
cisco Live!
Let's go
#CiscoLive

The bridge to possible

Scalable Real-time Actionable Insights From Network Telemetry and Video Data

Hugo Latapie, Principal Engineer

BRKETI-1002

cisco Live!

#CiscoLive

Cisco Webex App

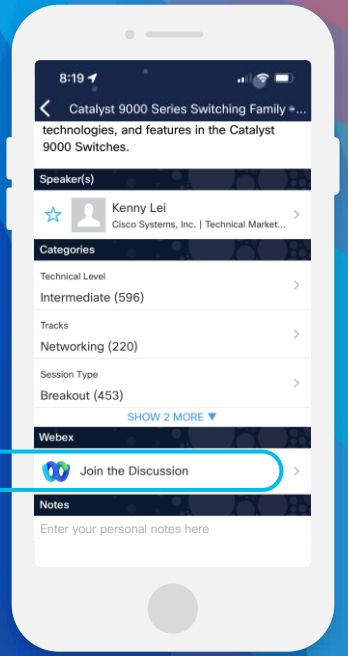
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 9, 2023.



https://ciscolive.ciscoevents.com/ciscolivebot/#BRKETI-1002

Agenda

- Intro to Intent Based Analytics
- Overview of Hybrid AI
- Networking Applications
- Video Applications
- LLM Applications
- Conclusion
- Q&A session

Intent-based Analytics

- An approach to analytics that focuses on understanding and adapting to the goals and objectives of the system's users
- As analytics applications become more complex and varied, tailoring analytics to user intent can lead to more efficient and effective solutions.
- By adapting to user intent, analytics systems can provide more relevant and accurate results, enhancing user experience and decision making

Context-Based Intent Suggestions

- Utilizing the power of responsible hybrid AI, intent-based analytics systems can predict and suggest intents based on the context in which they are being used.
- Examples of context-based intent suggestions:
  - Retail: Customer traffic analysis, product placement optimization, theft prevention
  - Airports: Passenger flow management, luggage tracking, monitoring
  - Education: Classroom engagement analysis, safety monitoring, facility usage optimization
- Benefits
  - Proactive analytics, enhanced user experience, time-saving

Key Componenets and Technologies Involved

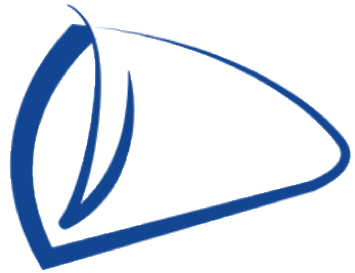
- Hybrid AI
- Multi-variate Time Series Analysis
- Self Supervised Learning
- Computer Vision
- Natural Language Processing
- Large Language Models
- Data fusion

Challenges and Opportunities

- Challenges
  - Developing accurate, reliable, intent recognition
  - Handling diverse user inputs
  - Ensuring privacy and ethical considerations
- Opportunities
  - Personalized analytics solutions, improved user experience, more efficient use of system resources, potential for new applications and market growth

Synergistic University Research

Hyperdimensional Computing – UCI
Self Sustaining Vibration Sensing & AI Through Symbiosis – Georgia Tech
VQPy – UCLA
Freeview-point Rendering from Videos, Photos, and Beyond – University of Washington
3D monocular detection/tracking – University of Washington
Crowd Mood Monitoring through Hybrid Sensing and Learning – University of Michigan
Interactive Video Exploration – University of Washington
Reasoning Engines: OpenNARS, ONA, and AERA – Universities of Temple/Stockholm/Reykjavik
Semantic Slam – University of Texas at Austin
Comparative Analysis of Interval Reachability for Robust Implicit and Neural Networks – Georgia Tech / UCSB



cisco DeepVision

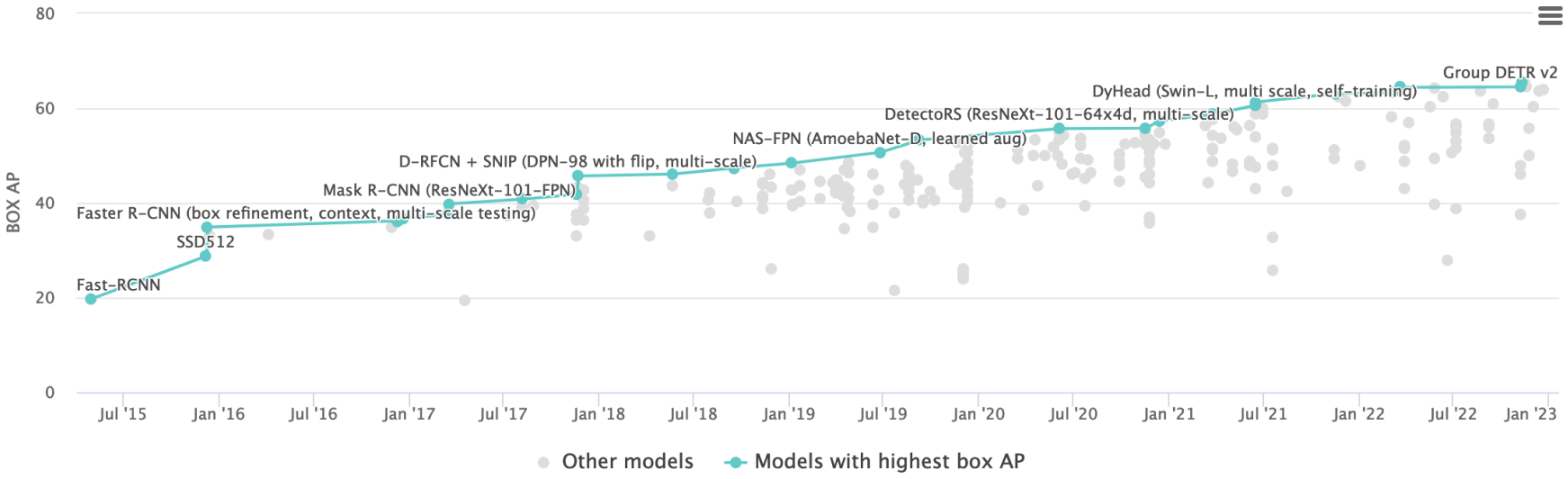
General Problem Space

- Today's AI continues to dramatically improve but we'll show that much remains to be done
- Creating trustworthy and accurate computer vision requires a cross-disciplinary team of experts

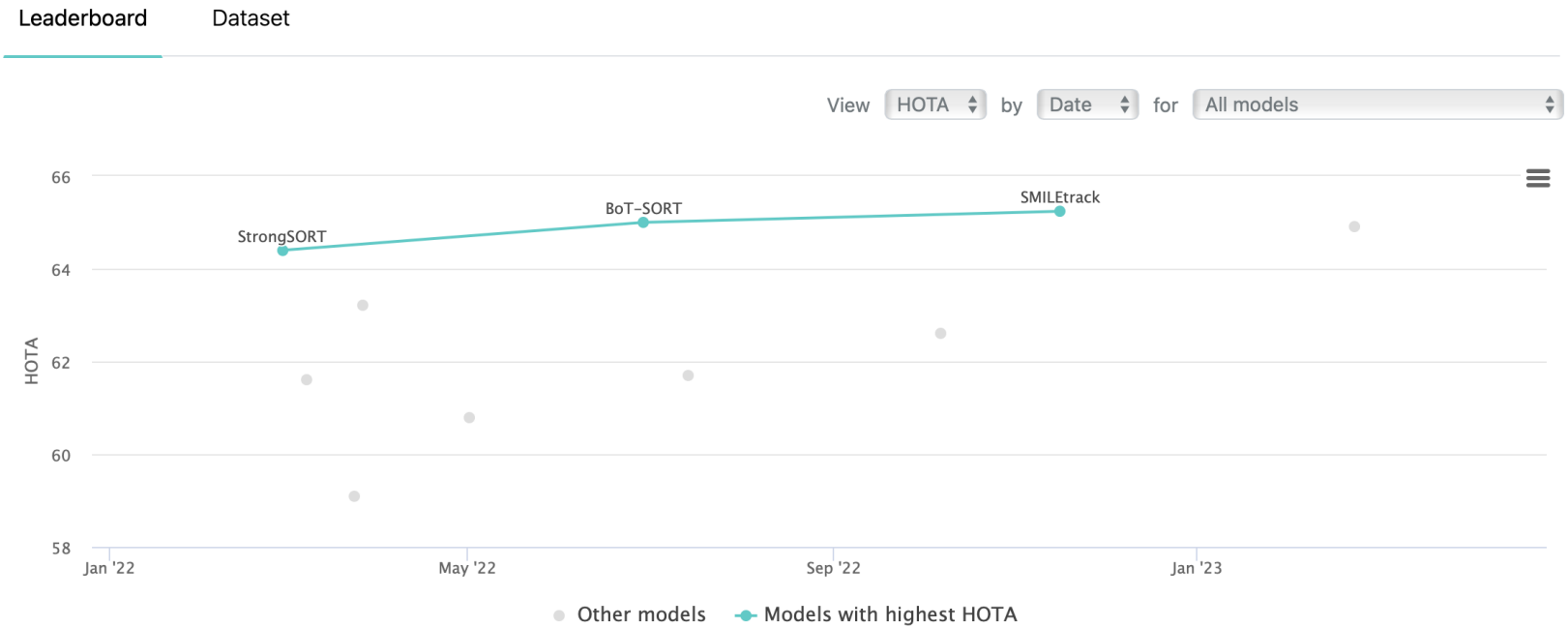
Current state of the art object detection

Leaderboard Dataset

View box AP by Date for All models

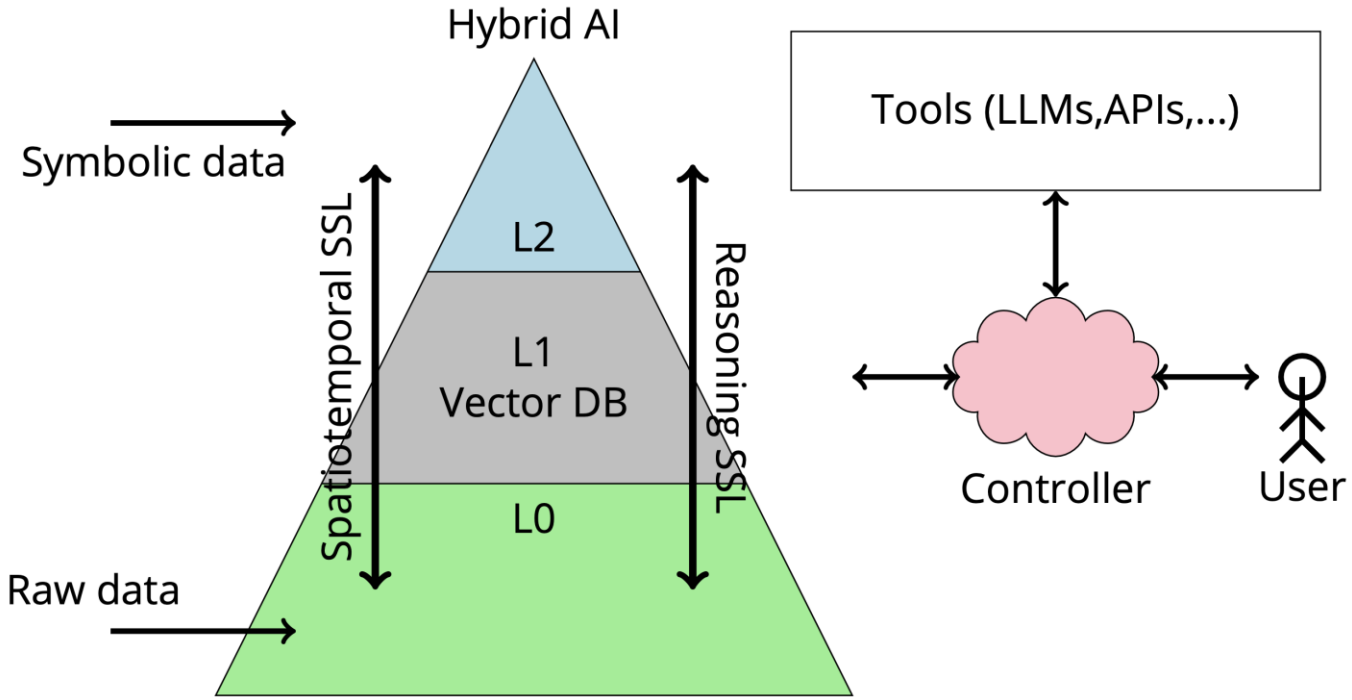


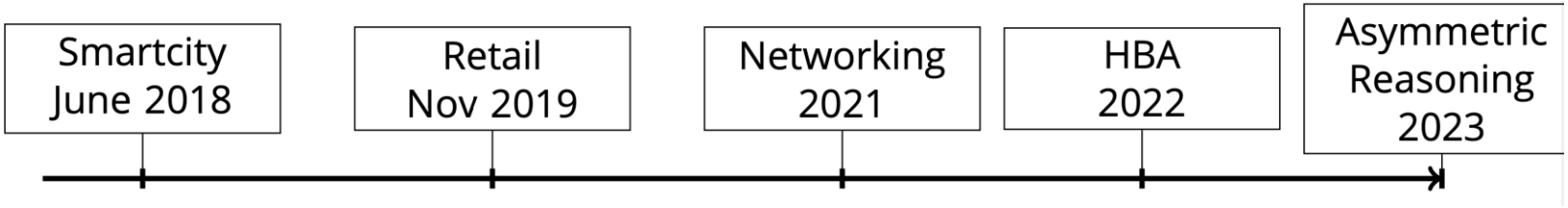
Current state of the art tracking



Question Answering on TruthfulQA

Leaderboard Dataset





cisco DeepVision

Evolution of Hybrid AI

Project	Objective	Results
Deep Fusion (2018)	Detect and map transportation safety levels and events in cities	• Works across camera types and in real-time • Calibrates in 1 minute vs manual camera by camera with weekly readjustment • Accurate: near-miss >90%, jay-walking >90%, speeding >90% (wrong way was unstable)
DFRE4Retail (2019)	Self-supervised auto-identification of products on shelves in Retail	• +Adds ability to detect 1000s of objects on shelves vs 30/40 in v1 • Accurate: >95% accurate
CX Intelligent Edge Video Analytics (2020)	CX 1 year 10-person effort for fully productized on-prem solution for people counting, face rec and license plate rec	• Fully productized version in demand by ~100s leads + ~10 initial deployments – launch tanked by privacy violations • SOTA products (no reasoning): 95% people counting, zone and flow analytics
Alert Storm (2020)	Enhance an existing rule-based approach i.e., mapped alerts to network	• Learning on just alerts is a bad approach – cannot tell if alert is good or bad • Cannot recommend actions without info on what happened before the alert (i.e., garbage in garbage out)
Kronos (2021)	Compressing, identifying, predicting network telemetry.	• +Adds explicit multivariate time series detection • 95% accuracy event extraction and prediction across 300k time series • Requires time series to be from one system with potential for interactions
SensorDog (2022)	Learning Human Behavior, For Sports & Entertainment	• +Adds ability to turn video tracking from DFRE4Retail into time series and take same approach of multivariate time series detection • Predictions off single time series made possible
Deep Vision (2023)	New hybrid AI framework with unique asymmetric reasoning / logic / causality	• +Adds configuration of AI/reasoning elements for optimal results • Moves results from current 65% or less to in the 80s% (live testing in process)

Recent Hybrid AI Publications

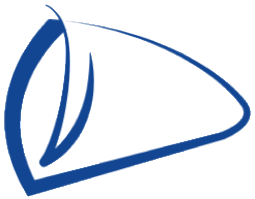
- Addressing the Unsustainability of Deep Neural Networks With Next-Gen AI – AGI 23
- Explicit Goal-Driven Autonomous Self-Explanation - AGI 23
- Comparative Reasoning for Intelligent Agents – AGI 23
- Automated Spatiotemporal Modeling for Real-Time Data-Driven Actionable Insights – IntelliSys 23
- Hybrid AI for IoT Actionable Insights & Real-Time Data-Driven Networks – IWSSL 22

Hybrid AI Mainstream Adoption

- Illusion of false necessity
- Lack of good ground truth datasets for AGI
- Plan is to address existing long standing unsolved challenges that include:
  - Datasets
  - Leaderboards
  - Impact
- We are starting with MOT and Object Detection
- Already have initial MOT results that appear to improve on the state of the art, still undergoing some development before we get into metrics

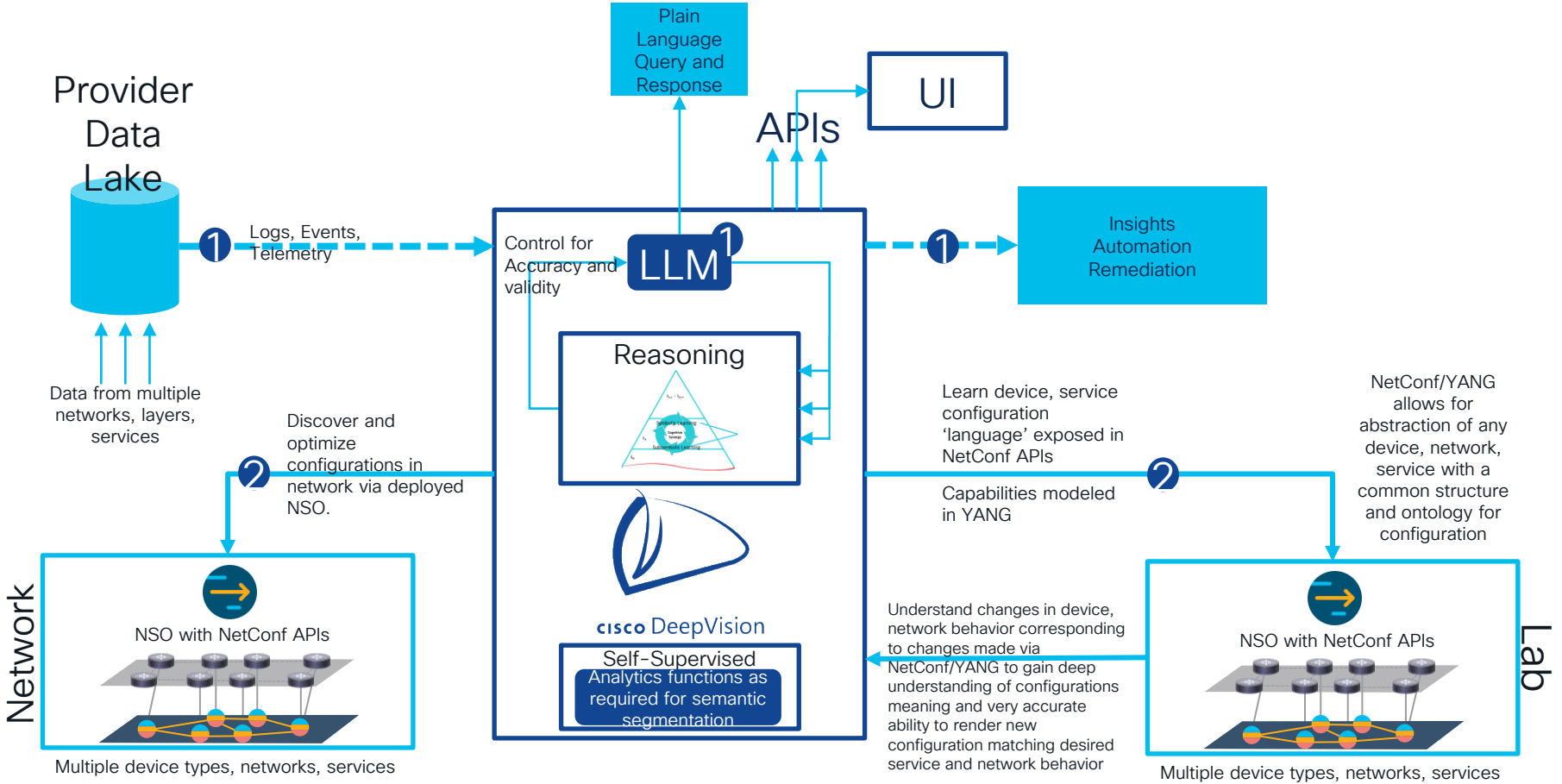
Features

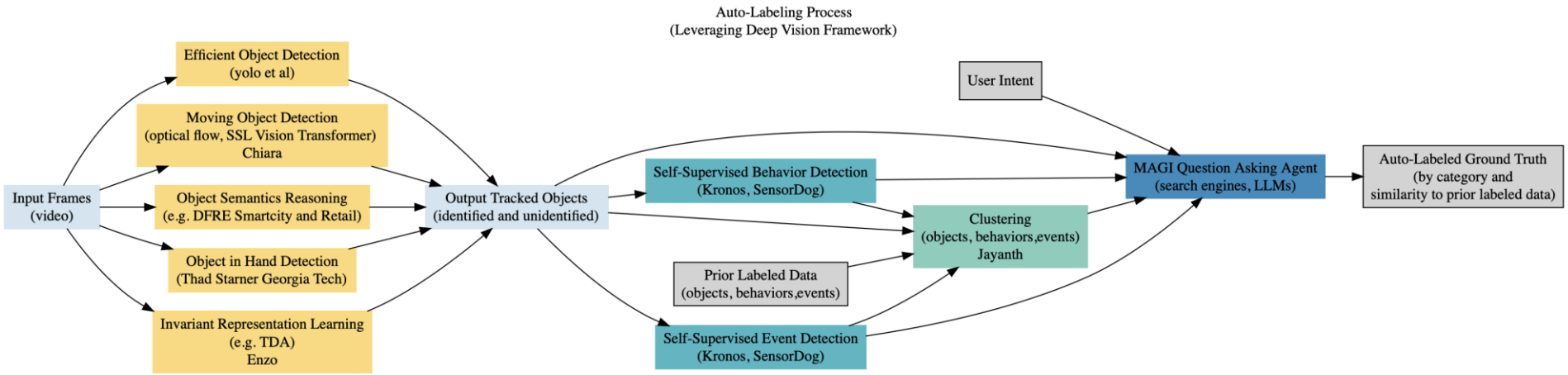
- Scalable and modular serverless open-source framework
- State-of-the-art object detectors, trackers, behavior detectors
- New types of reasoning engines
- 3D semantics projects
- Multi-modal data stream analytics
- Neurosymbolic integration
- Knowledge representation
- Self supervised learning
- Explainable Perception



cisco DeepVision

DeepVision SP Network Trial Phase 2





cisco DeepVision

Demo:
Hybrid AI Networking

Demo:
Hybrid AI Multi-Object Tracking

Demo:
Hybrid AI LLM

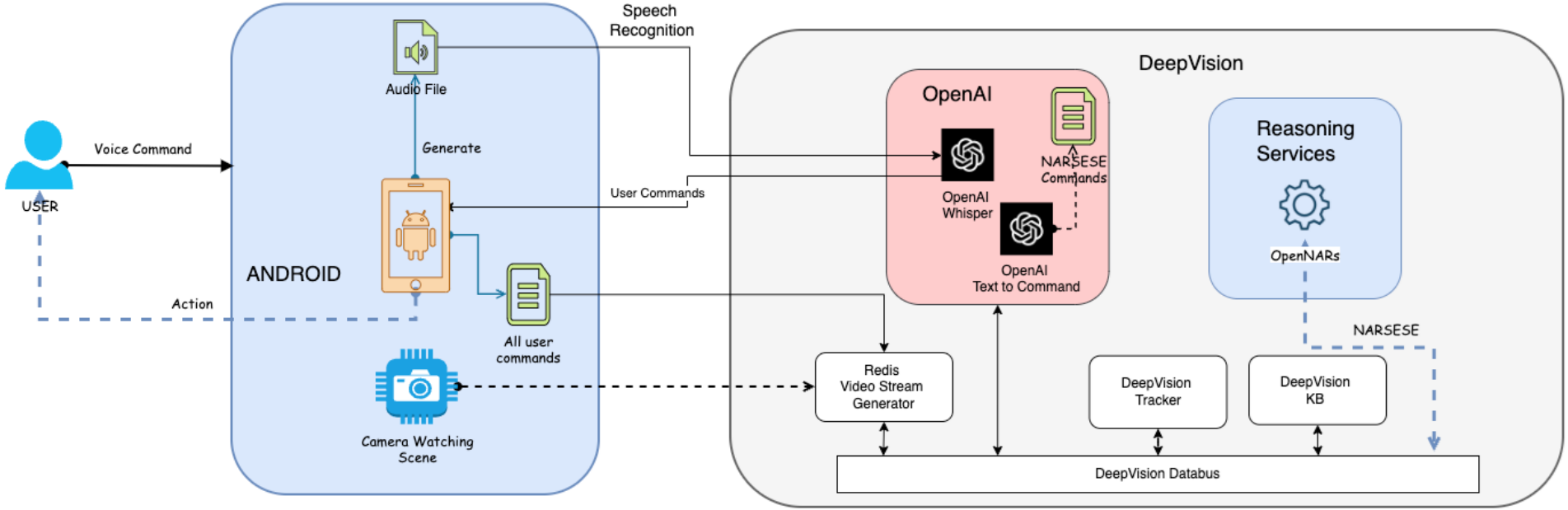
Demo:
Deep Vision Android App

Android Hybrid AI App

- Is an Android-based application that uses Computer-Vision, Speech-Recognition, and Non-Axiomatic Reasoning System* that mimic human brain capabilities by recognizing visual information, comparing them to predefined evidence of arguments, and thus supporting a conclusion.

- Major Components:
  - Android Platform.
  - DeepVision: Object detection and Recognition. (Tracking)
  - Speech Recognition: OpenAI Whisper (Speech to Text), OpenAI GPT-3 (Structured Text to Command)
  - OpenNARS for Application: Reasoner

Application Architecture



OpenNARS Reasoning Engine

- NARS (Non-Axiomatic Reasoning System) is an AGI (Artificial General Intelligence) system originated and developed by Dr. Pei Wang* for more than three decades.
- NARS is:
  - Finite: in time and space.
  - Real-time: new tasks can come and demand action at anytime, even if the system is busy.
  - Open: New premises comes and can always change the system beliefs.

Wang's definition of intelligence *

"Intelligence is the ability for a system to adapt to its environment while working with insufficient knowledge and resources"

OpenNARS for Application

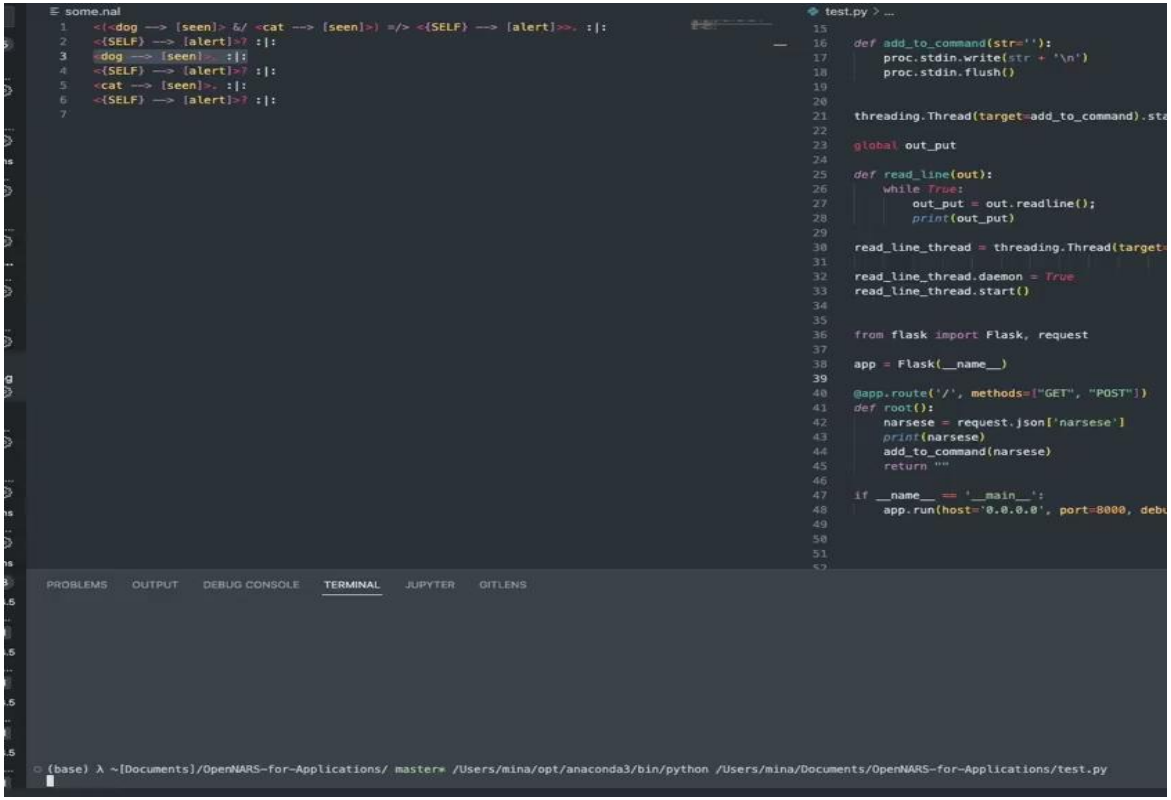
- **Non-Axiomatic Logic**: is a term logic with compound terms, supporting statements like:
  - **Inheritance**: you can say that *cat is an animal*: <cat → animal>
  - **Support set composition**: you can say *garfield is a cat, cats are animal*: <{garfield} → cat>, <cat → animal>
  - **Express Relationship**: you can express the sentence, *cats eat mice:* <(cat * mouse) → eat>.
  - **Temporal Relationship:** when we say that a leads to b: <a ⇒ b>

- Semantic Inference Example:
  - **Truth Deduction**: if term S and M is a P, we can conclude that S is a P
    - {(S → M), (M →P)} ⊢ (S → P)
  - **Truth Abduction**: if term A is B and A is C, we can conclude that C is a B
    - {(A → B), (A →C)} ⊢ (C → B)
  - **Truth Induction**: if term A is C and B is C, we can conclude that B is a A
    - {(A → C), (B →C)} ⊢ (B → A)

Fire Detection Full Example:

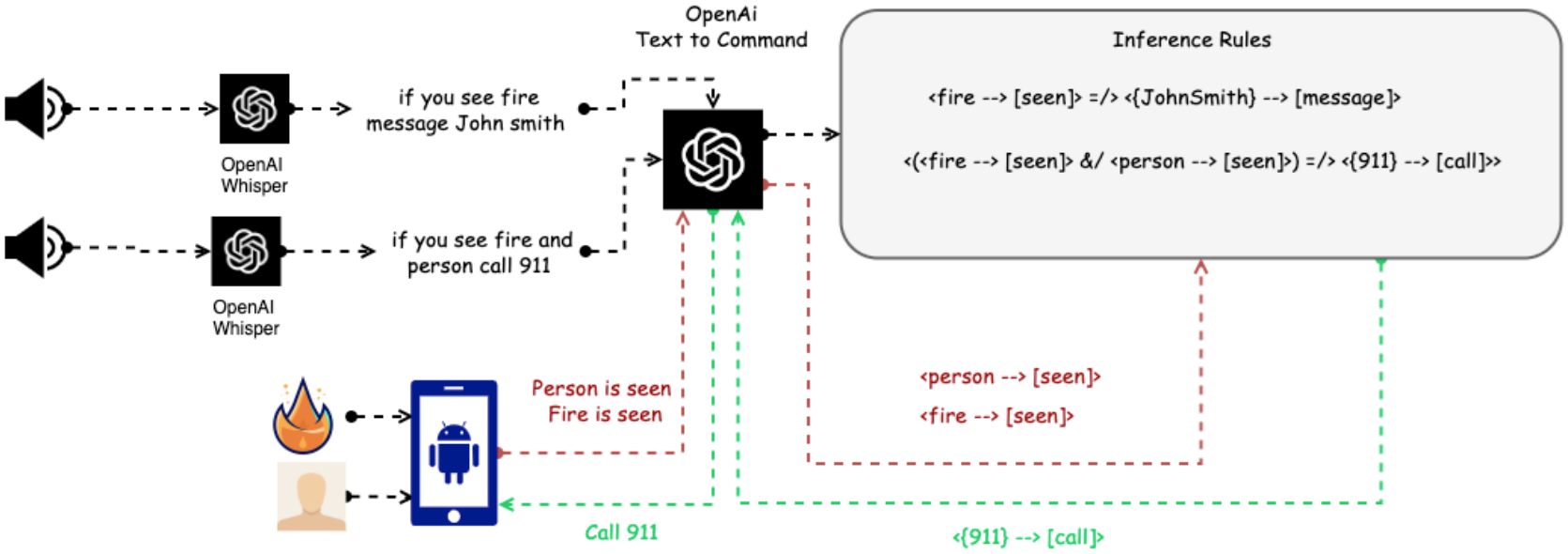
```
<<fire --> [seen]> =/> <{John Smith} --> [message]>>.
<(<fire --> [seen]> &/ <person --> [seen]>) =/> <{911} --> [call]>>.

<{John Smith} --> [message]>? :|:
//Answer: None.

<fire --> [seen]>. :|:
<{John Smith} --> [message]>? :|:
//Answer: <{John Smith} --> [message]>.

<{911} --> [call]>? :|:
//Answer: None.

<person --> [seen]>. :|:
<{911} --> [call]>? :|:
//Answer: <{911} --> [call]>.
```

OpenNARS in action



USE CASE

- If you see fire message John Smith
- If you see fire and person call 911

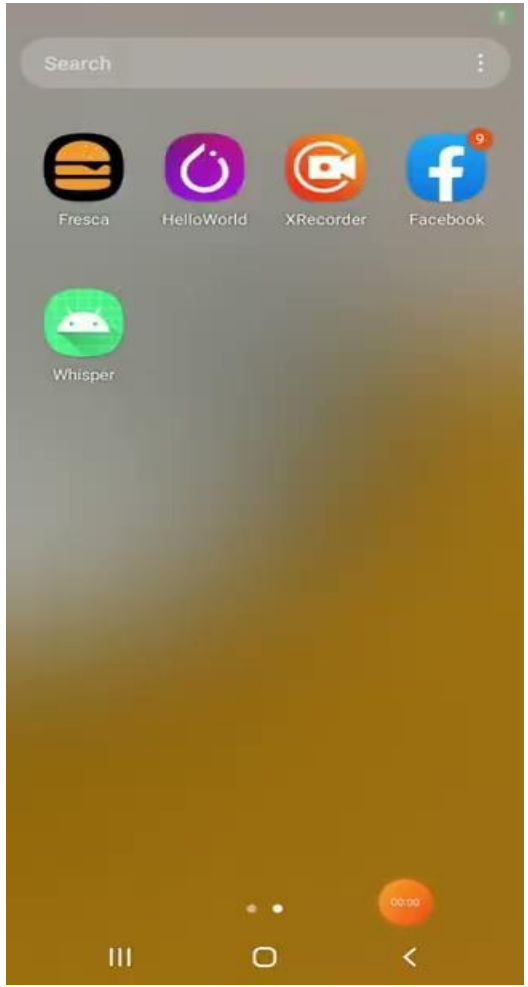


OpenAI Whisper

- Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web.
- Whisper is open-source model.
- Very robust to accents, background noise and technical languages.
- Whisper approaches human level robustness and accuracy on English speech recognition.

**Robust Speech Recognition via Large-Scale Weak Supervision. CoRR abs/2212.04356 (2022)*

Demo



Demo:
VQPy – Deep Vision

VQPy – An Object Oriented Language for Video Analytics

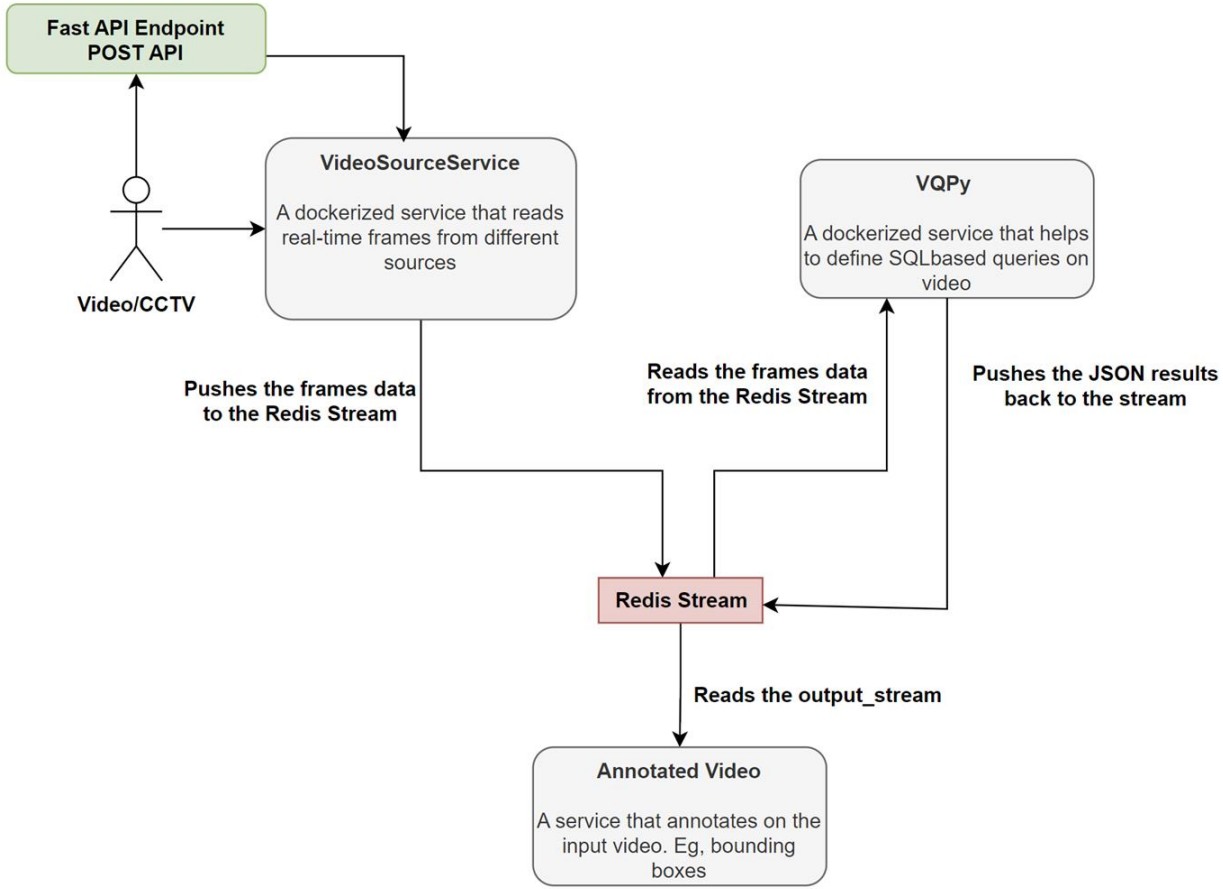
```
class Car(vqpy.VObjBase):

    @vqpy.property()
    @vqpy.stateful(2)
    def coordinate(self):

    @vqpy.property()
    def velocity(self):
```

VObj Class

```
class FindSpeedingCar(vqpy.QueryBase):

    @staticmethod
    def setting():
            filter_cons = {“__class__”: Car,
                           “velocity”: lambda
x: x > 50}
            select_cons = {“velocity”}
            return vqpy.Constraint(
                           select_cons,
                           filter_cons)
```

Query Class

Workflow



VQPy Demo – People Loitering

Region of Interest

Person of Interest

Sample JSON responses

Server response

Code	Details
200	Response body

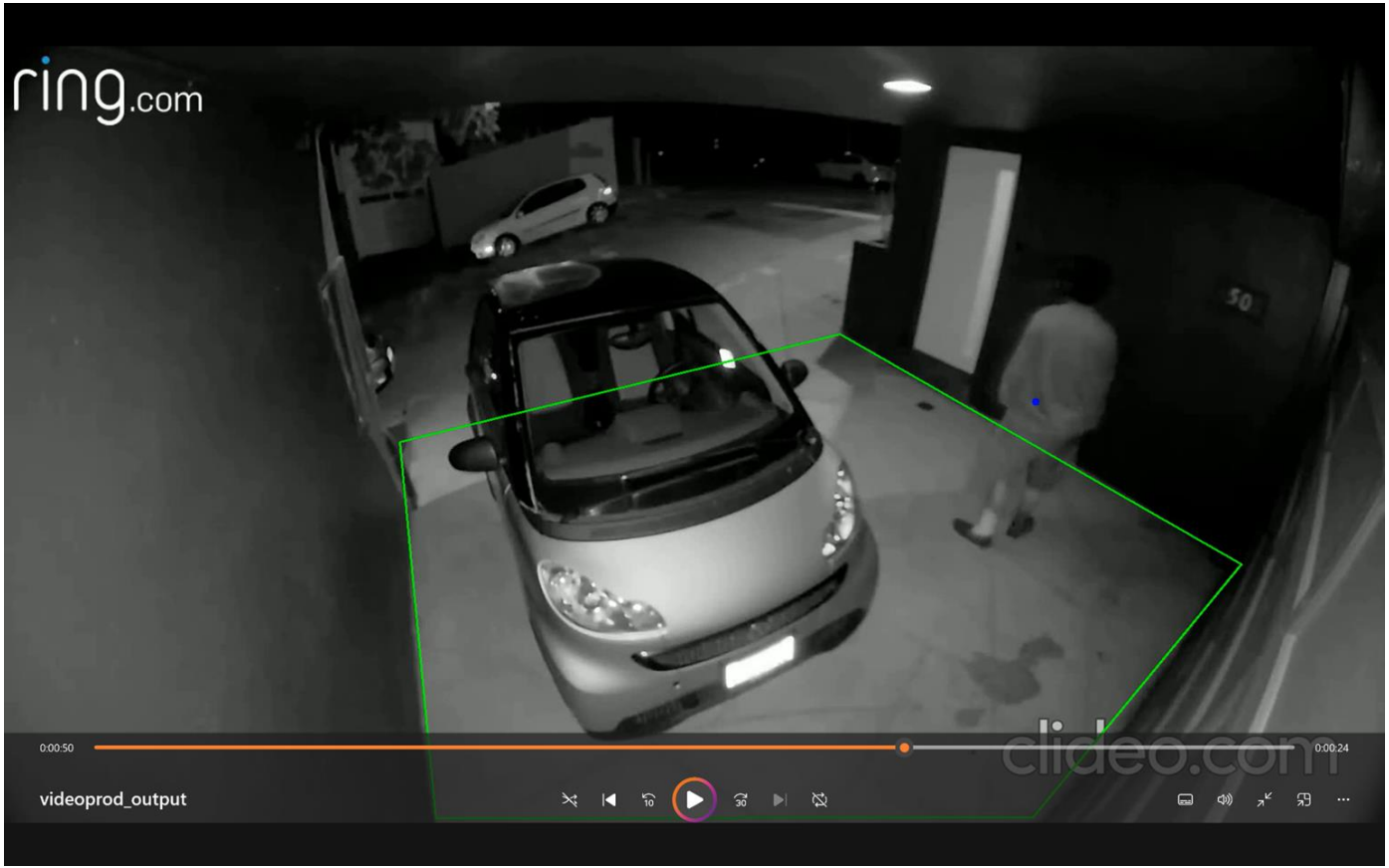

```
[
  {
    "frame_id": 295,
    "data": [
      {
        "track_id": 1,
        "coordinate": "[1386.    417.75]",
        "in_roi_periods": [
          [
            0,
            10
          ]
        ]
      }
    ]
  },
  {
    "frame_id": 296,
    "data": [
      {
        "track_id": 1,
        "coordinate": "[1386.    417.75]",
        "in_roi_periods": [
          [
            0,
            10
          ]
```


Download

Response headers

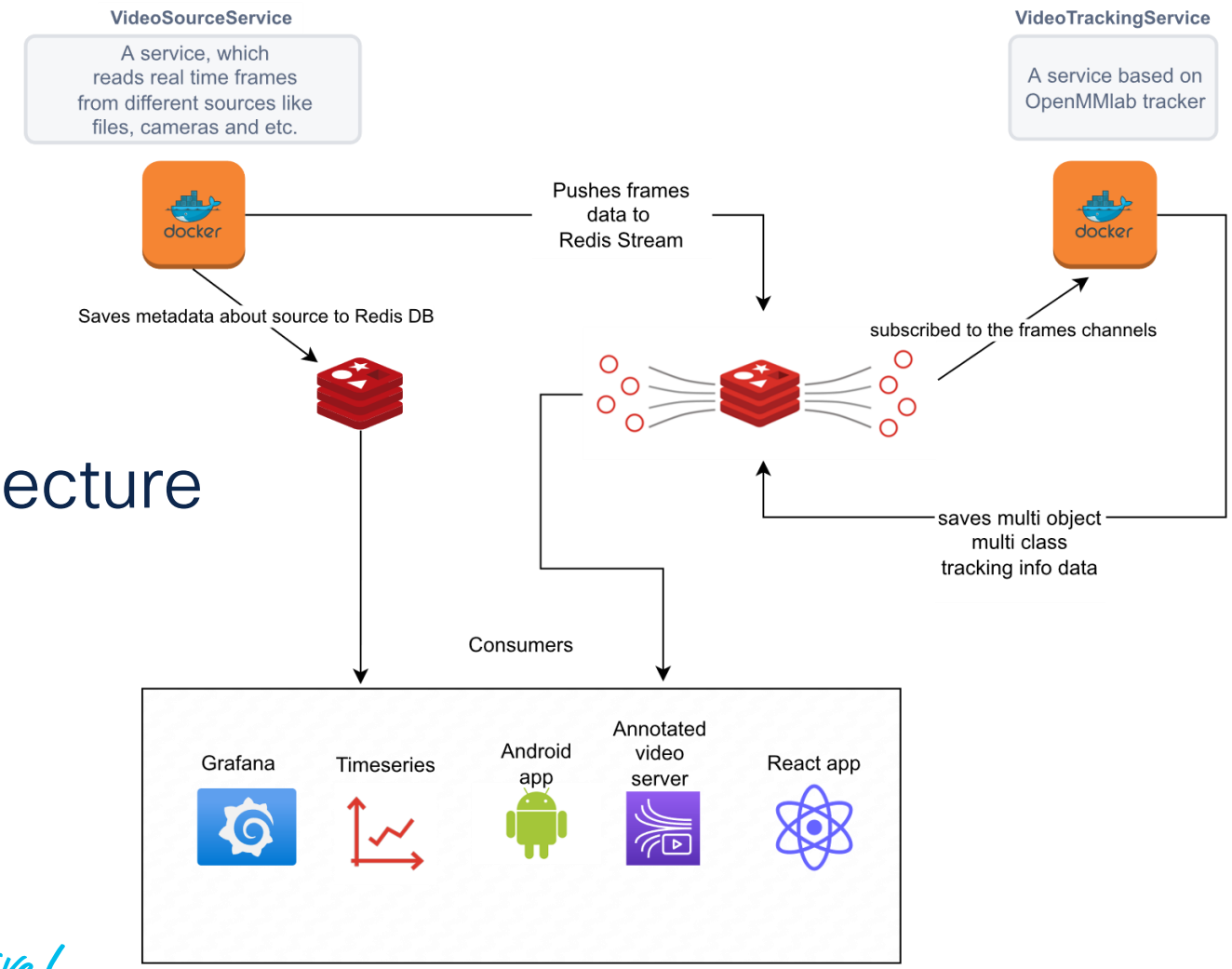
```
content-length: 126287
content-type: application/json
date: Thu,02 Mar 2023 03:42:02 GMT
ngrok-trace-id: f0e4f2d6ce57041f673222bdd0c557b0
server: uvicorn
```

Annotated Video



Demo:
Deep Vision

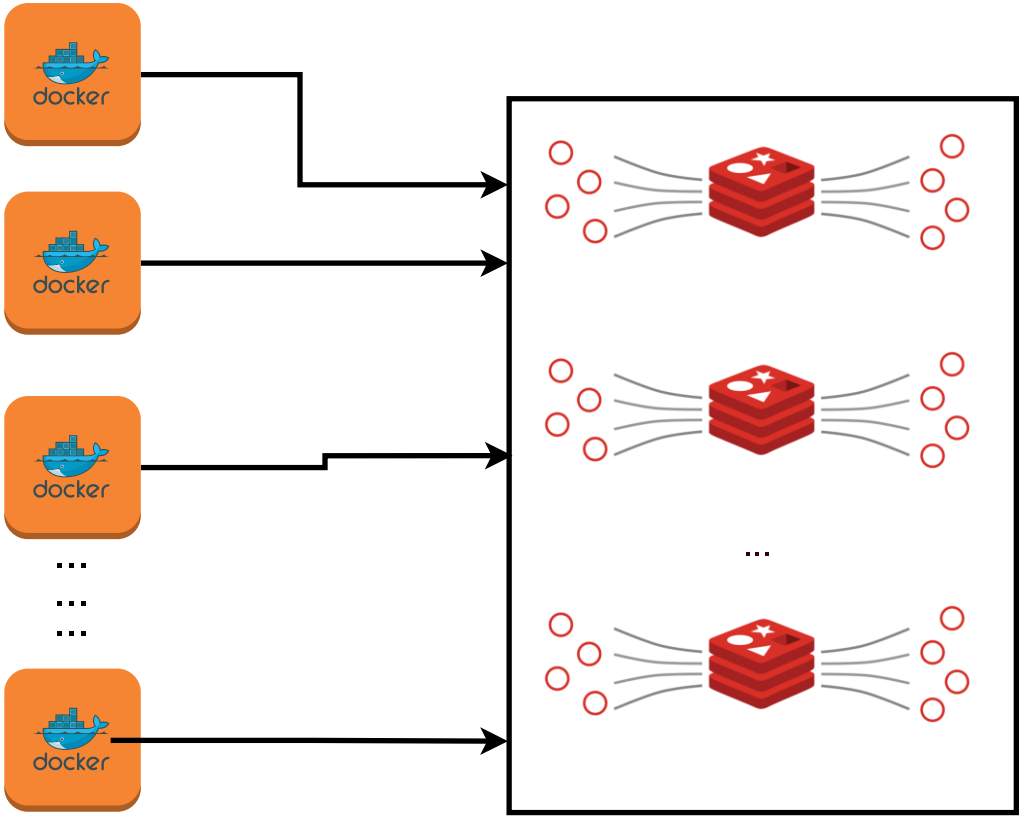
Architecture



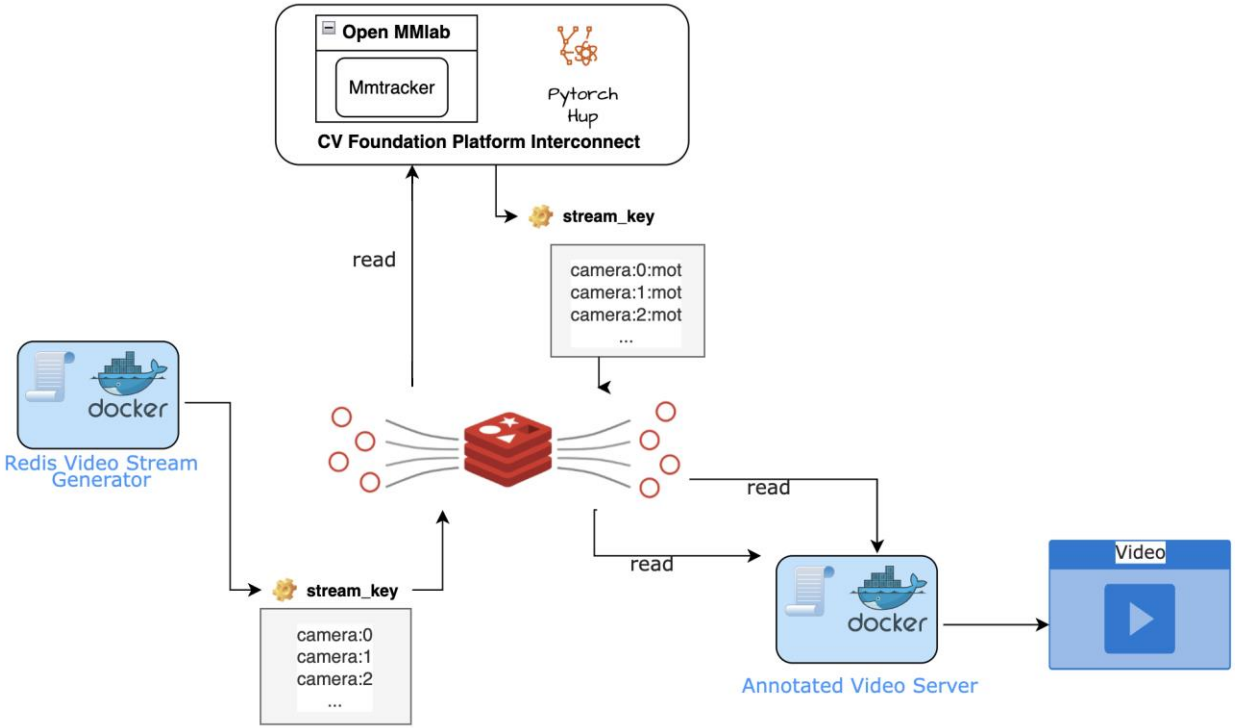
Video Source Service

Multi instance, multi container video frames producer from different type of sources

Multi instance, multi container data bus using Redis Streams



Annotated Video Server



OpenMMLab

- Based on PyTorch, OpenMMLab develops MMEngine to provide universal training and evaluation engine
- Provide high-quality libraries to reduce the difficulties in algorithm reimplementation
- Build a solid foundation for computer vision research and development
- Bridge the gap between academic research and industrial applications with full-stack toolchains

OpenMMLab

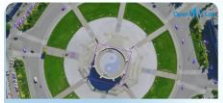
Deployment — MMDeploy

Computer Vision Libraries


MMClassification

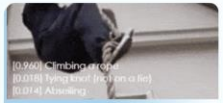

MMDetection


MMDetection3D


MMRotate


MMSegmentation


MMPose


MMAction2

MMOCR

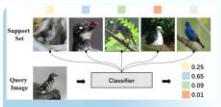

MMEditing


MMGeneration


MMFlow


MMTracking

MMHuman3D


MMFewshot

30+ Computer Vision Libraries

Foundational Libraries

MMCV — Neural Network Operators — Data Transforms

MMEngine — Training Engine — Evaluation Engine — Module Management

Deep Learning Framework

PyTorch

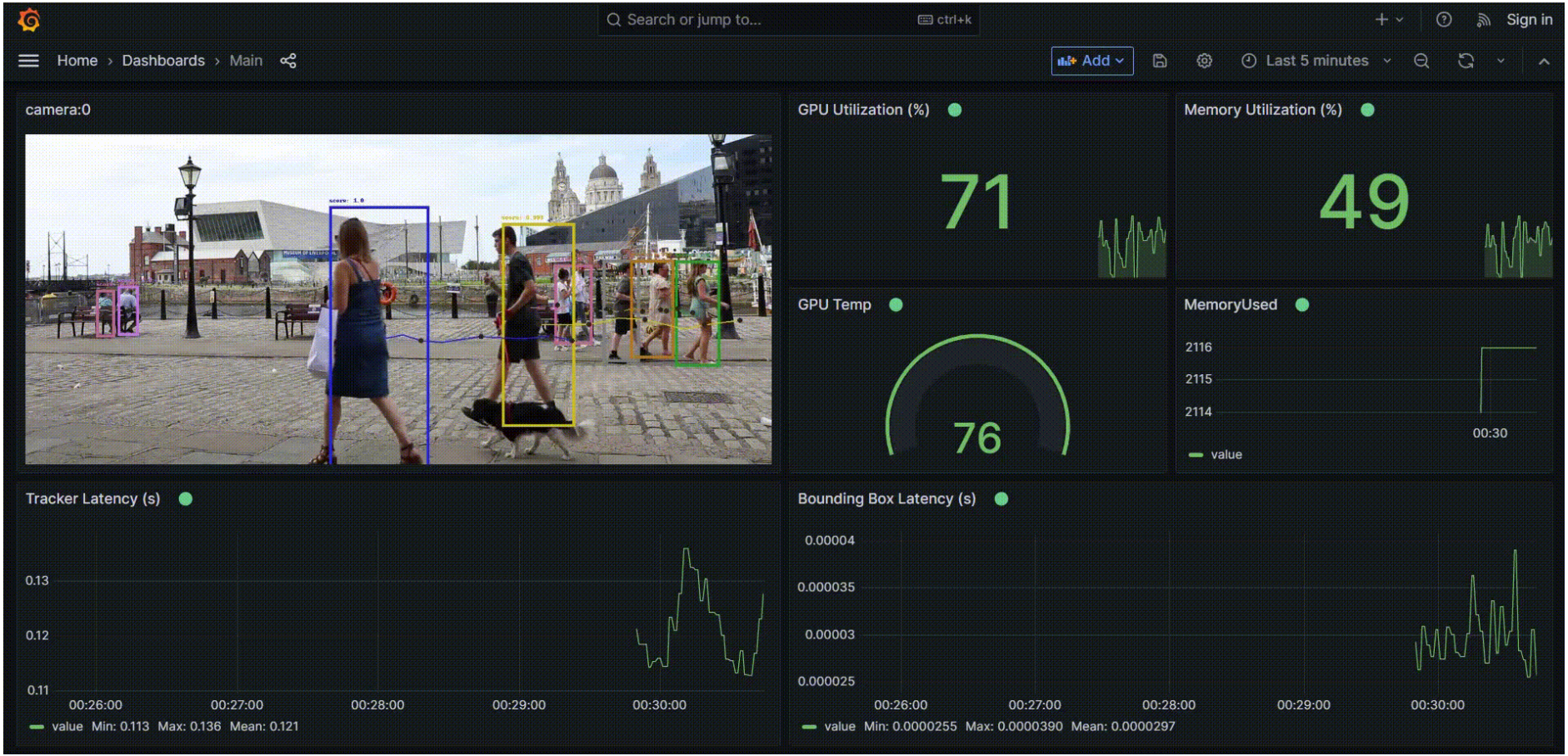
Features

OpenMMLab

- **Simple**: MMTracking interacts with other OpenMMLab projects. It is built upon MMDetection that we can capitalize any detector only through modifying the configs.
- **Fast**: All operations run on GPUs. The training and inference speeds are faster than or comparable to other implementations.

Deep Vision

- Real time streaming
- Stable containerization support

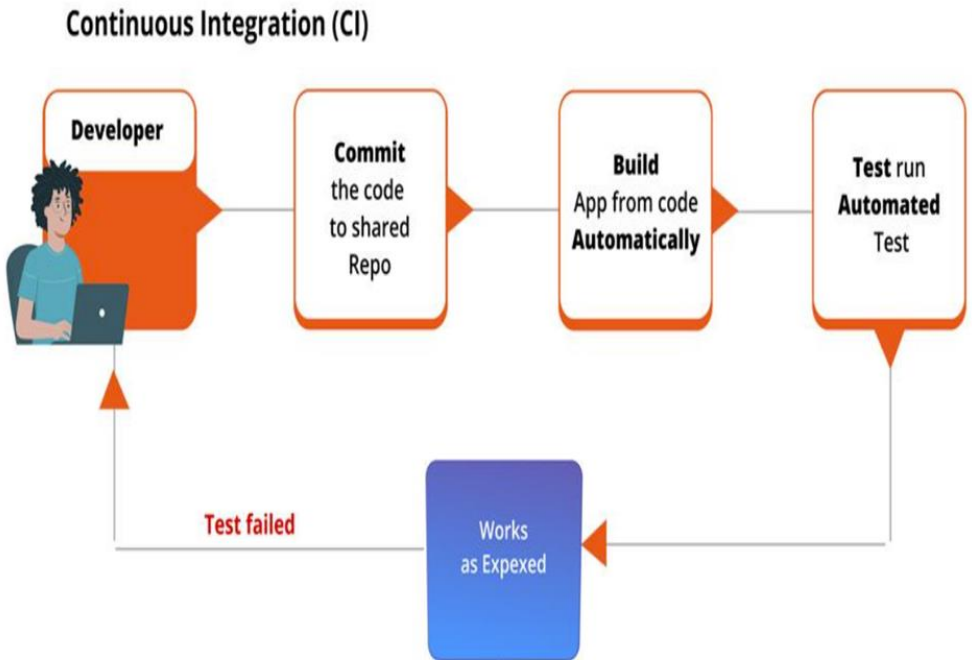


GitHub CI workflow

- CI – It refers to continuous integration

CI Steps:

- Tool Used : GitHub Actions.
- Configuration File : deepvision.yml (Located at .GitHub folder inside repo)



Steps followed:

- ❑ Set up Python 3.8

  - Installing python 3.8 version.

- ❑ Install dependencies

  - Using requirements.txt to install the libraries used i code repo.

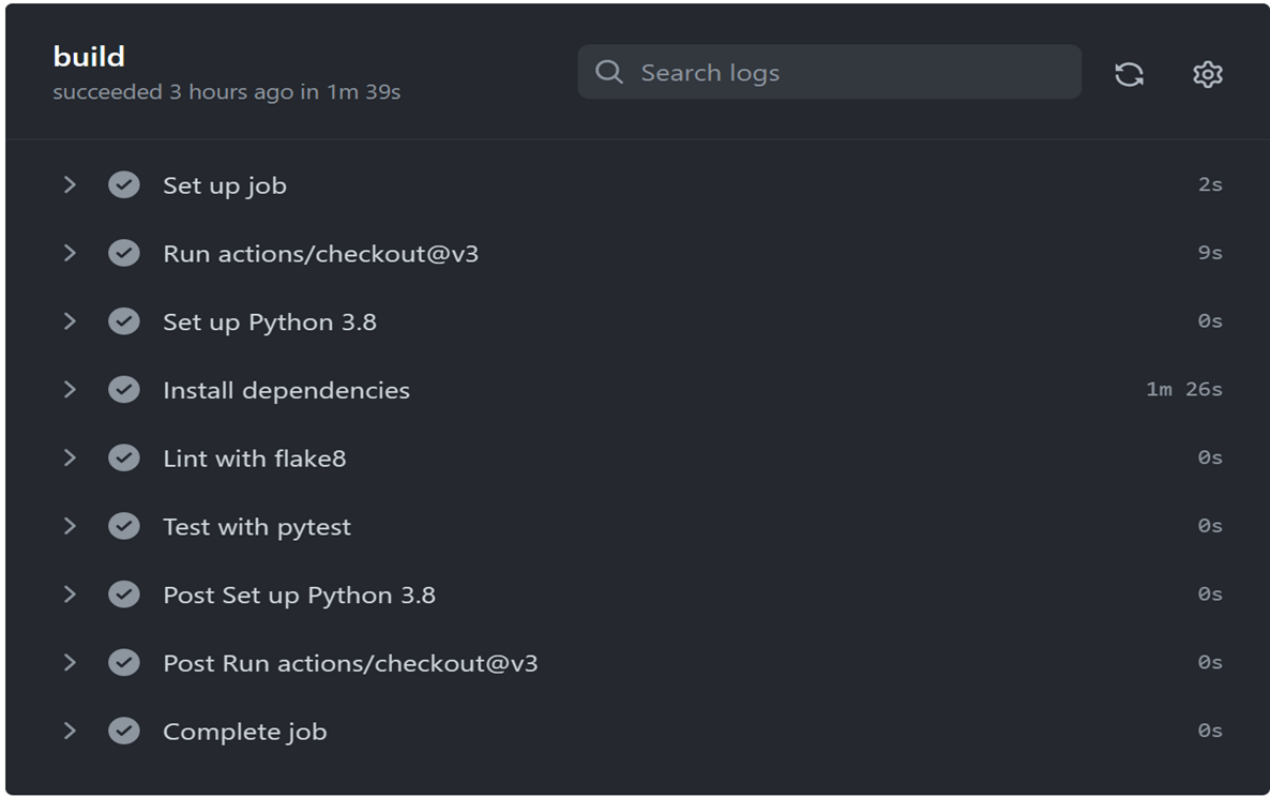
- ❑ Lint with flake8

  - Default

- ❑ Test with Pytest

  - Running all tests under unittests folder on repo.

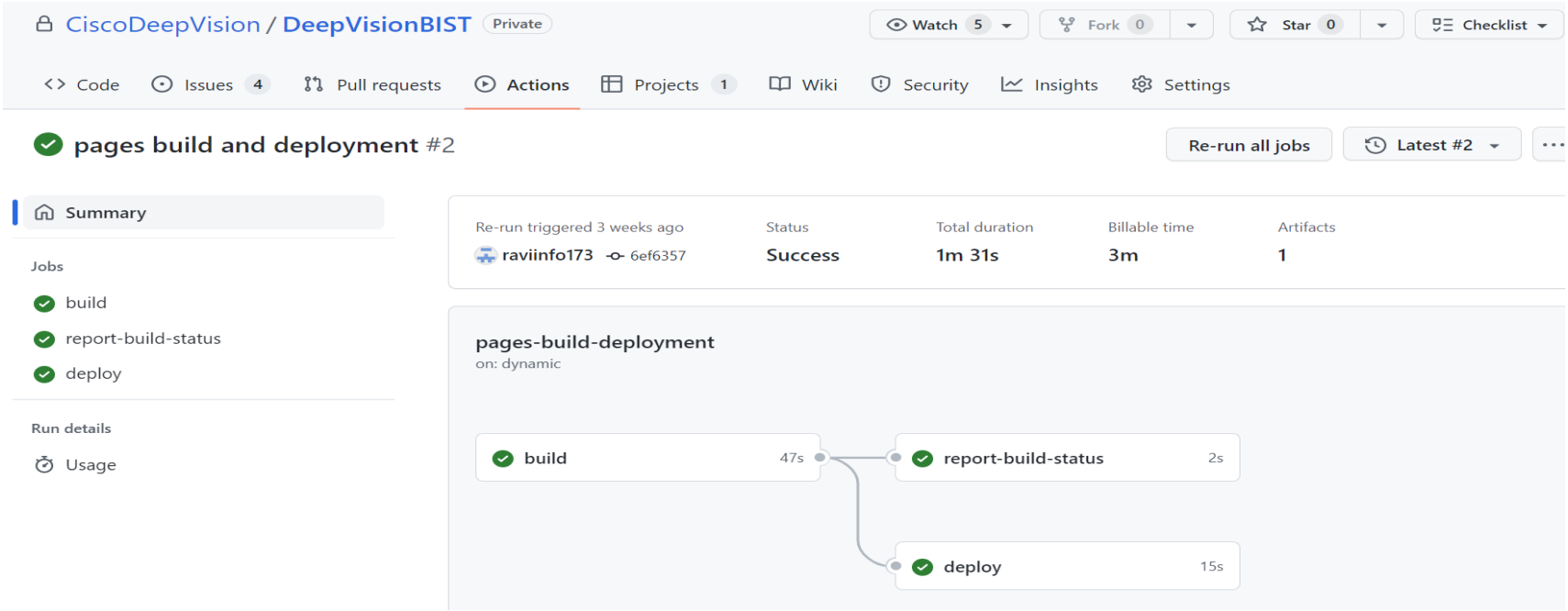
CI - test



Deploy to the server

- Deploy steps:
  - Tool Used : None (Used commands)
  - Configuration File: deploy.sh
- Steps Followed:
  - Do ssh into the desired server.
  - Place deploy.sh at the base location where repo needs to be deployed.
  - Run deploy.sh with branch name to deploy code from branch to server.(see command below)
- Steps Inside deploy.sh:
  - Cloning into repo directory from given branch in command.
- Command used:
  - ./deploy.sh -b <branch_name>

Deployment auto test



Vision Works

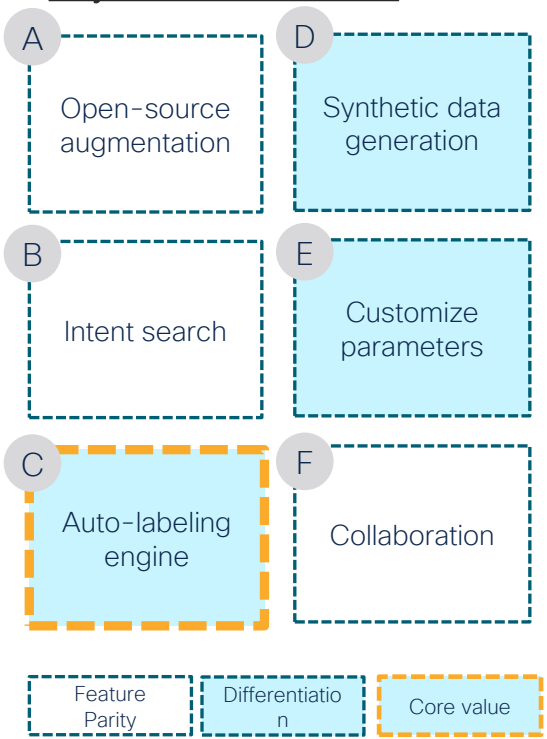
Our Mission Statement:

Create good data for the best possible training set across data modalities by focusing on 3 key steps…

1. Make raw data searchable
2. Support signal detection and model intent formation
3. Drive accurate inference

Vision Works: Product Overview

Key Product Features

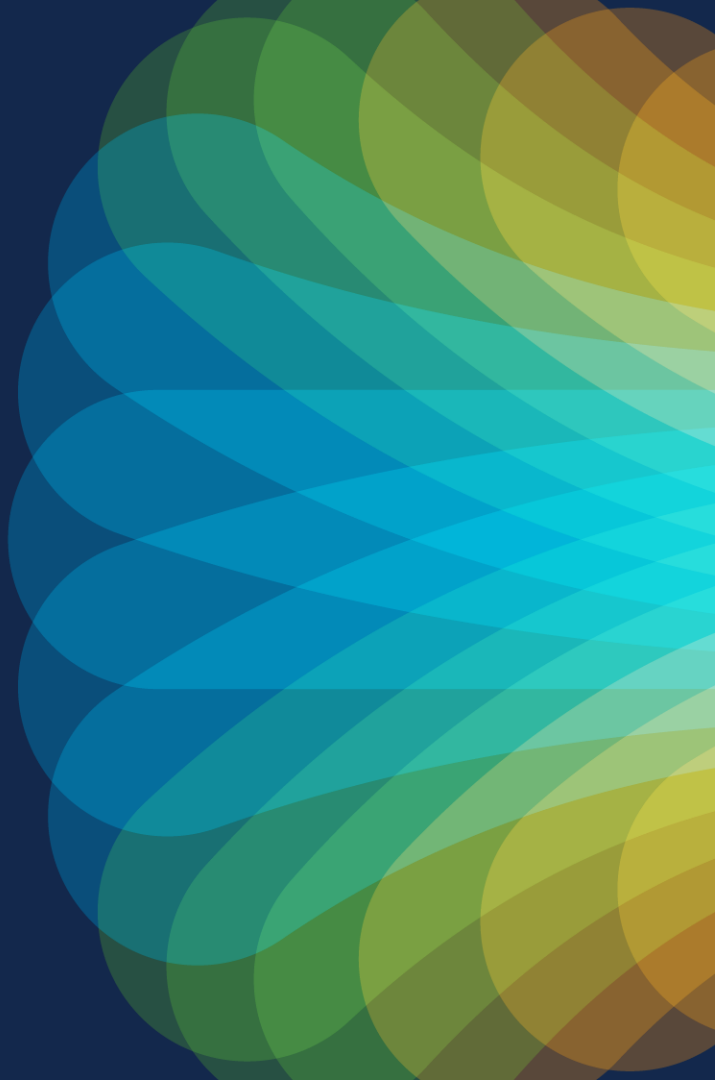


Core Value: Auto-Labeling Engine Components

Data exploration	Speed to model	Data prep accuracy
Benefit: Discovery Venture into unknown, new spaces which are prohibitively expensive to explore today, or assess new business problems. Component: Multi-modal embeddings drives clustering for the unknowns (i.e., unknown objects, behaviors, and events)	Benefit: Speed At least 5 times faster than todays labeling approaches, with potential for even speedier iteration cycles. Component: Auto-labeling engine i.e., Automatic “interesting” clip identification, Automatic cluster detection and label propagation	Benefit: Accuracy Similar to today’s levels i.e., Approach human levels of accuracy with average of 3 humans across data in bulk. Component: Integrating symbolic and sub-symbolic spaces allows us to reason and focus on the smallest amount of relevant information

Do you want to supercharge your data exploration, insights, and model preparation process?

Come talk to us!



Cisco Research Overview

Goal:

To conduct and foster research

in technology areas of strategic interest to Cisco

and generate business, technology and societal impact

University Engagements

Sponsored Research Partners

University Gift Program

100+ Universities
25+ Countries

University/Industry Consortiums

- MIT MLA
- MIT MediaLab
- Stanford ICME
- CMU Cylab
- Purdue CERIAS
- QEDC
- CQN
- UIDP

120 Total Projects Funded

Ethical AI (5)

Bias detection/mitigation, ethical design, privacy-preserving AI/ML, AI for ethics

Edge Computing(18)

Infrastructure, federated/distributed ML, CAVs, MLOps, serverless, 5G

Future of Work (10)

Productivity, worker wellness, smart home

Future Directions

Data Management,
Data Processing,
Sustainability,
Distributed Systems,
Edge Computing,
Metaverse,
Blockchain,
Security,
AI/ML,
Networking,
Cloud

Security (27)

Malware, pen-testing, privacy-preserving computation, SW supply chain, biometrics

Healthcare (13)

AI/ML for diseases, federated learning, mental healthcare, radiology, remote health monitoring

Supply Chain (4)

Anomaly prediction, supply-demand planning, RFID-based tracking, security

NLP (11)

NL understanding, language models, text summarization

AI/ML/CV (13)

Scene prediction, multimodal sensing, image reconstruction, AIOps

Others (19)

Storage switch, network verification, indoor localization, sustainability, hardware

Ramana Kompella
Head of Cisco Research

Research Team






Ops Team



University Research Engagements

Internal Research Projects

Thought Leadership

Visit Outshift in the World of Solutions!

outshift by CISCO

Take a picture of this slide and bring it to the Outshift booth in the World of Solutions. (#3307)

Get your badge scanned to be entered into our daily drawing for an Apple iPad!

Learn more about Great Bear!

More Learning Opportunities

Do Tell About OTel: An Introduction to OpenTelemetry and How AppDynamics is Embracing It
BRKAPP-1154

Wayne Brown
SE Manager, AMER Elite Partners,
Cisco Systems, Inc.

Application Performance, AppDynamics, Full Stack Observability
Thursday, Feb 92:15 PM - 3:15 PM CET

Location
TBU

OpenTelemetry is quickly becoming the de-facto vendor-neutral standard for collecting metrics, events, logs, and traces from a wide range of applications and systems. In this session, we will learn the history behind OpenTelemetry, the benefits that OpenTelemetry provides, the architecture of OpenTelemetry, and the future of the standard. Finally, we will talk about how AppDynamics is embracing the OpenTelemetry standard and supporting its developments.

Learn more about ET&I Cloud-Native Products

Panoptica
Simplified Cloud-Native Application Security for DevSecOps, Platform, and DevOps teams.

Calisti
Cisco Service Mesh Manager. An enterprise ready Istio platform for DevOps and SREs t automates lifecycle management and simplifies connectivity, security, and observability for microservice-based applications.

See them both in the Cisco Showcase

More Learning Opportunities

Real-Time Media in a Cloud Native World
BRKETI-2006

Giles Heron
Principal Engineer, Cisco Systems

Cloud
Thursday Feb 9, 2:30PM – 3:30PM CET

Location
Session Room G102

The cloud-native approach based on containerised micro-services has transformed the software landscape but is largely focussed on non-real time web-based applications, especially in the case of service meshes which use web proxies to interconnect workloads. Media Streaming Mesh uses real-time media proxies to observe, route, encrypt and protect north-south and east-west media traffic. You'll leave this session with an understanding of the Media Streaming Mesh architecture, of some of its key use cases, and how you can apply it to your own media workflows.

Data Security and Compliance in Cloud Native and On-Prem Applications
BRKETI-2414

Peter Bosch
Distinguished Engineer, Cisco Systems

Hybrid Cloud, Cloud
Thursday Feb 9, 4:00PM – 5:00PM CET

Location
Session Room G109

Application data represents the core to digital enterprises: data is handled by applications that are hosted on-prem, in-cloud, in containers and virtual machines, by API gateways, stored on in-VM databases and cloud storage resources. Losing, not knowing where it resides, not knowing if data is handled in a compliant manner, or not knowing someone copied the data can be a disaster to a digital enterprise. This talk presents Cisco's approach to data security and compliance. We present how Cisco tracks data in-flight and at-rest, and how to turn such data into information. We show where, in the application, information is vulnerable and the application is not compliant to the various compliancy standards.

Cloud Native Observability
BRKCLD-2158

Shannon McFarland
Distinguished Engineer, Cisco Systems

Hybrid Cloud, Cloud
Friday Feb 10, 9:15AM – 10:45AM CET

Location
Session Room E105

The need to have an end-to-end view of a complex microservices environment and the underlying infrastructure is growing in demand as customers migrate from legacy and monolithic workloads to Cloud Native environments. The days of scouring through endless logs and trying to correlate those logs with events, metrics, and even traces are coming to an end. Just 'seeing' the raw info does little to help one understand what is happening. Observability is so much more than 'seeing' the data but making sense of it and taking prescription actions based on the data. This session will go over common MELT (metrics, events, logging, and tracing)

Complete your Session Survey

- Please complete your session survey after each session. Your feedback is important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (open from Thursday) to receive your Cisco Live t-shirt.

- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Session Catalog and clicking the "Attendee Dashboard" at https://www.ciscolive.com/emea/learn/sessions/session-catalog.html

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

Continue Your Education

Visit the Cisco Showcase for related demos.

Book your one-on-one Meet the Engineer meeting.

Attend any of the related sessions at the DevNet, Capture the Flag, and Walk-in Labs zones.

Visit the On-Demand Library for more sessions at ciscolive.com/on-demand.

Fill out your session surveys!

Attendees who fill out a minimum of four session surveys and the overall event survey will get **Cisco Live-branded socks** (while supplies last)!

Attendees will also earn 100 points in the **Cisco Live Challenge** for every survey completed.

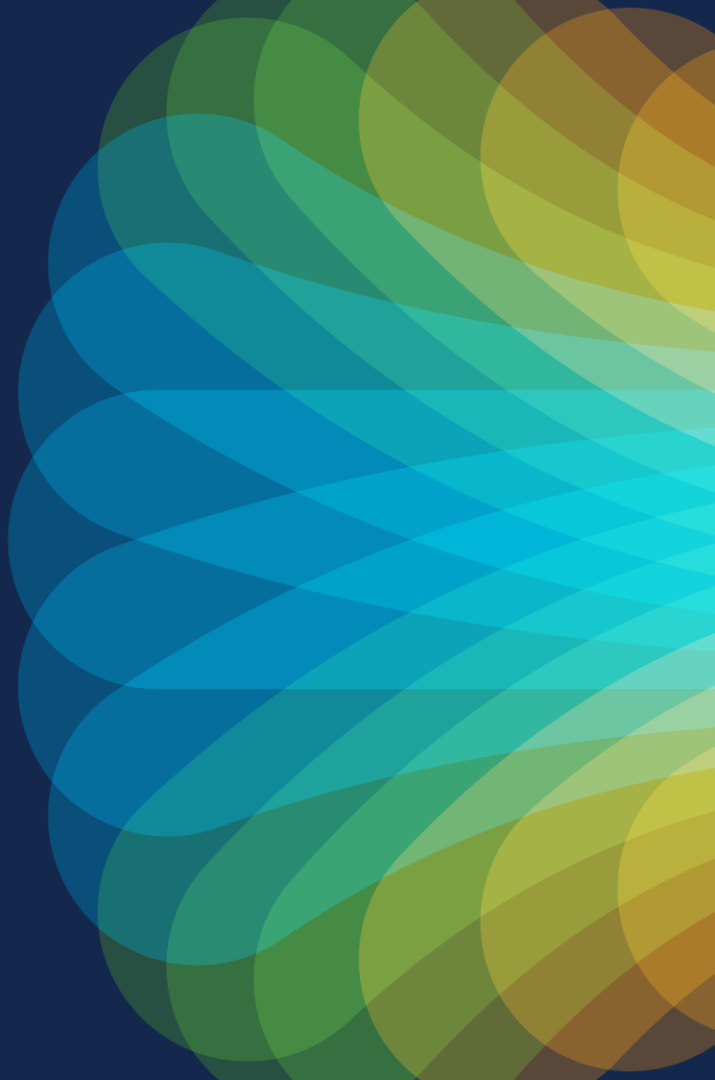
These points help you get on the leaderboard and increase your chances of winning daily and grand prizes

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

The bridge to possible

Thank you



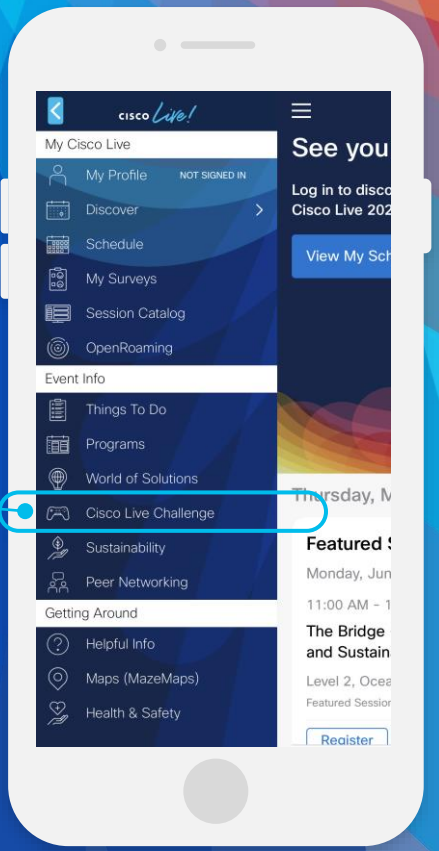
Cisco Live
Challenge

Gamify your Cisco Live experience!

Get points for attending this session!

How:

1. Open the Cisco Events App.
2. Click on 'Cisco Live Challenge' in the side menu.
3. Click on View Your Badges at the top.
4. Click the + at the bottom of the screen and scan the QR code:



cisco Live!


cisco Live!
Let's go
#CiscoLive