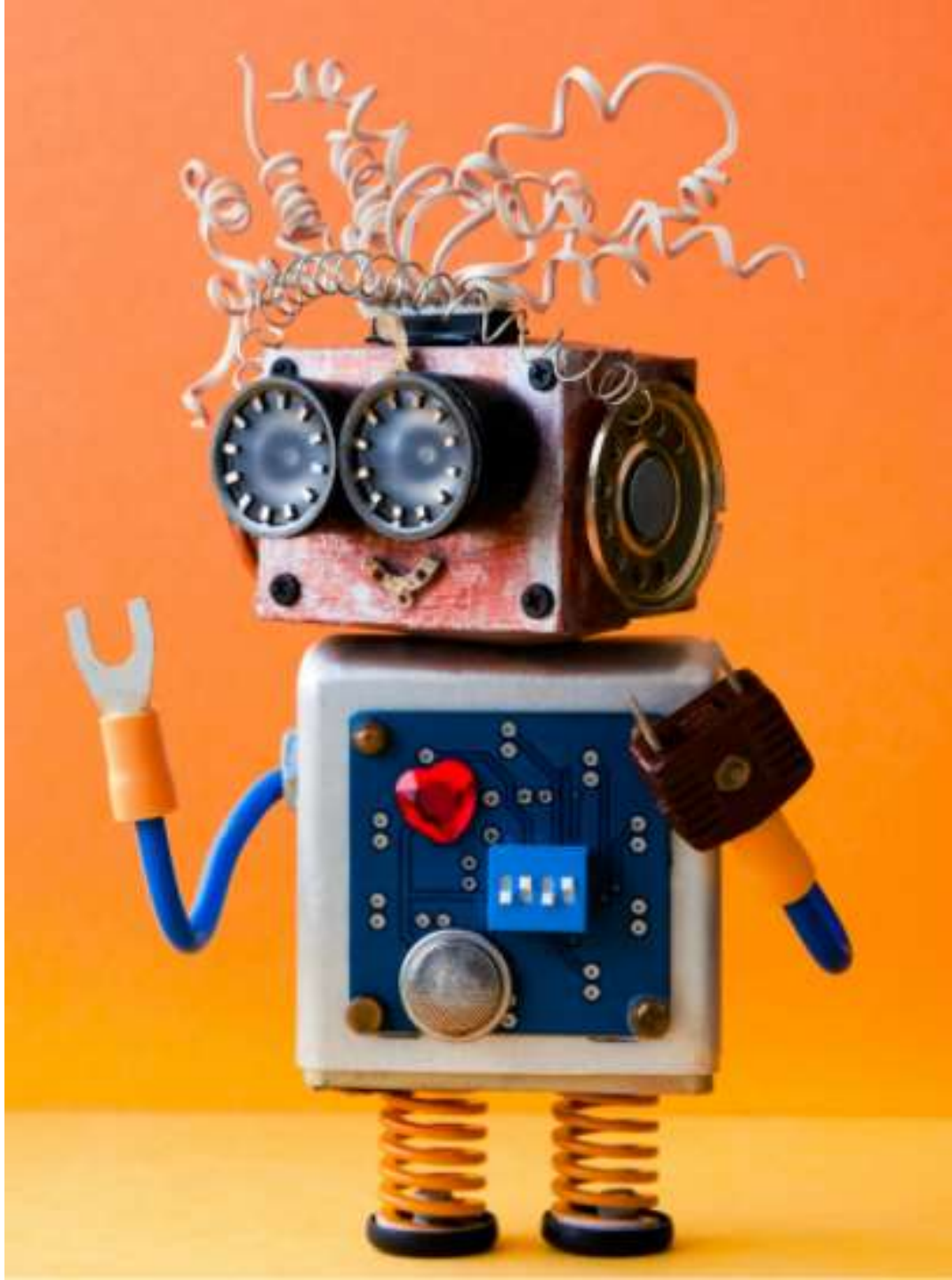
# PROJECT OVERVIEW & INTRO TO GROUNDTRUTH

EASY ADVANCED



# PROJECT OVERVIEW

- In this project, we will learn how to label images using Amazon SageMaker GroundTruth.
- These are the key learning outcomes:
  1. The need for labelled datasets
  2. Applications of supervised learning
  3. Challenges of obtaining labelled datasets
  4. Learn how to define a labeling job using Amazon SageMaker Groundtruth
  5. Learn how to label data in Amazon SageMaker Groundtruth
  6. Understand the concept of JsonL and manifest files

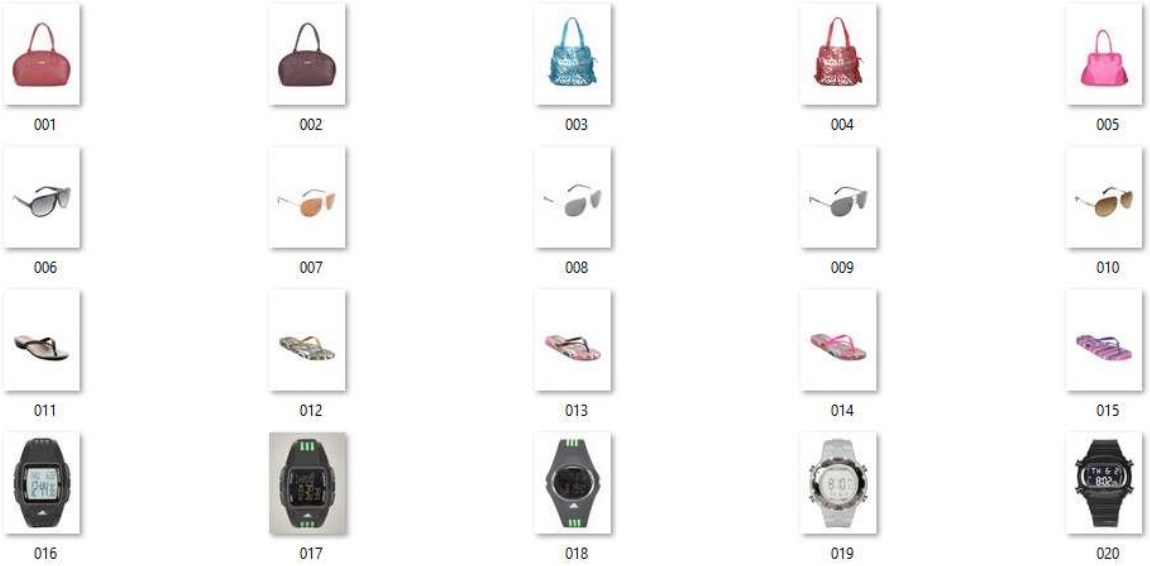
## 20 IMAGES BELONGS TO 4 CLASSES (BALANCED DATASET)

BAGS

EYEWEAR

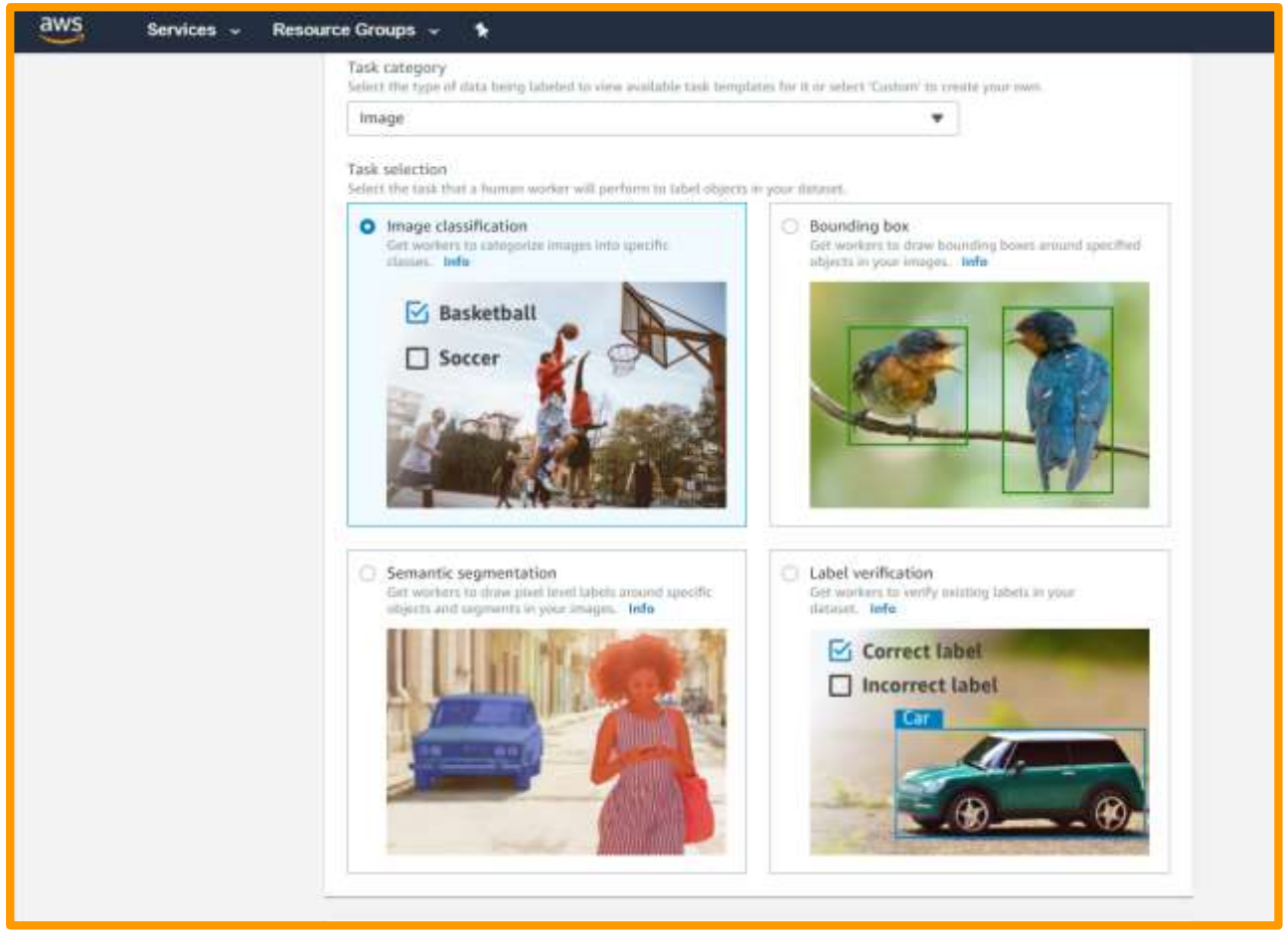
FLIPFLOPS

WATCHES



# SAGEMAKER GROUNDTRUTH 101

- AWS SageMaker GroundTruth is a service offered by AWS to label data.
- In machine learning terminology, Ground truth means “gold standard”!
- Ground Truth indicates the “true” or “real” class that you would like your model to learn how to predict.




**AMAZON SAGEMAKER GROUNDTRUTH**

# GROUNDTRUTH KEY LABELING TASKS AND FEATURES

- Several labeling tasks are available in SageMaker GroundTruth:
  - Bounding boxes
  - Image Classification
  - Semantic Segmentation
  - Text Classification
  - Custom Tasks
- Check this out: https://aws.amazon.com/sagemaker/data-labeling/features/

# AVAILABLE DATA LABELING WORKFORCES

**Public Mechanical Turks:** Amazon SageMaker facilitates the interaction between customers who require data labeling and an on-demand 24x7 global workforce of 500K contractors globally.

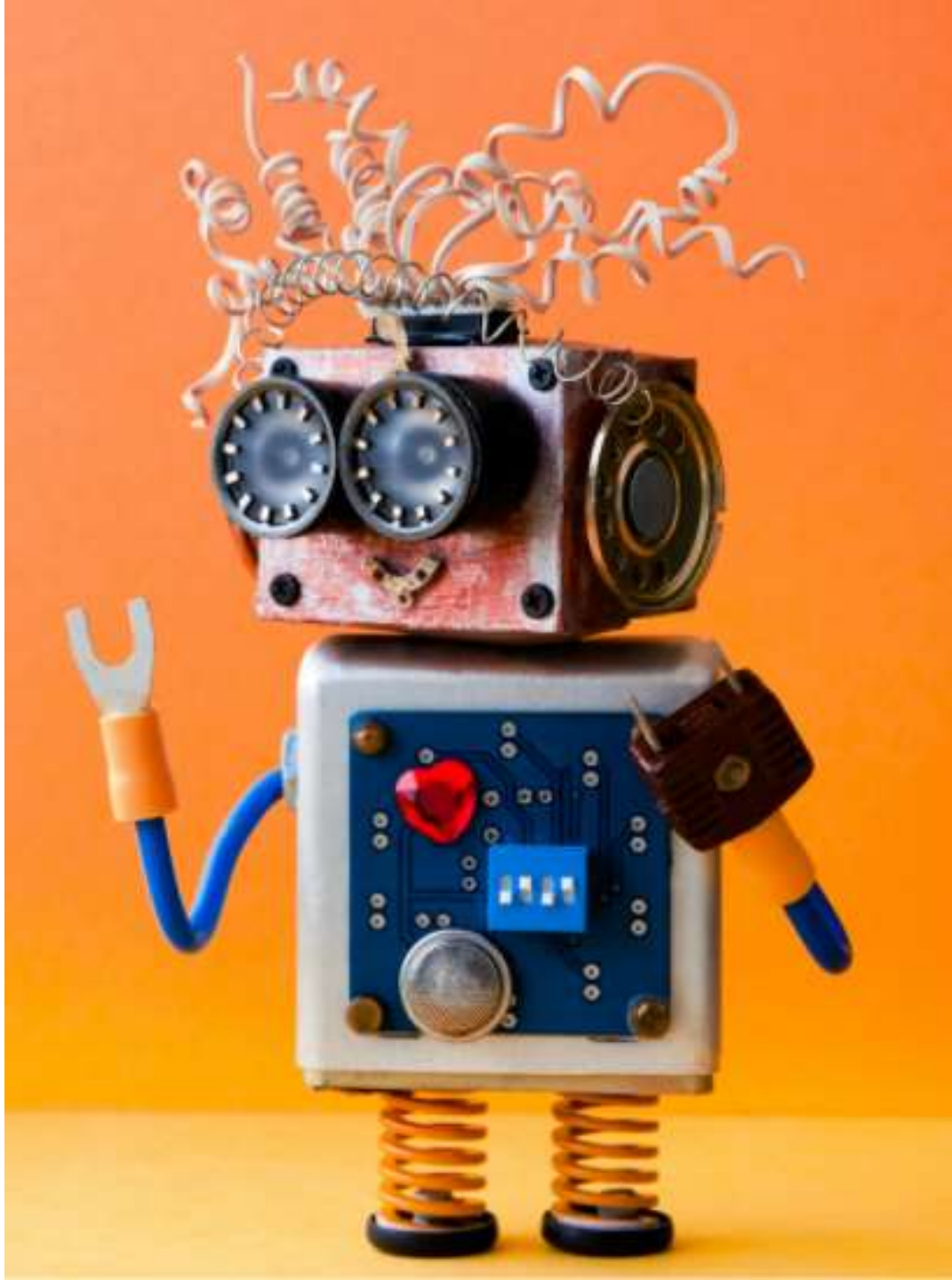
**Private:** A team of labelers can be specified by the customer including their own private labelers. Everything is managed through SageMaker Ground Truth. No need for labelers to have IAM or Amazon account.

**Vendors:** SageMaker Groundtruth provides a curated third-party vendors who can offer data labeling services.

# 21st CENTURY NEW GOLD

EASY

ADVANCED



# WHY DATA IS CONSIDERED THE NEW GOLD OF THE 21st CENTURY?

- The data revolution is here! Data is the new gold of the 21st Century.
- Companies nowadays have access to a massive amount of data and their competitive advantage lies in their ability to gain valuable insights from this data.
- Data can empower companies to boost their revenues, improve processes and reduce costs.
- Data could be leveraged in many industries such as finance, banking, healthcare, transportation, and technology sectors.

Image Source: https://www.flickr.com/photos/tao_zhyn/442965594

# DATA-DRIVEN SUCCESS STORIES

- Netflix leverages customer data to recommend new content to users using its data-driven recommender system which earns it ~$1B USD in customer retention.
- Amazon integrated data-driven recommendations at every stage of the purchasing process which resulted in 29% sales increase to $12.83 billion during its second fiscal quarter 2020, up from $9.9 billion during same time last year".

Source: https://medium.com/eleks-labs/4-powerful-use-cases-for-data-science-in-finance-35d50075ff80
Source: https://emerj.com/ai-sector-overviews/ai-in-banking-analysis/
Source: https://www.precisely.com/blog/big-data/beyond-big-data-examples-success
Photo: https://commons.wikimedia.org/wiki/File:Logonfx.png
Photo: https://commons.wikimedia.org/wiki/File:Amazon_PNG6.png

# DATA-DRIVEN SUCCESS STORIES

- “JP Morgan invests $11.5 billion/year in new data driven technologies. Its machine learning-based Contract Intelligence (COiN) platform reviews 12,000 commercial loan agreements in few hours compared to 360,000 man-hours it would take to do so manually.”
- “Banking institutions prevented $22 billion worth of fraudulent transactions in 2018 with the power of AI/ML”.
- “Bank of America introduced Erica chatbot that served 6 million users as of March 2019.”
- Electronic trades account for almost 45% of revenues in cash equities trading” U.K. research firm Coalition Report.

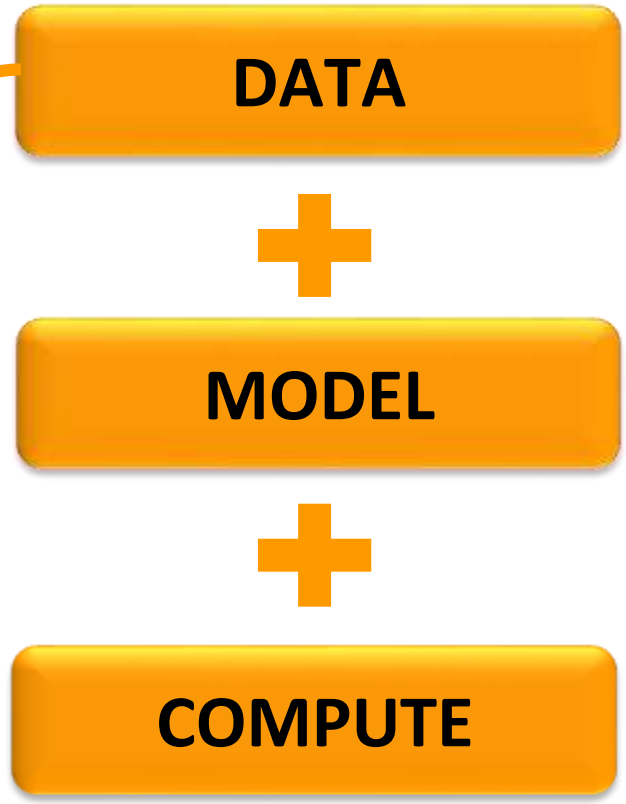


Source: https://medium.com/eleks-labs/4-powerful-use-cases-for-data-science-in-finance-35d50075ff80
Source: https://emerj.com/ai-sector-overviews/ai-in-banking-analysis/
Source: https://www.precisely.com/blog/big-data/beyond-big-data-examples-success
Photo: https://www.flickr.com/photos/bensutherland/178395814
Photo: https://www.flickr.com/photos/moneyblognewz/5280927344

# CORPORATE INTELLIGENCE WITH AI

- Artificial intelligence is the science that empowers computers to mimic human intelligence such as decision making, text processing, and visual perception.
- In order to train Artificial intelligence and Machine Learning models, companies need a massive amount of data.

THIS WILL BE OUR FOCUS! WE WILL LEARN HOW TO LABEL, CLEAN, CONCATENATE, MERGE, AND VISUALIZE THE DATA USING PANDAS AND THEN LEARN HOW TO DO THIS IN AWS SAGEMAKER

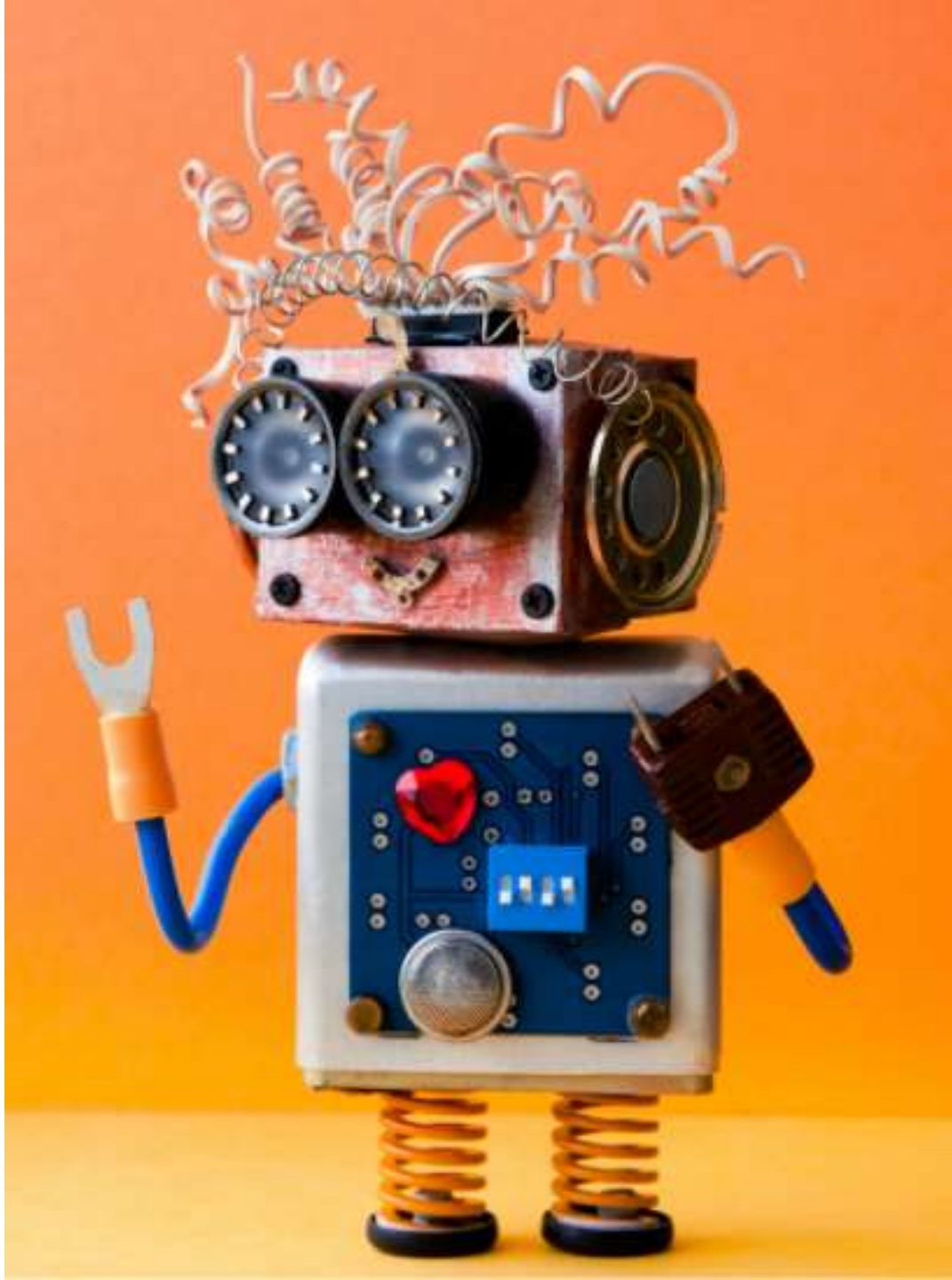



https://www.flickr.com/photos/mikemacmarketing/30212411048

# DATA SOURCES AND TYPES

EASY

ADVANCED



# DATA SOURCES

- Data can come from so many sources and forms such as images, audio, video, and text.
- Collecting, structuring and analysing this data is critical for companies to gain customers insights and set their marketing and product strategies.
- “Data preparation accounts for about 80% of the work of data scientists”, Forbes.

**IMAGE/VIDEO**

**TEXT (CORPUS)**

**AUDIO/SOUND**

**TIMESERIES/SIGNALS**

*Source: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=410bbd6a6f63*

Photo Credit: https://pxhere.com/en/photo/1454351
Photo Credit: https://www.flickr.com/photos/29881930@N00/2086641598
Photo Credit: https://commons.wikimedia.org/wiki/File:Mobile_phone_text_messages.jpg
Photo Credit: https://en.wikipedia.org/wiki/File:Messages_Yosemite.svg
Photo Credit: https://www.pexels.com/photo/blue-and-yellow-graph-on-stock-market-monitor-159888/

# DATA TYPES: LABELED VS. UNLABELED DATASET

- There are generally two types of data that we could use to train AI models.

**UNLABELED DATASET**

*Unlabeled data consists of data that does not have explanation (class or tag) associated with it.*

**LABELED DATASET**

**LABEL = "CAT"**

**LABEL = "DOG"**

*Labeled data consists of unlabeled data but with a "class" or "tag" associated with it.*

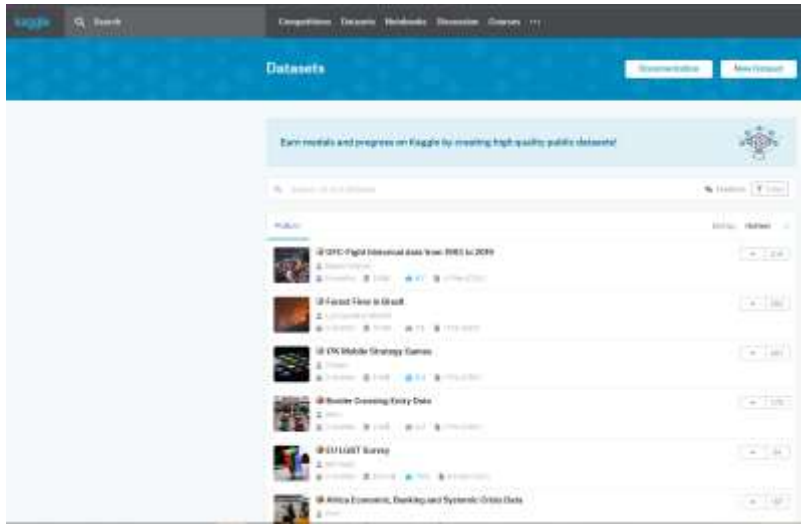
Photo Credit: https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/
Photo Credit: https://www.pexels.com/photo/portrait-of-a-dog-257540/

# GOOD Vs. BAD DATA

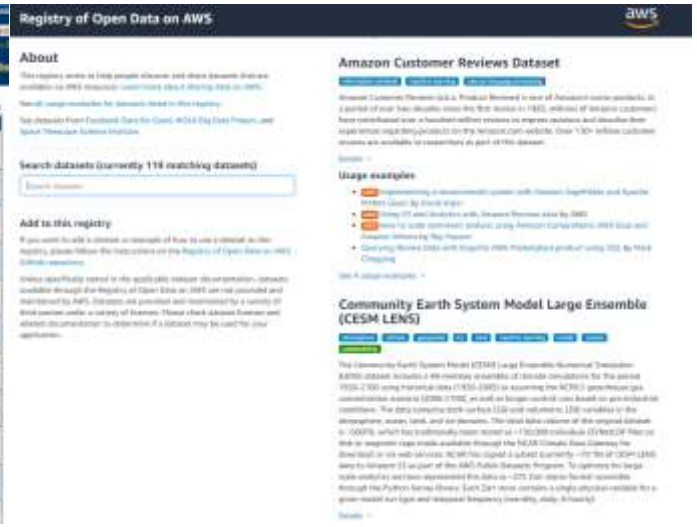
| GOOD DATA | BAD DATA |
| --- | --- |
| Many samples (large number of data points) | Few samples (small number of data points) |
| Not Biased | Biased |
| Does not contain missing data points | Contains missing data points |
| Only contains (relevant) important features | Contains many irrelevant (useless) features |
| Does not contain duplicate samples | Contains duplicate samples |

# WHERE DOES THIS DATA COME FROM?

- Data could also come from multiple sources such as Kaggle, UCI, and AWS Dataset.
- ImageNet is an open source repository of images consisting of 21,841 subcategories (classes) and over 14 million images.
- AWS dataset registery: https://registry.opendata.aws/

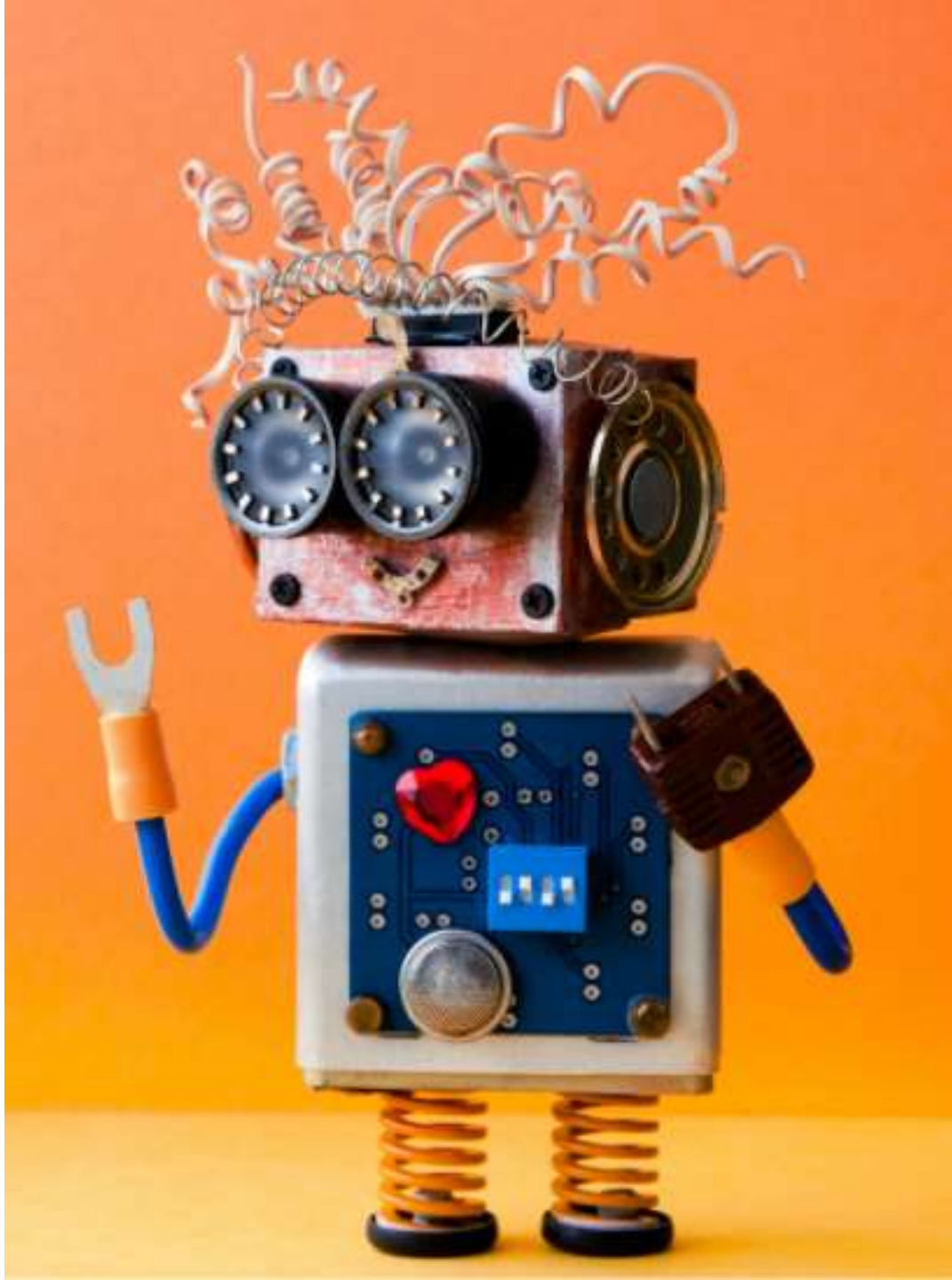







[Check out website here: https://archive.ics.uci.edu/ml/datasets.php](https://archive.ics.uci.edu/ml/datasets.php)
[Check out website here: https://www.kaggle.com/datasets](https://www.kaggle.com/datasets)

# WHY DO WE NEED LABELED DATA?

# ML TRAINING STRATEGIES

## Supervised Learning

- Used in cases where large dataset with known labels (outputs) are available.
- The learning algorithm evaluates output (i.e.: makes predictions), compares output against the label, and adjust model weights/parameters and repeat.

## Unsupervised Learning

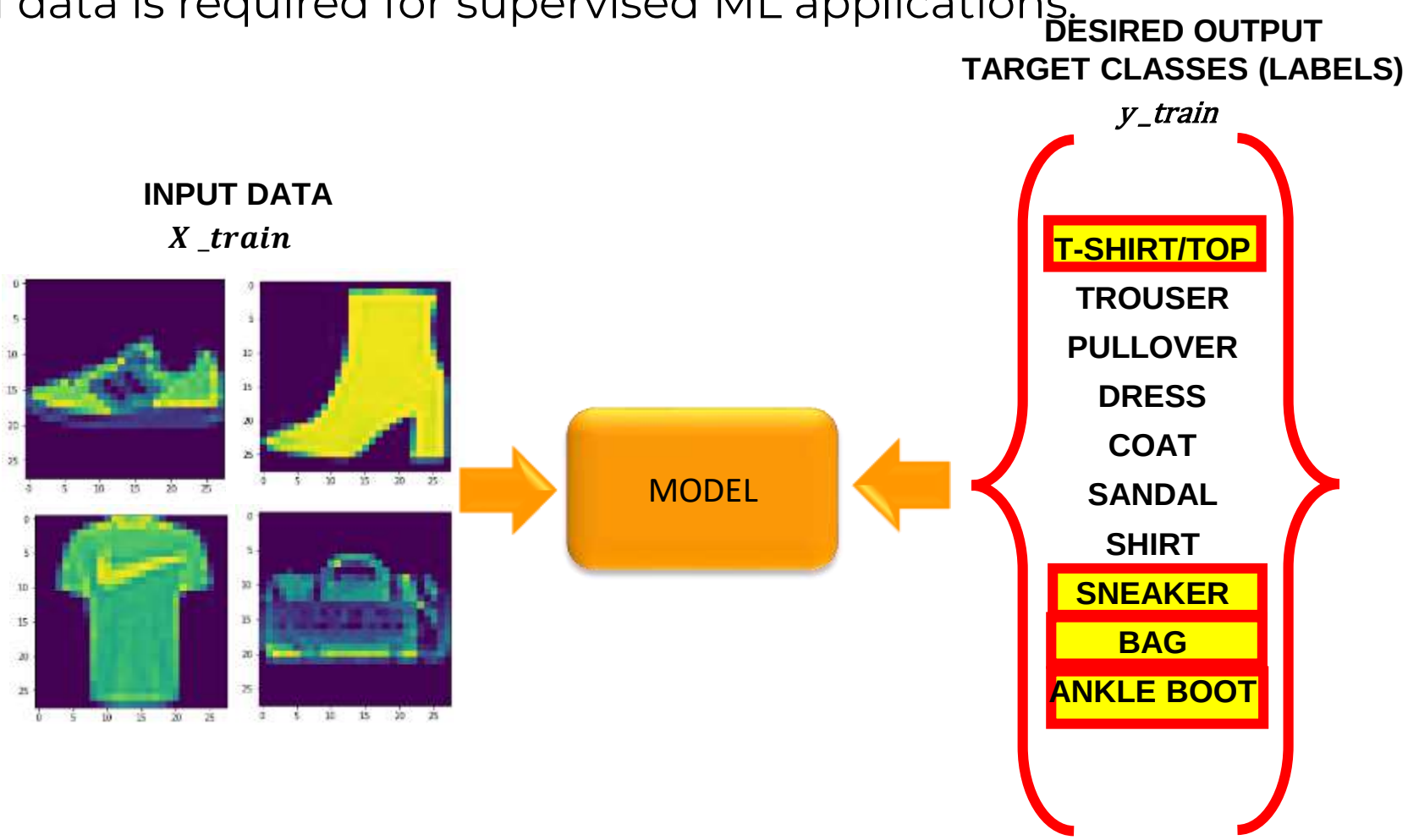
- Used with "unlabeled" data (not categorized) (Ex: k-means clustering).
- Since learning algorithm works with unlabeled data, there is no way to assess the accuracy of the structure suggested by the algorithm

## Reinforced Learning

- Learning algorithm takes actions that maximizes cumulative reward.
- Over time, ML models learn to prefer the right kind of action and avoid the wrong one.

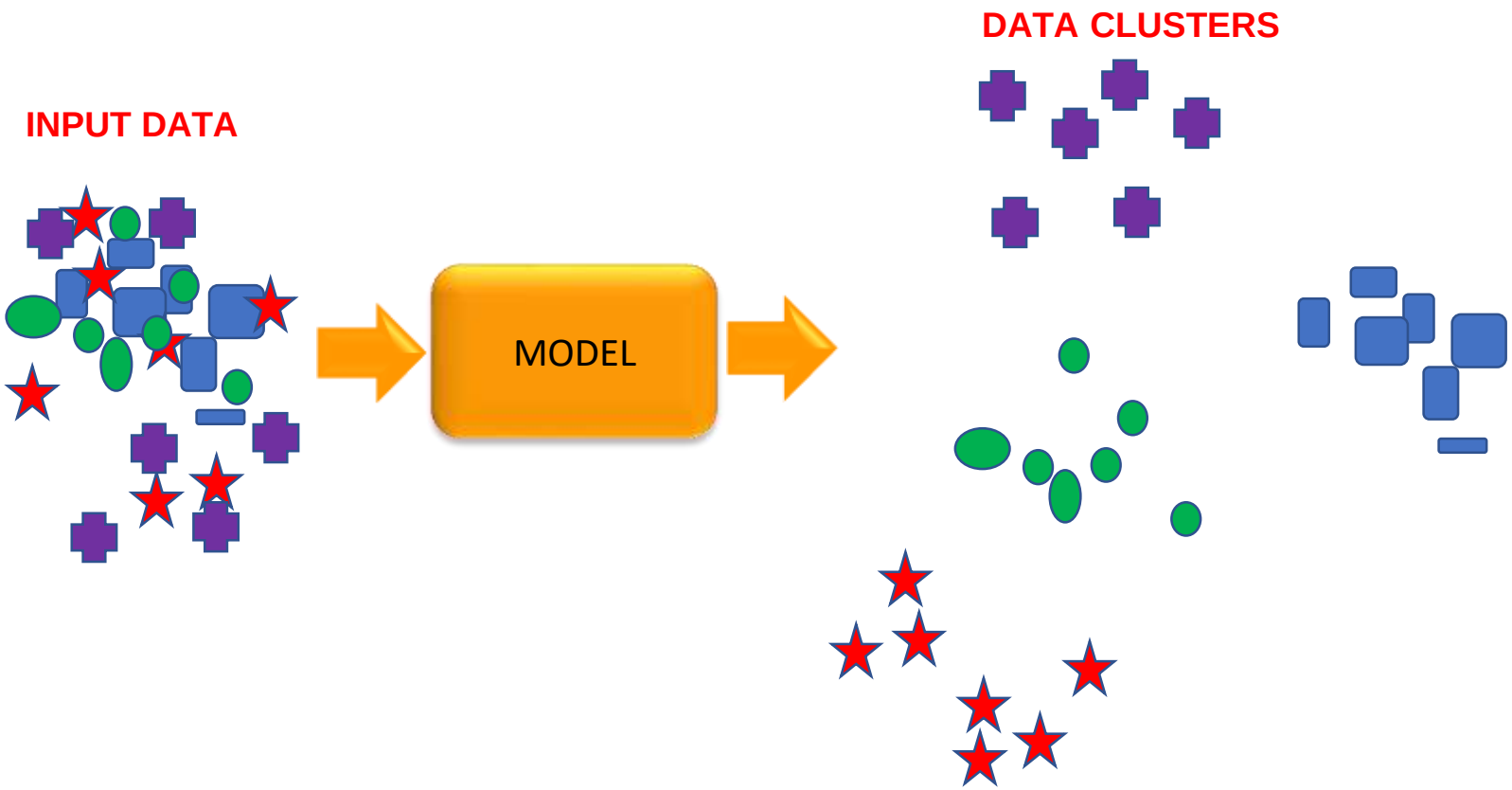
# MACHINE LEARNING: SUPERVISED LEARNING

- Supervised: used to train algorithms using labeled input and output data.
- Performance is assessed by comparing trained model prediction vs. real output.
- Labeled data is required for supervised ML applications.



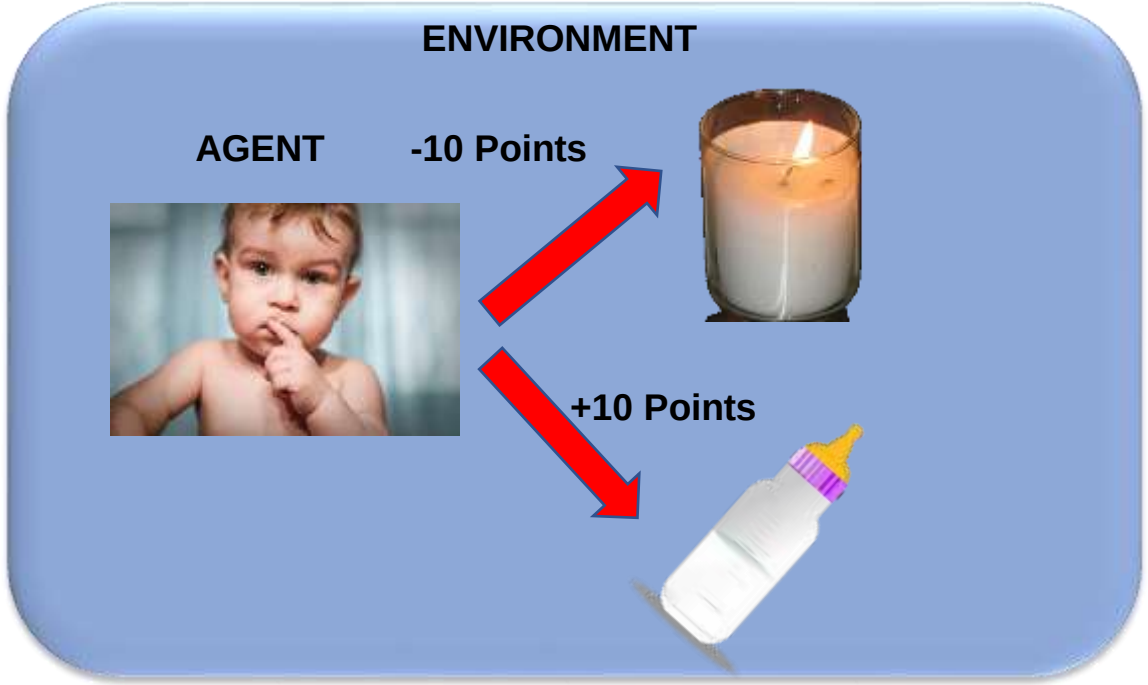
# MACHINE LEARNING: UNSUPERVISED LEARNING

- Unsupervised learning: provides the algorithm with no labeled data.
- The algorithm attempts at discovering hidden patterns within the training data.
- Unsupervised learning methods can analyze complex data that humans might find difficult to interpret.
- No feedback!



# MACHINE LEARNING: REINFORCEMENT LEARNING

- Reinforcement learning allows machines take actions to maximize cumulative reward.
- Reinforcement algorithms learn by trial and error through reward/penalty.
- Two elements: environment and learning agent.
- The environment rewards the agent for correct actions.
- Based on the reward or penalty, agent improves its environment knowledge to make better decision.



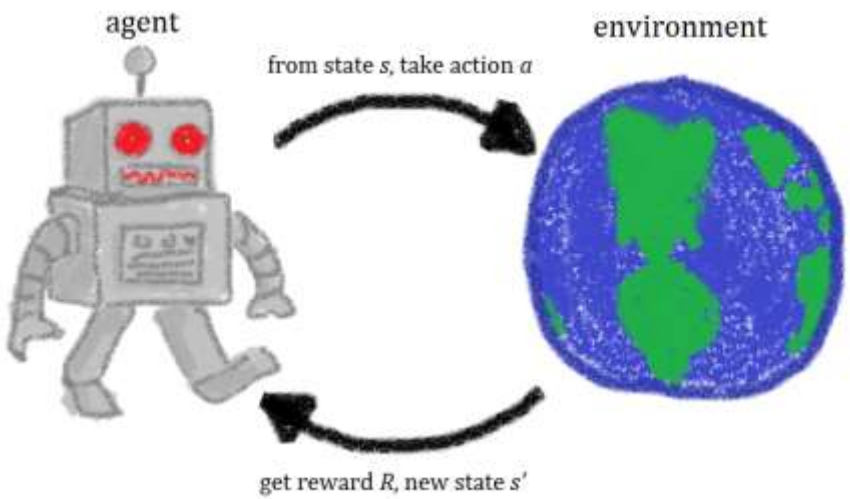
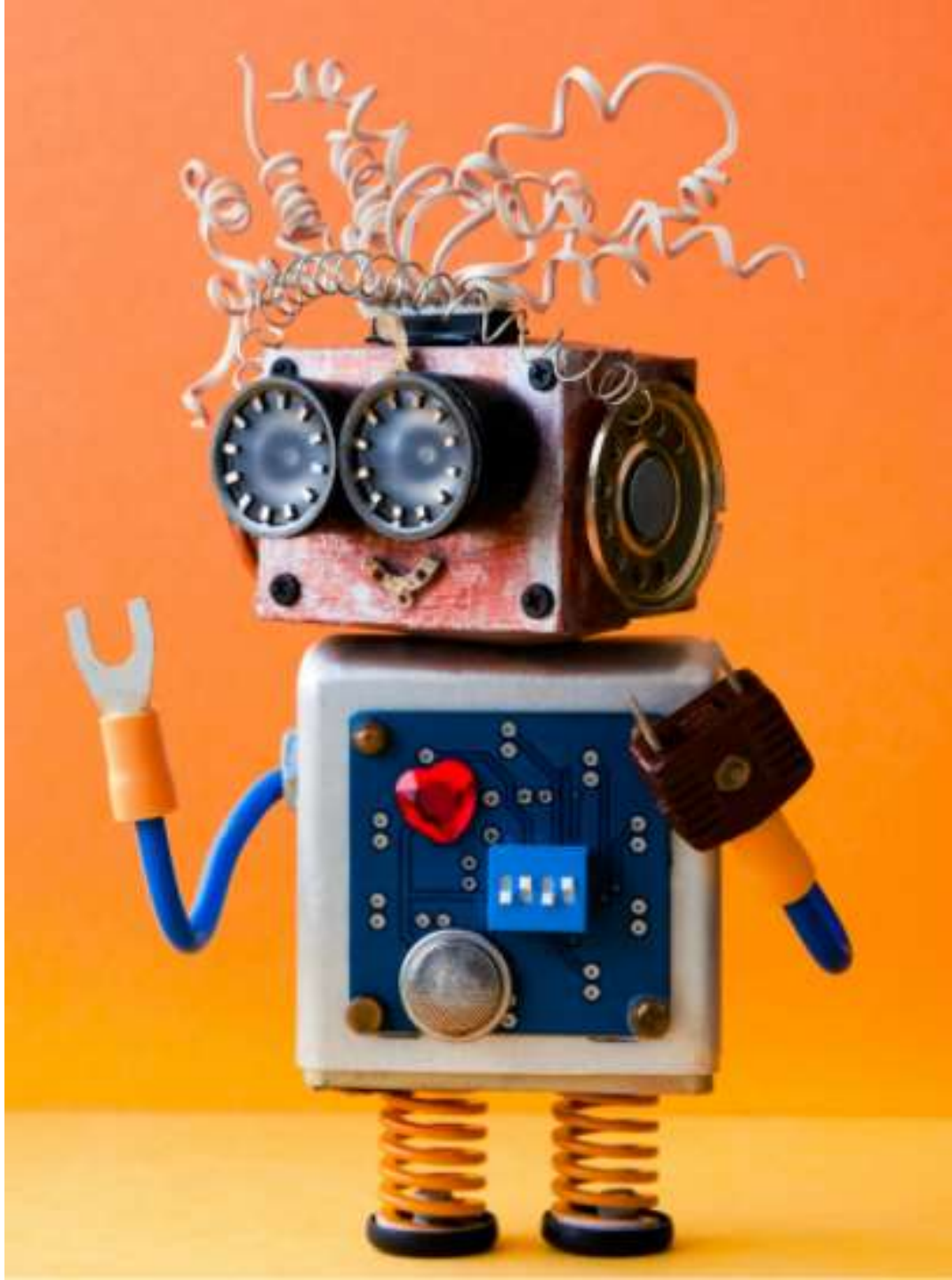




**Photo Credit:** https://commons.wikimedia.org/wiki/File:RI_agent.png

# DATA LABELING CHALLENGES & APPLICATIONS

EASY

ADVANCED



# DATA LABELING CHALLENGES

Modern state of the art deep learning models require a massive amount of labeled datasets

Large team of Human labelers are required to do the labeling

Inherent biases in human labelers require controls/standardization

Most data scientists' time is spent curating and labeling data

# USE CASES OF DATA LABELING

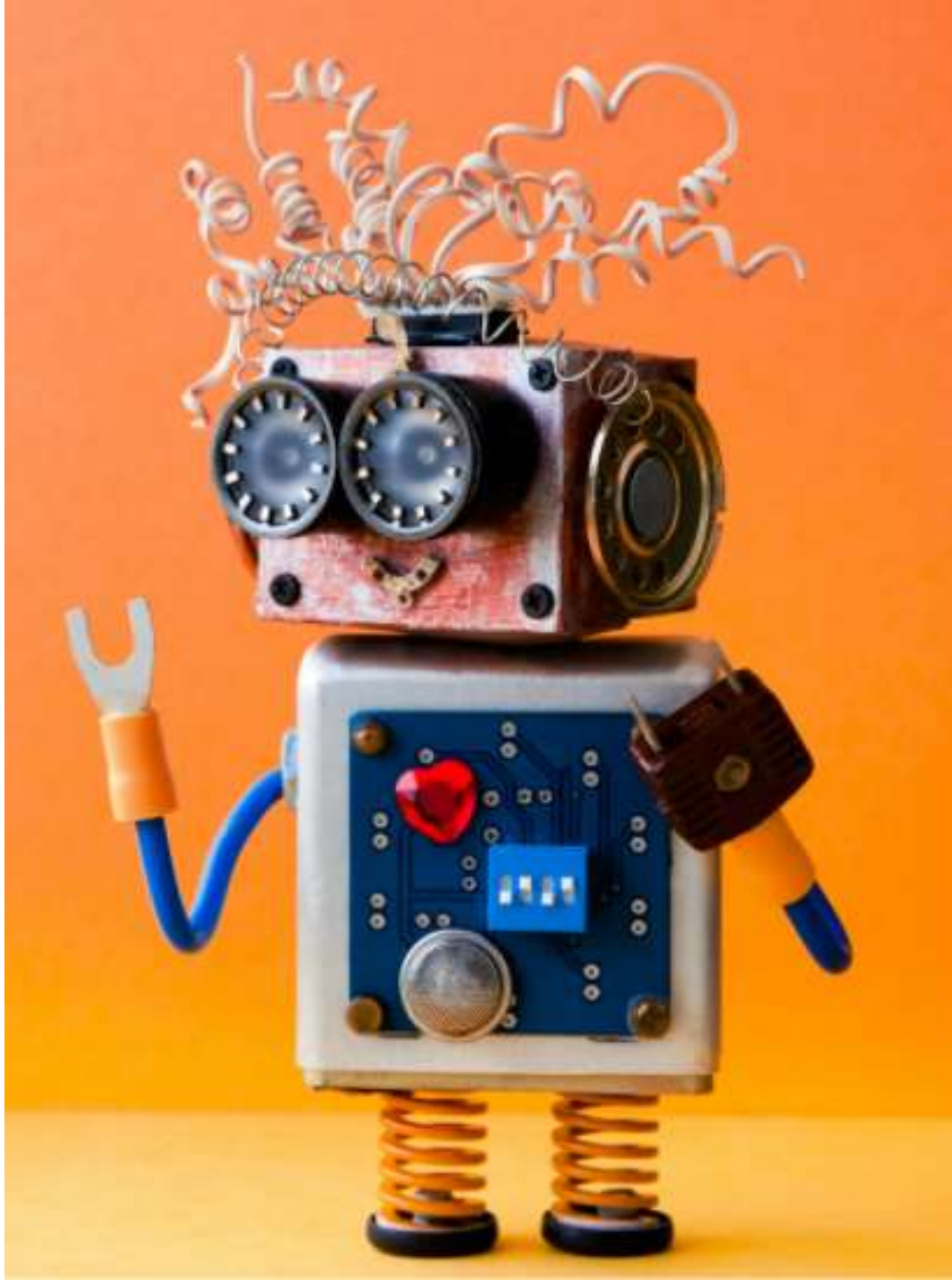
Text Data Analysis (News and Social Media Feeds)

Manufacturing Services: identify defects from images

Self-Driving Cars: label pedestrians, vehicles and traffic devices

AI-powered precision Agriculture: Identify precise locations of crops vs. weeds from images

# JSON LINES FORMAT & MANIFEST FILES

# JSON LINES (JSONL)

- JSON Lines text format is a format used for storing structured data that could be processed one record at a time.
- JSON Line format is generally used to store data labels.

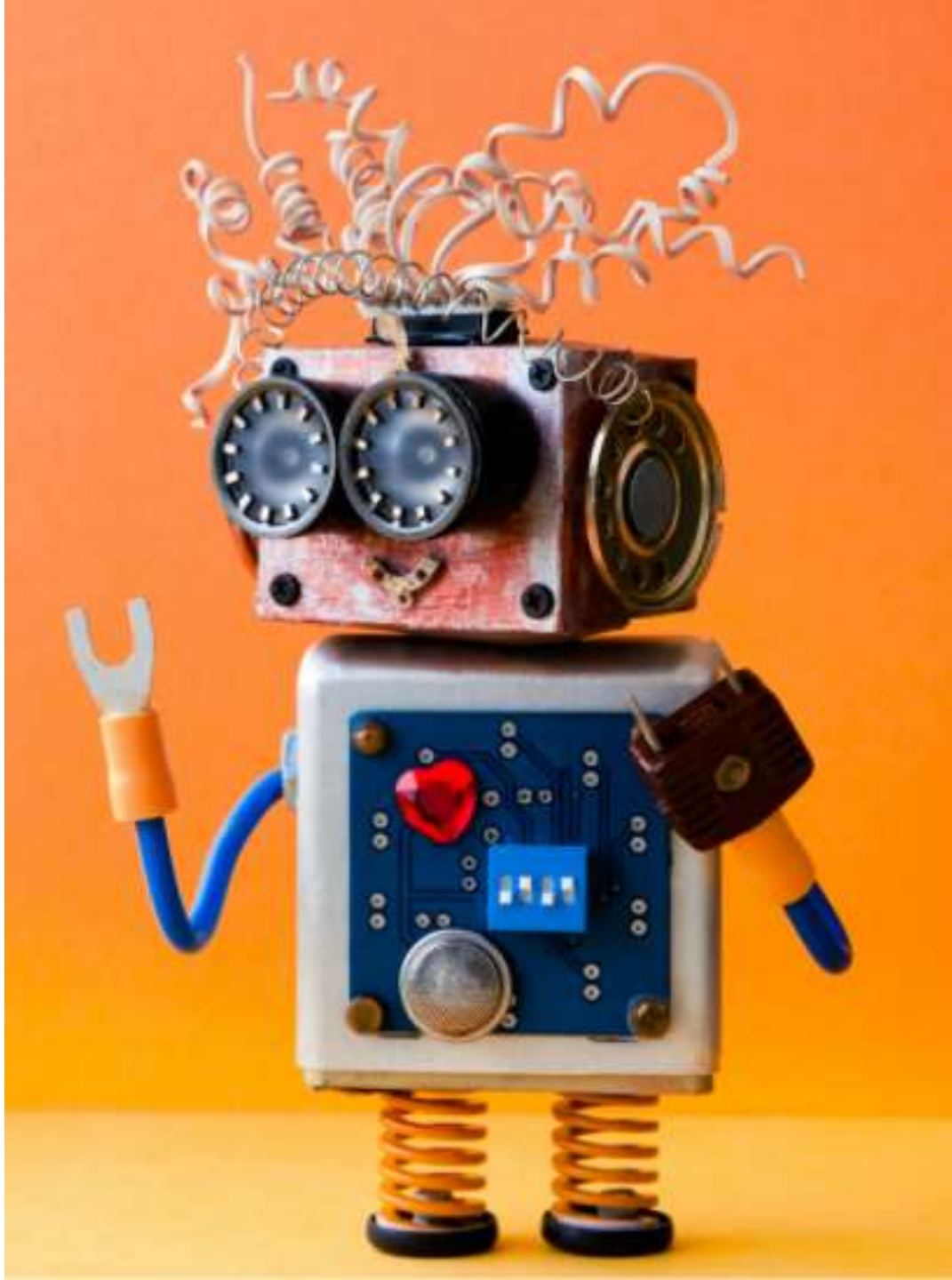
```
{“Image 1”: “Cat”}
{“Image 2”: “Dog”}
{“Image 3”: “Lion”}
```

# MANIFEST FILES 101

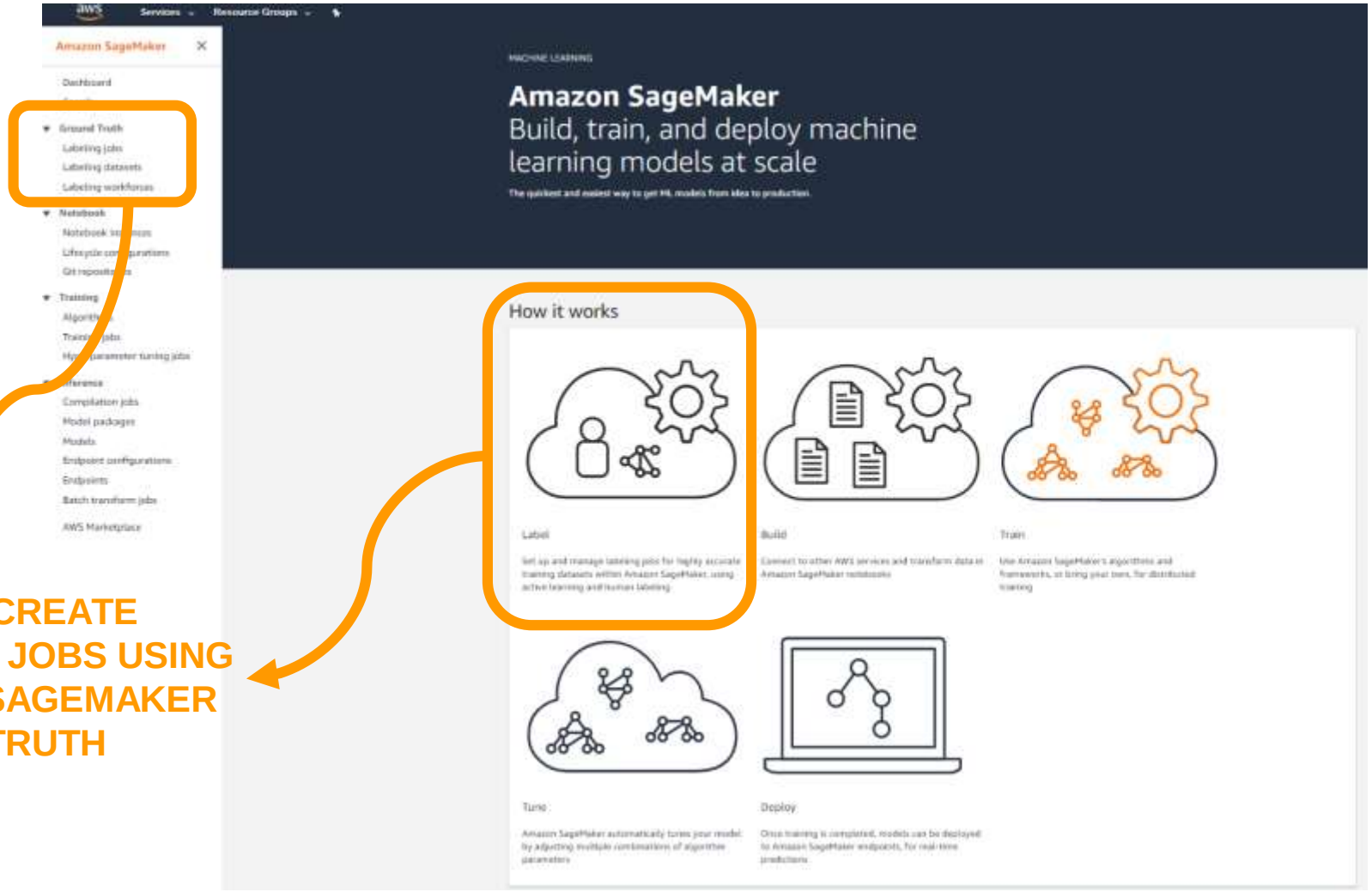
- In computer vision classification-type problems (Supervised ML applications), a manifest file can be useful since it contains data about the inputs (images) and outputs (labels).
- When you run an Amazon SageMaker GroundTruth Job, the output from this process includes a manifest file.
- Manifest files are in JSON lines format where each line is a complete JSON object representing the labeling information for an image.
- Each manifest file itself contains N JSON objects, where N is the number of images that are used in this dataset.


```
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/severe/c3cd0200df79.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "severe", "human-annotated":
"yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/severe/913490237ad4.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "severe", "human-annotated":
"yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/severe/a80dab8eddf4.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "severe", "human-annotated":
"yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/severe/910bfd38e2f5.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "severe", "human-annotated":
"yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/severe/f6f433f3306f.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "severe", "human-annotated":
"yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
{"source-ref": "s3://aws-ml-engineer/diabetic-retinopathy/train/healthy/3b2b91590590.png", "auto-label": 1, "auto-
label-metadata": {"confidence": 1, "job-name": "labeling-job/auto-label", "class-name": "healthy", "human-
annotated": "yes", "creation-date": "2017-03-01", "type": "groundtruth/image-classification"}}
```

# DATA LABELING IN SAGEMAKER GROUNDTRUTH DEMO – PART 1

# HOW TO OBTAIN LABELED DATA USING AWS? SAGEMAKER GROUNDTRUTH

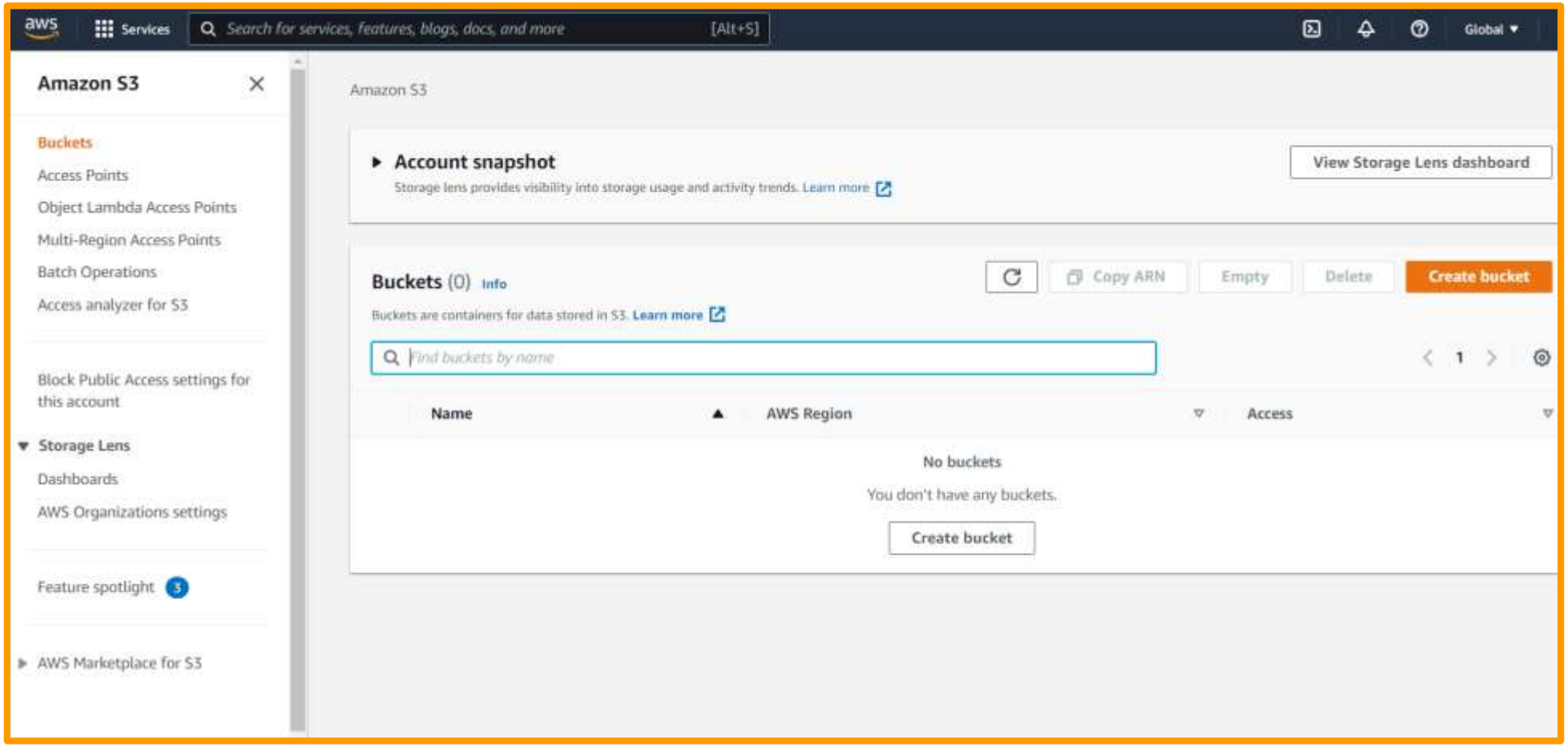



YOU CAN CREATE LABELING JOBS USING AMAZON SAGEMAKER GROUND TRUTH SERVICE

https://aws.amazon.com/sagemaker/groundtruth/

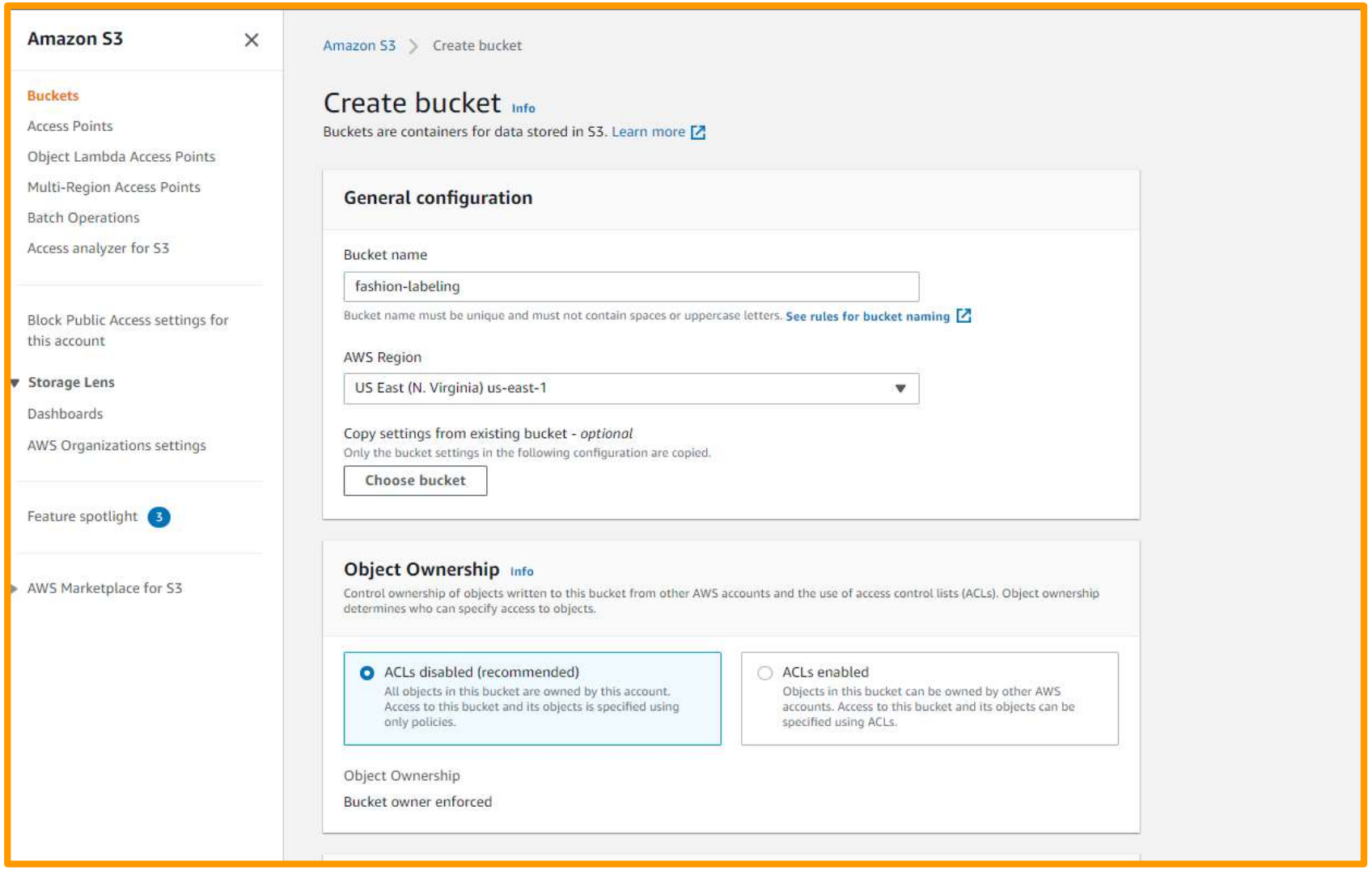
# SAGEMAKER GOUNDTRUTH DEMO

UPLOAD IMAGES TO S3, GO TO S3 AND
CLICK ON CREATE BUCKET



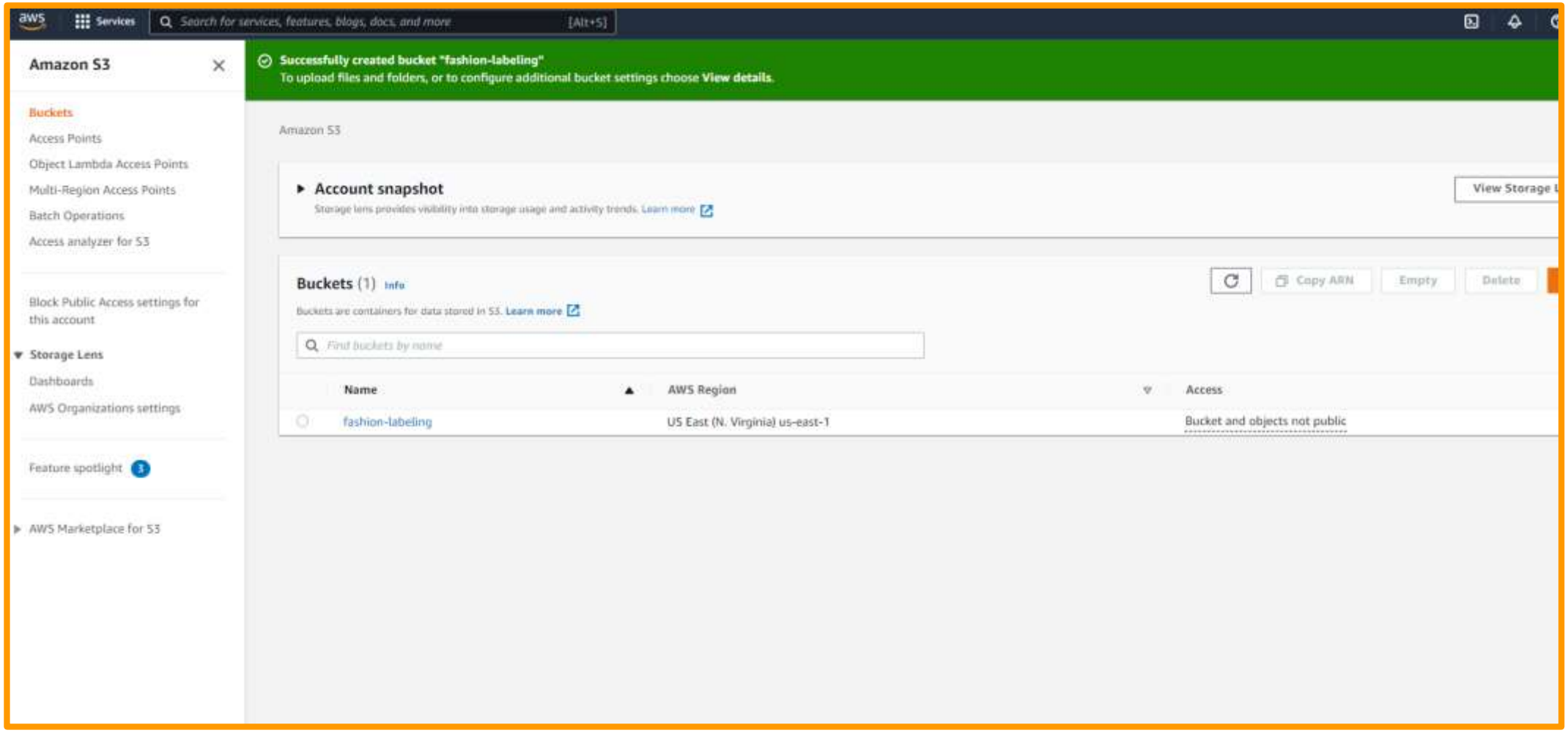
# SAGEMAKER GOUNDTRUTH DEMO

## GIVE IT NAME AND CLICK CREATE BUCKET THEN UPLOAD IMAGES TO S3



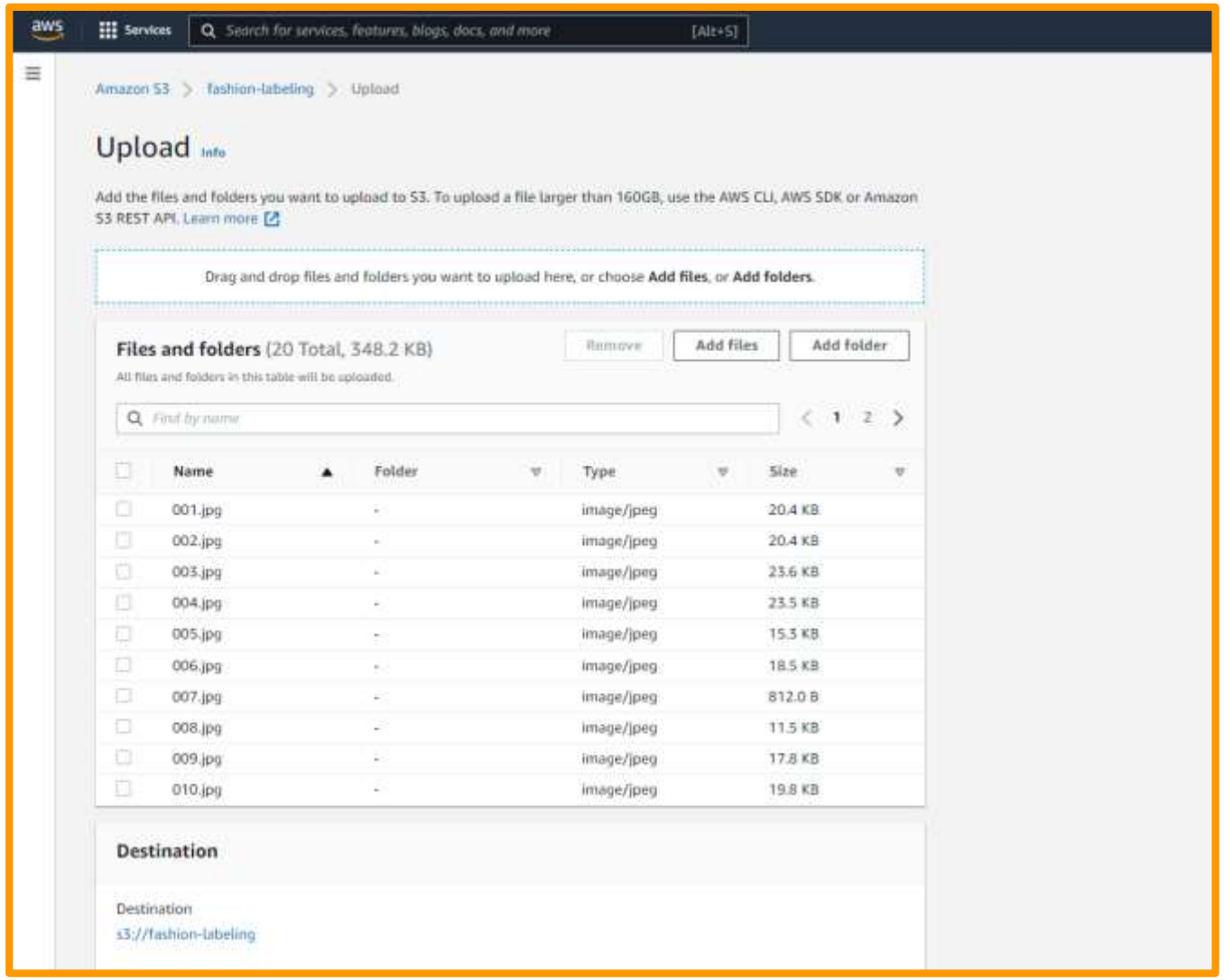
# SAGEMAKER GOUNDTRUTH DEMO

CLICK ON THE NEWLY CREATED BUCKET
AND UPLOAD THE IMAGES



# SAGEMAKER GOUNDTRUTH DEMO

## CLICK UPLOAD



# SAGEMAKER GOUNDTRUTH DEMO

## IMAGES ARE NOW UPLOADED TO S3

aws Services Search for services, features, blogs, docs, and more [Alt+S]

Upload succeeded
View details below.

The information below will no longer be available after you navigate away from this page.

### Summary

| Destination | Succeeded | Failed |
|---|---|---|
| s3://fashion-labeling | 20 files, 348.2 KB (100.00%) | 0 files, 0 B (0%) |

Files and folders | Configuration

### Files and folders (20 Total, 348.2 KB)

Find by name

| Name | Folder | Type | Size | Status | Error |
|---|---|---|---|---|---|
| 001.jpg | - | image/jpeg | 20.4 KB | Succeeded | - |
| 002.jpg | - | image/jpeg | 20.4 KB | Succeeded | - |
| 003.jpg | - | image/jpeg | 23.6 KB | Succeeded | - |
| 004.jpg | - | image/jpeg | 23.5 KB | Succeeded | - |
| 005.jpg | - | image/jpeg | 15.3 KB | Succeeded | - |
| 006.jpg | - | image/jpeg | 18.5 KB | Succeeded | - |
| 007.jpg | - | image/jpeg | 812.0 B | Succeeded | - |
| 008.jpg | - | image/jpeg | 11.5 KB | Succeeded | - |
| 009.jpg | - | image/jpeg | 17.8 KB | Succeeded | - |
| 010.jpg | - | image/jpeg | 19.8 KB | Succeeded | - |

# SAGEMAKER GOUNDTRUTH DEMO

## 20 IMAGES BELONGS TO 4 CLASSES (BALANCED DATASET)

BAGS


001


002


003


004


005

EYEWEAR


006


007


008


009


010

FLIPFLOPS


011


012


013


014


015

WATCHES


016


017

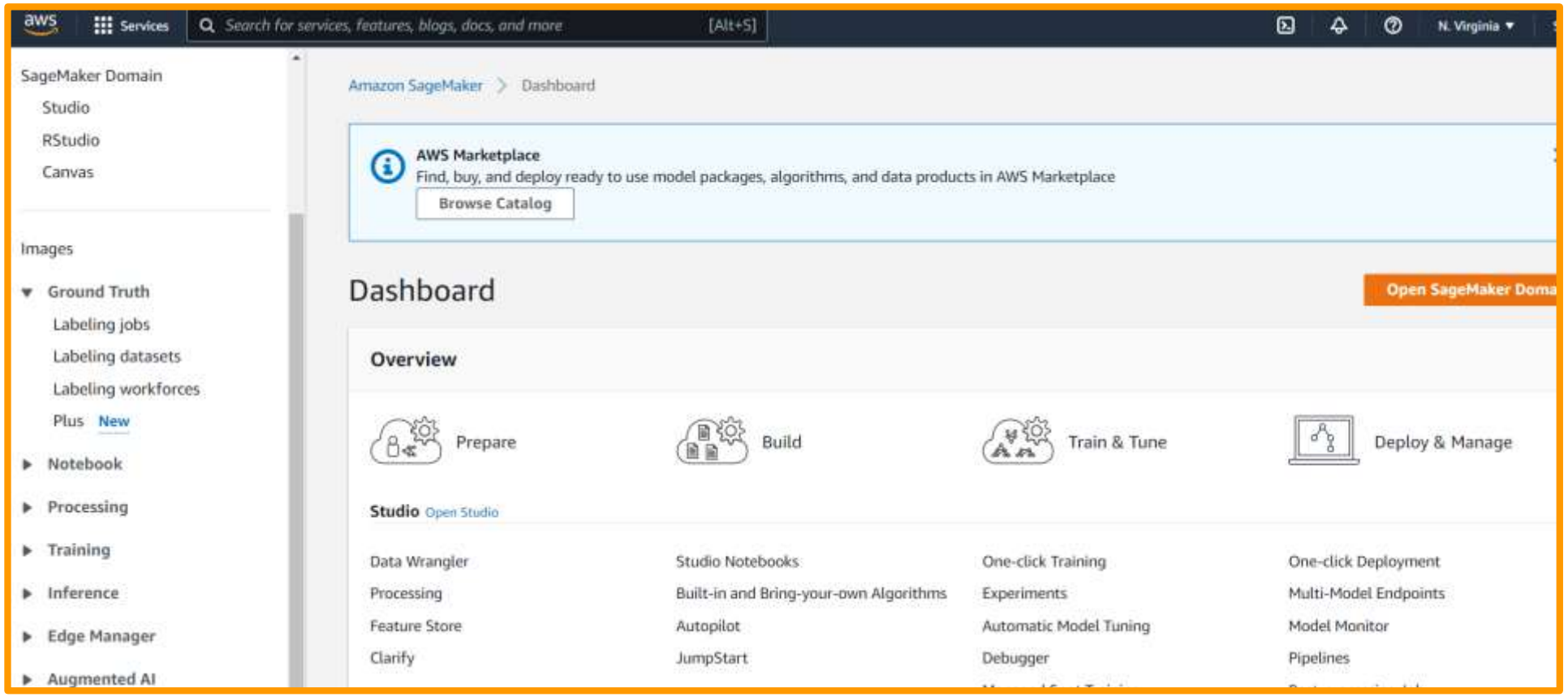

018


019


020

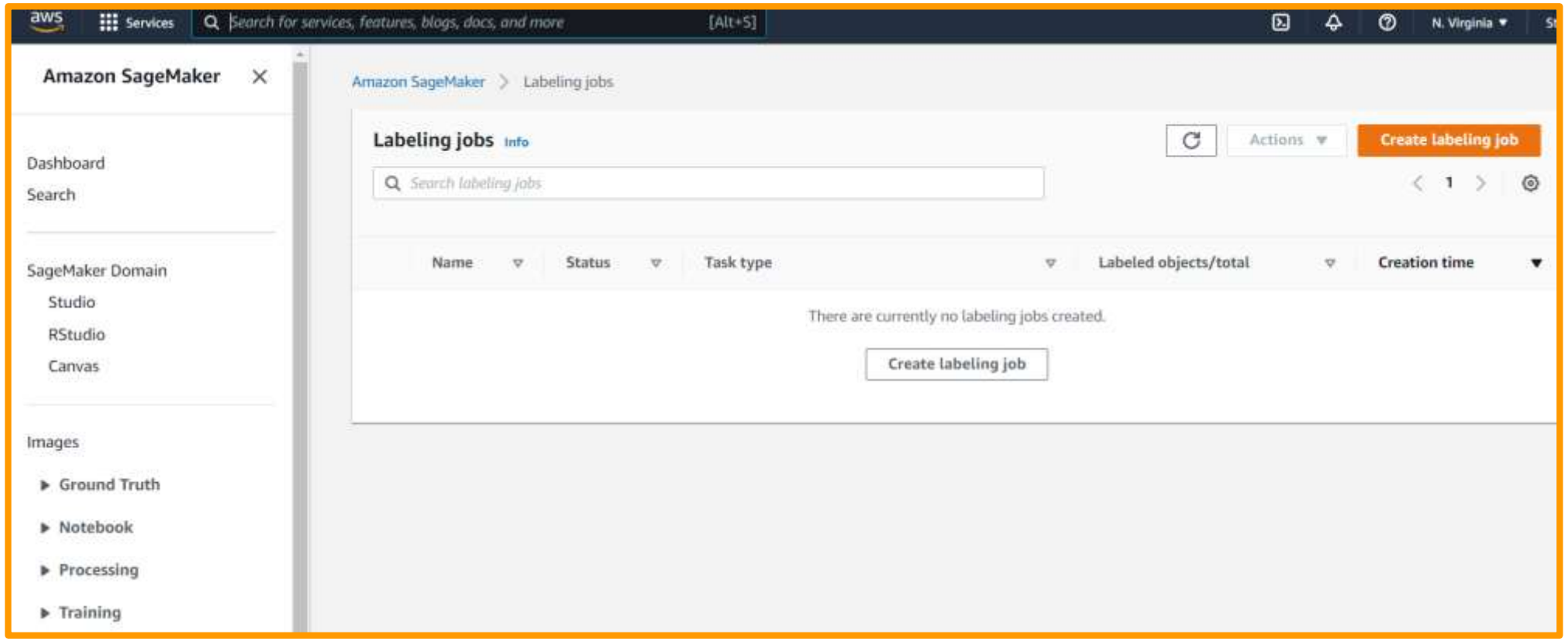
# SAGEMAKER GOUNDTRUTH DEMO

## OPEN SAGMAKER AND CLICK ON LABELING JOBS



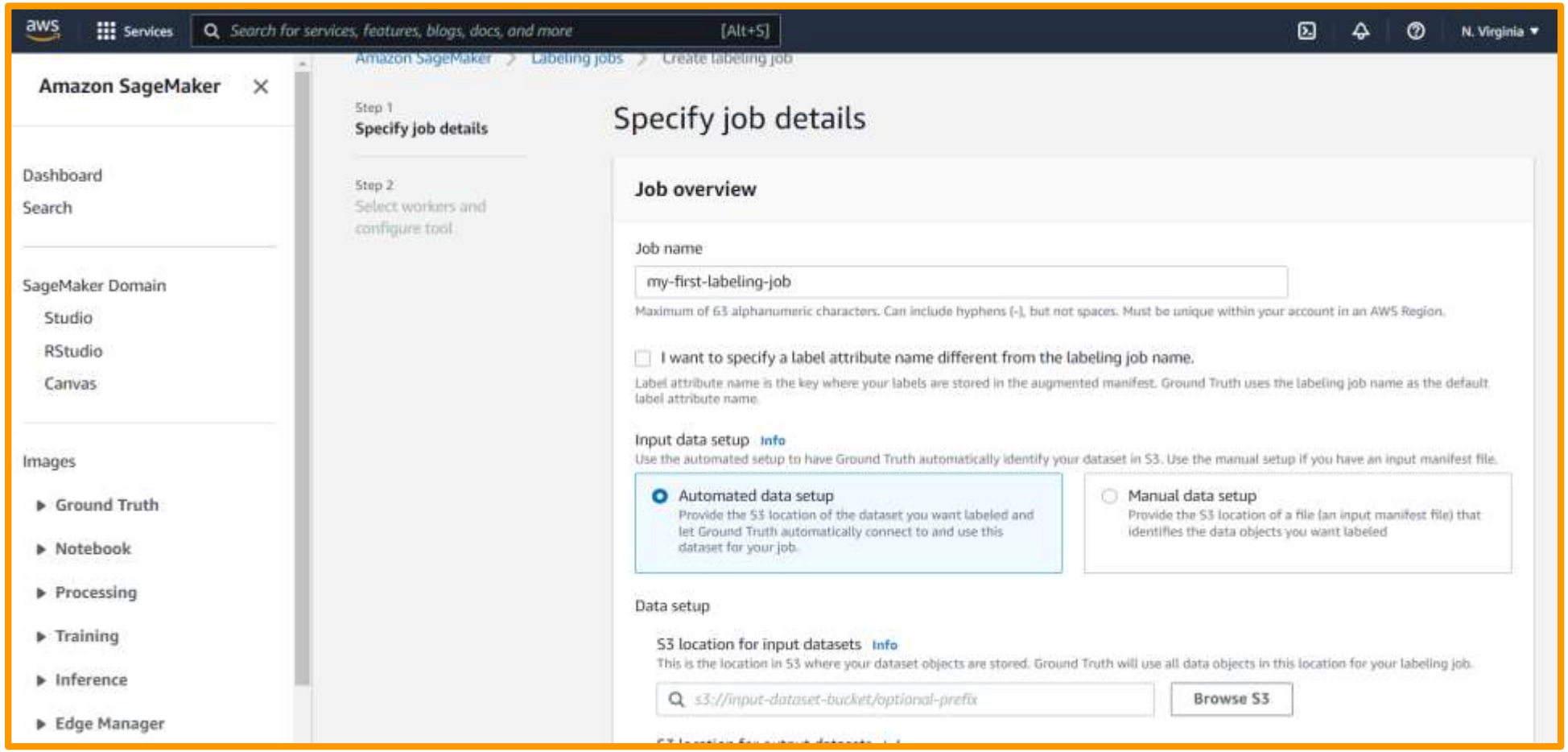
# SAGEMAKER GOUNDTRUTH DEMO

## CLICK ON CREATE LABELING JOB



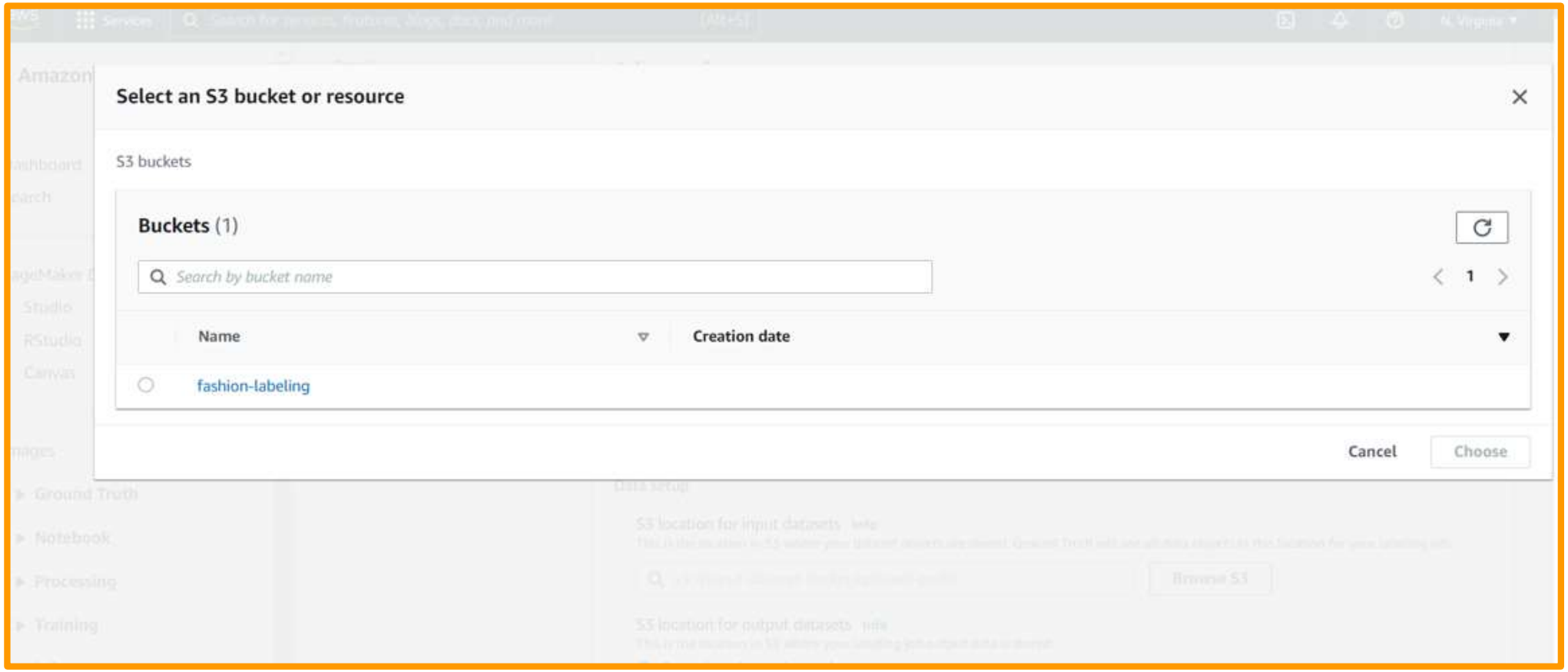
# SAGEMAKER GOUNDTRUTH DEMO

## GIVE A NAME TO THE JOB AND CLICK ON BROWSE S3



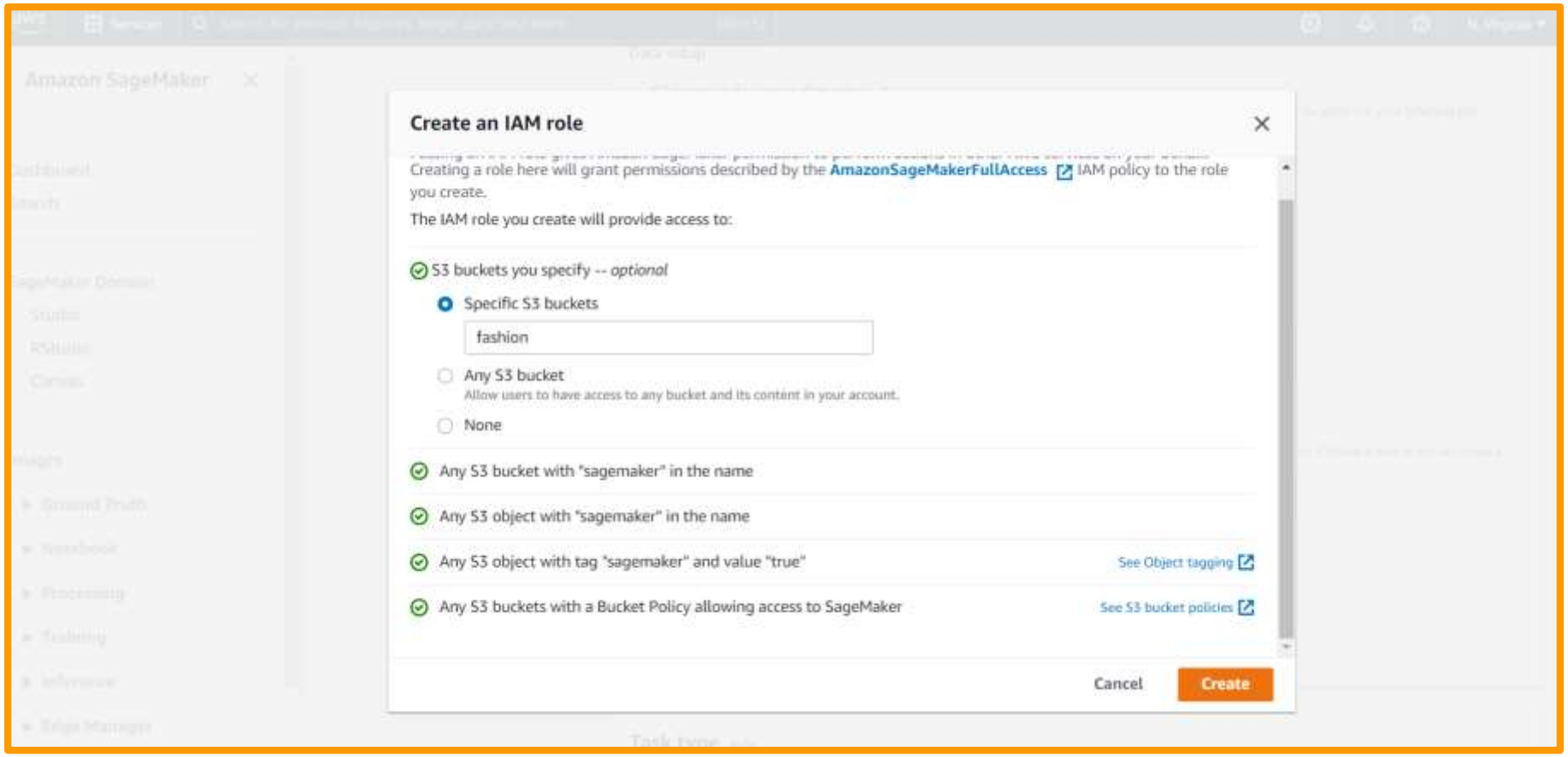
# SAGEMAKER GOUNDTRUTH DEMO

CLICK ON BROWSE S3 AND
POINT TO THE DATASET



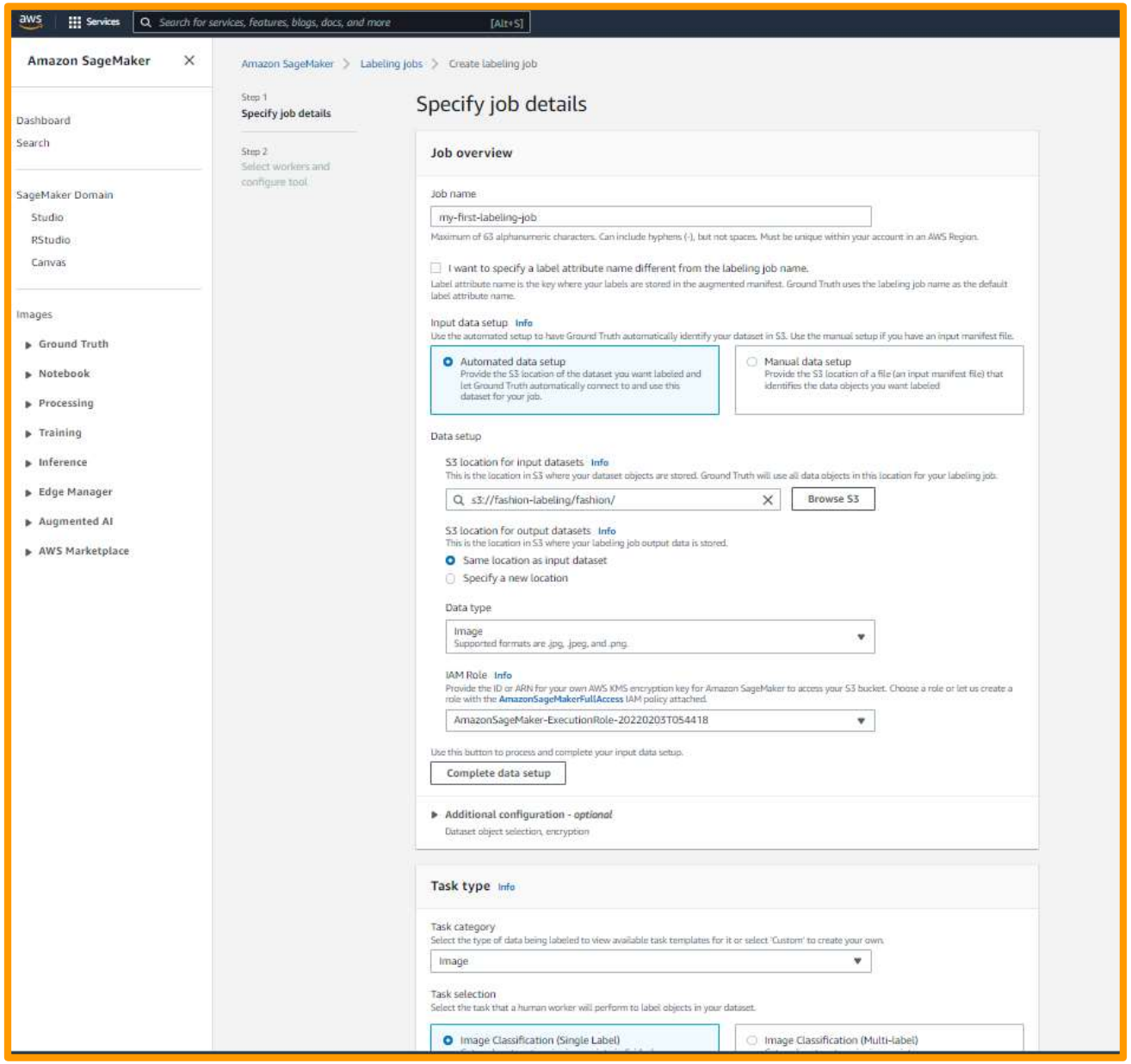
# SAGEMAKER GOUNDTRUTH DEMO

## CREATE A NEW IAM ROLE AND GIVE ACCESS TO A SPECIFIC BUCKET



# SAGEMAKER GOUNDTRUTH DEMO

## A SUMMARY OF THE LABELING JOB



# SAGEMAKER GOUNDTRUTH DEMO

## DON'T FORGET TO CLICK ON "COMPLETE DATA SETUP"

### Job overview

Job name

my-first-labeling-job

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

☐ I want to specify a label attribute name different from the labeling job name.

Label attribute name is the key where your labels are stored in the augmented manifest. Ground Truth uses the labeling job name as the default label attribute name.

Input data setup Info

Use the automated setup to have Ground Truth automatically identify your dataset in S3. Use the manual setup if you have an input manifest file.

- (●) Automated data setup — Provide the S3 location of the dataset you want labeled and let Ground Truth automatically connect to and use this dataset for your job.
- ( ) Manual data setup — Provide the S3 location of a file (an input manifest file) that identifies the data objects you want labeled

Data setup

S3 location for input datasets Info

This is the location in S3 where your dataset objects are stored. Ground Truth will use all data objects in this location for your labeling job.

s3://fashion-labeling/fashion/ [Browse S3]

S3 location for output datasets Info

This is the location in S3 where your labeling job output data is stored.

- (●) Same location as input dataset
- ( ) Specify a new location

Data type

Image

Supported formats are .jpg, .jpeg, and .png.

IAM Role Info

Provide the ID or ARN for your own AWS KMS encryption key for Amazon SageMaker to access your S3 bucket. Choose a role or let us create a role with the **AmazonSageMakerFullAccess** IAM policy attached.

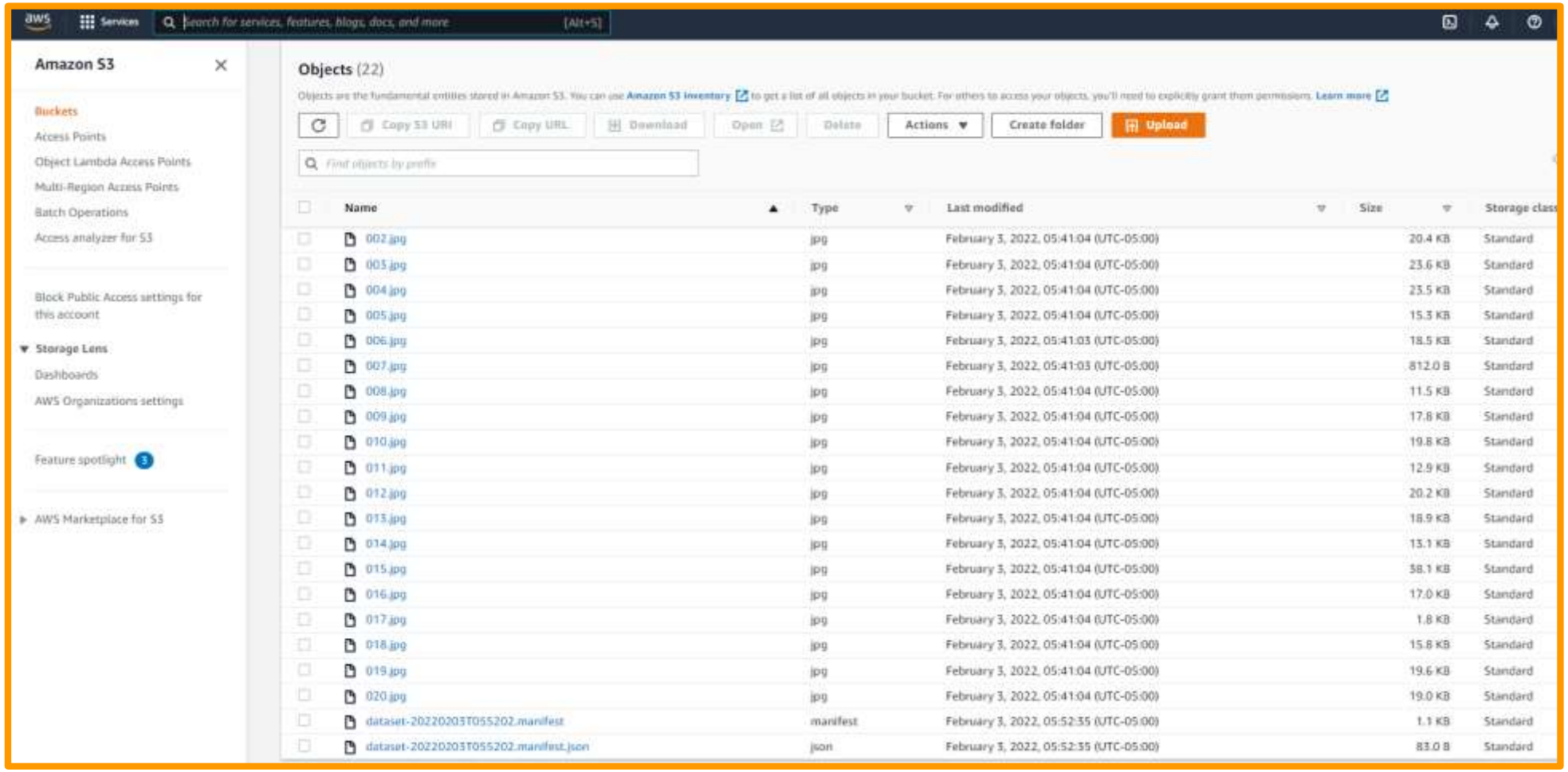
AmazonSageMaker-ExecutionRole-20220203T055170

Use this button to process and complete your input data setup.

[Complete data setup]

# SAGEMAKER GOUNDTRUTH DEMO

## WHEN YOU CLICK ON “COMPLETE DATA SETUP”, A MANIFEST FILE WILL BE GENERATED IN S3



# SAGEMAKER GOUNDTRUTH DEMO

## INPUT MANIFEST FILE SHOULD LOOK LIKE THIS

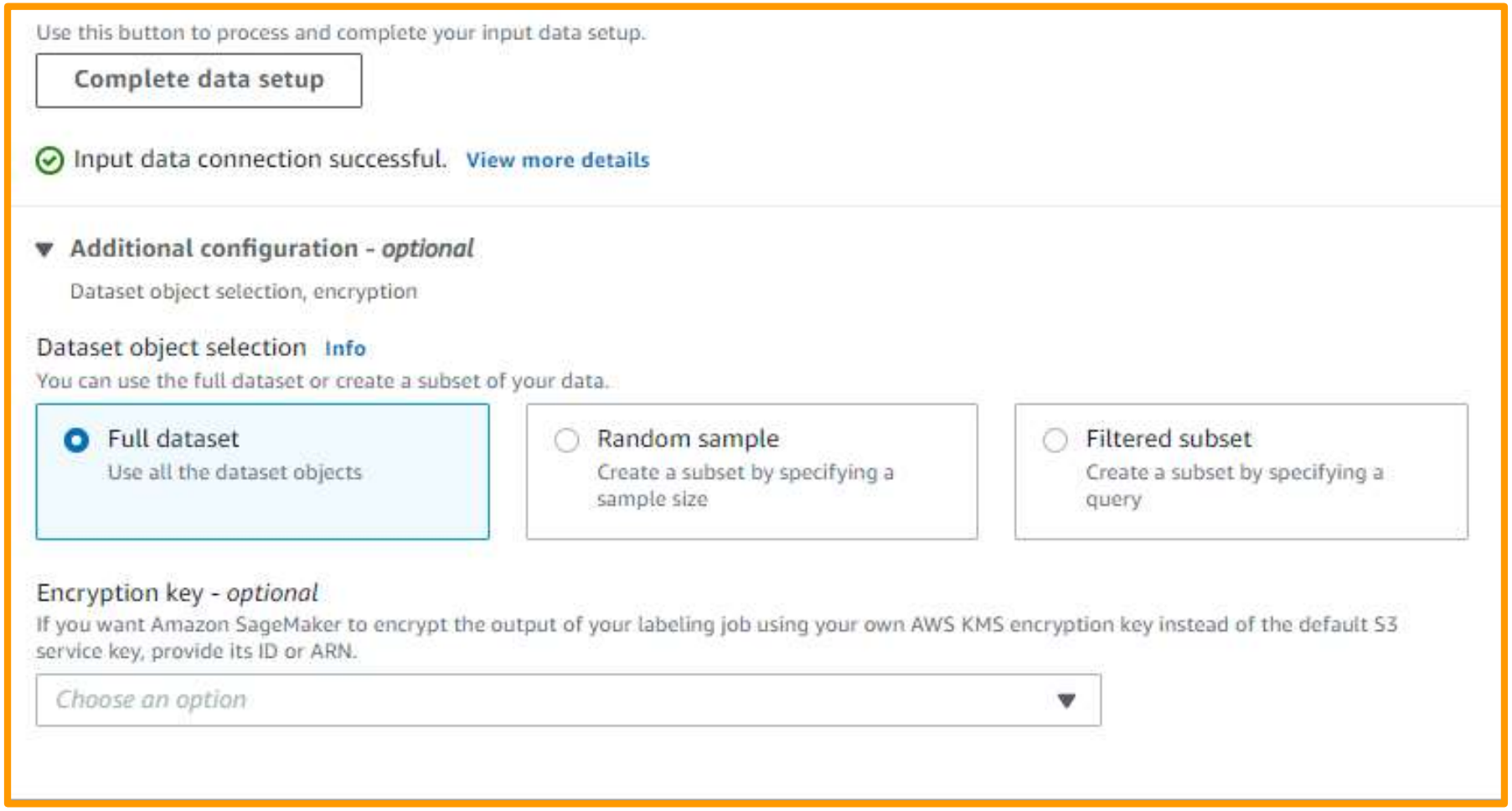
dataset-20220203T055202.manifest - Notepad

File Edit Format View Help

```
{"source-ref":"s3://fashion-labeling/fashion/001.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/002.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/003.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/004.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/005.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/006.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/007.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/008.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/009.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/010.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/011.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/012.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/013.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/014.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/015.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/016.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/017.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/018.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/019.jpg"}
{"source-ref":"s3://fashion-labeling/fashion/020.jpg"}
```

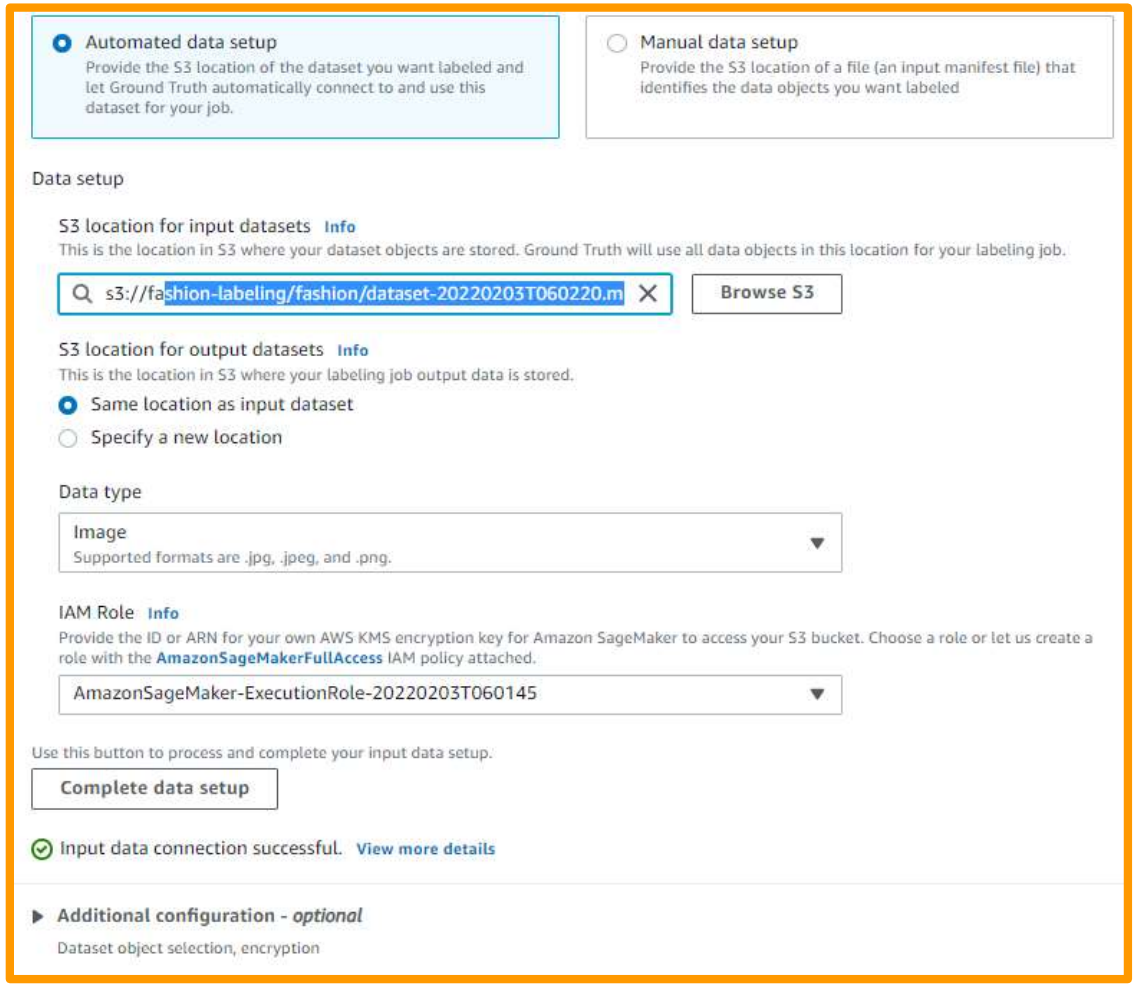
# SAGEMAKER GOUNDTRUTH DEMO

NOTE THAT YOU CAN EITHER LABEL THE ENTRE DATASET OR CHOOSE A RANDOM SAMPLE



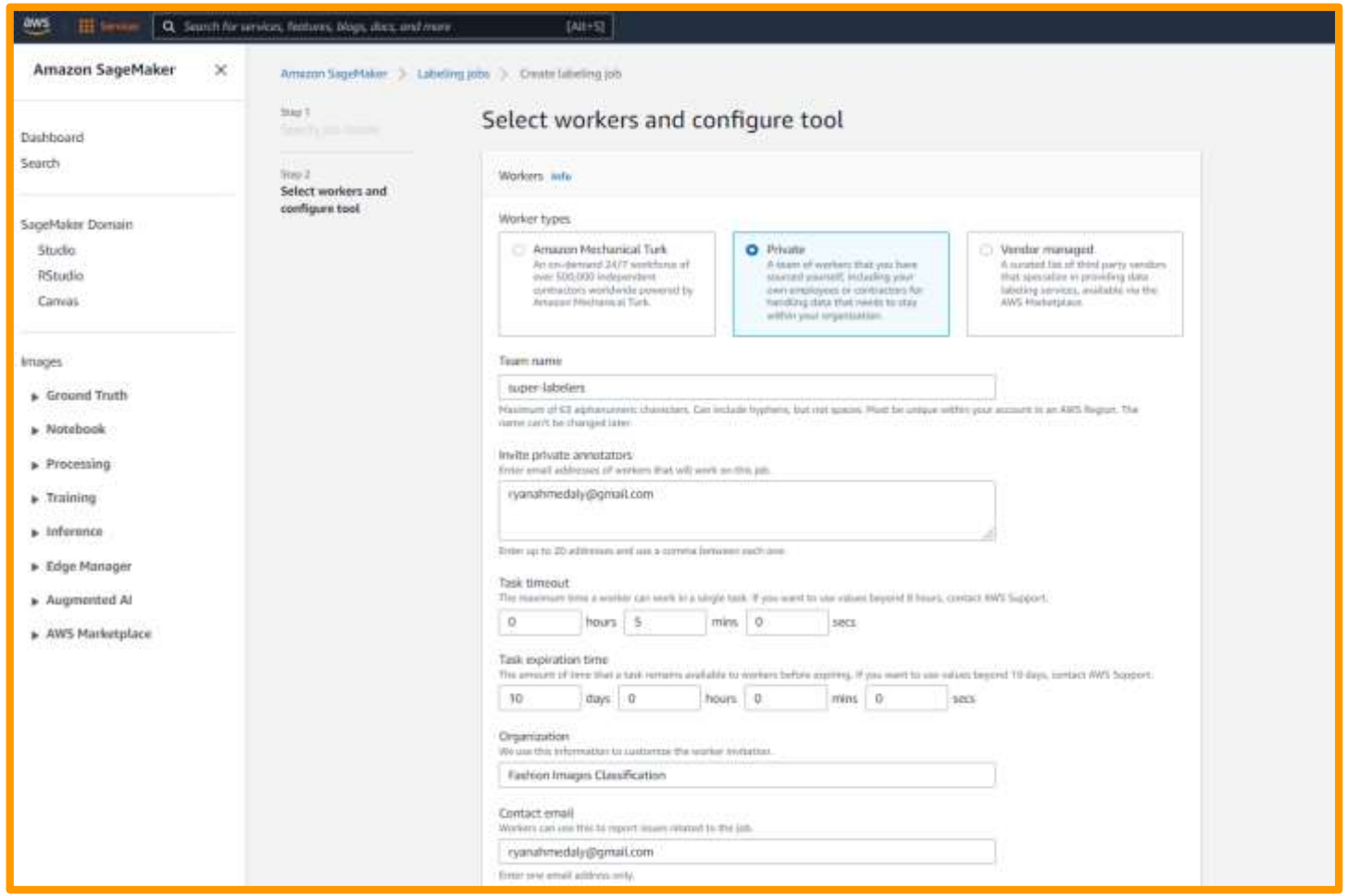
# SAGEMAKER GOUNDTRUTH DEMO

NOTE THAT ONCE YOU ESTABLISH THE DATA CONNECTION, THE MANIFEST FILE PATH IS UPDATED AS SHOWN BELOW

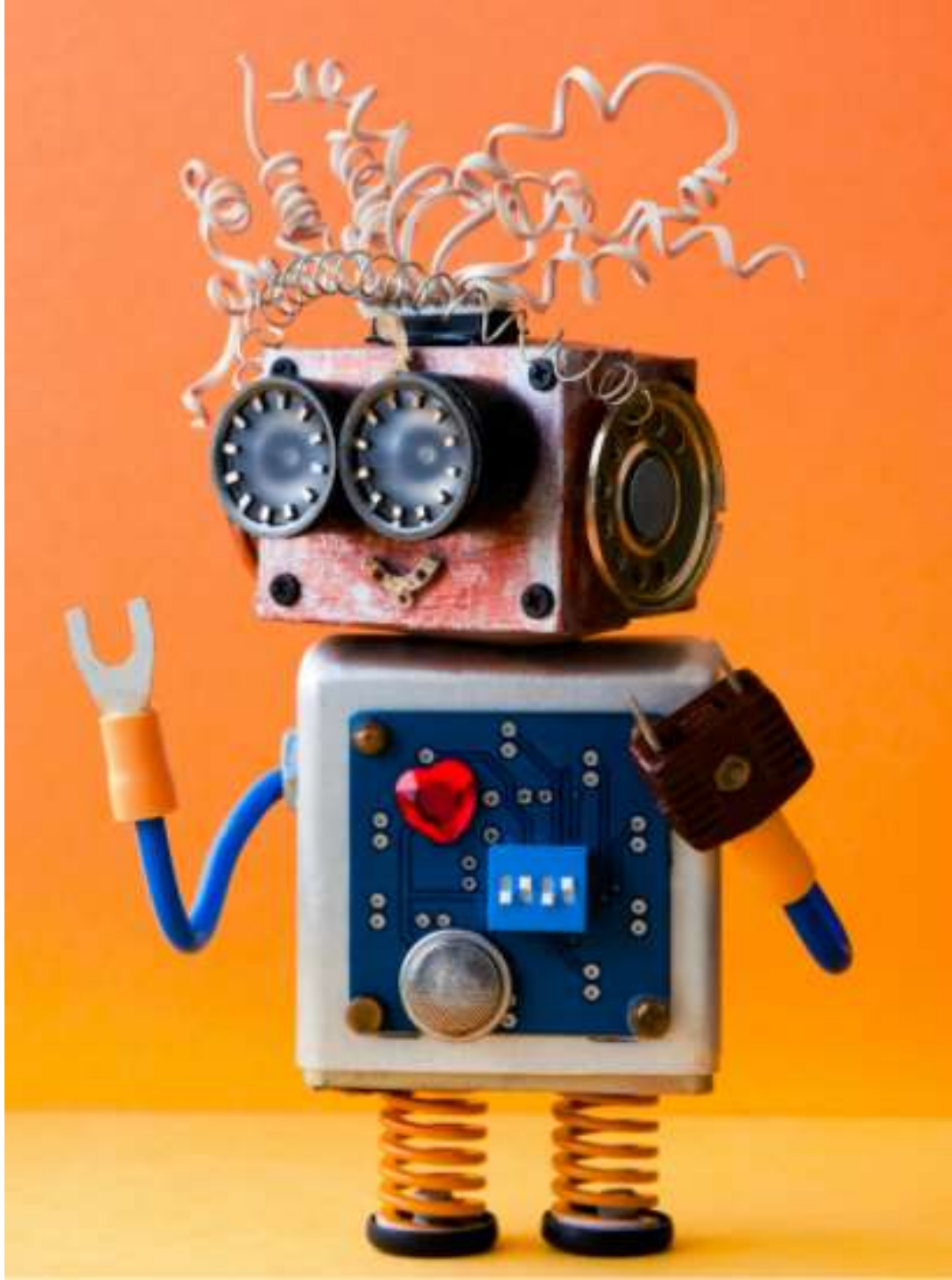


# SAGEMAKER GOUNDTRUTH DEMO

## LET’S KICK START A PRIVATE DATA LABELING JOB

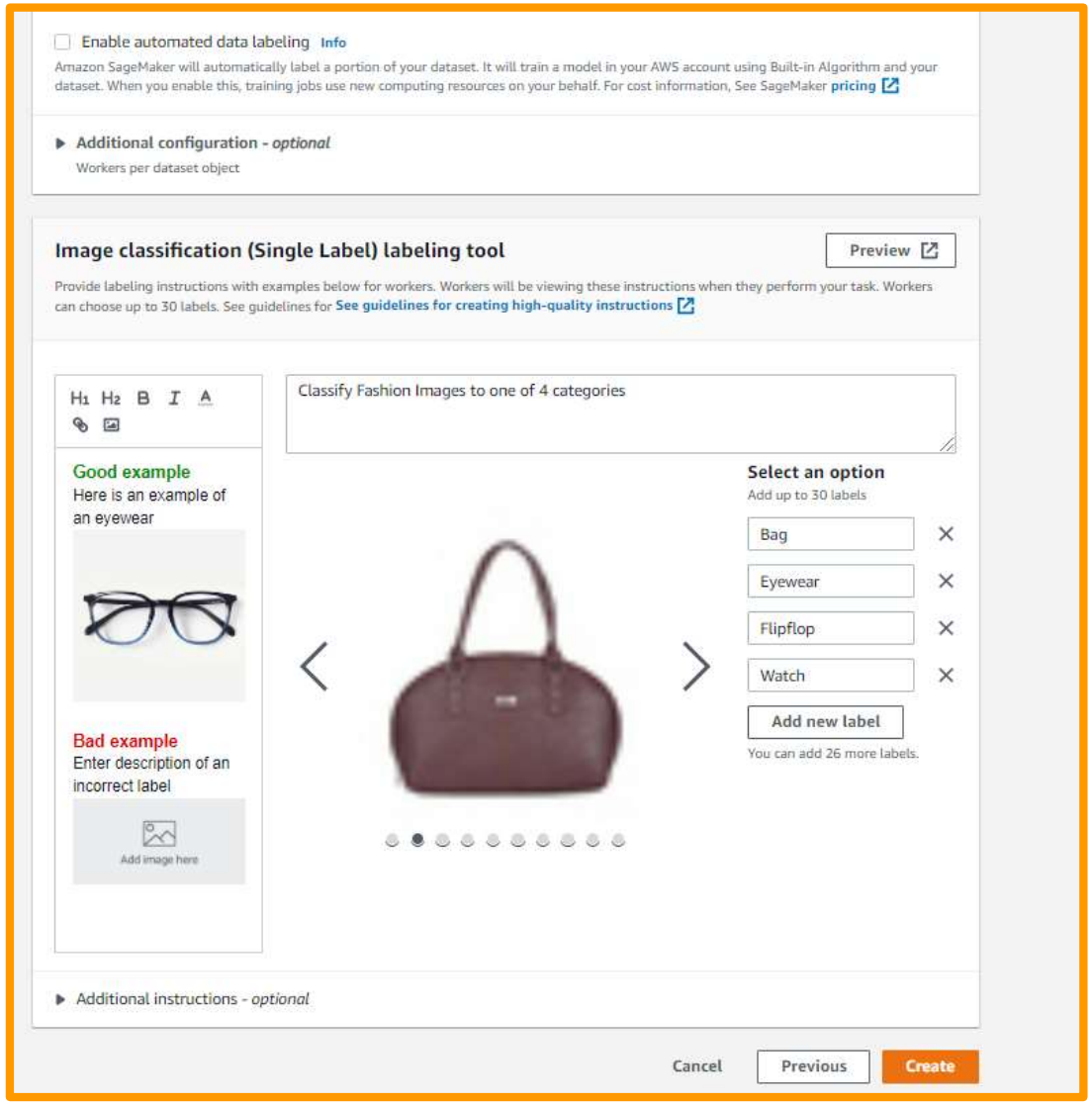


# DATA LABELING IN SAGEMAKER GROUNDTRUTH DEMO – PART 2

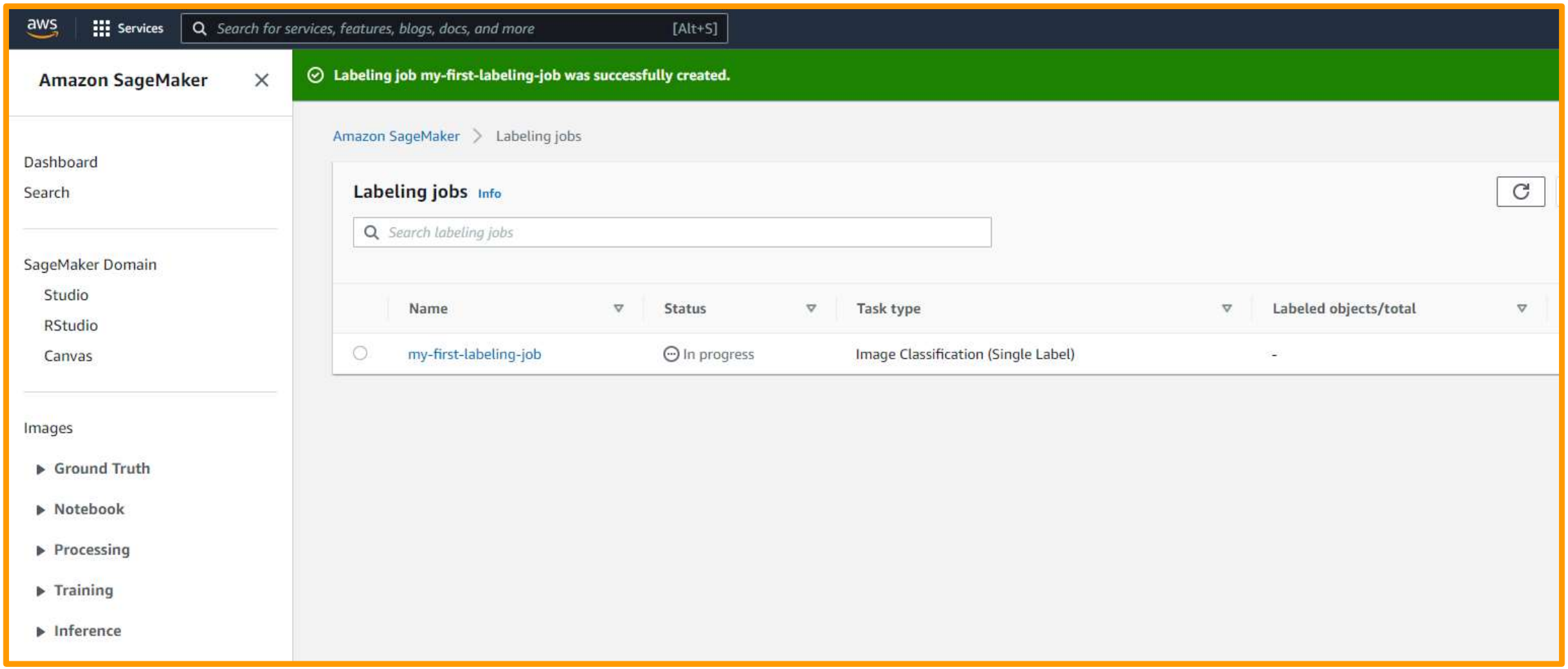
# SAGEMAKER GOUNDTRUTH DEMO

## SPECIFY THE LABELING JOB REQUIREMENTS



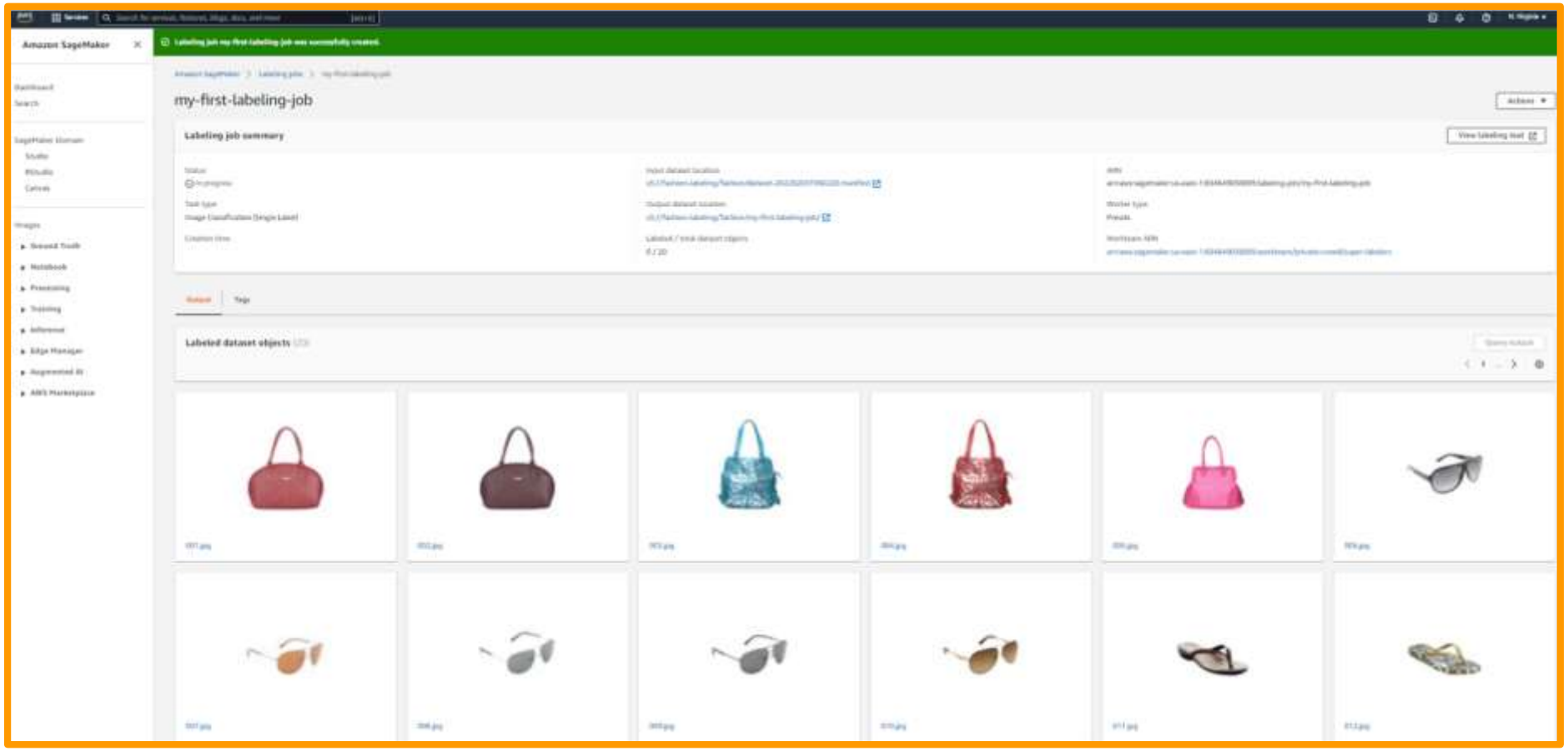
# SAGEMAKER GOUNDTRUTH DEMO

NOTE THAT THE LABELING JOB HAS BEEN SUCCESSFULLY CREATED



# SAGEMAKER GOUNDTRUTH DEMO

CLICK ON THE LABELING JOB TO CHECK THE STATUS AND SEE IF IT MAKES SENSE!



# SAGEMAKER GOUNDTRUTH DEMO

## THIS IS THE INVITE THAT LABELERS WILL RECEIVE IN THEIR EMAILS

You are invited by ryanahmedaly@gmail.com from Fashion Images Classification to work on a labeling project Inbox ×

**no-reply@verificationemail.com** 6:17
to me

Hi,

**You are invited by ryanahmedaly@gmail.com from Fashion Images Classification to work on a labeling project**

Click on the link below to log into your labeling project.

https://ikl06k96wa.labeling.us-east-1.sagemaker.aws

You will need the following username and temporary password provided below to login for the first time.

User name: **ryanahmedaly@gmail.com**

Temporary password: **B4zX/8FA**

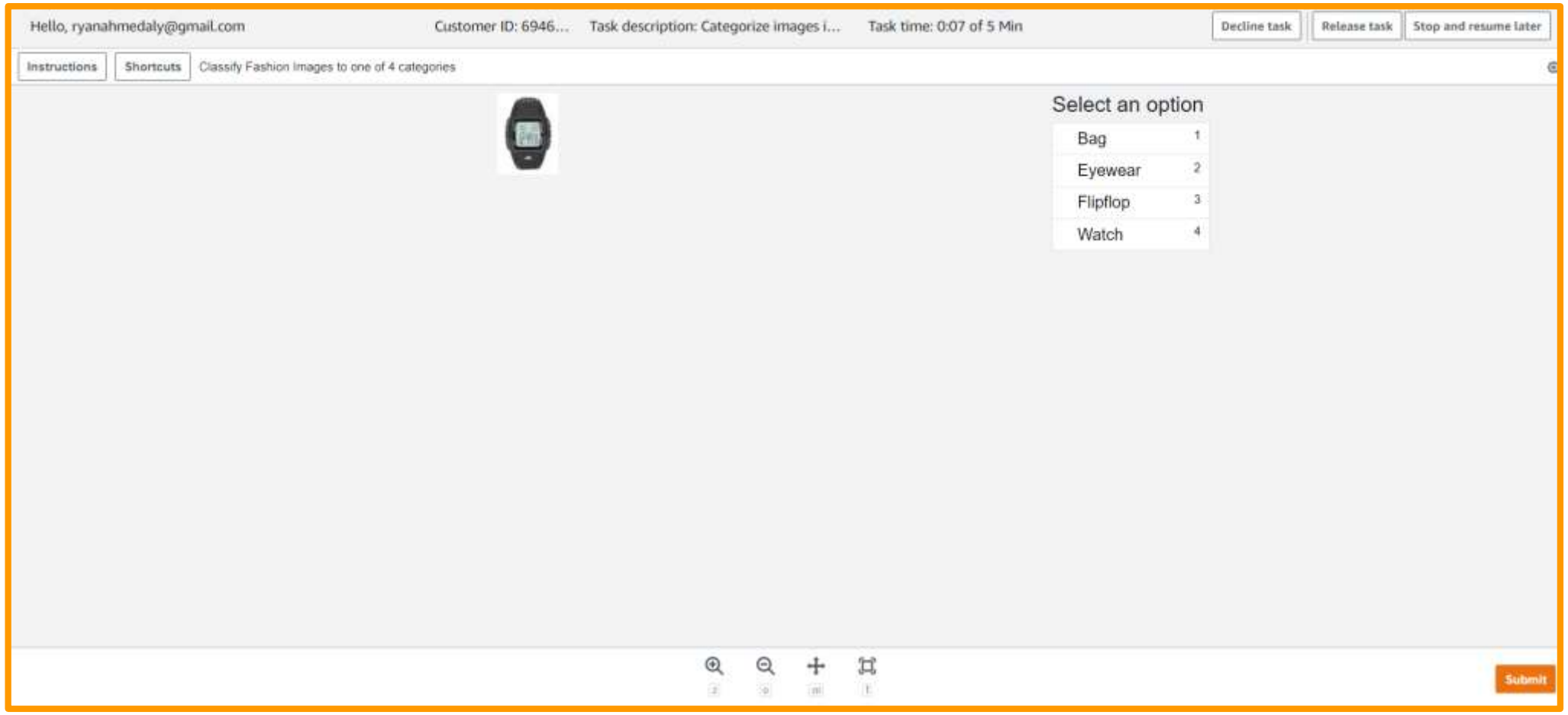
Once you log in with your temporary password, you will be required to create a new password for your account.

After creating a new password, you can log into your private team to access your labeling project.

If you have any questions, please contact us at ryanahmedaly@gmail.com.

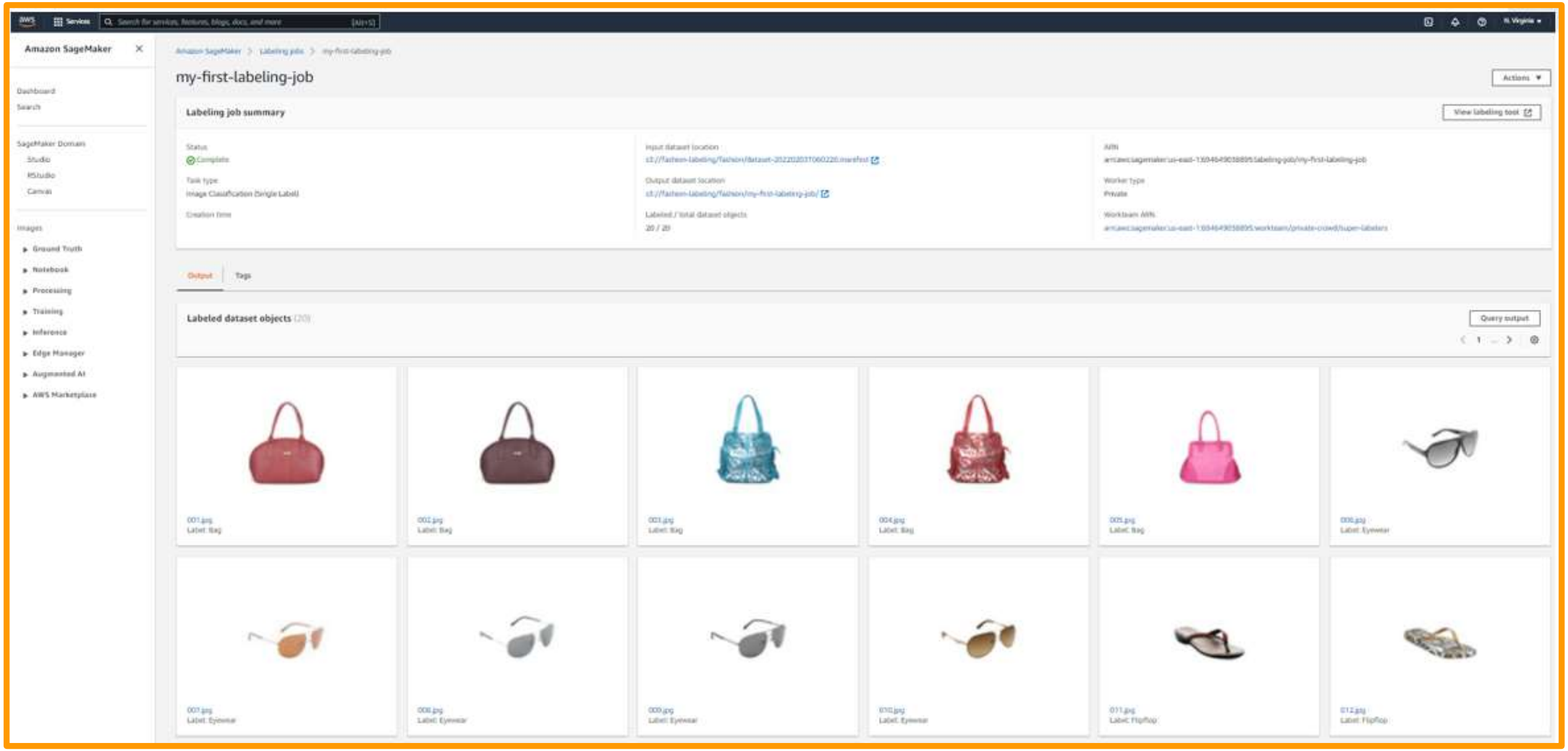
# SAGEMAKER GOUNDTRUTH DEMO

## THIS IS WHAT THE TASK LOOK LIKE!



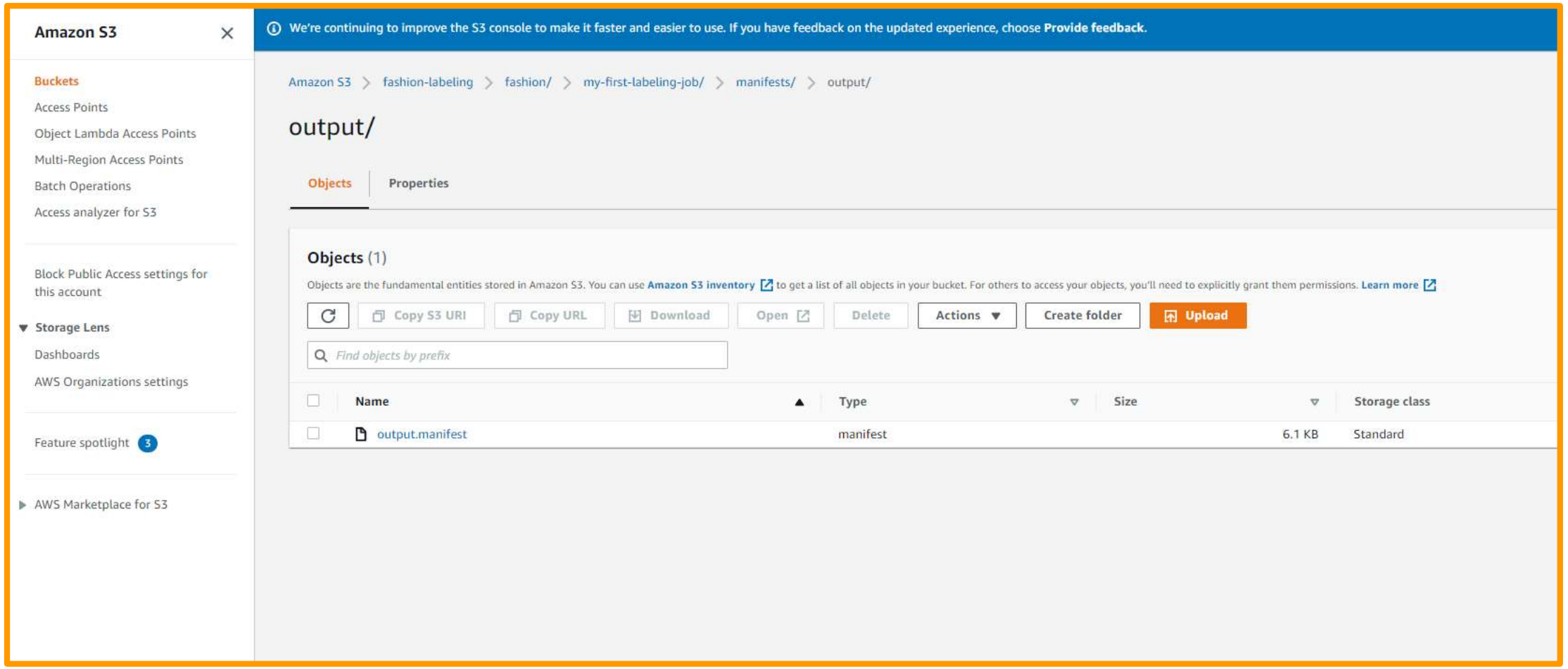
# SAGEMAKER GOUNDTRUTH DEMO

THE LABELING JOB IS NOW COMPLETE, YOU CAN PROCEED WITH EXPLORING THE LABELS



# SAGEMAKER GOUNDTRUTH DEMO

## GO TO S3 AND CHECK OUT THE MANIFEST OUTPUT FILE



# SAGEMAKER GOUNDTRUTH DEMO

CHECK OUT THE OUTPUT MANIFEST FILE. NOTE THAT CONFIDENCE IS SHOWN ZERO SINCE WE DON’T HAVE MULTIPLE LABELERS DOING THE SAME JOB AND WE DIDN’T LEVERAGE THE AUTOLABEL


output.manifest - Notepad

File Edit Format View Help

```
{"source-ref":"s3://fashion-labeling/fashion/001.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","conf
{"source-ref":"s3://fashion-labeling/fashion/002.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","conf
{"source-ref":"s3://fashion-labeling/fashion/003.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","conf
{"source-ref":"s3://fashion-labeling/fashion/004.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","conf
{"source-ref":"s3://fashion-labeling/fashion/005.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","conf
{"source-ref":"s3://fashion-labeling/fashion/006.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","
{"source-ref":"s3://fashion-labeling/fashion/007.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","
{"source-ref":"s3://fashion-labeling/fashion/008.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","
{"source-ref":"s3://fashion-labeling/fashion/009.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","
{"source-ref":"s3://fashion-labeling/fashion/010.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","
{"source-ref":"s3://fashion-labeling/fashion/011.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job",
{"source-ref":"s3://fashion-labeling/fashion/012.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job",
{"source-ref":"s3://fashion-labeling/fashion/013.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job",
{"source-ref":"s3://fashion-labeling/fashion/014.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job",
{"source-ref":"s3://fashion-labeling/fashion/015.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job",
{"source-ref":"s3://fashion-labeling/fashion/016.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","co
{"source-ref":"s3://fashion-labeling/fashion/017.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","co
{"source-ref":"s3://fashion-labeling/fashion/018.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","co
{"source-ref":"s3://fashion-labeling/fashion/019.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","co
{"source-ref":"s3://fashion-labeling/fashion/020.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","co
```

output.manifest - Notepad

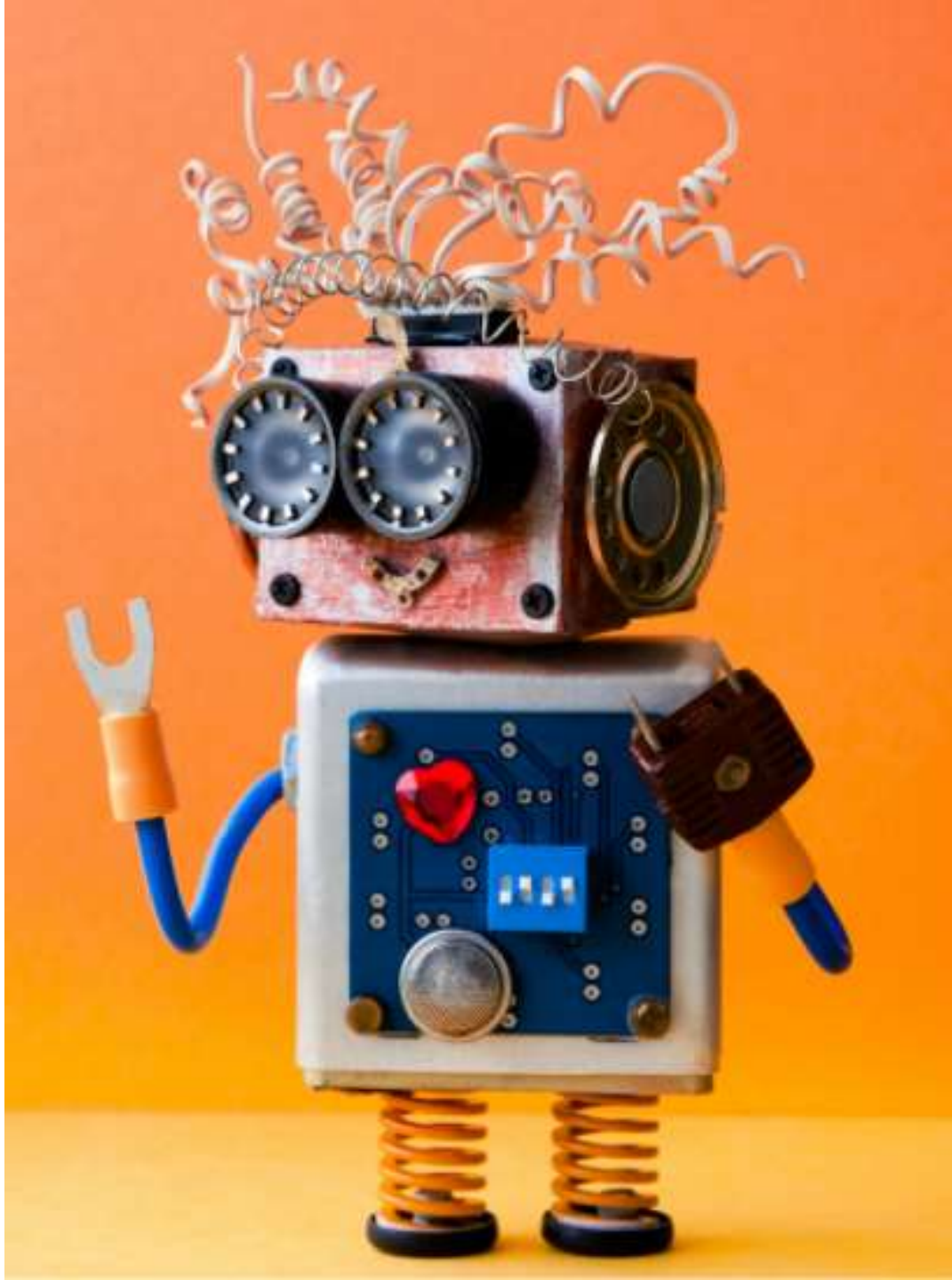
File Edit Format View Help

```
ling/fashion/001.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"gr
ling/fashion/002.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"gr
ling/fashion/003.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"gr
ling/fashion/004.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"gr
ling/fashion/005.jpg","my-first-labeling-job":0,"my-first-labeling-job-metadata":{"class-name":"Bag","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"gr
ling/fashion/006.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type"
ling/fashion/007.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type"
ling/fashion/008.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type"
ling/fashion/009.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type"
ling/fashion/010.jpg","my-first-labeling-job":1,"my-first-labeling-job-metadata":{"class-name":"Eyewear","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type"
ling/fashion/011.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type
ling/fashion/012.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type
ling/fashion/013.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type
ling/fashion/014.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type
ling/fashion/015.jpg","my-first-labeling-job":2,"my-first-labeling-job-metadata":{"class-name":"Flipflop","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type
ling/fashion/016.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"
ling/fashion/017.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"
ling/fashion/018.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"
ling/fashion/019.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"
ling/fashion/020.jpg","my-first-labeling-job":3,"my-first-labeling-job-metadata":{"class-name":"Watch","job-name":"labeling-job/my-first-labeling-job","confidence":0,"type":"
```

# FINAL END-OF-DAY CAPSTONE PROJECT

EASY

ADVANCED



# FINAL CAPSTONE PROJECT

1. Using the Traffic signs datasets included in the course package.
2. Create a labeling job using Amazon SageMaker GroundTruth
3. Review the input manifest file
4. Perform the labeling job
5. Review the output manifest file and ensure that the labeling job was successful

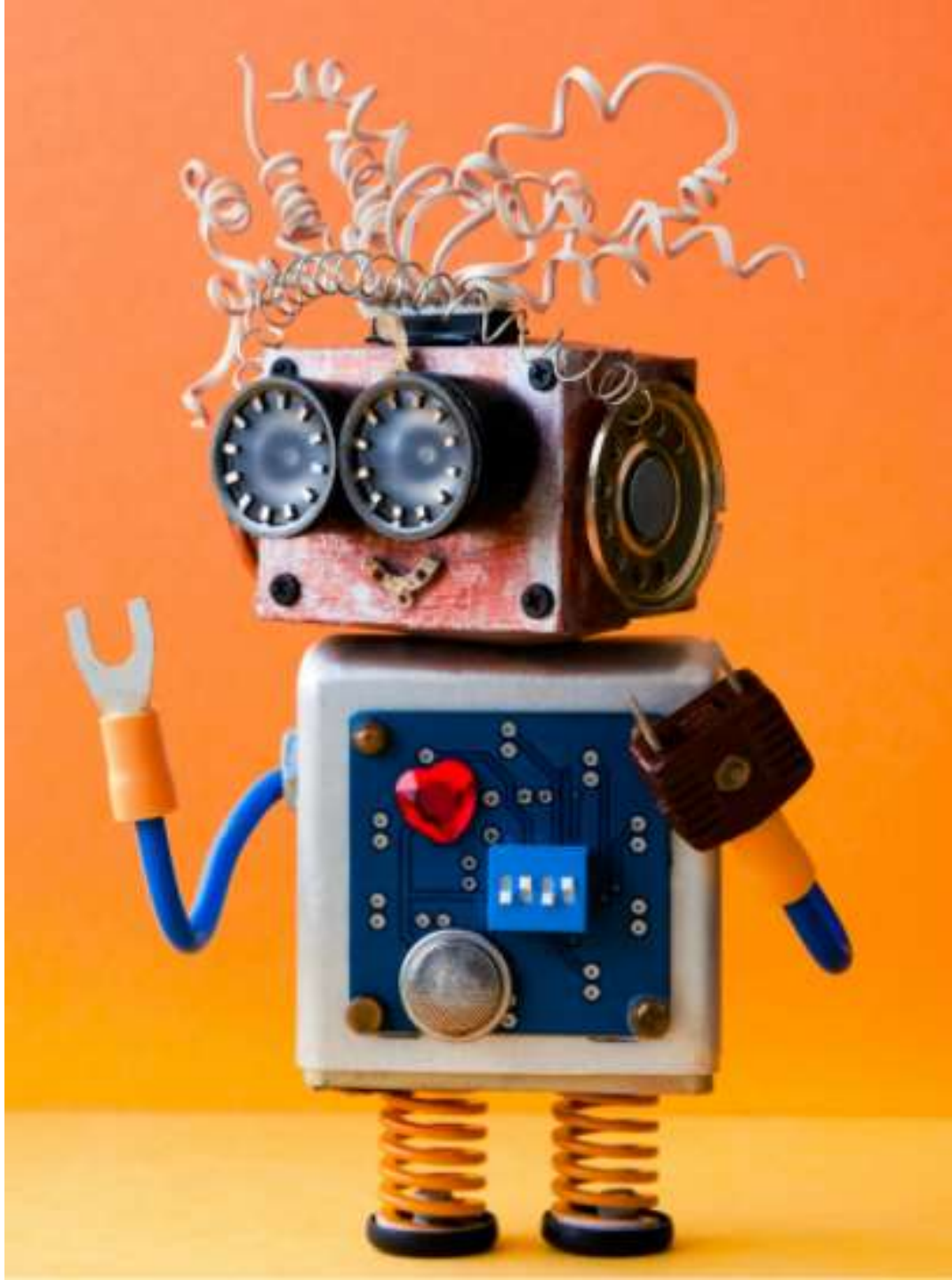

1 2 3 4 5 6 7 8 9

AWS SageMaker Pricing Examples: https://aws.amazon.com/sagemaker/data-labeling/pricing/?nc=sn&loc=3&refid=ps_a134p0000078pqxaae&trkcampaign=acq_paid_search_brand

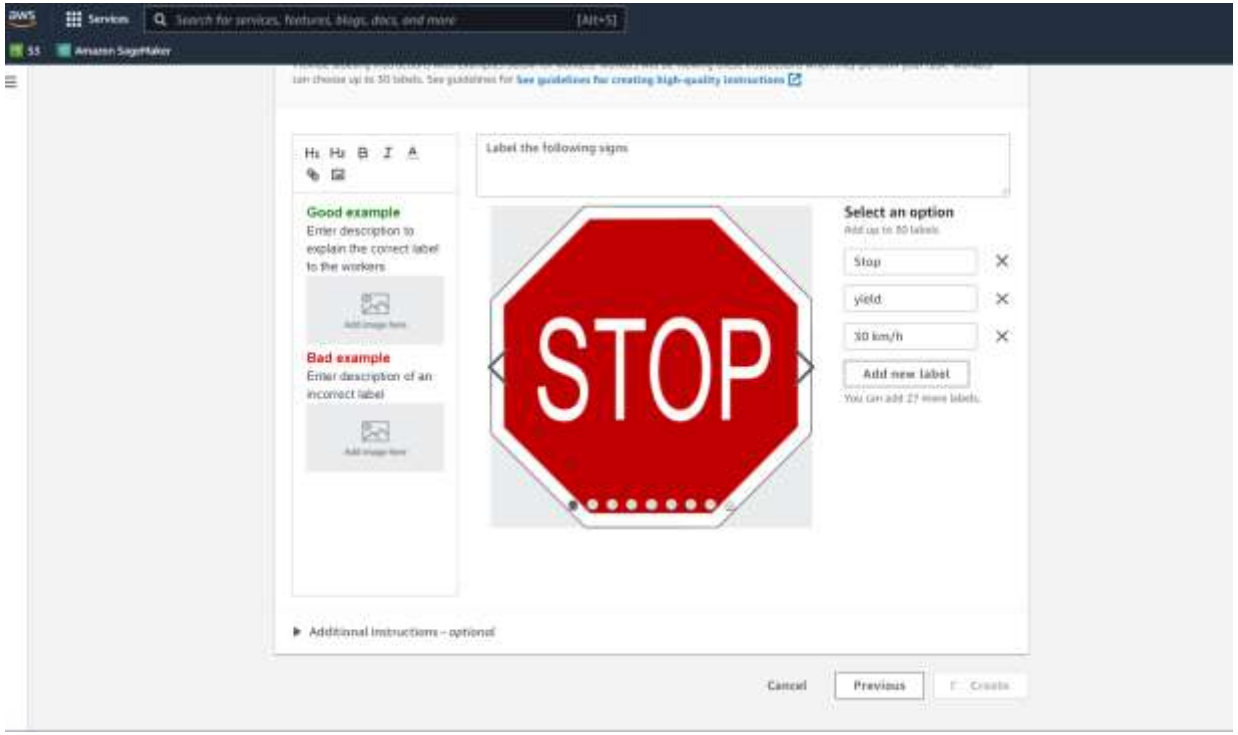
# FINAL END-OF-DAY CAPSTONE PROJECT SOLUTION

EASY

ADVANCED



# PROJECT SOLUTION



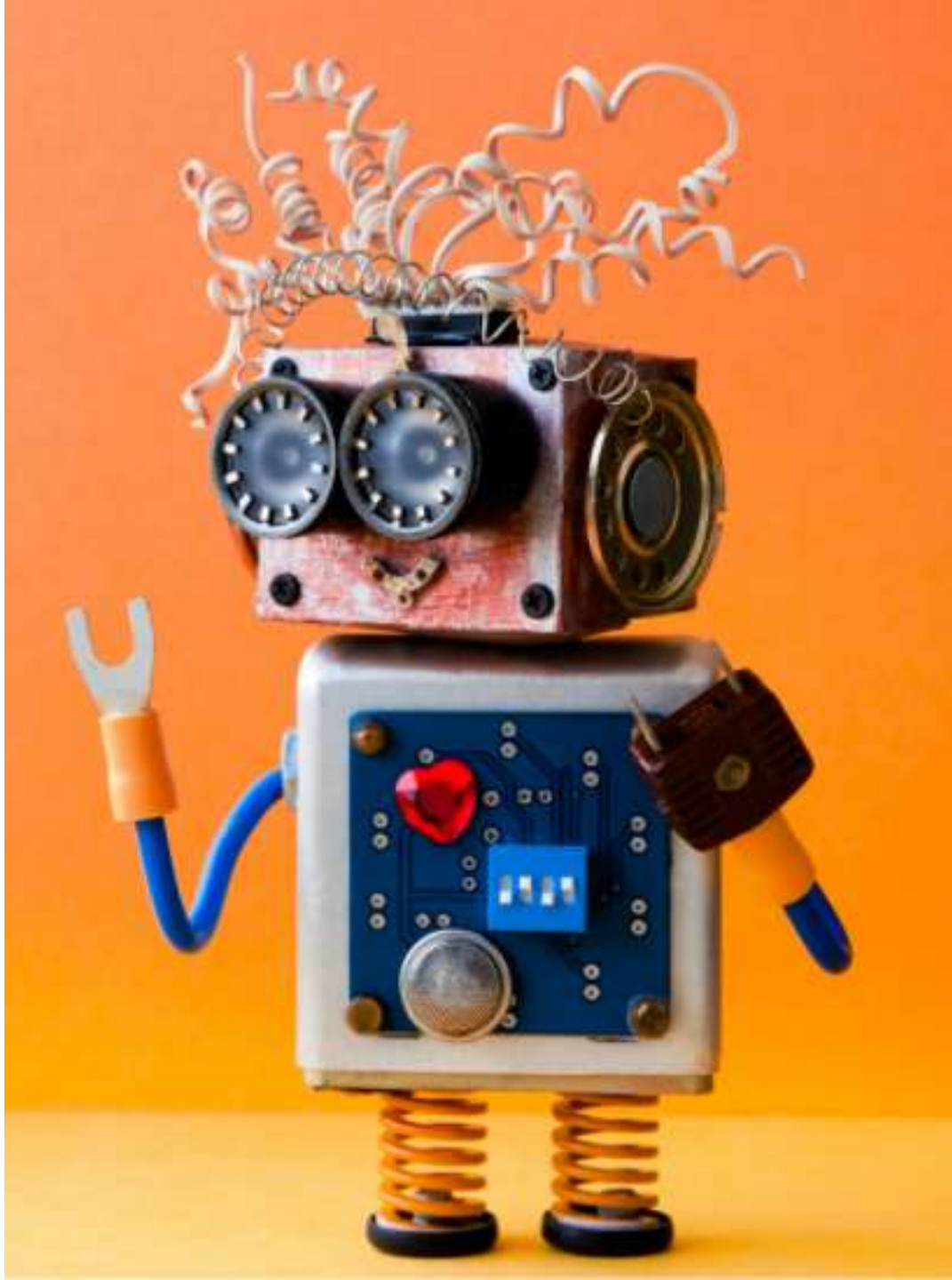

File Edit View


```
[{"datasetObjectId":"4","consolidatedAnnotation":{"content":{"labeling-job-project":1,"labeling-job-project-metadata":{"class-name":"yield","job-name":"labeling-job/labeling-job-project","confidence":0.0,"type":"gro
":{"content":{"labeling-job-project":0,"labeling-job-project-metadata":{"class-name":"Stop","job-name":"labeling-job/labeling-job-project","confidence":0.0,"type":"groundtruth/image-classification","human-annotated"
eling-job-project-metadata":{"class-name":"30 km/h","job-name":"labeling-job/labeling-job-project","confidence":0.0,"type":"groundtruth/image-classification","human-annotated":"yes","creation-date":"2022-03-25T20:05
```

# EXTRAS!

EASY

ADVANCED



# GROUNDTRUTH VS. GROUNDTRUTH PLUS

## AMAZON SAGEMAKER GROUND TRUTH

- Amazon SageMaker ground truth empowers companies and individuals to build and manage data labeling workflows.
- You need to manage human annotators such as mechanical turks, third-party vendors, or your own human labelers.

## SAGEMAKER GROUND TRUTH PLUS

- Amazon SageMaker Ground truth plus creates and manages the workflow on your behalf.
- There is no need to manage workforces. Instead, a trained ML team will handle all data security, labeling, privacy and compliance.
- It reduces data labeling costs by 40%
- All what you need to do is to upload your datasets.

# AUTOMATED DATA LABELING

- Auto labeling uses active learning to determine if input data is well understood or not (Easy or hard).
- Input unlabeled images are used with labeled output images to continuously train machine learning model which in turn reduces the number of human labelers required overtime (~70% reduction).
- If there is no enough dataset, auto labeling is not recommended (threshold 1000 images)

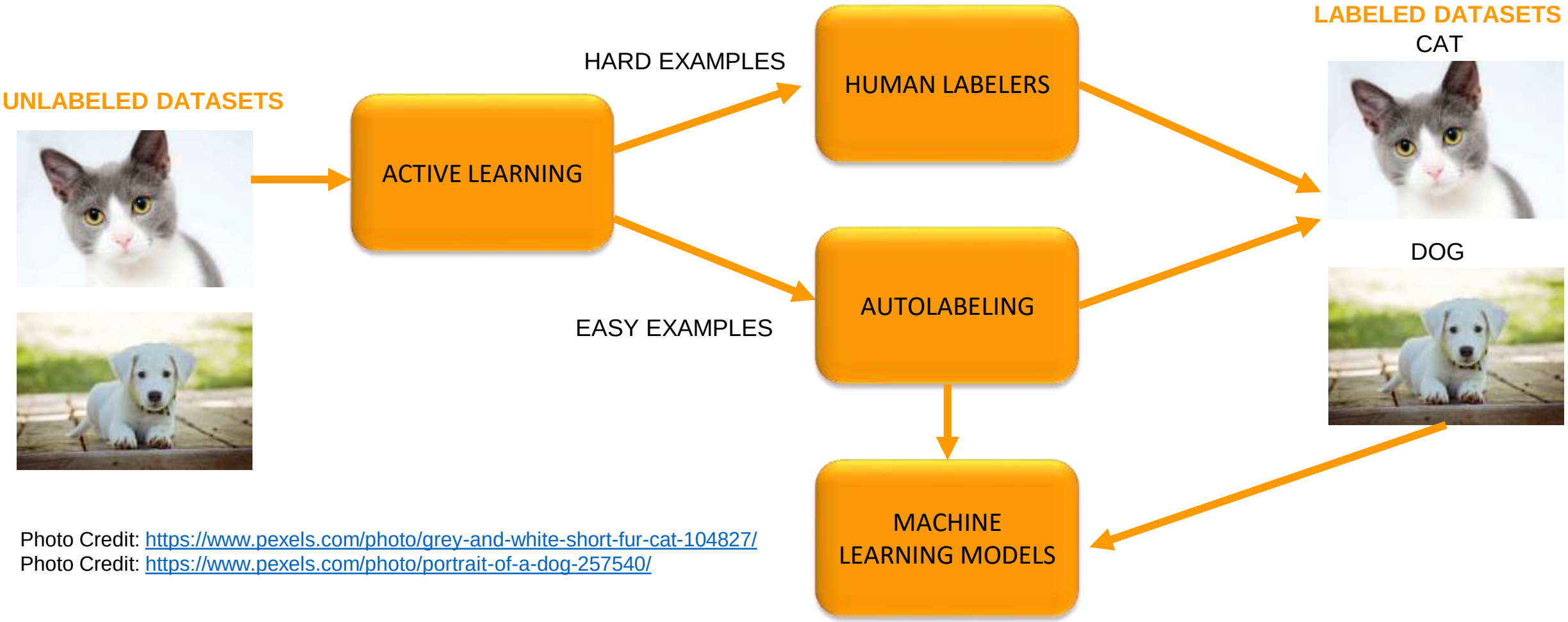



Photo Credit: https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/
Photo Credit: https://www.pexels.com/photo/portrait-of-a-dog-257540/