AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

AIM235

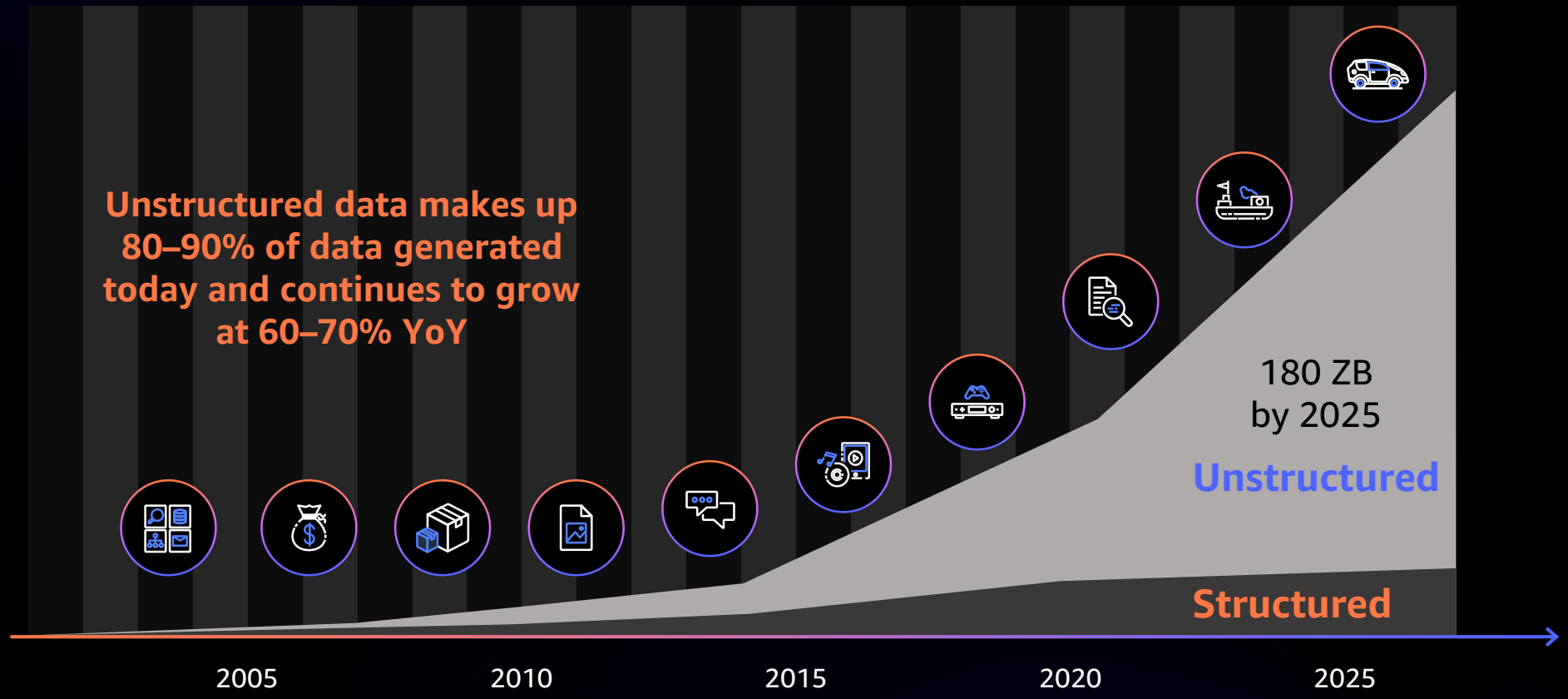
Introducing Amazon SageMaker Ground Truth Plus

Kumar Chellapilla
General Manager, Human-In-Loop ML/AI Services, AWS

Rick Carragher
Vice President of Engineering, Samsara

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Growth of unstructured data

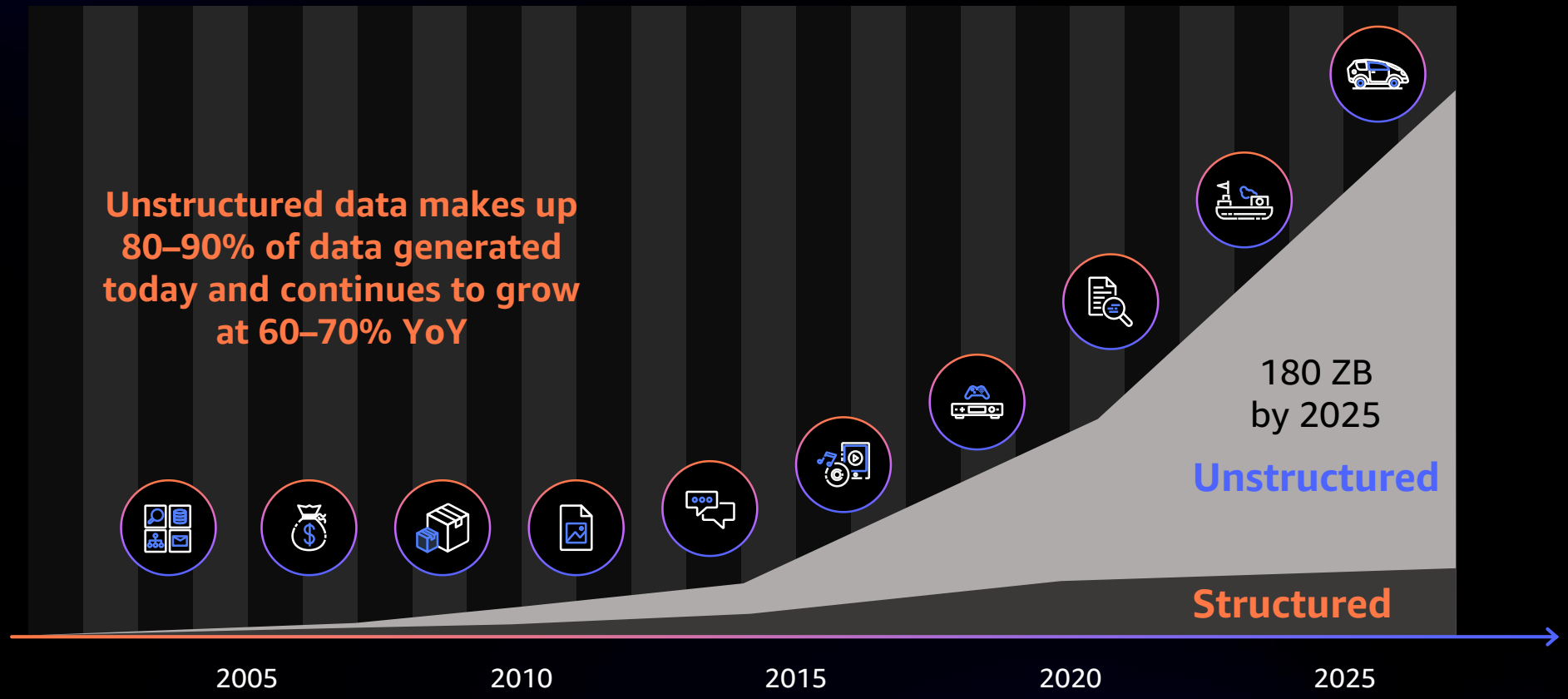


Unstructured data makes up **80–90%** of data generated today and continues to grow at **60–70%** YoY

(Source: Huawei)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

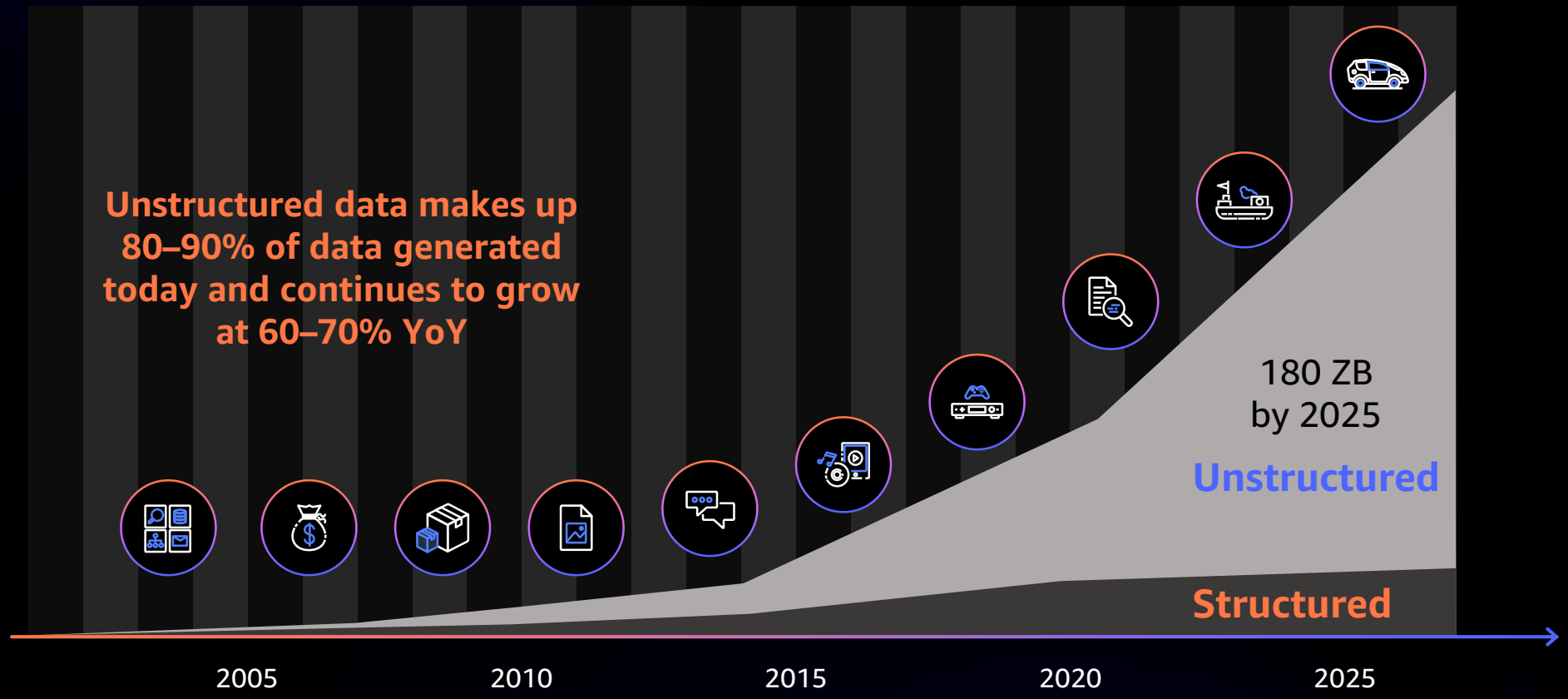
Growth of unstructured data



Machine learning (ML) models and artificial intelligence (AI) Systems need to scale to **unstructured data**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

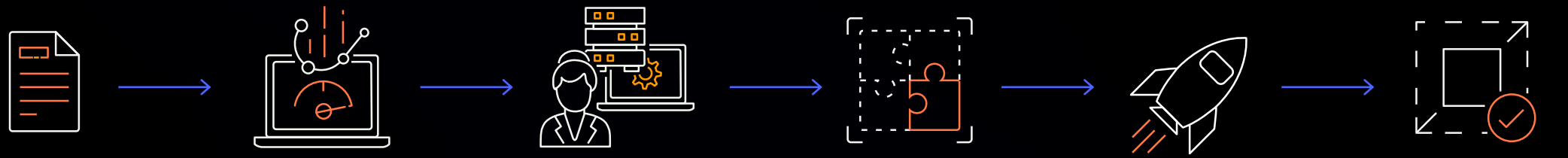
Growth of unstructured data



Data labeling unlocks value in **unstructured data**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Typical machine learning process



Collect, prepare, and label training data

Choose and optimize ML algorithm

Set up and manage environments for training

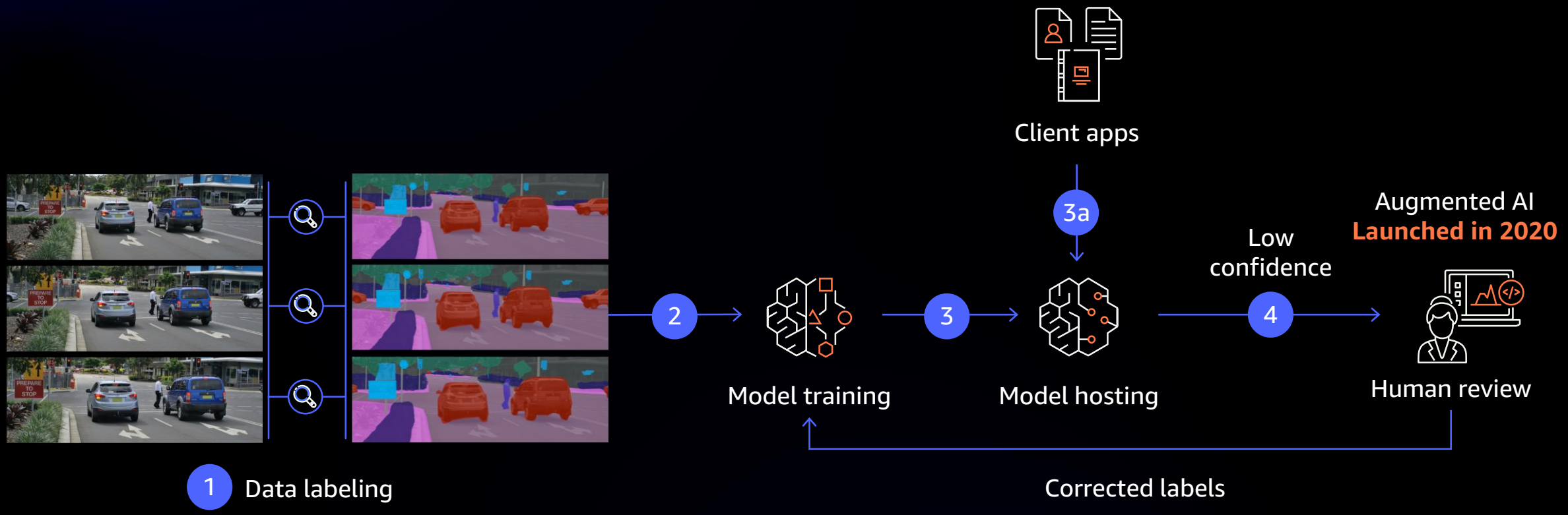
Train and tune ML models

Deploy models in production

Scale and manage the production environment

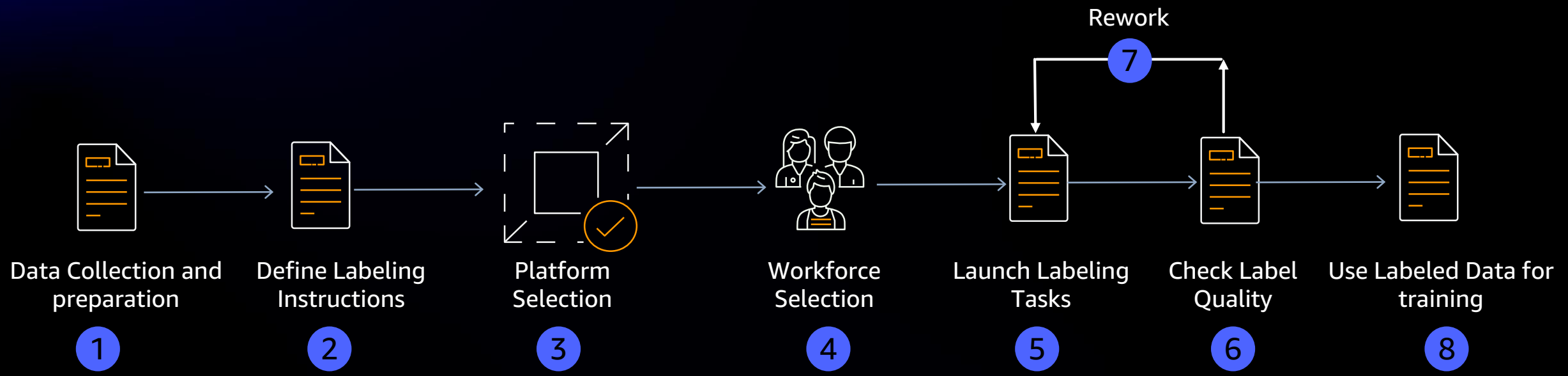
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Human in Loop in machine learning



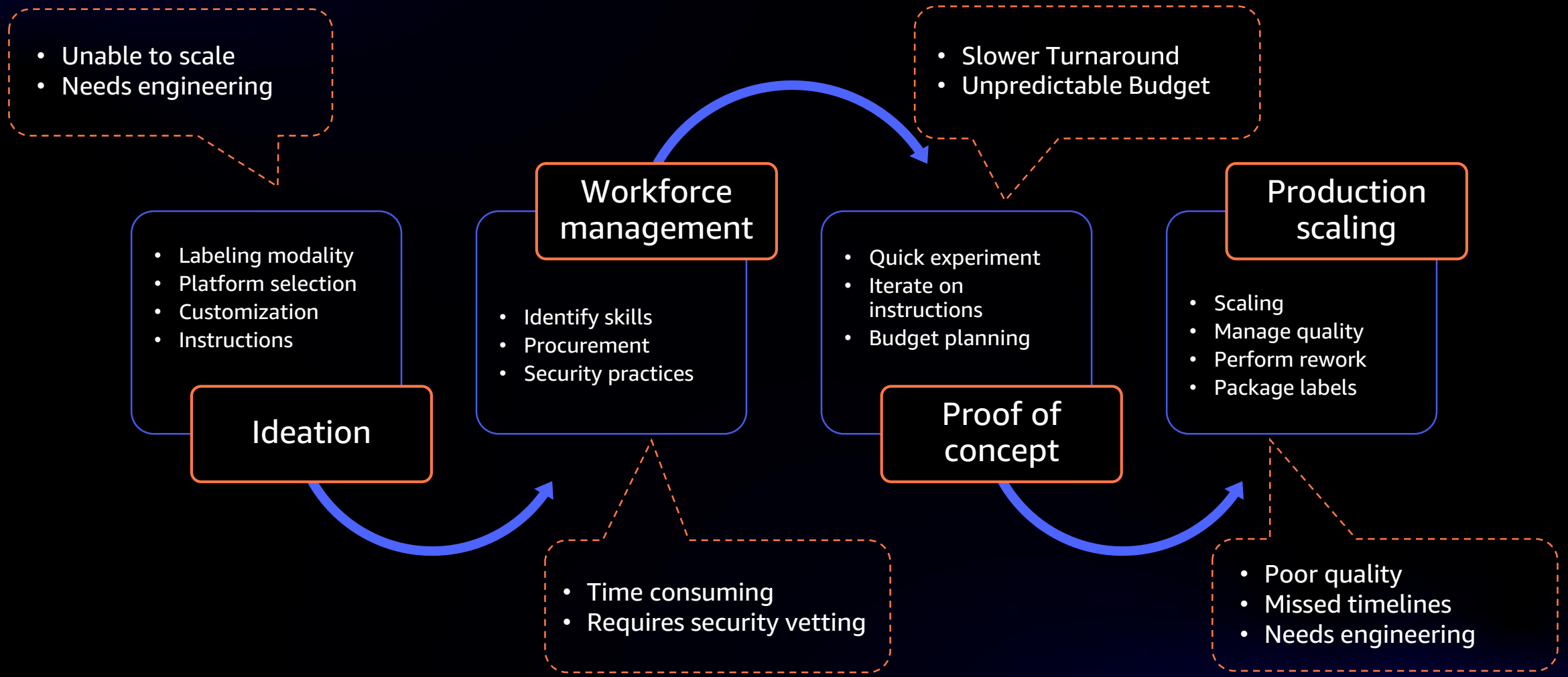
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Typical data labeling process



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges: Operating data labeling projects



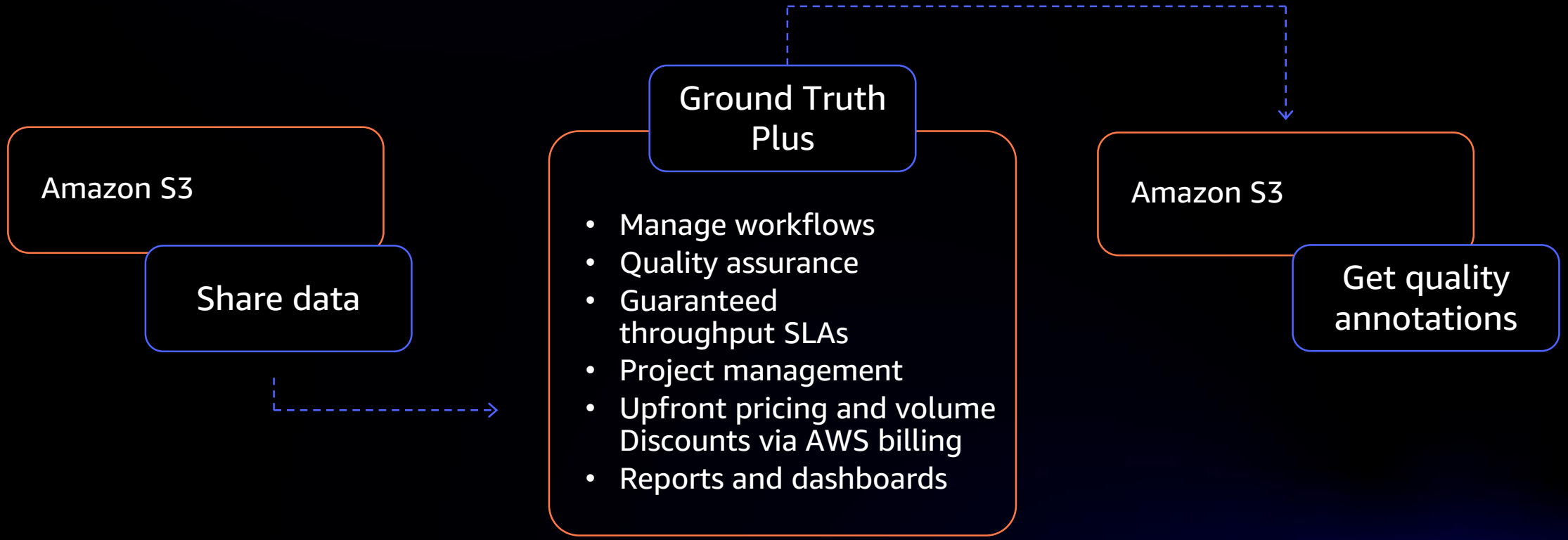
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing Ground Truth Plus

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Ground Truth Plus: Data labeling as a service

TURNKEY OFFERING THAT COMBINES TECHNOLOGY, PROCESS, AND SERVICES TO DELIVER HIGH-QUALITY LABELS



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Expert Workforce

HIGHLY SKILLED, DIVERSE, AND ELASTIC AWS MANAGED WORKFORCE

M-Turk workforce (GT)



Subjective surveys

Low-skilled tasks

Managed workforce (GT Plus)



Highly skilled trained workforce

AWS employees, security cleared

Project management | Training | Process improvement | Quality management

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Ground Truth Plus Value Proposition

Combine technology, process, and services to deliver high quality labels



Turnkey solution

Dedicated PM, Custom features, AI/ML consulting, label quality & meet throughput



High Quality

Multi-step quality workflow includes ML model to validate human labels



Expert workforce

Labeling is performed by highly-skilled workforce trained on a variety of ML tasks



Cost Reduction

ML based assistive tooling, ML models for Active Learning & Pre-labeling reduces costs by up to 40%



Transparency

Metrics dashboards & Review UI to track progress of training datasets



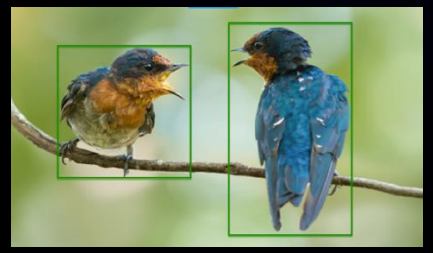
Secure

Data never leaves AWS, Authorized Amazon employees have access to data

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Labeling UI and Workflow Customizations

70+ WORKFLOWS



Bounding boxes



Image classification



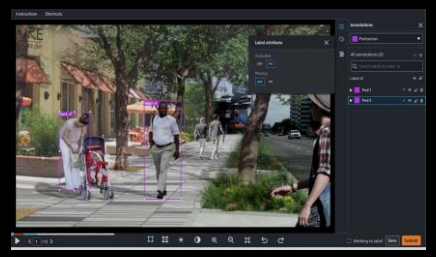
Segmentation



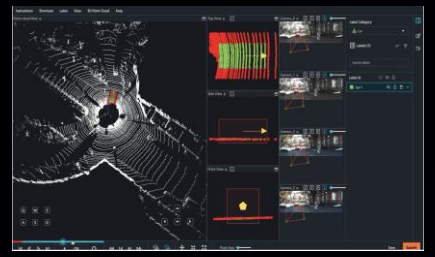
Label verification



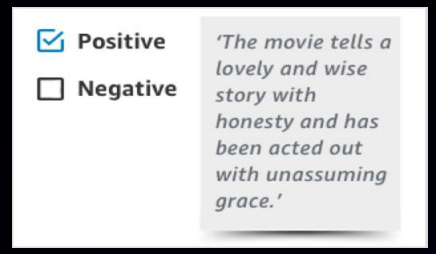
Custom template



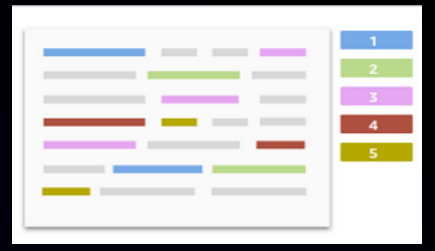
Video



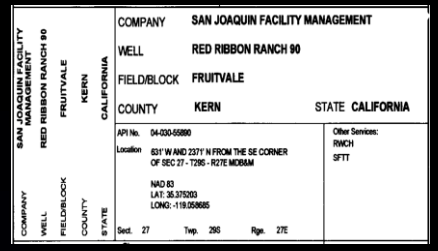
LIDAR 3D Point Cloud



Text classification



Named entity recognition



OCR, form, table

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Ground Truth Plus Customers

"At the NFL, we continue to look for new ways to use machine learning to help our fans, broadcasters, coaches, and teams benefit from deeper insights. Football is a fast-moving sport where plays can happen in a split second. While coaches and referees carefully watch the game, it can be difficult to watch all players on a field for safety. Computer vision allows us to accurately detect player safety incidents, but developing these algorithms requires expertly labeled data. Now with Amazon SageMaker Ground Truth Plus, we have custom workflows and user interfaces for sophisticated labeling tasks, which helps us improve player safety."

Jennifer Langton, SVP, Player Health and Innovation, NFL

"At Airbnb, we are increasingly integrating ML across all aspects of our business. As a result, our teams consistently need to generate and maintain high-quality data in order to train and test ML models. We were looking for a way to generate high-quality text classification data results on one hundred thousand paragraphs of customer service logs in Mandarin so we can better serve our customers and reduce dependencies on our customer service team. With Amazon SageMaker Ground Truth Plus, the AWS Team built a customized data labeling workflow, which included a customized ML model that was able to achieve 99% classification accuracy."

Wei Luo, Data Scientist, Airbnb China

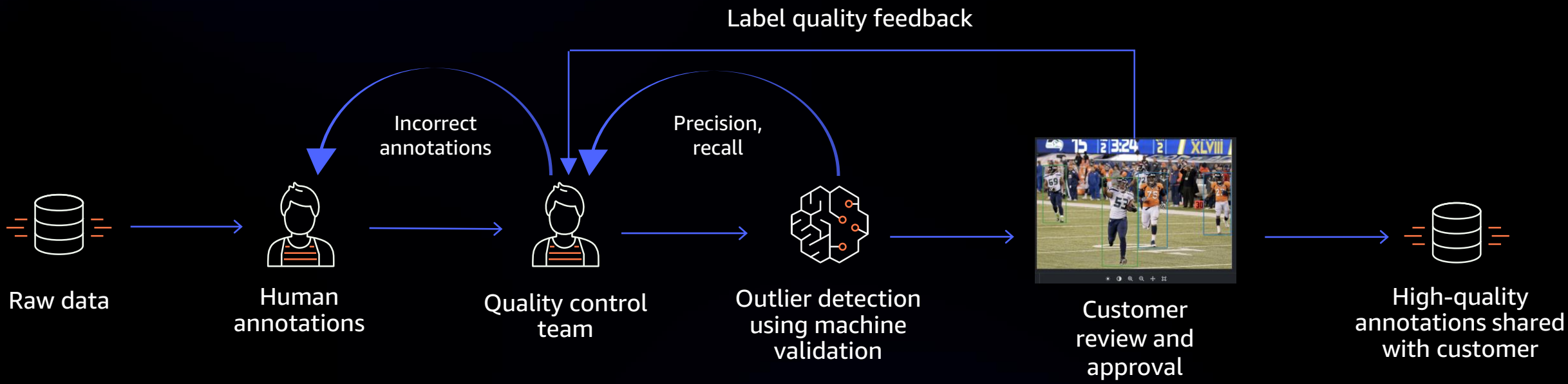
"At Samsara, we're driving the digital transformation of physical operations. With our Connected Operations Cloud, companies that depend on physical operations can harness IoT data, analytics, and AI to develop actionable insights and improve their safety, efficiency, and sustainability. With Amazon SageMaker Ground Truth Plus, we're able to receive high-quality labeled data and gain access to purpose-built tooling designed to further improve our ability to detect and address safety risks."

Rick Carragher, VP of Engineering, Samsara

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Improve label quality using machine validation



Workforce training | Assistive tooling | Consensus algorithm | Policy management

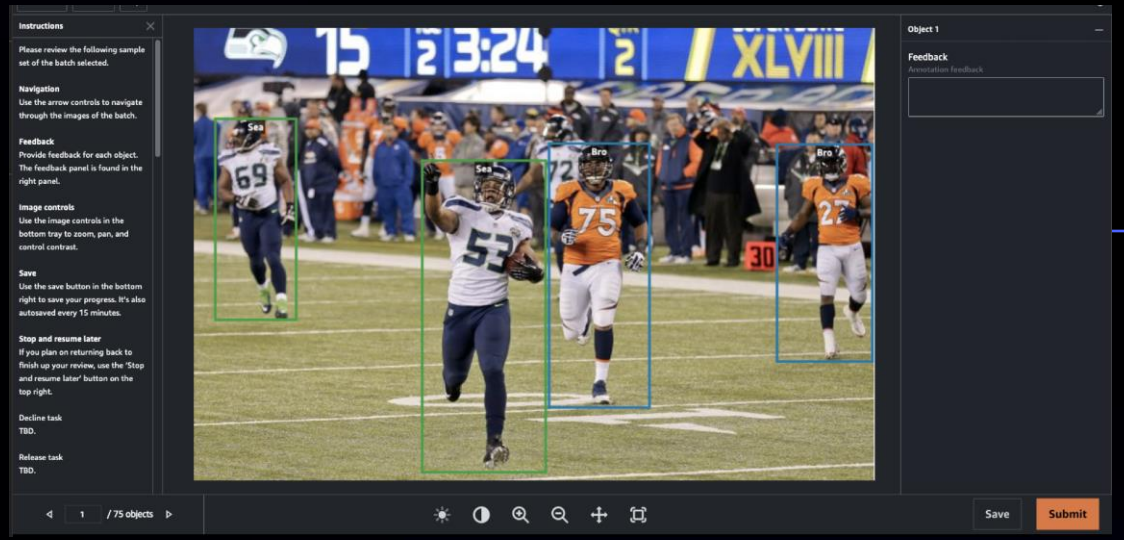
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Guaranteed quality service-level agreement

SLA

Quality SLA

- ✓ Bounding Box tightness
- ✓ Misclassified attributes
- ✓ Missed objects
- ✓ …



Review and adjust randomly sampled labels, provide feedback

Reject

Free rework

Quality metrics

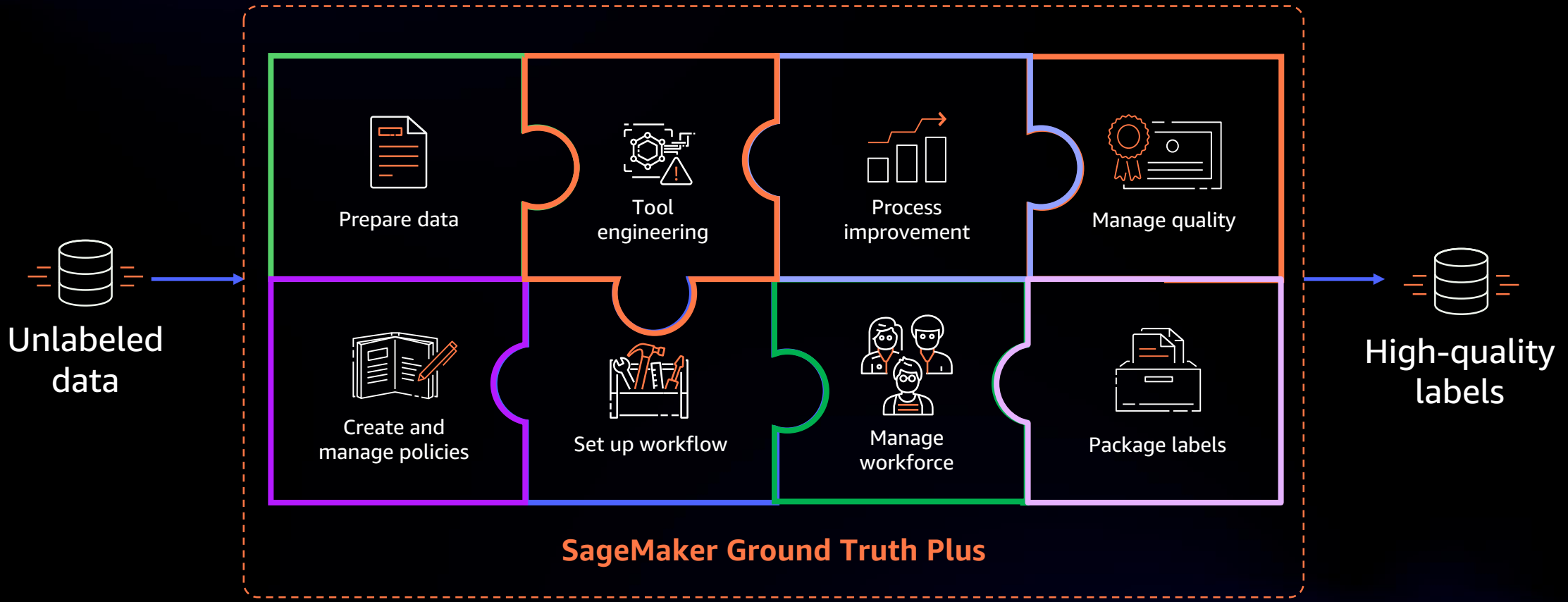
Approve

Package labels

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Turnkey data labeling service

PROCESS EXCELLENCE



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tracking and Visibility

Centrally monitor all labeling projects

Filter projects

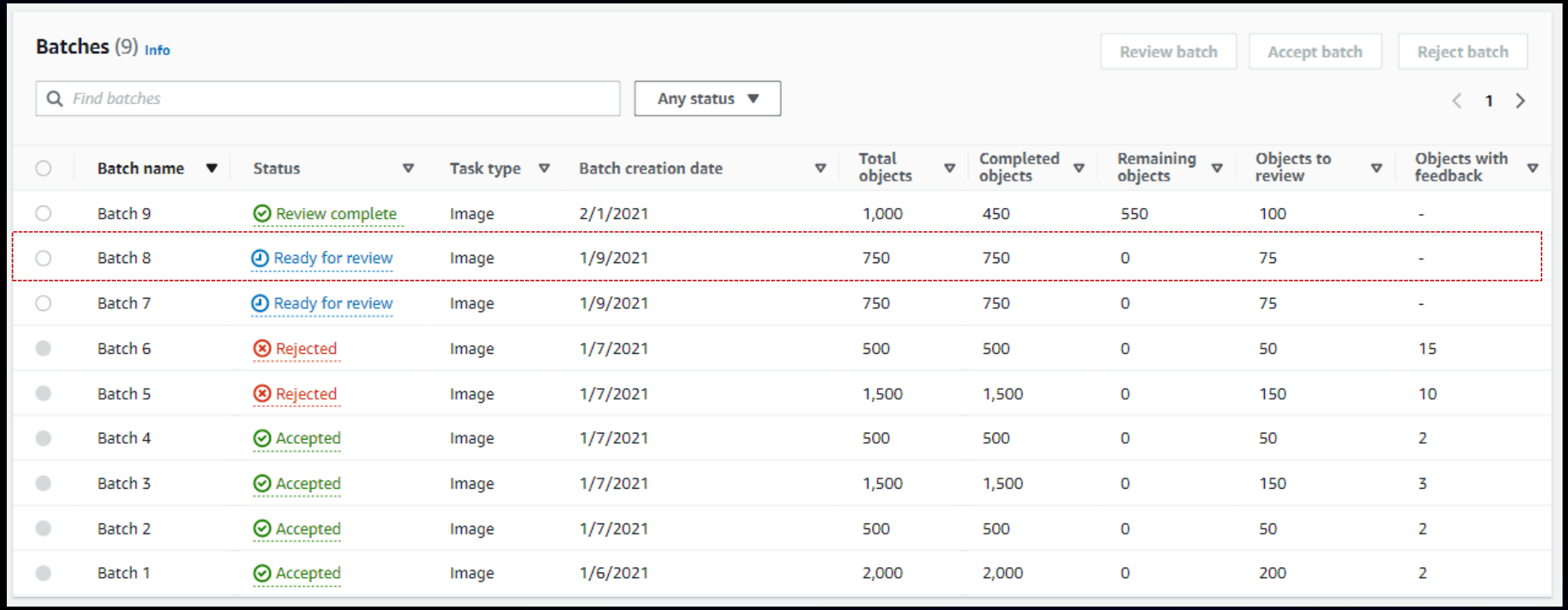
1

Project name	Status	Task type	Start date	Est. end date	Total objects	Completed objects	Remaining objects	Objects with feedback
beaconprojectname-26485	Consultation	Video	-	-	-	-	-	-
client-projecttest-654	Setup in progress	Text	-	-	-	-	-	-
client-projecttest-845	POC	Video	-	-	-	-	-	-
internal-projecttest-641	POC	Text	-	-	-	-	-	-
beacon-projecttest-876	POC complete	Image	-	-	-	-	-	-
client-projecttest-586	Pricing complete	Point cloud	-	-	-	-	-	-
beaconprojectname-26485	Contract executed	Image	-	-	-	-	-	-
client-projecttest-654	Production data pending	Point cloud	-	-	-	-	-	-
client-projecttest-845	Production	Text	1/6/2021	1/20/2021	2,000	550	1,450	542
internal-projecttest-641	Production	Point cloud	1/7/2021	1/16/2021	500	125	375	123
beacon-projecttest-876	Production complete	Video	1/7/2021	1/16/2021	1,500	1,023	477	1,018
client-projecttest-586	Production complete	Point cloud	1/6/2021	1/20/2021	190	190	0	185
client-projecttest-586	Paused	Image	1/6/2021	1/20/2021	1,500	1,500	0	1,486

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tracking and Visibility

Select a project and monitor batch-level progress



Batches (9) Info

Review batch | Accept batch | Reject batch

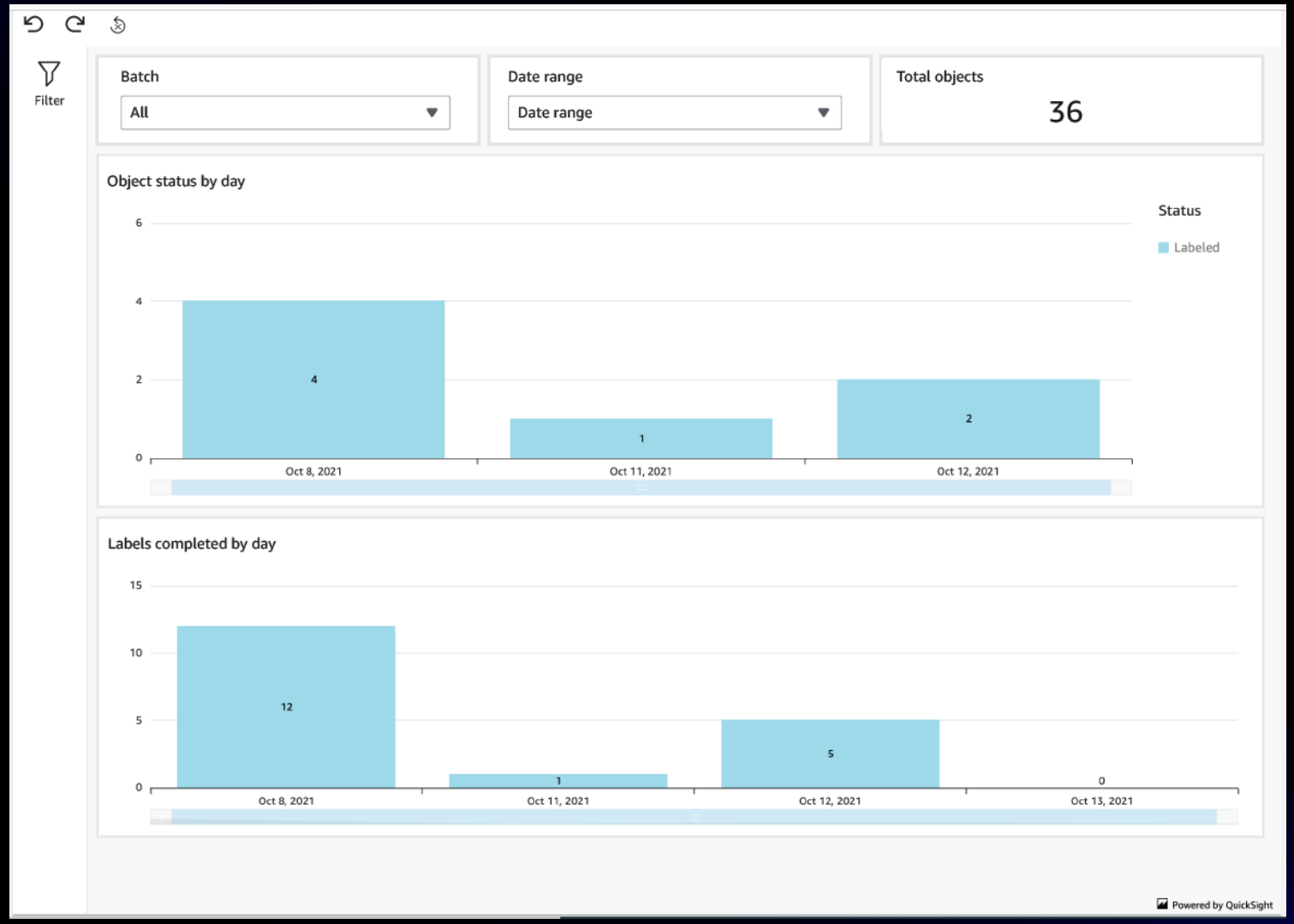
Find batches | Any status | 1

	Batch name	Status	Task type	Batch creation date	Total objects	Completed objects	Remaining objects	Objects to review	Objects with feedback
○	Batch 9	Review complete	Image	2/1/2021	1,000	450	550	100	-
○	Batch 8	Ready for review	Image	1/9/2021	750	750	0	75	-
○	Batch 7	Ready for review	Image	1/9/2021	750	750	0	75	-
○	Batch 6	Rejected	Image	1/7/2021	500	500	0	50	15
○	Batch 5	Rejected	Image	1/7/2021	1,500	1,500	0	150	10
○	Batch 4	Accepted	Image	1/7/2021	500	500	0	50	2
○	Batch 3	Accepted	Image	1/7/2021	1,500	1,500	0	150	3
○	Batch 2	Accepted	Image	1/7/2021	500	500	0	50	2
○	Batch 1	Accepted	Image	1/6/2021	2,000	2,000	0	200	2

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tracking and Visibility

Track granular throughput metrics



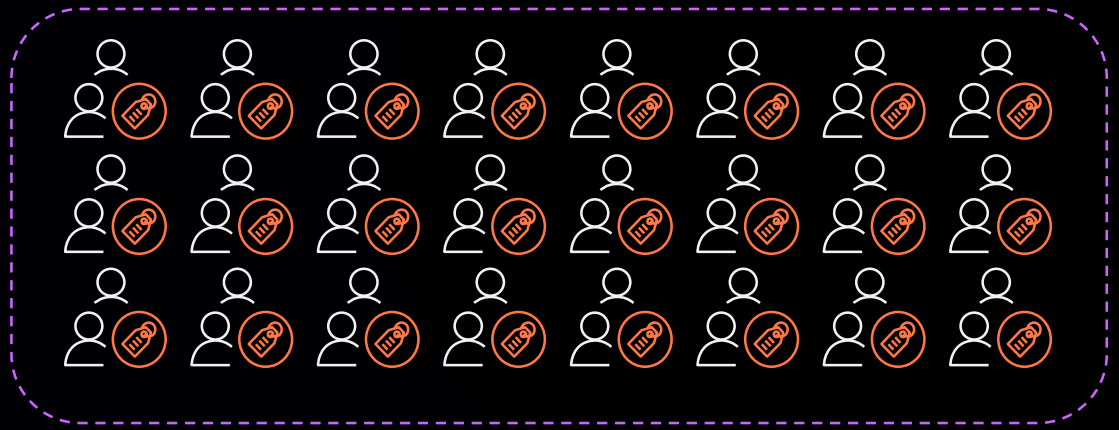
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Predictable throughput with dynamic scaling



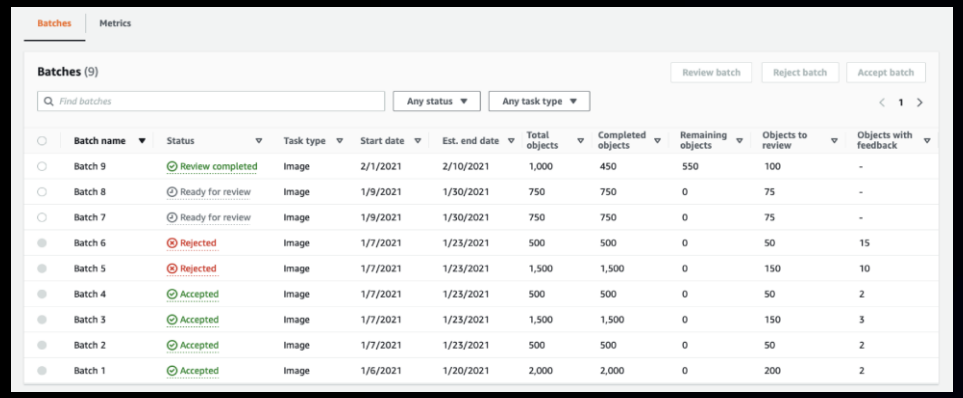
Experimentation team

Faster turnaround

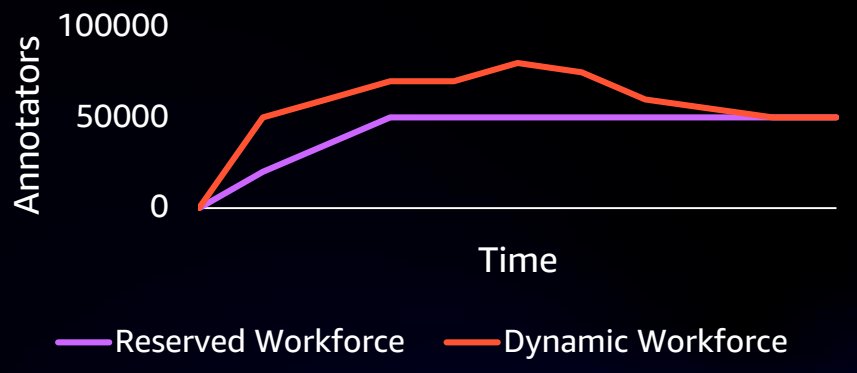


Production team

Reserved or Dynamic Workforce



Mutually agreed schedule

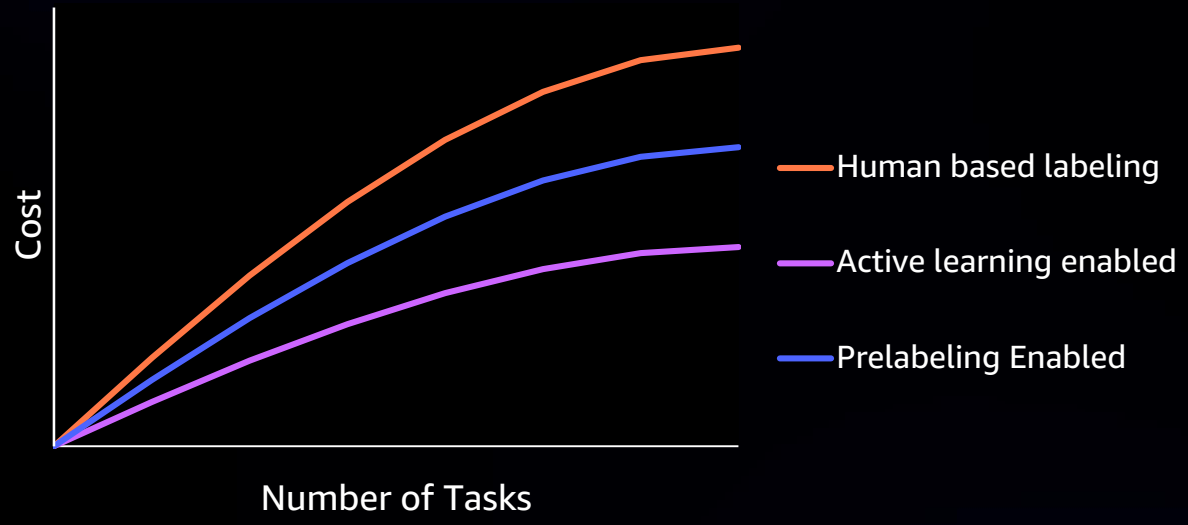


Workforce ramp-up

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lower the cost of labeling by up to 40%

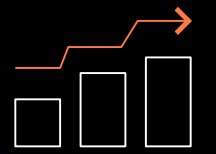
ML techniques



Process improvements



Continuous workforce training



Process Improvement

Interpolation | Autosegmentation | Predict next | Cuboid snapping

Assistive tooling

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Ground Trut | Amazon SageMaker Ground Trut | Amazon SageMaker Ground Trut

ce8c57gtz7.labeling.us-east-1.sagemaker.aws/#/projects/reInvent-Demo

Batches | Metrics

Batches (20) Info

Review batch | Reject batch | Accept batch

Find batches | Any status | 1

Batch name	Status	Task type	Batch creation date	Total objects	Completed objects	Remaining objects	Failed objects
batch1	Accepted	Video object tracking	11/16/2021	3	3	0	0
batch2	In-progress	Video object tracking	11/17/2021	3	1	2	0
batch3	In-progress	Video object tracking	11/17/2021	3	0	3	0
batch4	In-progress	Video object tracking	11/17/2021	3	0	3	0
batch5	In-progress	Video object tracking	11/17/2021	3	2	1	0
batch6	Review complete	Video object tracking	11/18/2021	3	3	0	0
batch7	In-progress	Video object tracking	11/18/2021	3	1	2	0
batch8	In-progress	Video object tracking	11/18/2021	3	0	3	0

Batch status

The following describes the status type for a batch.

Waiting for data
Awaiting input data from you.

Data received
Data to be labeled received.

In-progress
Data labeling is in progress.

Ready for review
Data labeling is complete. A subset of labeled objects from the batch are ready for you to review. This is an optional step.

Review submission in-progress
Review feedback is currently being processed.

Review complete
You have successfully reviewed the batch. Next, you have to accept or reject it. This action can not be undone.

Accepted
You have accepted the labeled data and will receive it in your



samsara

Rick Carragher
Vice President of Engineering, Samsara

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Connected Operations Platform



The Connected Operations Platform: Capture Data and Gain Insights



Leading IoT Data Advantage

2 Trillion+ DATA POINTS/YR

38 Billion+ MINUTES OF VIDEO FOOTAGE/YR

1 Billion+ AI-BASED DETECTIONS/YR

7 Billion+ API CALLS/YR

1 Million+ IOT DEVICES SHIPPED

Video-Based Safety

- Real-time HD In-Cab Video Stream
- Real-time HD Road-facing Video Stream
- Multicamera Systems
- AI detections
- Speed and traffic
- Insurance Data

Vehicle Telematics

- GPS
- Harsh Turns (Gyroscope)
- Harsh Braking (Accelerometer)
- Ignition Status
- Vehicle Idle Time
- Mileage
- Tire Pressure
- Geofencing
- Engine RPM
- Pedal Position
- Airflow Rate
- Engine Coolant Temperature

Apps and Driver Workflows

- Electronic Documents
- Third-Party Data Via Integrations
- Messaging
- Routing and Delivery
- Compliance (HOS, Tachograph)
- Driver Vehicle Inspection Report
- Vehicle and Equipment Inspection

Equipment Monitoring and Control

- Vibration
- Temperature
- GPS
- Geofencing
- Water Level
- DTC codes
- Barometric pressure
- Generator data
- Refrigerant data

Site Visibility

- IP Camera Stream
- Motion detection
- Person detection
- Vehicle detection
- Floor plans
- Person/vehicle attribute search

Risk-adjusted Insurance Premiums | Predictive Maintenance
Fleet Benchmarking | Workflow Automation

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Samsara Effect: Smarter Operations



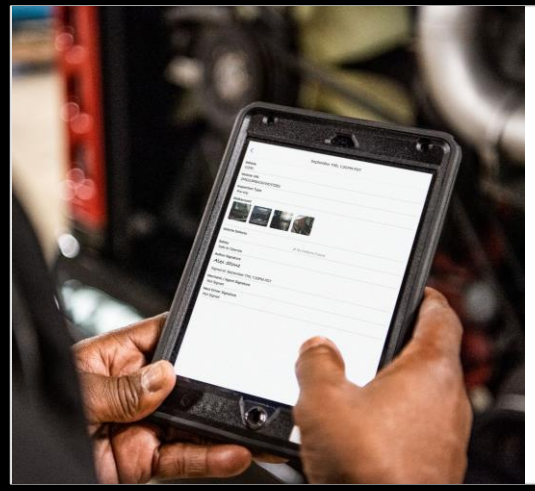
Decreased accidents by 43% with video-based coaching and alerts

BRAGG CRANE SERVICE



76% increase in dispatch productivity over one year and improving customer service and their bottom line

Uniti Fiber



Saved $600K annually on operational and administrative costs by going paperless

UFP INDUSTRIES



150,000 gallons of fuel saved resulting in 1,400 ton CO_2 reduction

DOHRN TRANSFER COMPANY

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

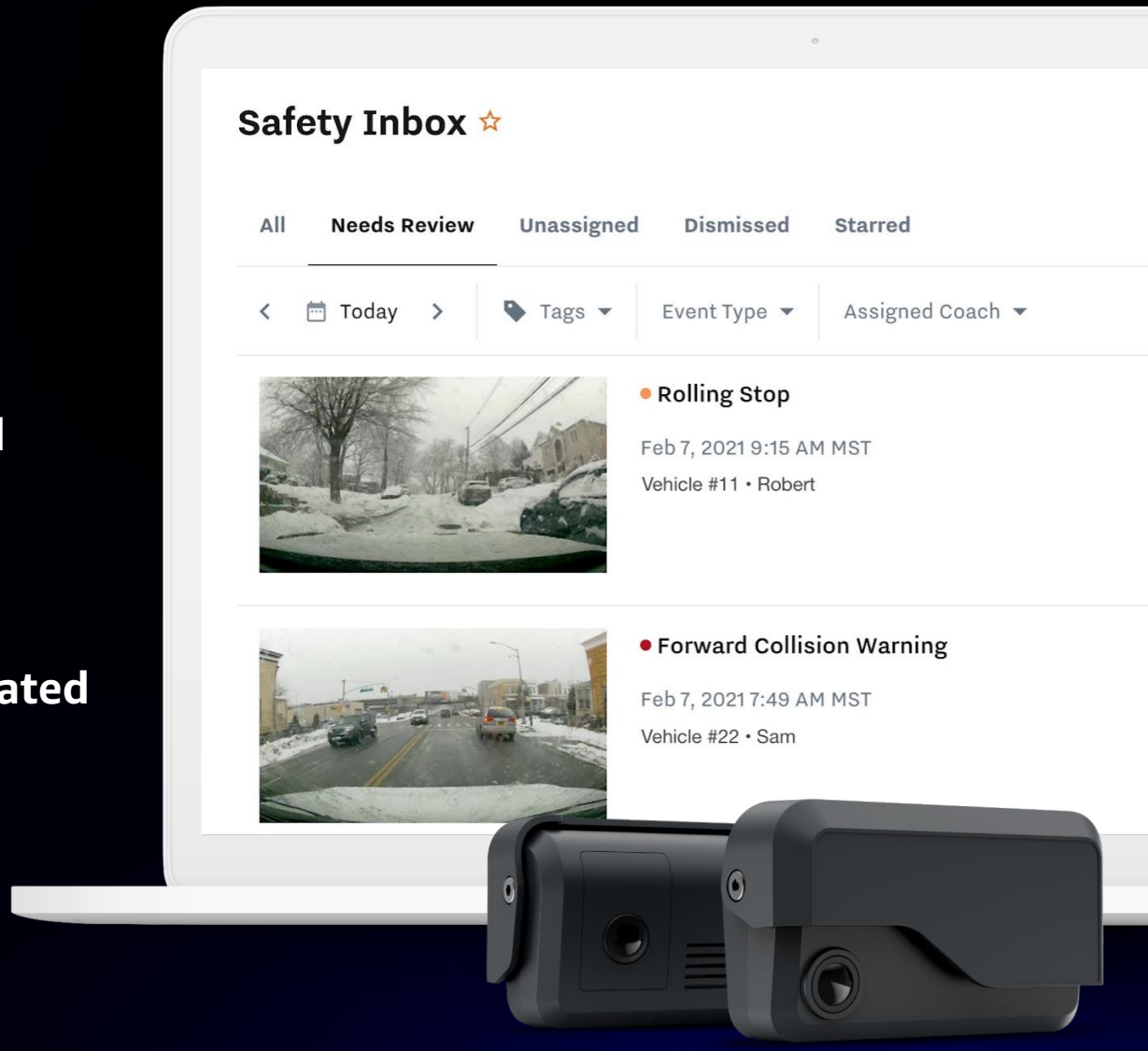
Products: Video-Based Safety

Internet-connected vehicle cameras and cloud-based safety dashboards for driver safety and coaching

- ✓ Gain real-time visibility with live video streaming and automatic incident detection
- ✓ Coach drivers effectively with preventative alerts and manager-led coaching workflows
- ✓ Streamline incident review and reporting with automated driver scorecards and reporting tools

Sprint Waste Services

Learn how Sprint Waste reduced costs by 50%



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Computer vision use cases that improve safety



AI Dash Cameras

Driver facing camera

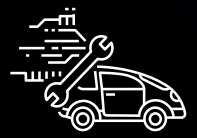
- Distracted driving
- Mobile phone usage
- Seatbelt compliance

Road facing camera

- Following distance
- Collision warning

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges with Training Data Generation



Collect data for variety of use cases

Lack of Open source Domain Specific Dataset



Annotation Tooling

Engineering efforts to build custom annotation tooling



Data Privacy & Security

Ensure Data privacy and security is maintained during labeling

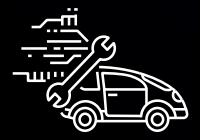


ML Training Infrastructure

Requires Seamless integration of labeled data with ML Training

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Ground Truth Plus helped address our challenges cost effectively



Collect data for variety of use cases

Managed variety of labeling use cases to meet the demand

Collaboration and developed custom annotation tooling



Annotation Tooling



Data Privacy & Security

Native SOC2 compliance and Data Privacy Controls

Packaged labels for seamless integration with data lake for ML training



ML Training Infrastructure



On Time Delivery 100K+ Labels per project with 6 projects running concurrently



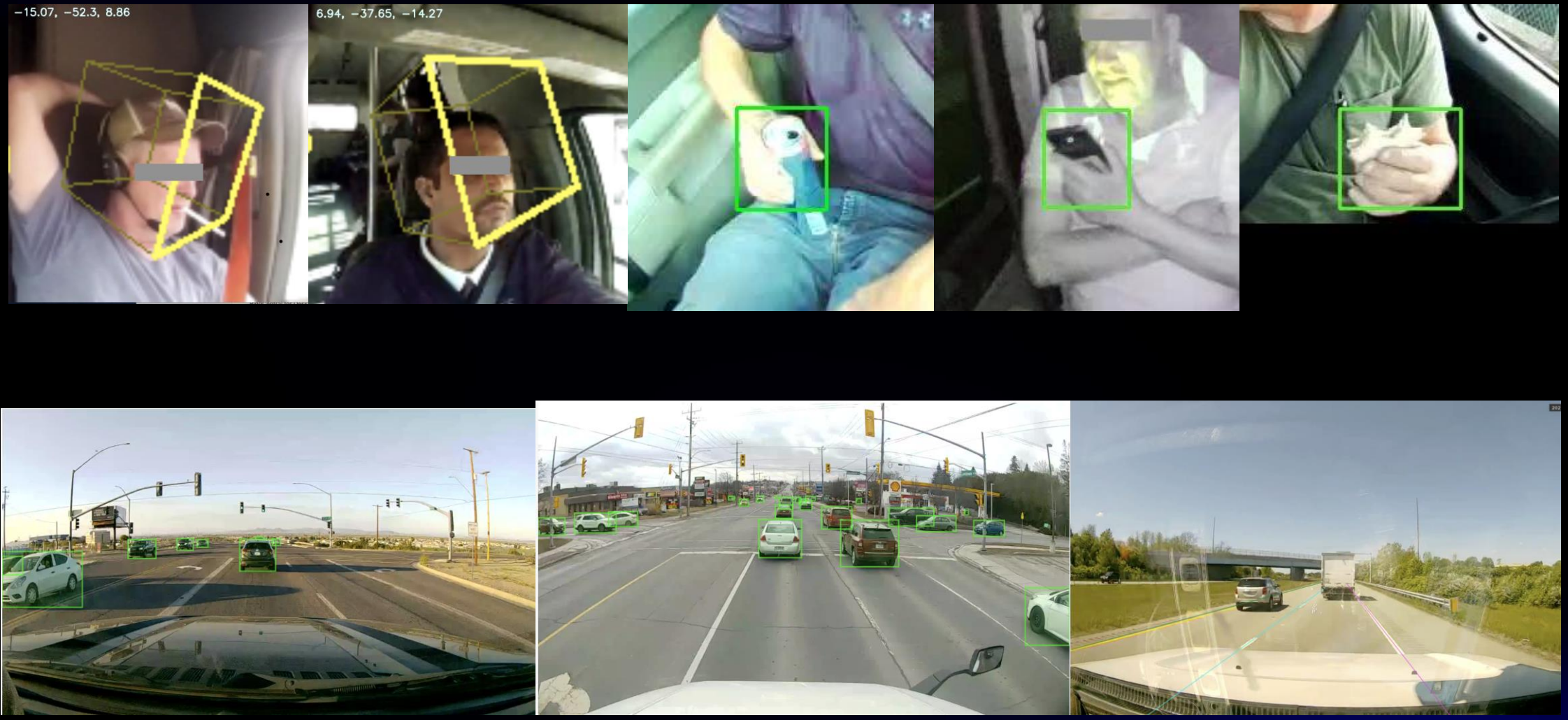
Address Quality issues with tight Feedback loop with Operations Manager



Volume Discount with standardized rate card and AWS Billing

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Labeled data samples



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting Started

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting started with Ground Truth Plus

Getting started Info

Ground Truth Plus is a turnkey data labeling service that enables you to easily create high-quality training datasets without having to build labeling applications and manage the labeling workforce on your own. Ground Truth Plus sets up data labeling workflows and manages them on your behalf. Learn more



Step 1. Request a pilot

Fill out the project requirement form and our team of AWS experts will schedule a call with you to discuss and set up your data labeling project.

Request a pilot



Step 2. Share data

After the consulation call, share raw data in an Amazon S3 bucket.



Step 3. Create project team

Create a project team after AWS experts set up your data labeling project. The project team helps you ensure that the right users have access to the project portal.

Create project team



Step 4. Access project portal

The project project portal allows you and your team to monitor project progress, track metrics, and provide feedback on the output data.

Open project portal



Step 5. Receive labeled data

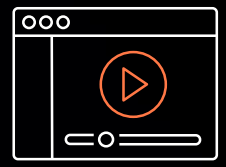
Collect high-quality labeled data securely in an Amazon S3 bucket.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Next steps



Schedule a consultation call



Execute a pilot workload at no cost

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.