

---

# AutoML CI/CD/CT: Continuous Training and Deployment Pipeline

**Organization:** Bayes Studio Inc.

**Mentor:** Hedayat Zarkoob

**Team Members:** Sepehr Heydarian, Tien Nguyen,  
Elshaday Yoseph, Archer Liu



---

# Our Team



Sepehr Heydarian



Archer Liu



Elshaday Yoseph



Tien Nguyen

# Summary

01

## Introduction

- Wildfires! ----- 5
- Who are Bayes Studio ----- 6
- The problem ----- 7
- Our solution ----- 8
- Why is it important ----- 9

02

## Our Product

- Product High Level Overview ----- 11
- Pipeline breakdown & challenges ----- 12-20

03

## Timeline

- Weekly plan of pipeline development and delivery ----- 22



---

**01**

# Introduction





# Our Partner

- A start-up based in Vancouver focusing on advanced AI solutions to tackle environmental disasters
- Develop wildfire detection AI tools and monitoring tools
- Harness the power of artificial intelligence for proactive disaster management.



# The Problem

- Over 5 million images, growing rapidly
- Manual labelling and retraining is slow and inefficient
- Delays in updating the model result in delays on insights during emergencies
- Require an automated, scalable pipeline





# Our Solution

- Create automated pipeline
  - Pre-labeling
  - Human-in-the-loop review
  - continuous training and deployment
- Utilize pre-trained models
- Utilize open source tools



# Why is it important

- Faster and more accurate wildfire detection
- Saves time, cost, and human effort
- Scalability results in better services for the government and other agencies
- Strengthens Bayes Studio's missions and market position

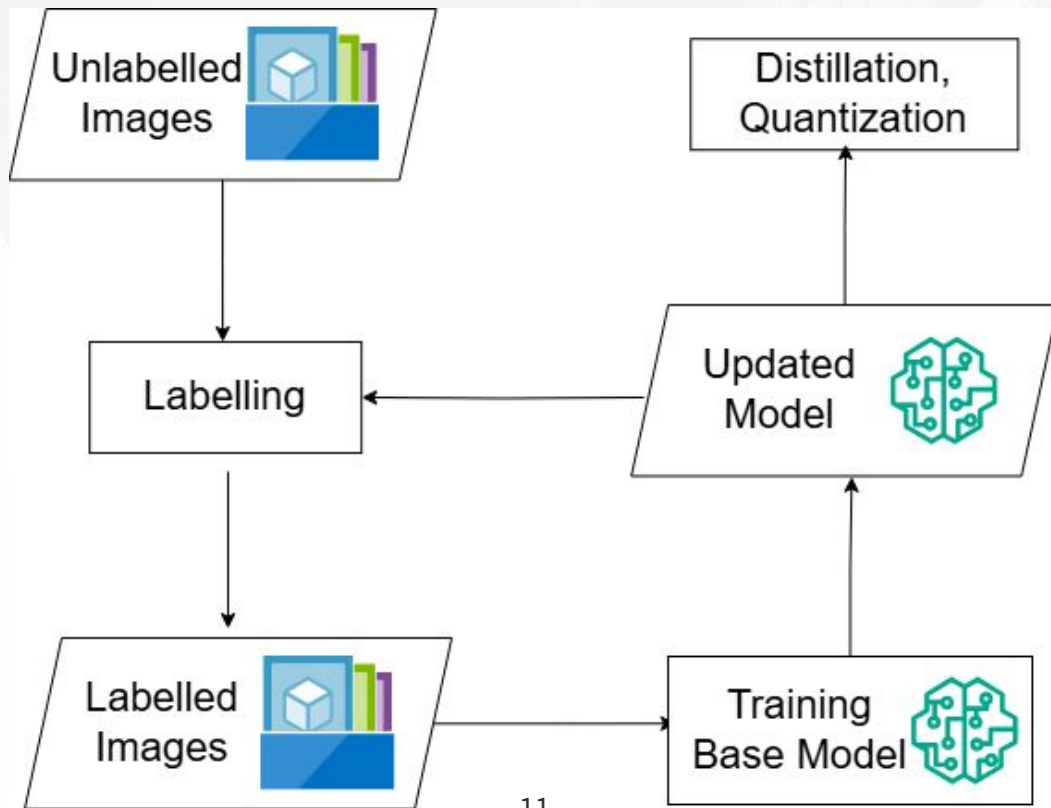


---

**02**

# **Our Product**

# Product High Level Overview



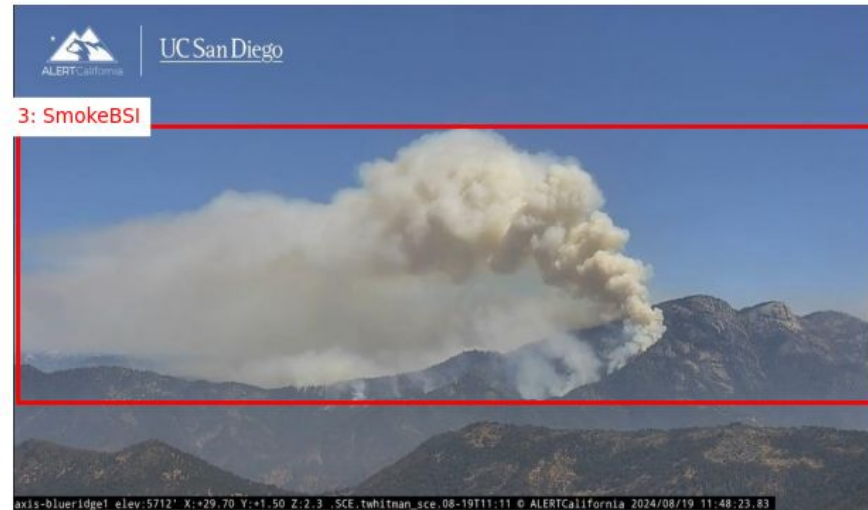
---

# The Input to our Pipeline ~ The Data

- Image data from Roboflow.com + manual collection
- 3M+ labeled images (6M+ annotations), 2M+ unlabeled
- 5 object detection classes: **Fire, Smoke, Lightning, Human, Vehicle**
- ~ 500 new unlabelled images monthly



# Labelling example

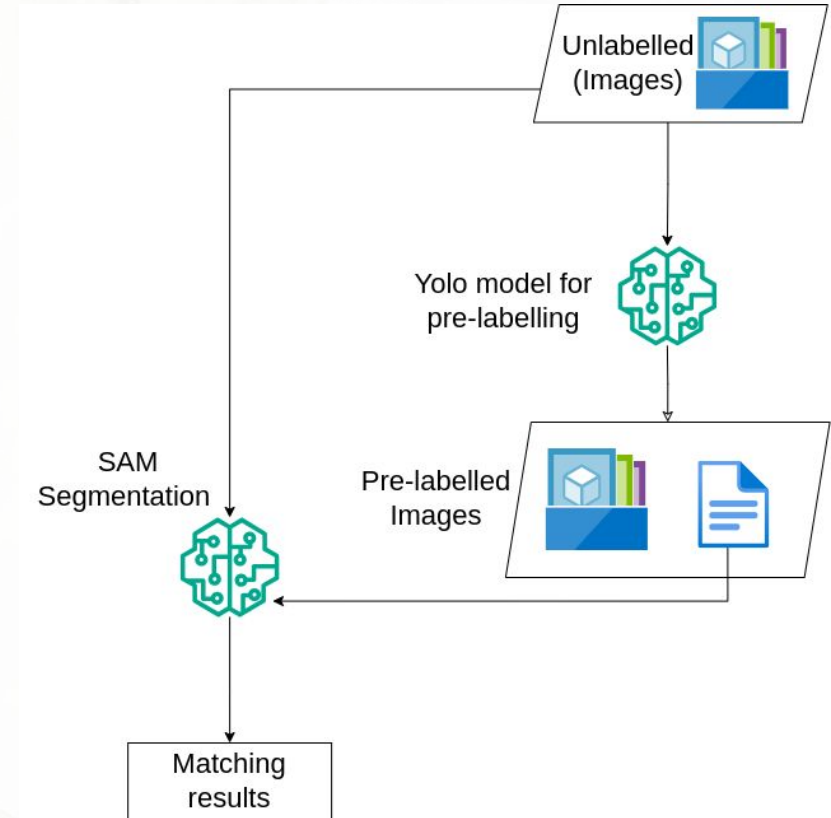


3 0.5024765625 0.5107361111111112 0.9950468750000001 0.5445416666666667

# Pipeline Breakdown

## Labelling

- **Input:** Unlabelled images
- **Process:**
  - YOLO → boxes + labels
  - SAM → segmentation
  - Matching
- **Output:**
  - Labelled or flagged images



# Pipeline Breakdown - Cont.

## Challenge: Matching YOLO and SAM Outputs

- YOLO gives bounding boxes
- SAM gives detailed masks
- But... **what counts as a match?**

Yolo



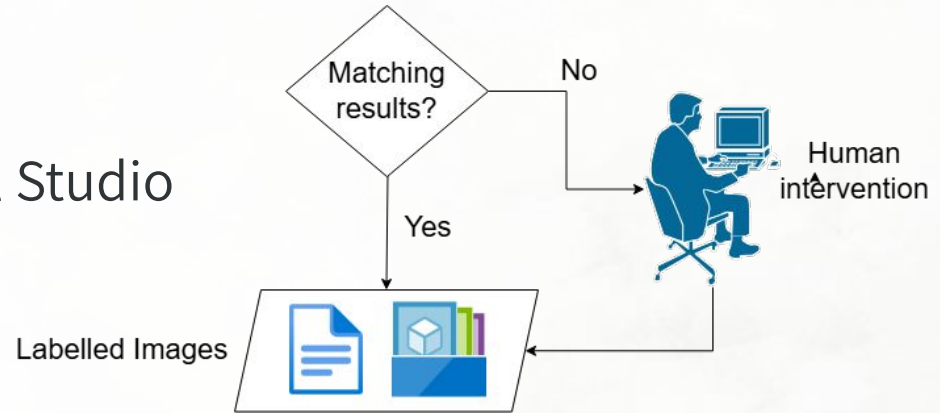
SAM



# Pipeline Breakdown - Cont.

## Human-in-the-loop

- **Input:** Flagged predictions
- **Process:** Reviewed in Label Studio
- **Output:** HQ labelled data





# Pipeline Breakdown - Cont.

## Challenge: Integrating Human Review Interface

- **Open-source interfaces** (Label Studio)
  - **Display:** Disputed image + pre-labels
  - **Task:** Human reviewer validates labels

**How to integrate Label Studio into our product without disrupting the pipeline?**

Please verify the unmatched label(s)



YOLO-predicted labels: Vehicle

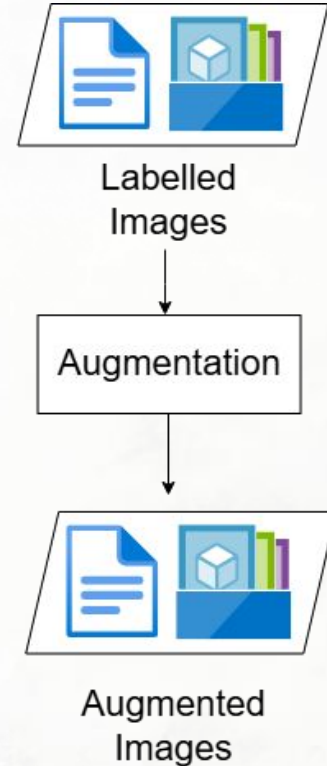
Fire<sup>[1]</sup> Smoke<sup>[2]</sup> Person<sup>[3]</sup> Lightning<sup>[4]</sup> **Vehicle<sup>[5]</sup>**

Reference: Tkachenko, M., Malyuk, M., Holmanyuk, A., & Liubimov, N. (2020–2025). Label Studio: Data labeling software. GitHub

# Pipeline Breakdown - Cont.

## Augmentation

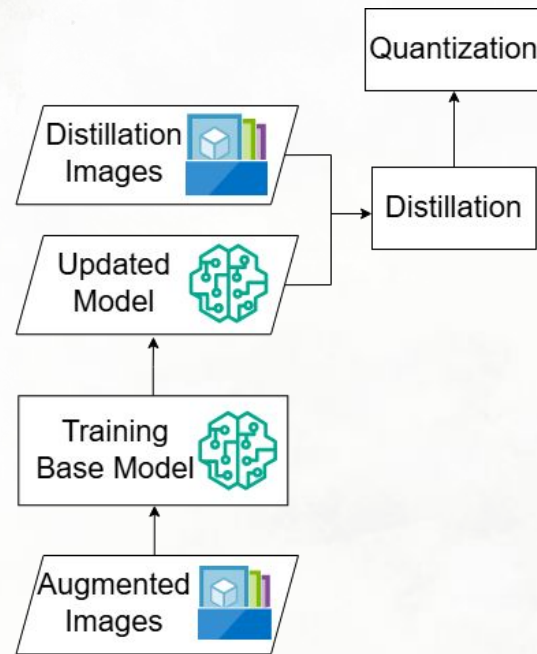
- **Input:** Labelled images
- **Process:** Apply flips, brightness, noise, etc.
- **Output:** Augmented dataset



# Pipeline Breakdown - Cont.

## Training, Distillation, Quantization

- **Input:** Final dataset + Distillation subset
- **Process:**
  - Train base model
  - Distill & quantize for deployment
- **Output:** Lightweight, deployable model



# Pipeline Breakdown - Cont.

## Challenge: Pipeline Complexity

- **System:** End-to-end pipeline
  - **Flow:** Pre-labeling → Human-in-the-loop → Augmentation  
→ Training → Distillation → Quantization → Registry → ...

**How to connect modular pipeline stages while ensuring robustness ?**





---

03

# Timeline

Weekly plan of pipeline development and delivery

# Timeline

Task	Description	Date
<b>Task 1</b>	Project setup, create the overall pipeline. Submit final proposal report.	May 5 - May 9
<b>Task 2</b>	Integrate pre-labeling + SAM check into the pipeline. Design & implement human review interface.	May 12 - May 15
<b>Task 3</b>	Augmentation. Integrate training into the pipeline for model updates.	May 19 - May 23
<b>Task 4</b>	Integrate distillation and quantization into the pipeline for deployment.	May 26 - May 30
<b>Task 5</b>	Run full pipeline test to ensure the pipeline runs successfully.	June 2 - June 6
<b>Task 6</b>	Submit runnable data product. Prepare and deliver presentation.	June 9 - June 11
<b>Task 7</b>	Finalize data product and written report based on feedback.	June 14 - June 25

# Thanks!

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)