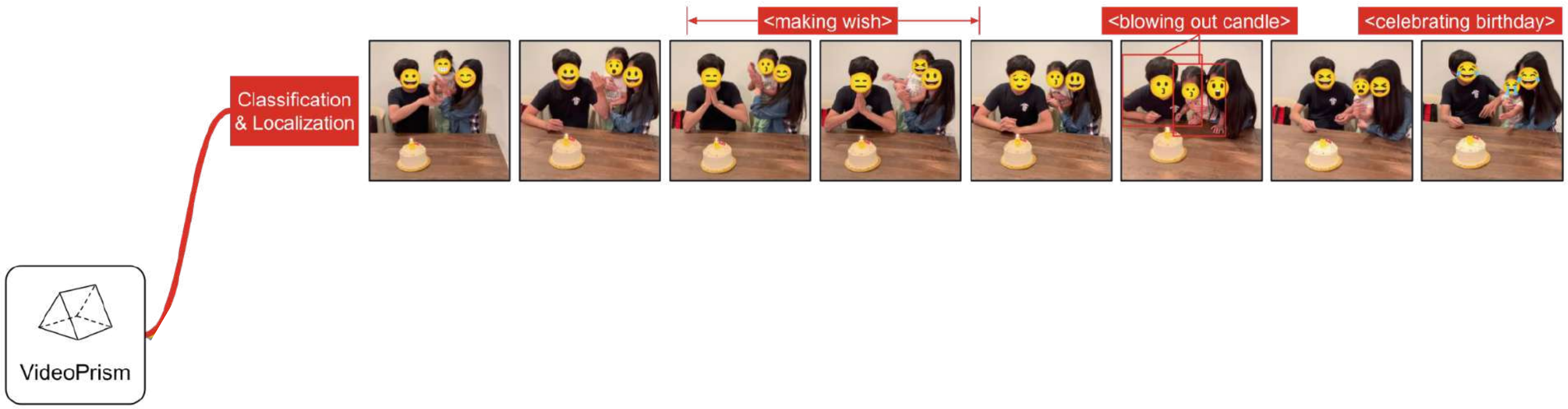


What is VideoPrism?

A foundational **video encoder** that enables **state-of-the-art** performance for **video understanding**

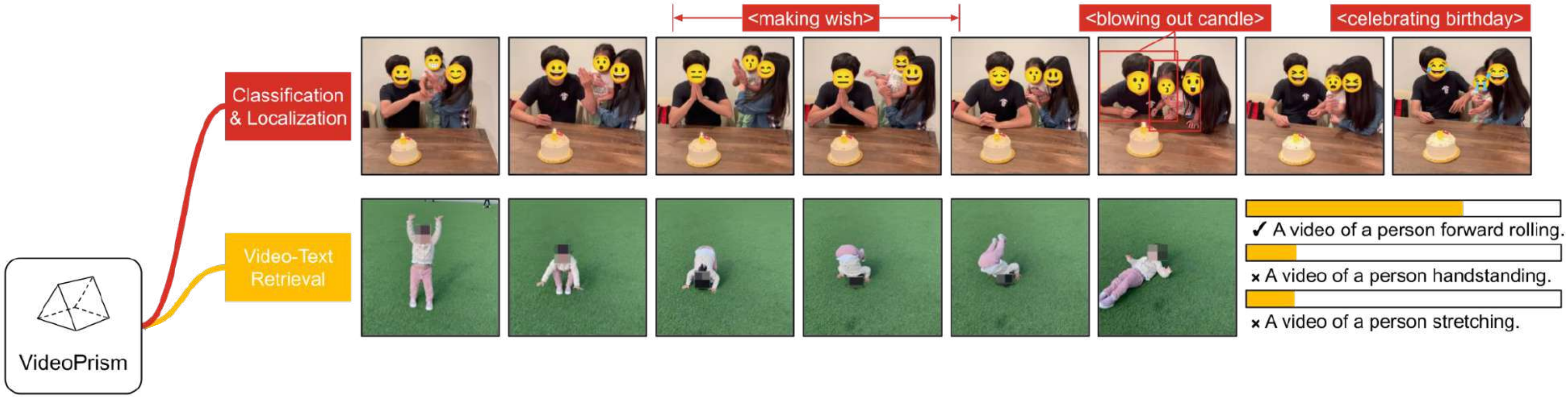
What is VideoPrism?

A foundational **video encoder** that enables **state-of-the-art** performance for **video understanding**



What is VideoPrism?

A foundational **video encoder** that enables **state-of-the-art** performance for **video understanding**

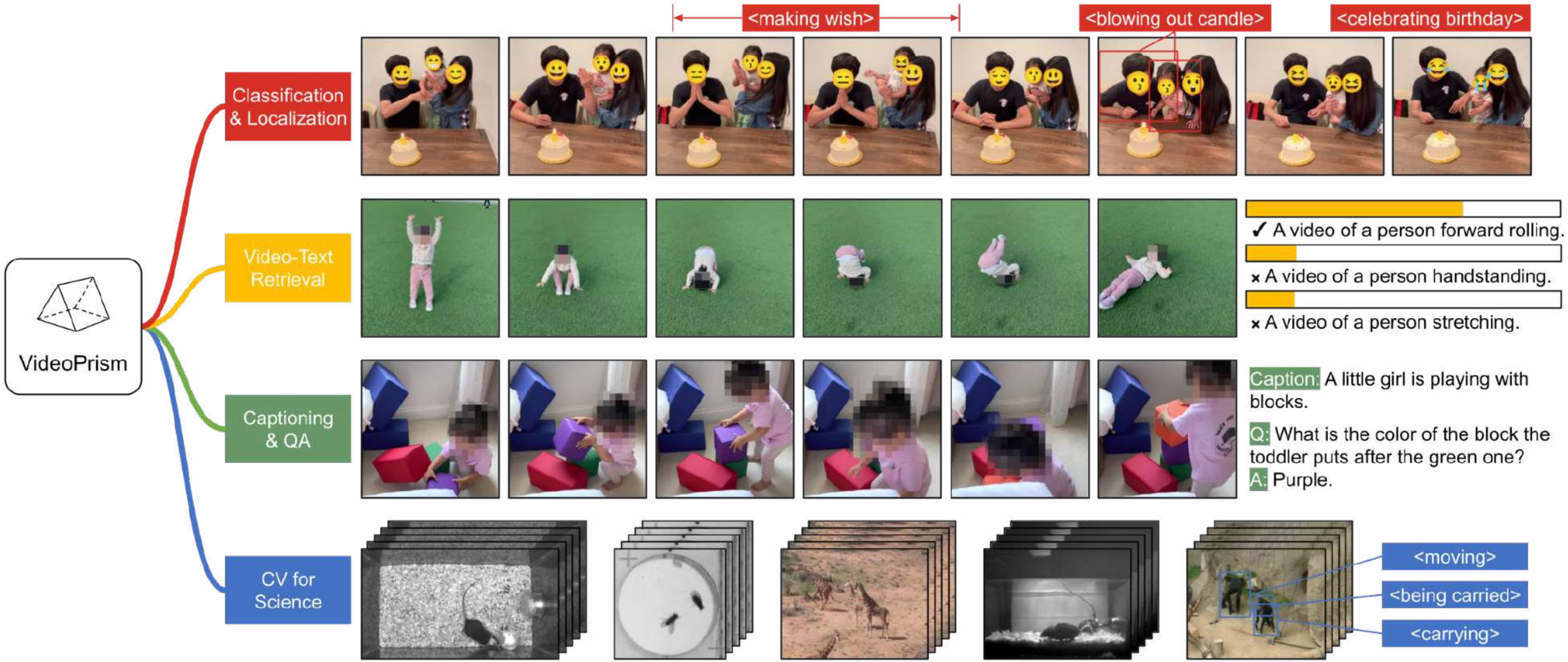


What is VideoPrism?

A foundational **video encoder** that enables **state-of-the-art** performance for **video understanding**

What is VideoPrism?

A foundational **video encoder** that enables **state-of-the-art** performance for **video understanding**

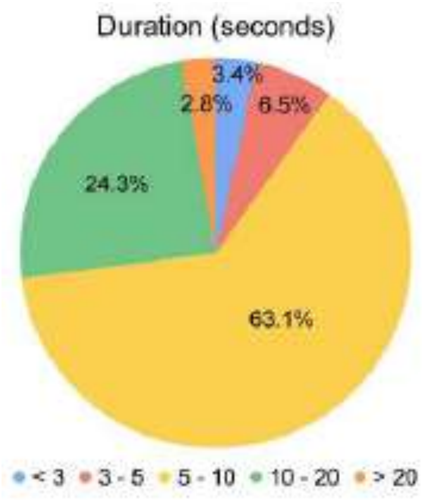


How is VideoPrism trained?

Large scale training data: **619M** video-text pairs:
(36M with high-quality captions + 583M with noisy parallel text).

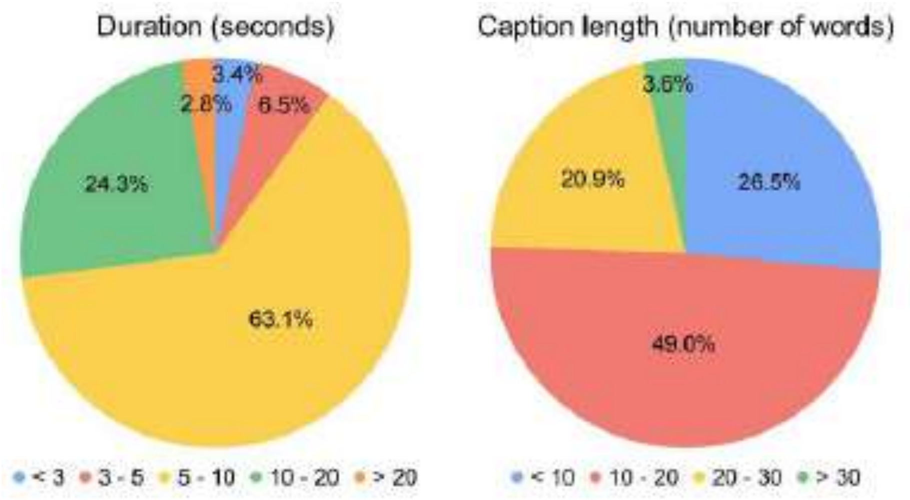
How is VideoPrism trained?

Large scale training data: **619M** video-text pairs:
(36M with high-quality captions + 583M with noisy parallel text).



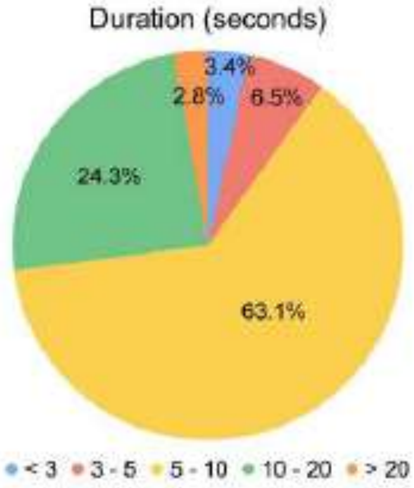
How is VideoPrism trained?

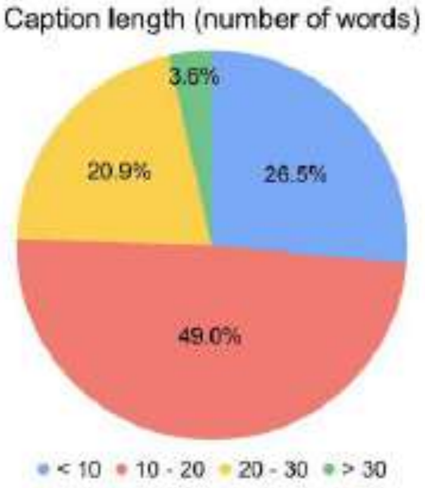
Large scale training data: **619M** video-text pairs:
(36M with high-quality captions + 583M with noisy parallel text).

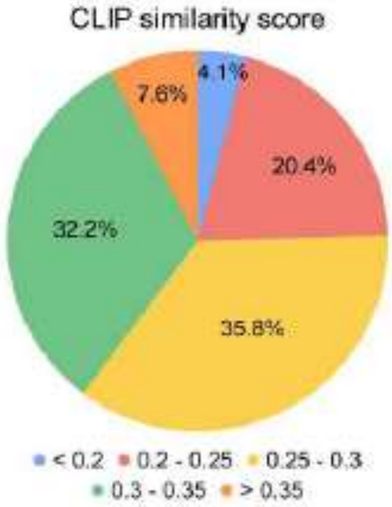


How is VideoPrism trained?

Large scale training data: **619M** video-text pairs:
(36M with high-quality captions + 583M with noisy parallel text).



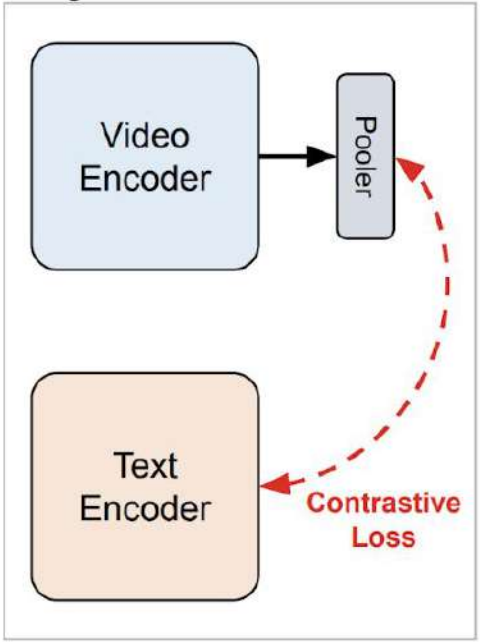




How is VideoPrism trained?

Two stage training:

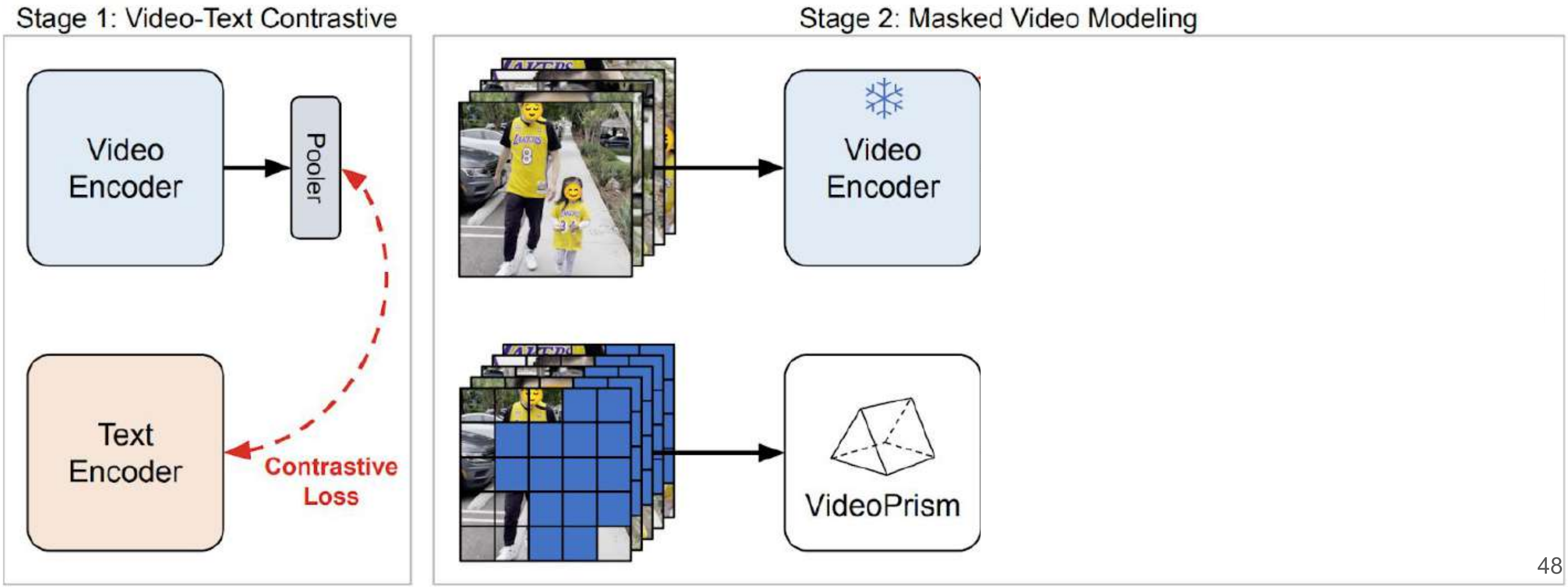
Stage 1: Video-Text Contrastive



Stage 2: Masked Video Modeling

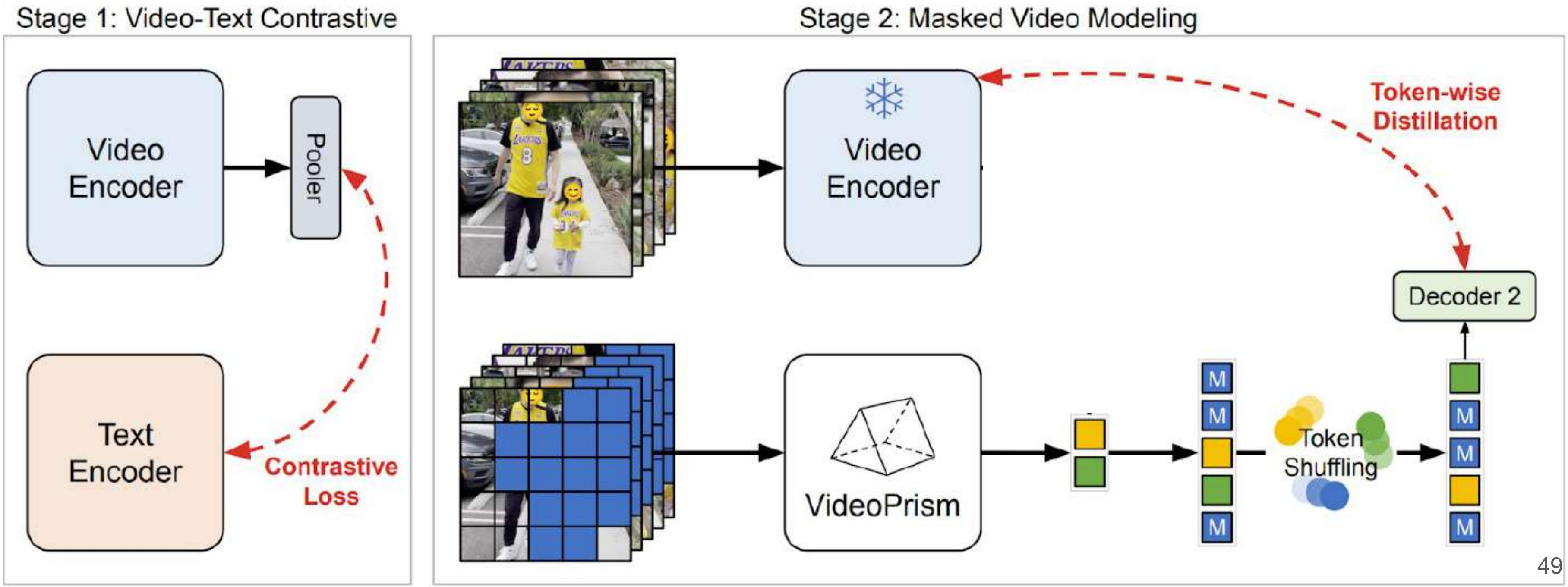
How is VideoPrism trained?

Two stage training:



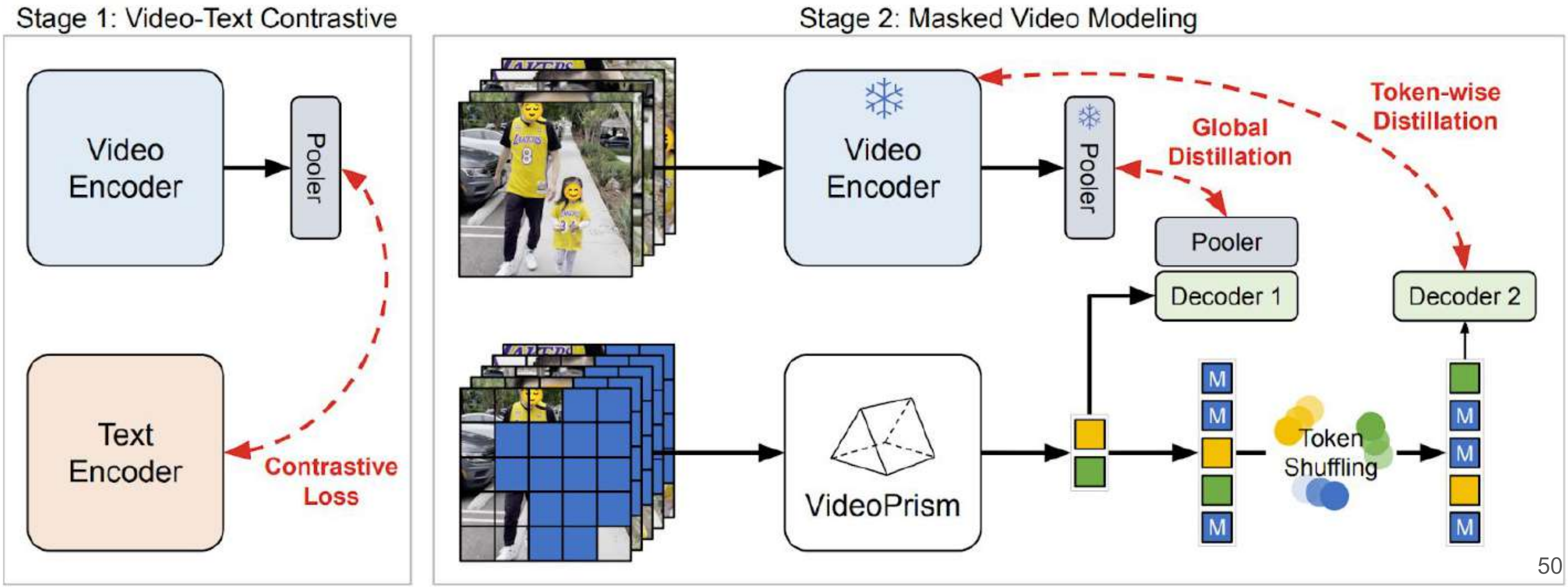
How is VideoPrism trained?

Two stage training:

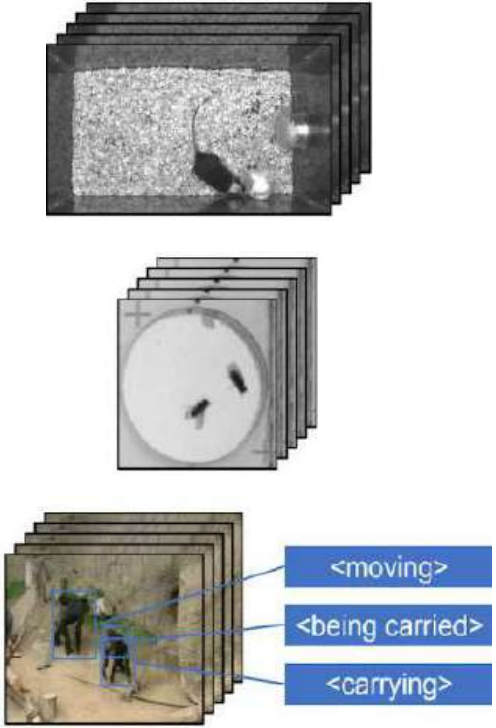


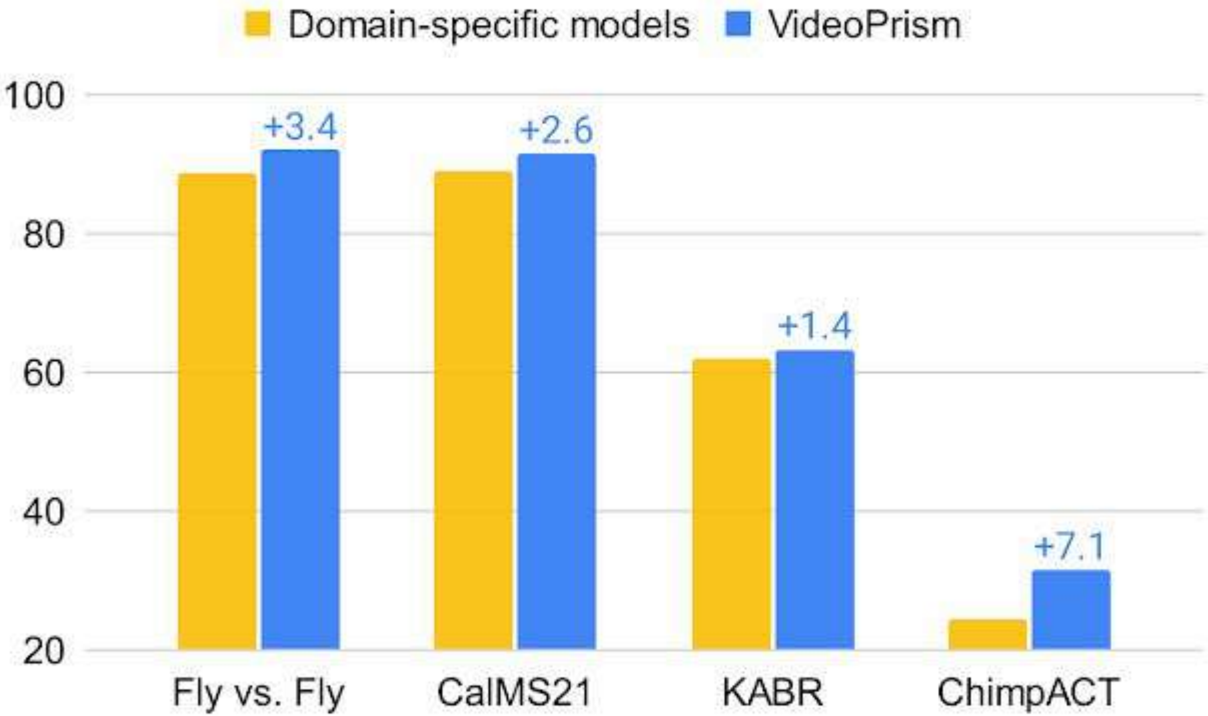
How is VideoPrism trained?

Two stage training:



VideoPrism for science






Long Zhao


Nitesh Bharadwaj Gundavarapu


Liangzhe Yuan


Hao Zhou

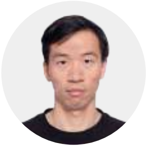

Shen Yan


Jennifer Sun


Luke Friedman


Rui Qian


Tobias Weyand


Yue Zhao


Rachel Hornung


Florian Schroff


Ming-Hsuan Yang


David Ross


Huisheng Wang


Hartwig Adam


Mikhail Sirotenko


Ting Liu


Boqing Gong

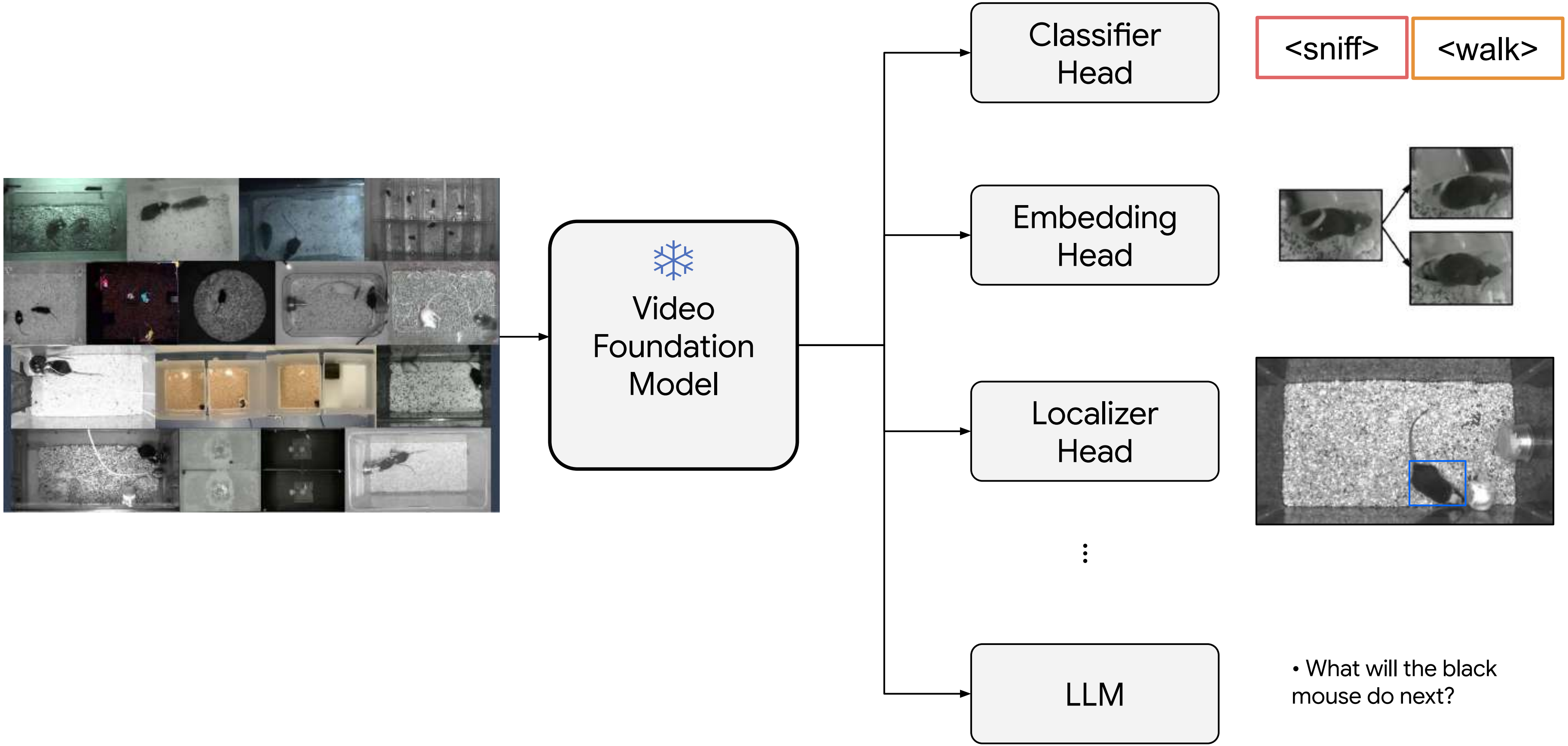

David Hendon


Alex Siegman

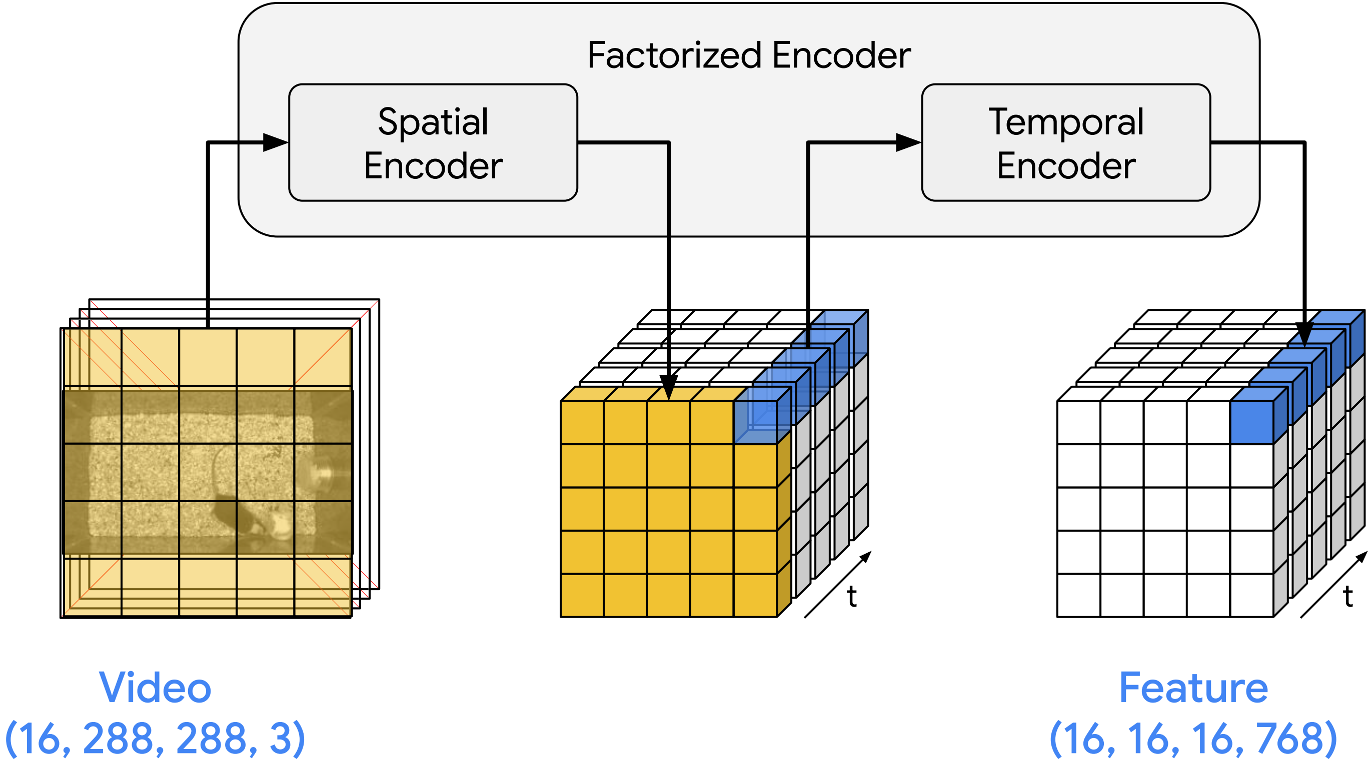
Perception

- Why is it important to extract symbolically interpretable representations?
- Can we have a general-purpose foundation model for learning representations?
- Can they extract symbols from domain-specific data?

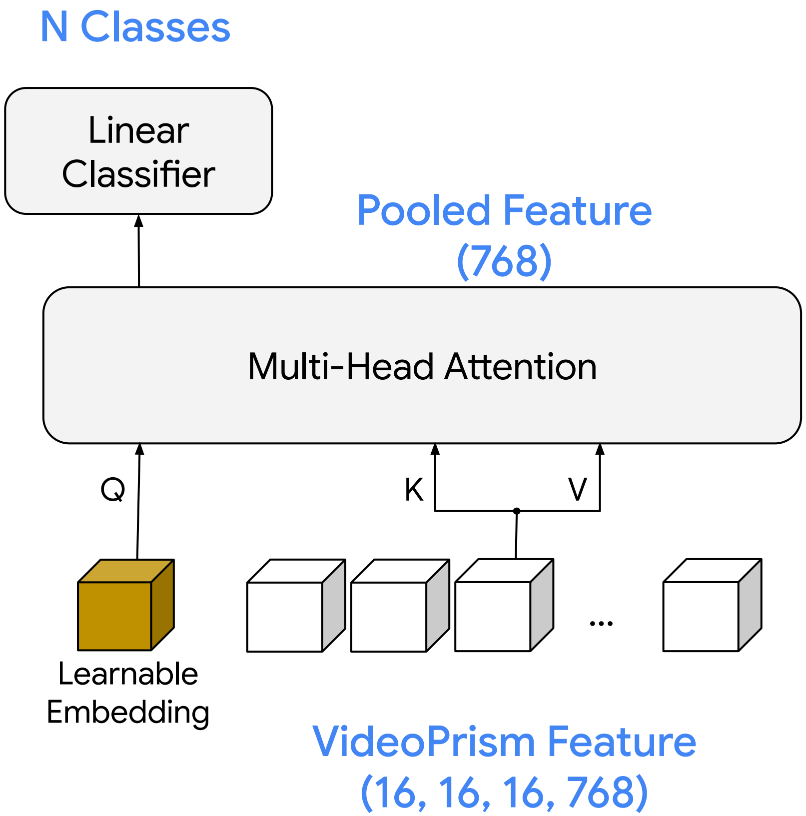
Foundation Model Approach



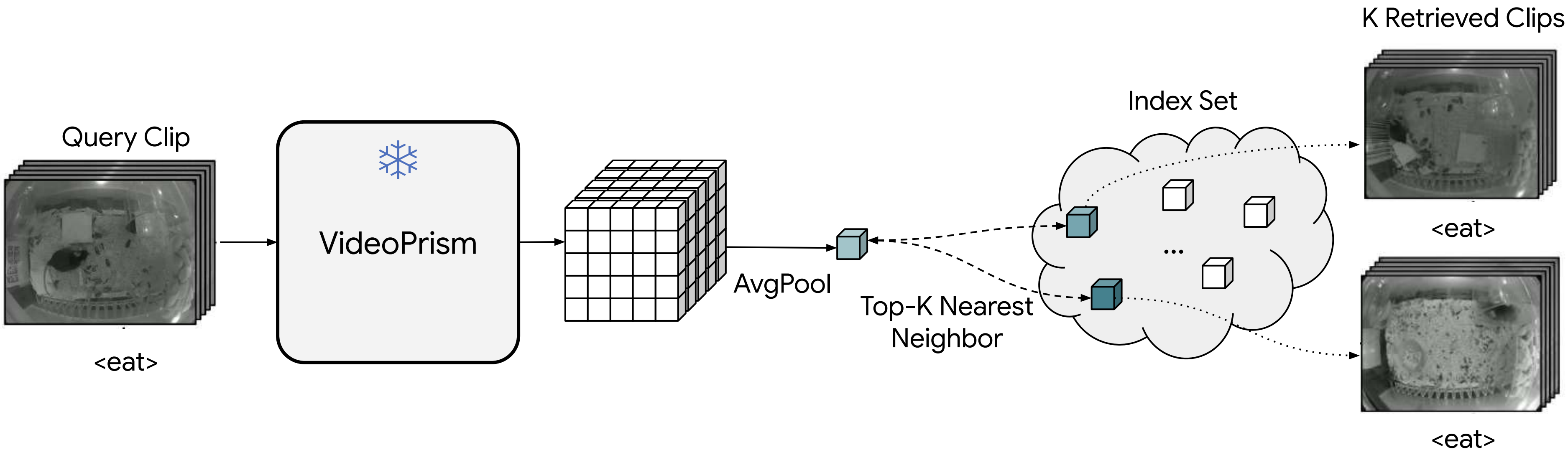
VideoPrism Architecture



Classification

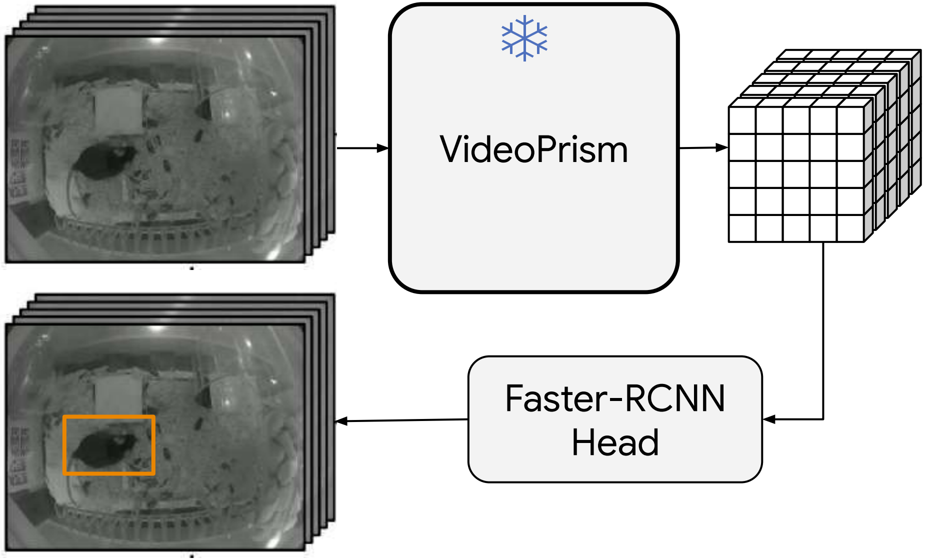


Retrieval



Query Clip
<eat>
VideoPrism
AvgPool
Top-K Nearest
Neighbor
Index Set
...
K Retrieved Clips
<eat>
<eat>

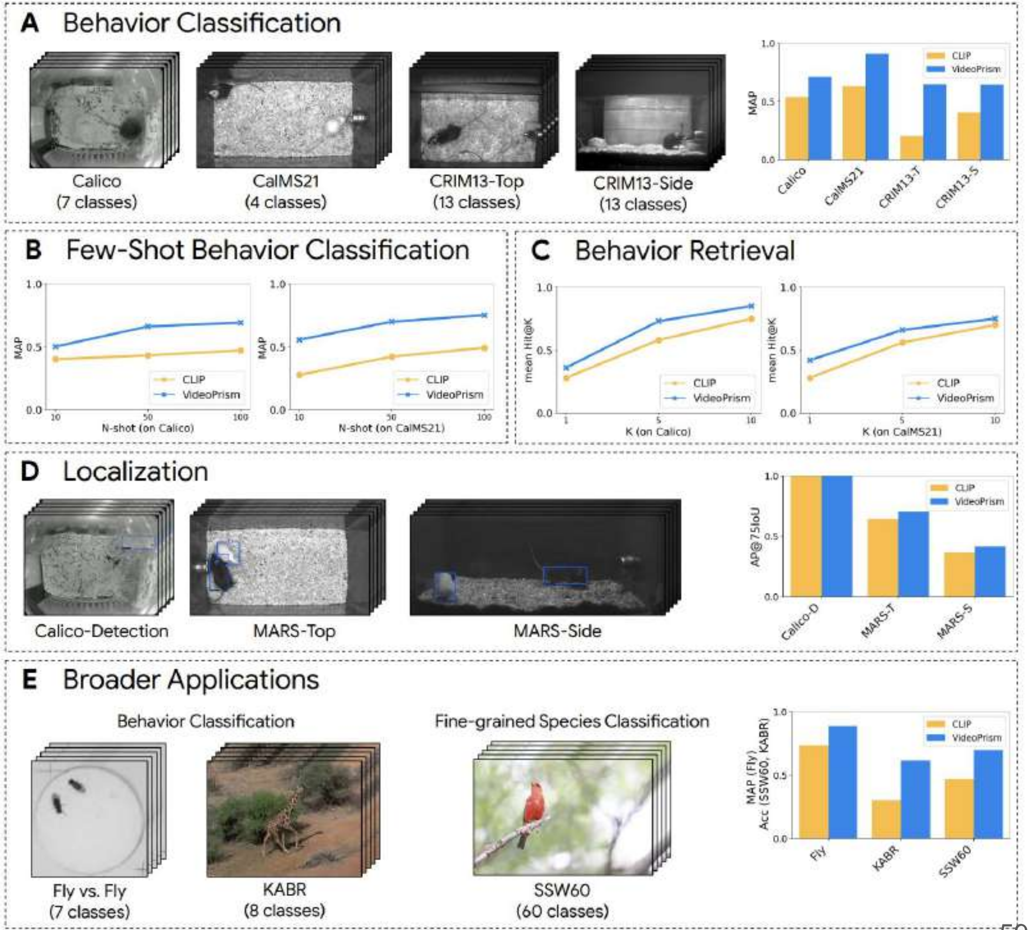
Localization

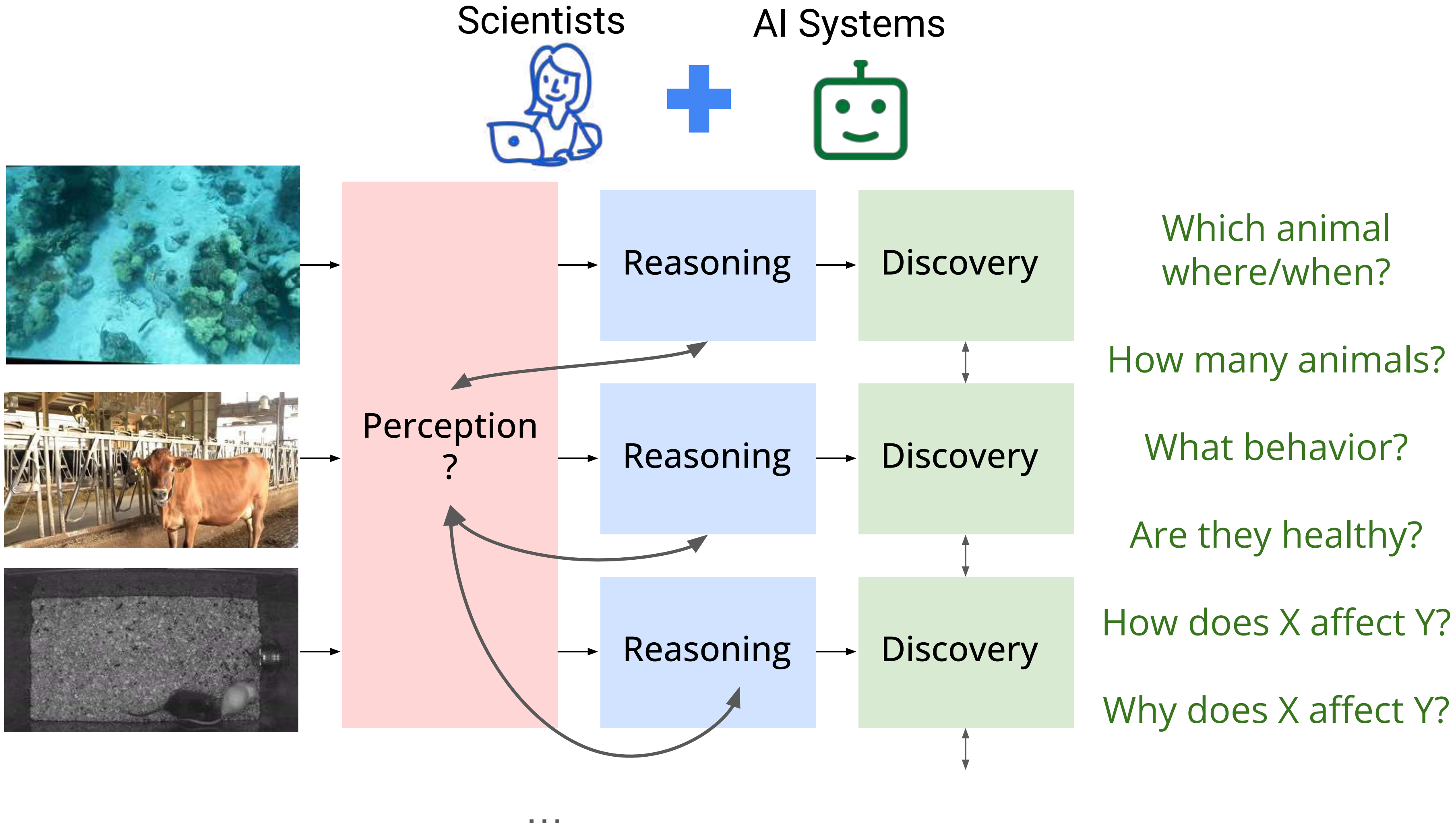


Video Foundation Models for Animal Behavior Analysis

Jennifer J. Sun*, Hao Zhou, Long Zhao, Liangzhe Yuan, Bryan Seybold, David Hendon, Florian Schroff, David A. Ross Hartwig Adam, Bo Hu†, Ting Liu†*

[1]Google.



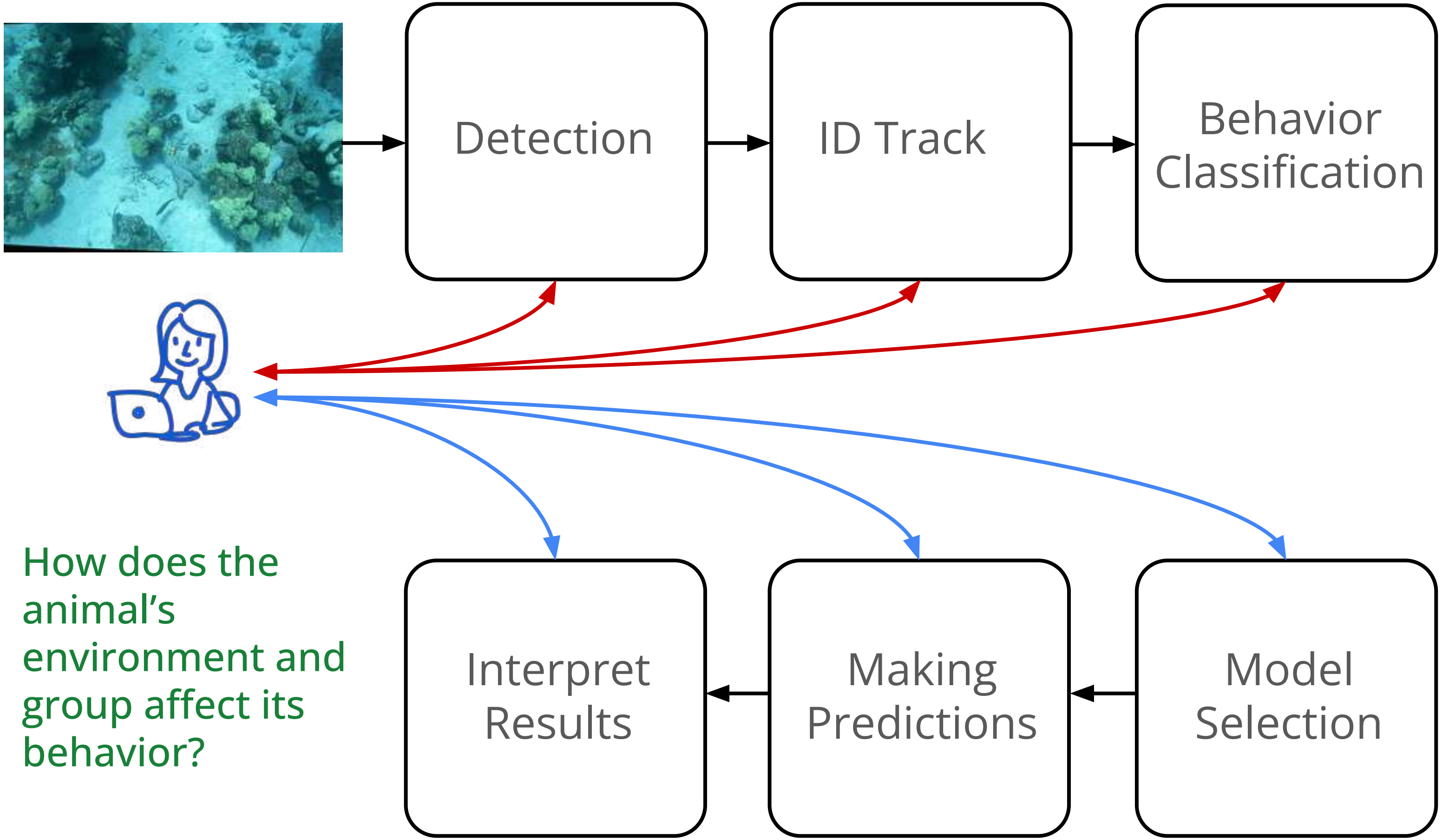

Scientists
AI Systems
Perception
?
Reasoning
Discovery
Reasoning
Discovery
Reasoning
Discovery
Which animal
where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?
...

Perception & Reasoning

- Why is it so hard to build effective scientific workflows?

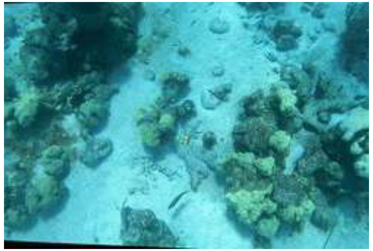
Current Analysis Pipelines

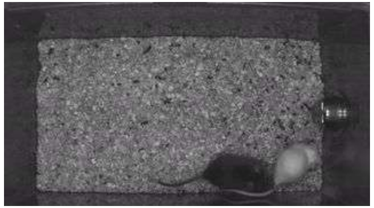
Vast model space w/ feedback



How does the animal's environment and group affect its behavior?

Which tool (if any) will work out-of-the-box?





YOLOv9

DETR

YOLOv10

SAM

SAM2

YOLOv11

OWL-VIT

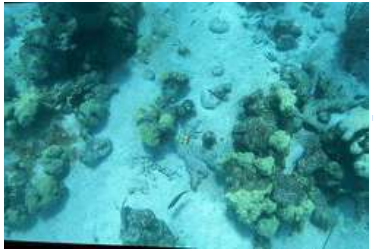
Cotracker

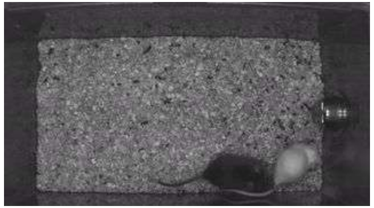
OWLv2

GroundingSAM/
DINO

2026 & beyond…

Which tool (if any) will work out-of-the-box?





YOLOv9

DETR

SAM

YOLOv10

SAM2

YOLOv11

OWL-VIT

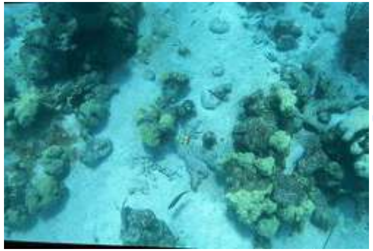
Cotracker

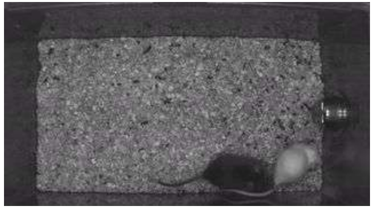
OWLv2

GroundingSAM/
DINO

2026 & beyond…

Which tool (if any) will work out-of-the-box?





YOLOv9

DETR

SAM

YOLOv10

SAM2

YOLOv11

OWL-ViT

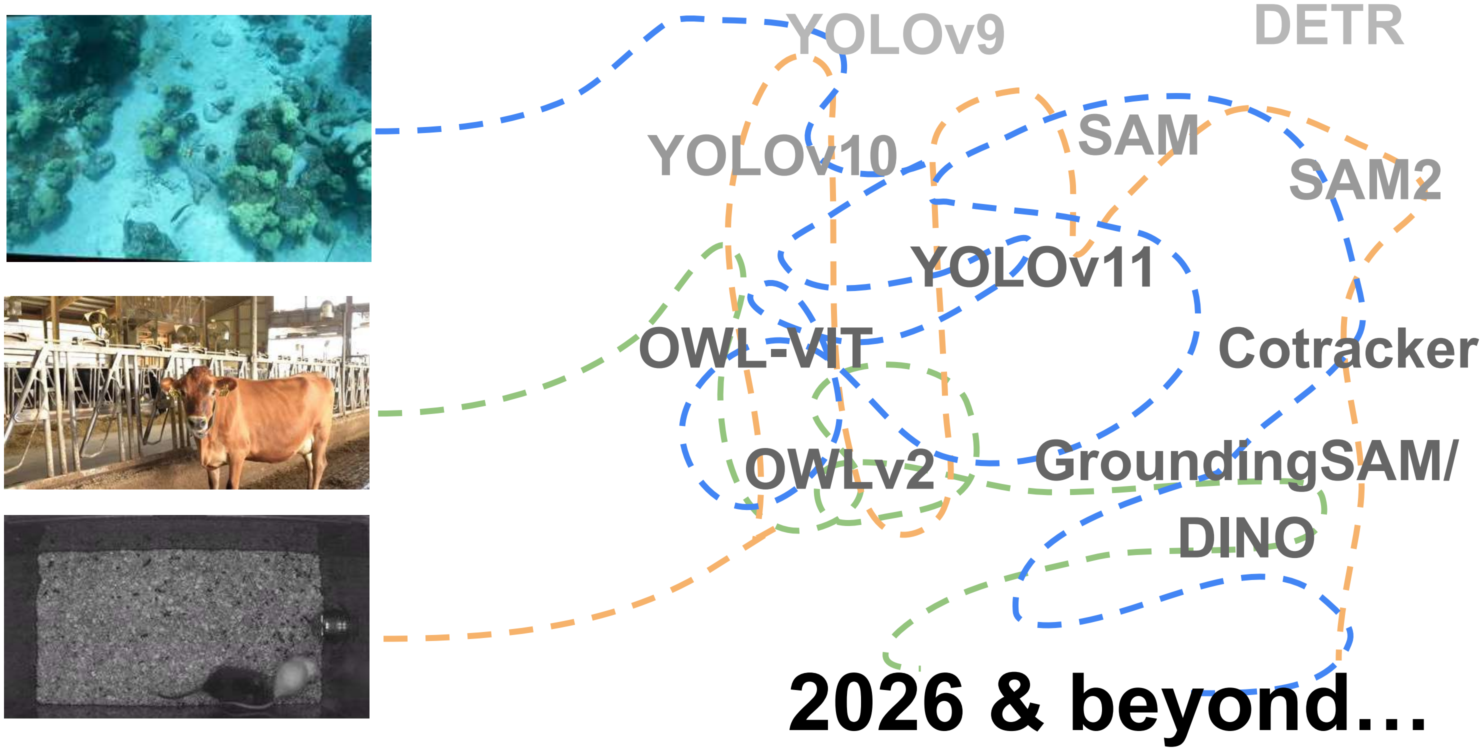
Cotracker

OWLv2

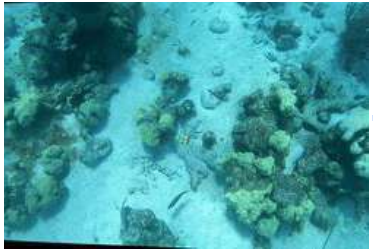
GroundingSAM/
DINO

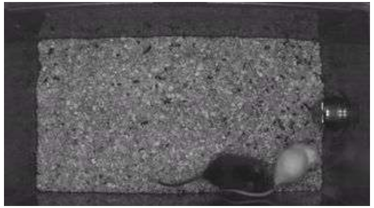
2026 & beyond…

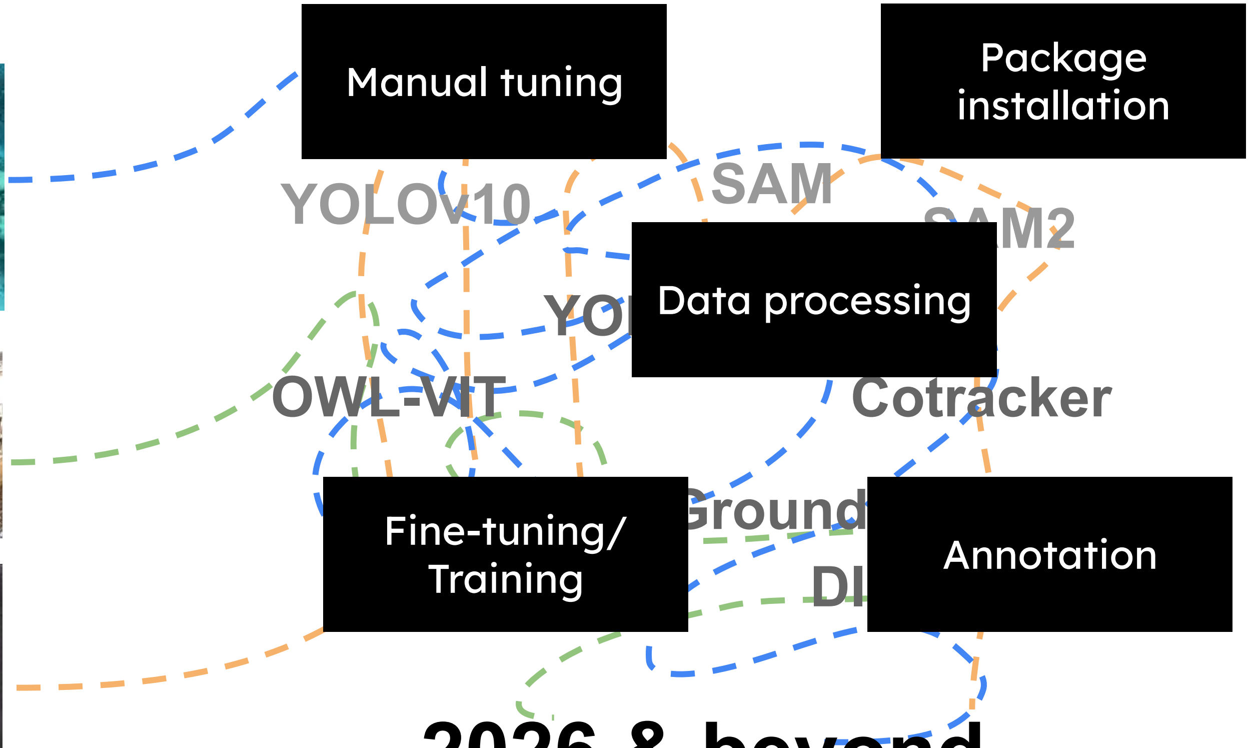
Which tool (if any) will work out-of-the-box?



Which tool (if any) will work out-of-the-box?





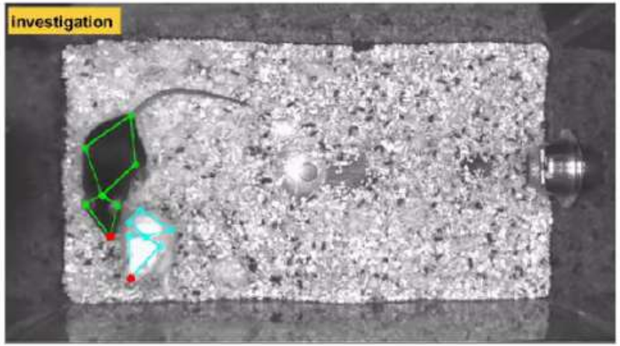


2026 & beyond…

Perception & Reasoning

- Why is it so hard to build effective scientific workflows?
- Instead of manual effort, can we have an AI agent discover an optimal workflow for us?

Program Synthesis



IF (**distance between noses**) < A AND

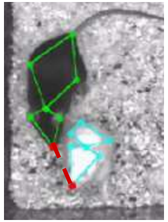
(**facing angle**) < B

THEN **investigation** IF

(**acceleration of mouse 1**) > C

ELSE **investigation** IF

(**distance from nose 1 to centroid 2**) < D



Features defined by experts
(or language models)

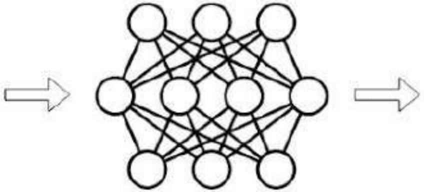
Superoptimization in Program Synthesis

Find “better” programs
(e.g. better = faster)

Human code

```cpp
#include <iostream>
using namespace std;

int main(){
    int n;
    cin >> n;
    int sum = 0;
    for (int i = 1; i <= n; i++) {
        sum += i;
    }
    cout << sum << endl;
    return 0;
}
```



Synthesized code

```cpp
#include <iostream>
using namespace std;

int main(){
    int n;
    cin >> n;
    cout << n*(n+1)/2 << endl;
    return 0;
}
```

[Shypula., et. al. ICLR (2024)]

Can we superoptimize scientific analysis workflows?

Find “better” programs
(e.g. better = more accurate for analysis)

Human code

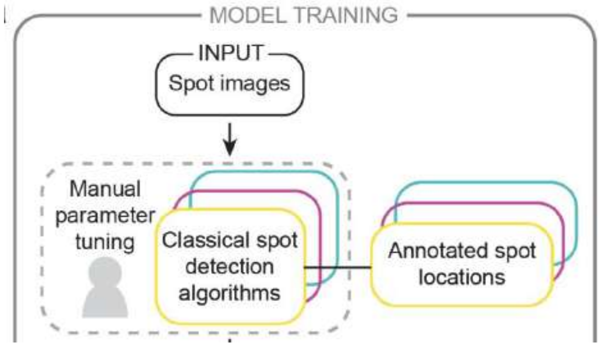
Synthesized code

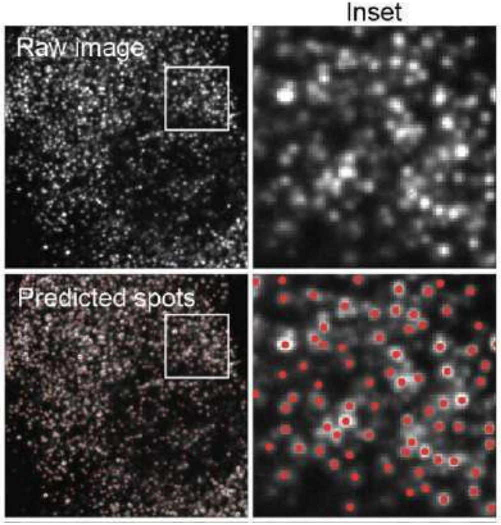
Can we superoptimize scientific analysis workflows?

Find “better” programs
(e.g. better = more accurate for analysis)

Accurate single-molecule spot detection for image-based spatial transcriptomics with weakly supervised deep learning

Emily Laubscher[1], Xuefei (Julie) Wang[2], Nitzan Razin[2], Tom Dougherty[2], Rosalind J. Xu[3,4,5], Lincoln Ombelets[1], Edward Pao[2], William Graf[2], Jeffrey R. Moffitt[3,4,6], Yisong Yue[7], and David Van Valen[2]





Can we superoptimize scientific analysis workflows?

Find “better” programs
(e.g. better = more accurate for analysis)

Human code

```python
def min_max_normalize_clipping (image):
    image_processed = []
    for img in images.raw:
        img = np.clip(img,
a_min=np.percentile(img, 0.01),
a_max=np.percentile(img, 99.9))
        min_val = np.min(img)
        max_val = np.max(img)
        normal_image = (img - min_val) /
(max_val - min_val)
        image_processed.append(normal_image)
    return np.array(image_processed)
```

Expert function
F1 Score: 0.841
Time: Weeks/Months

Synthesized code

```python
def blurred_laplacian_of_gaussian(images):
    processed_images_list = []
    for img_array in images:
        img = np.copy(img_array)
        img_float32 = cv.normalize(img, None, 0, 1,
cv.NORM_MINMAX).astype(np.float32)
        bilateral = cv.bilateralFilter(img_float32, d=5,
sigmaColor=0.09, sigmaSpace=9)
        gauss = cv.GaussianBlur(bilateral, (3,3), 0)
        lap = cv.Laplacian(gauss, cv.CV_32F, ksize=3)
        abs_lap = np.abs(lap)
        lap_norm = cv.normalize(abs_lap, None, 0, 1,
cv.NORM_MINMAX).astype(np.float32)
        if img_array.ndim == 3 and img_array.shape[2] == 1:
            lap_norm = lap_norm[:, :, np.newaxis]
        processed_images_list.append(lap_norm)
    return np.array(processed_images_list, dtype=np.float32)
```

Agent function
F1 Score: 0.902
Time: 10 hours

Can we superoptimize scientific analysis workflows?

Find “better” programs
(e.g. better = more accurate for analysis)

Human code

```
def min_max_normalize_clipping(image):
    image_processed = []
    for img in images.raw:
        img = np.clip(img,
a_min=np.percentile(img, 0.01),
a_max=np.percentile(img, 99.9))
        min_val = np.min(img)
        max_val = np.max(img)
        normal_image = (img - min_val) /
(max_val - min_val)
        image_processed.append(normal_image)
    return np.array(image_processed)
```

Expert function
F1 Score: 0.841
Time: Weeks/Months

Synthesized code

Deployed into real workflow!

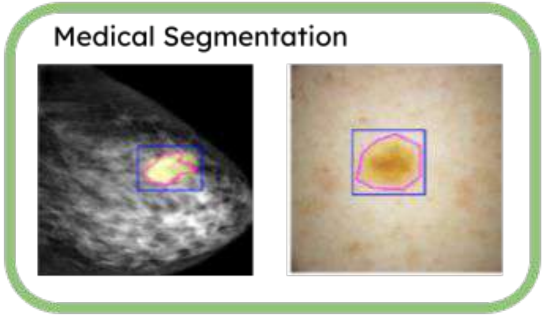
Agent function
F1 Score: 0.902
Time: 10 hours

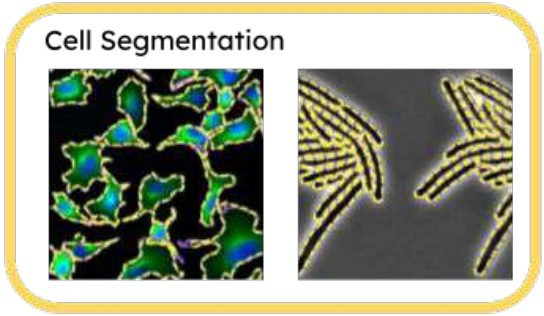
Agentic Superoptimization of Scientific Analysis Workflows

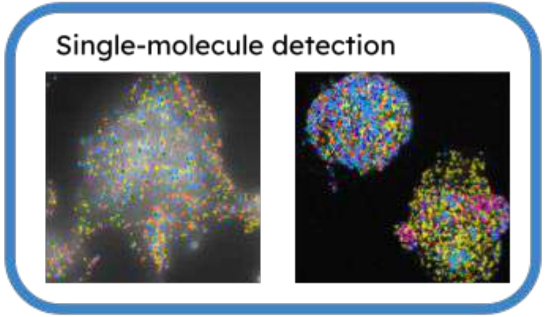

Julie Wang


Jonathan Chen

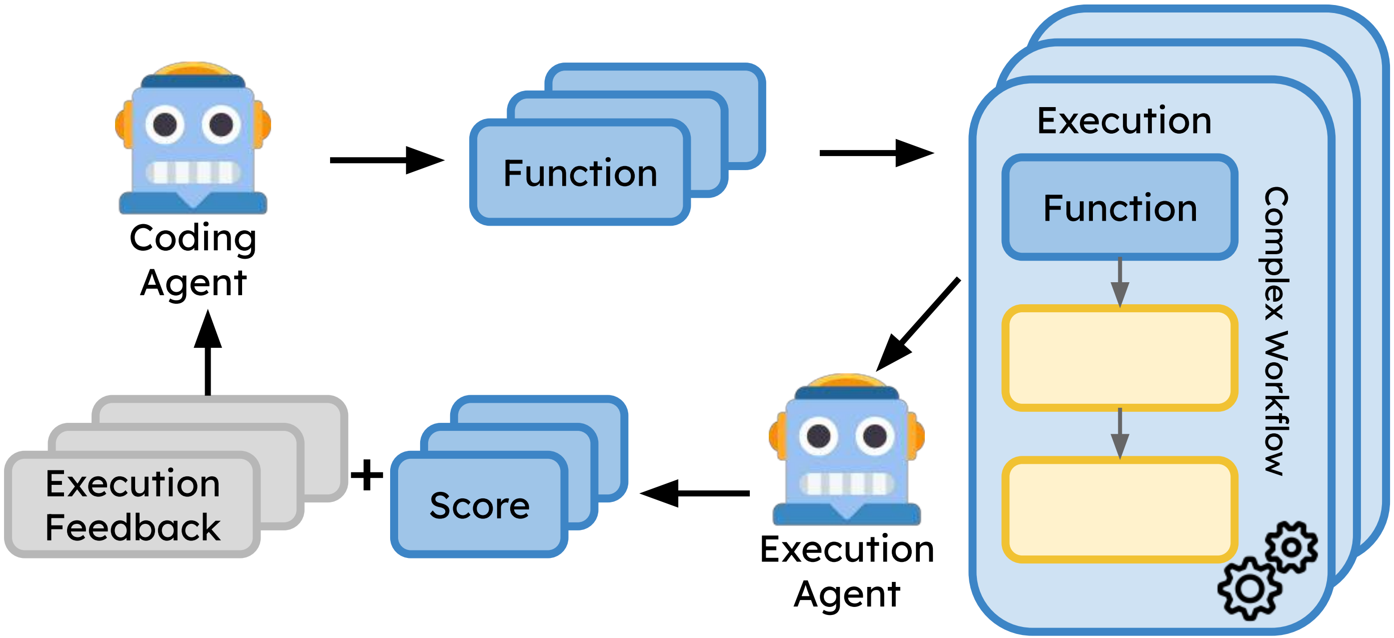

Yisong Yue



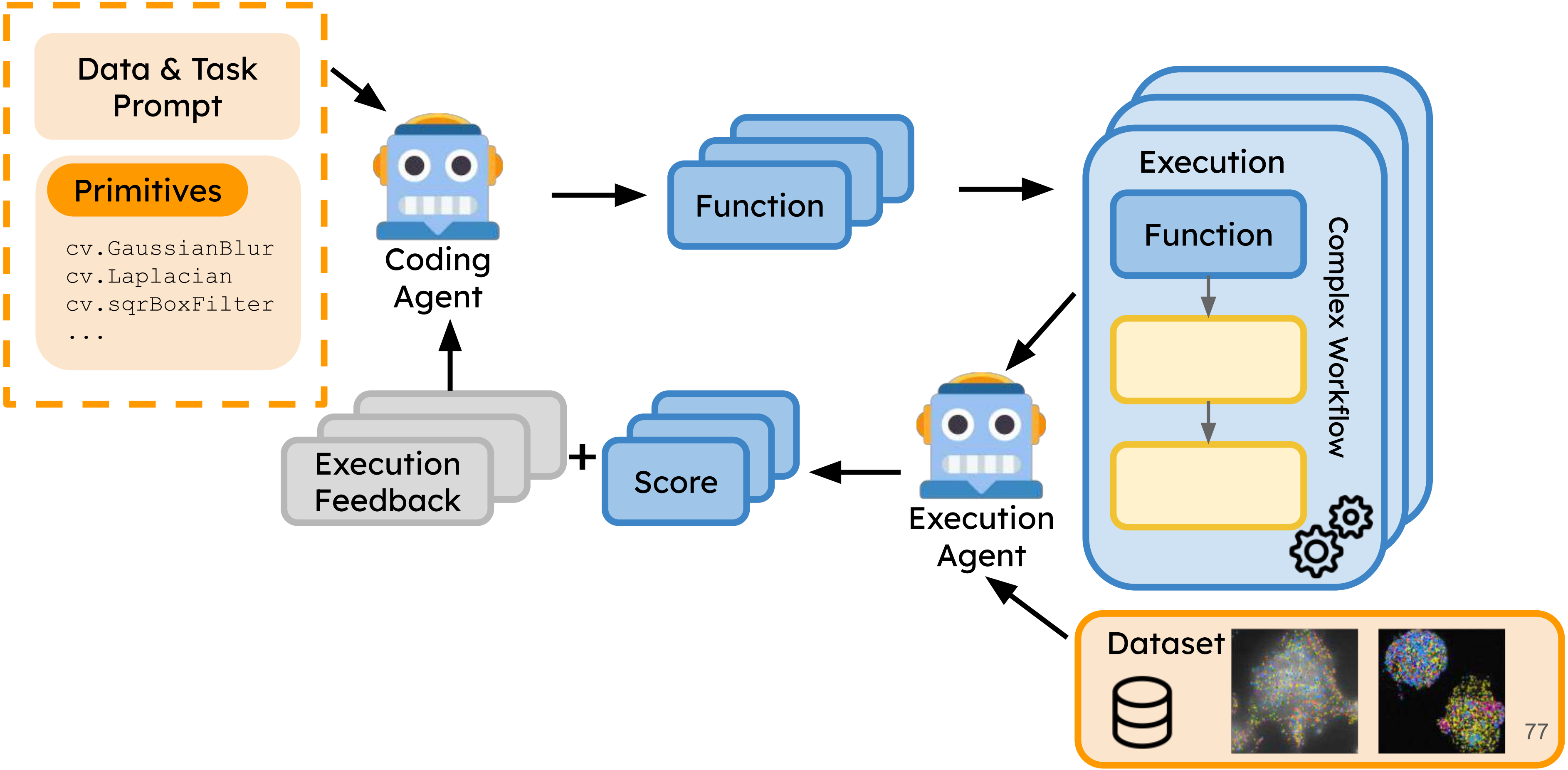




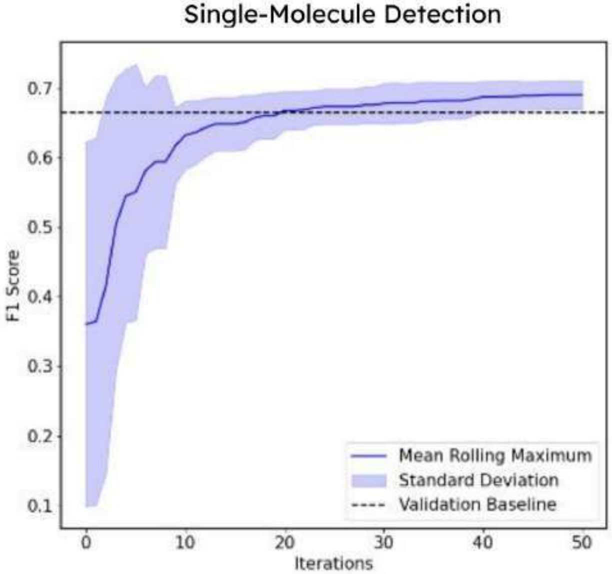
Proof-of-concept Agent System

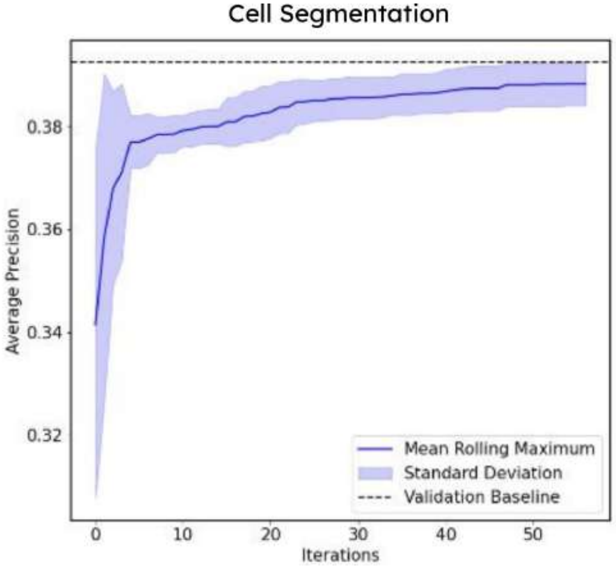


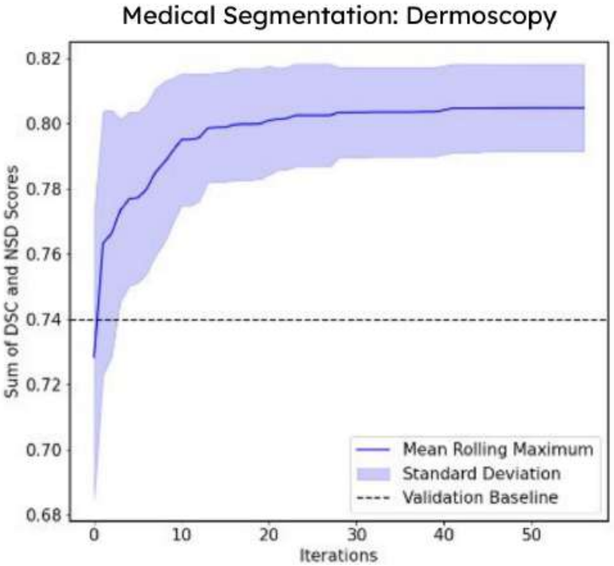
Proof-of-concept Agent System



Agentic Superoptimization Results

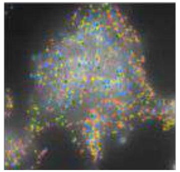
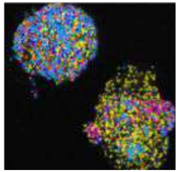






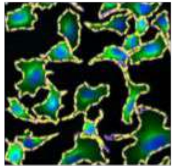
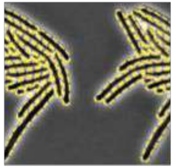
DeepCell Spots

build no status | docs failing | coverage 77% | license Modified Apache 2 | pypi package 0.4.2 | downloads 33/month
python 3.7 | 3.8 | 3.9

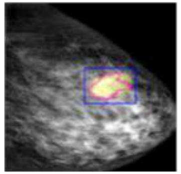
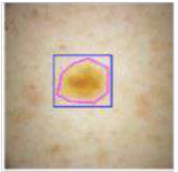
Cellpose

docs passing | tests passing | codecov 47% | pypi package 4.0.4 | downloads 1M
downloads/month 39k | python 3 | license BSD-3-Clause | contributors 61 | website up
forum 543 topics | repo size 116.7 MiB | Stars 1.7k | Forks 470

MedSAM

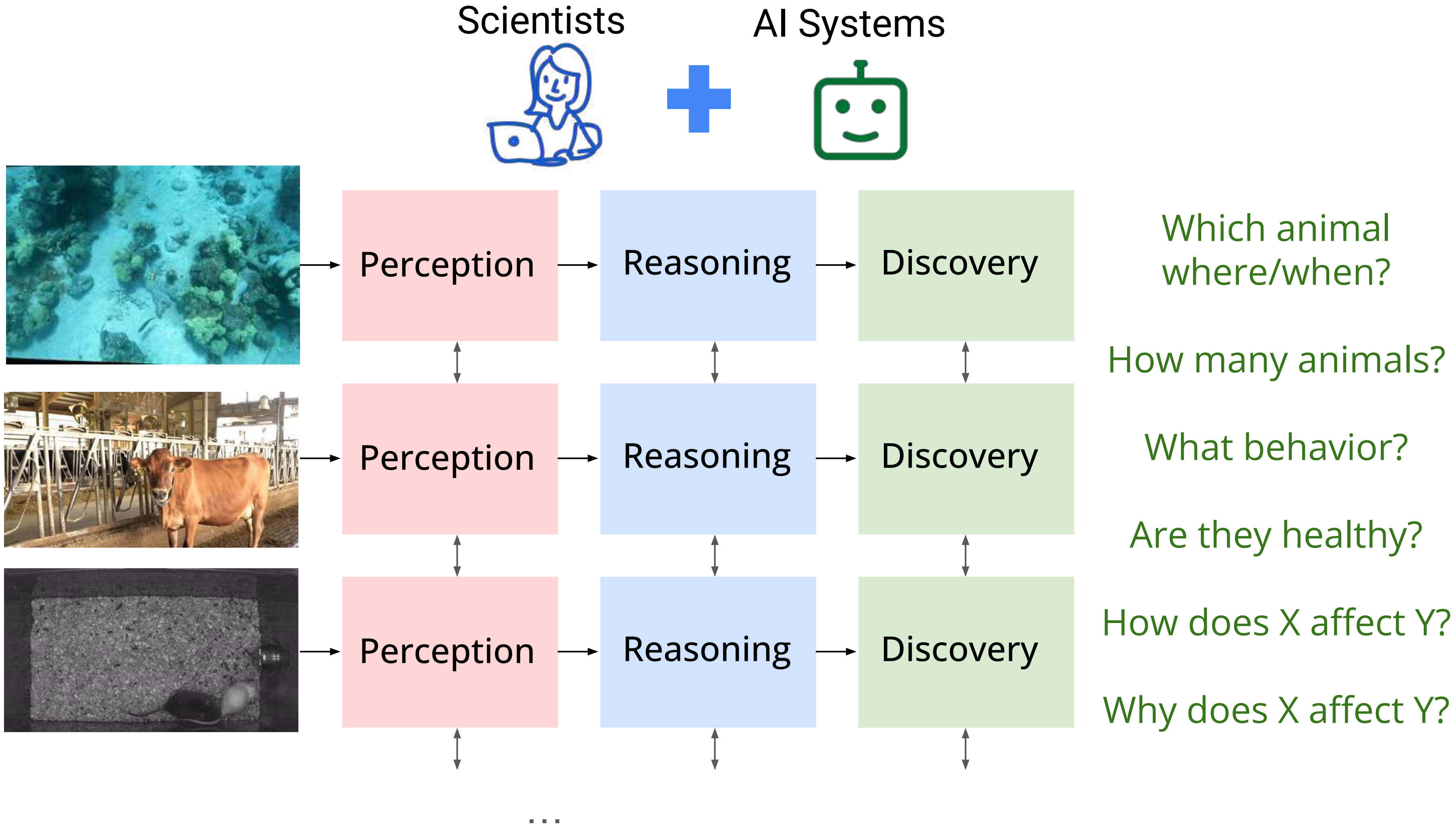
This is the official repository for MedSAM: Segment Anything in Medical Images.

Perception & Reasoning

- Why is it so hard to build effective scientific workflows?
- Instead of manual effort, can we have an AI agent discover an optimal workflow for us?
- Can we accelerate the discovery process?

Discovery

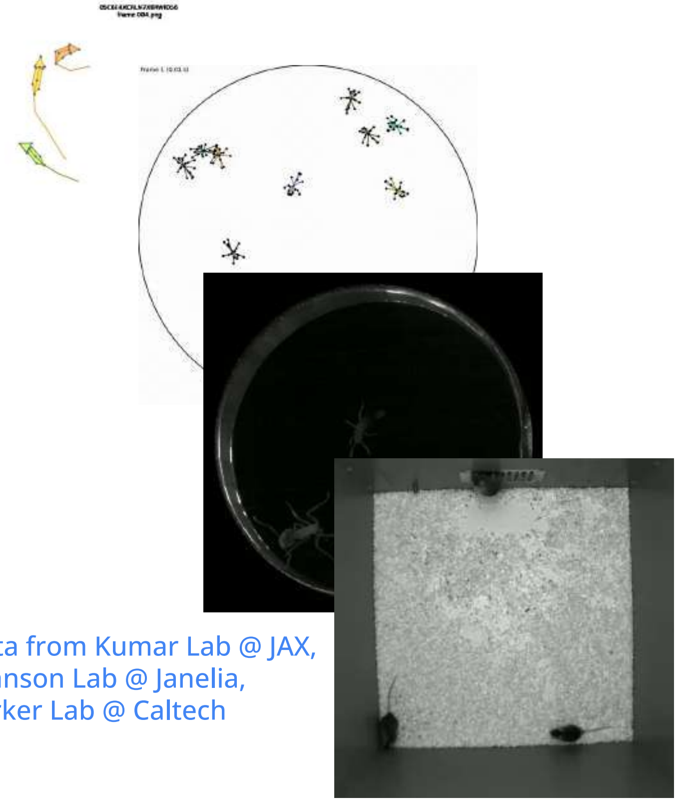

Scientists
AI Systems
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
Perception
Reasoning
Discovery
...
Which animal where/when?
How many animals?
What behavior?
Are they healthy?
How does X affect Y?
Why does X affect Y?

Call to Action

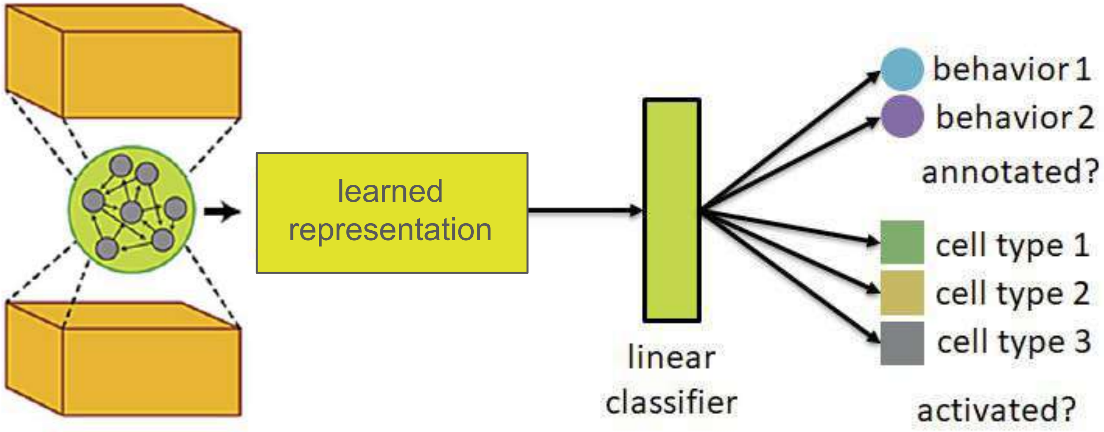
- Representative datasets & benchmarks

Benchmarking Animal Behavior (in the lab)

Ann
Kennedy



Data from Kumar Lab @ JAX,
Branson Lab @ Janelia,
Parker Lab @ Caltech



secret tasks

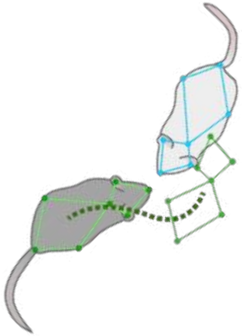
MABe22: A Multi-Species Multi-Task Benchmark for Learned Representations of Behavior, Sun, …, Branson, Kennedy, 2023

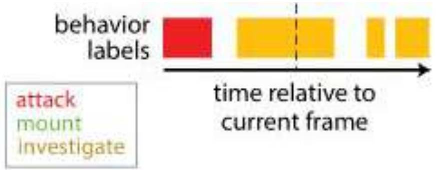
Call to Action

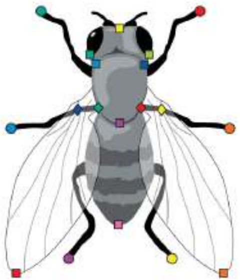
- Representative datasets & benchmarks
- Quantifying discovery

Accuracy Problem

Given a way to measure success,
I want to get the number as high as possible

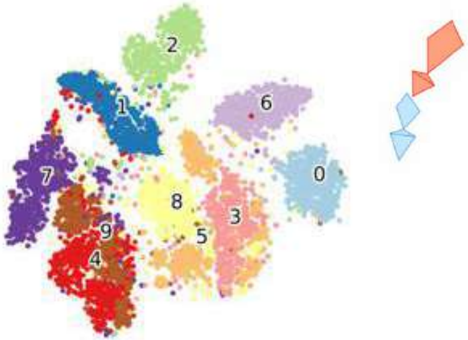






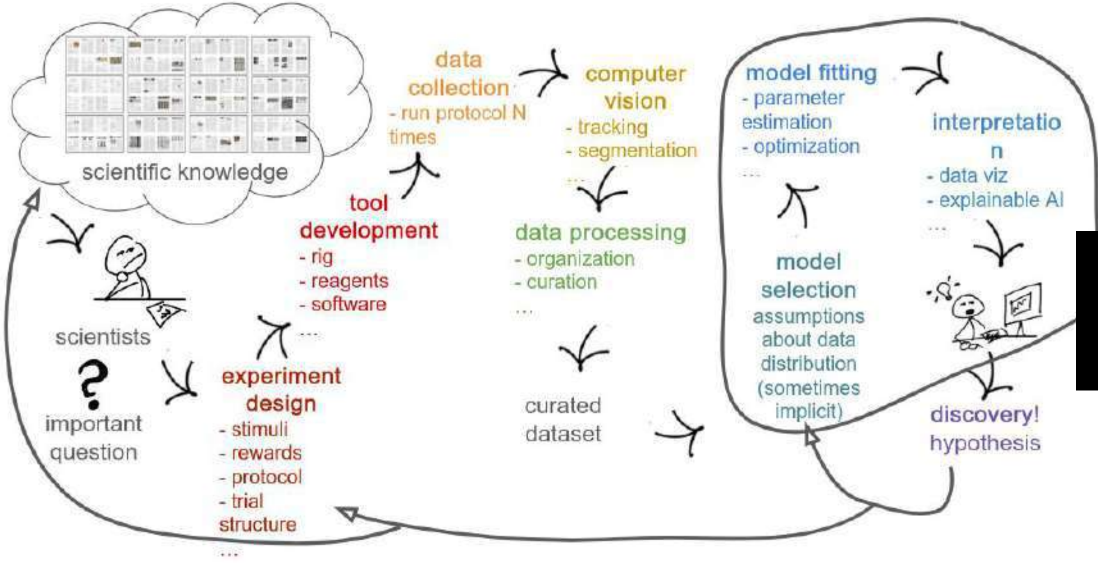
Discovery Problem

I want the model to lead to new & true insights
(typically hard to measure)



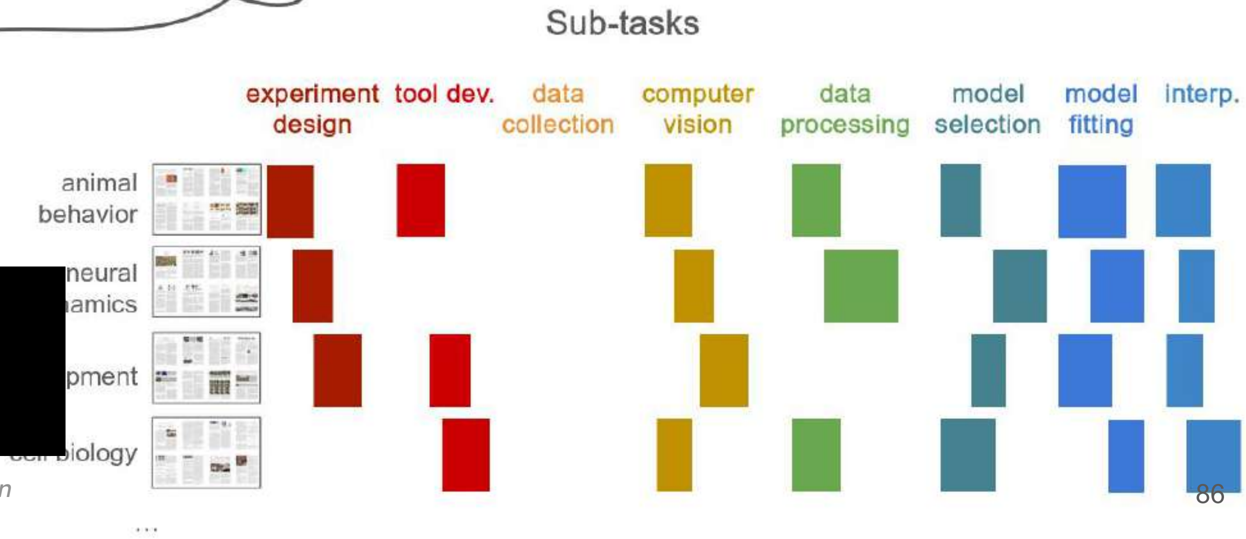
Call to Action

- Representative datasets & benchmarks
- Quantifying discovery
- Collaborations across fields



Kristin Branson

Which subtask do we want AI to automate most?



- Learn to tune the tools
- Learn to standardize data
- Learn to use new tools

Adapted from our joint talk at *Grounding Cognition in Mechanistic Insight Conference*, Janelia, May 2025

Acknowledgements


Andrew Hein


Atharva Sehgal


Kilian Weinberger


Yoav Artzi


Julie Wang


Ling-Wei Kong


Abby Grassick

CORNELL UNIVERSITY FOUNDED A.D. 1865

Google

Caltech

FDA


Yisong Yue


Ann Kennedy


Swarat Chaudhuri


David J. Anderson


Kristin Branson


Pietro Perona

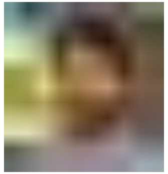

Tomomi Karigo


Cristina Segalin


Brady Weissbourd


Ting Liu


Jonathan Chen


Alex Farhang


Sophia Stiles


Kai Horstmann


Renata Ivanek


Linxi Zhao


Yijia Dai


Xinyu Yang