A top-down view of a variety of fresh ingredients arranged on a dark surface. The items include sliced lemons, whole and sliced tomatoes, almonds, cubed mango, red quinoa, yellow corn, a halved avocado, blueberries, kiwi slices, ginger root, and various other vegetables and fruits. The text is overlaid in the center.

# AI Technology Solutions *to* Identify and Grade Food-Images

End-to-End Solutions Built on AWS Platform



# Image Processing architecture based on Multi-modal LLM performed better over CNN Sequential model across many metrics

Many business problems related to image processing remain unsolved due to the extensive effort required for obtaining the ground truth data, processing large datasets, and building complex Deep Neural Network (DNN) models.

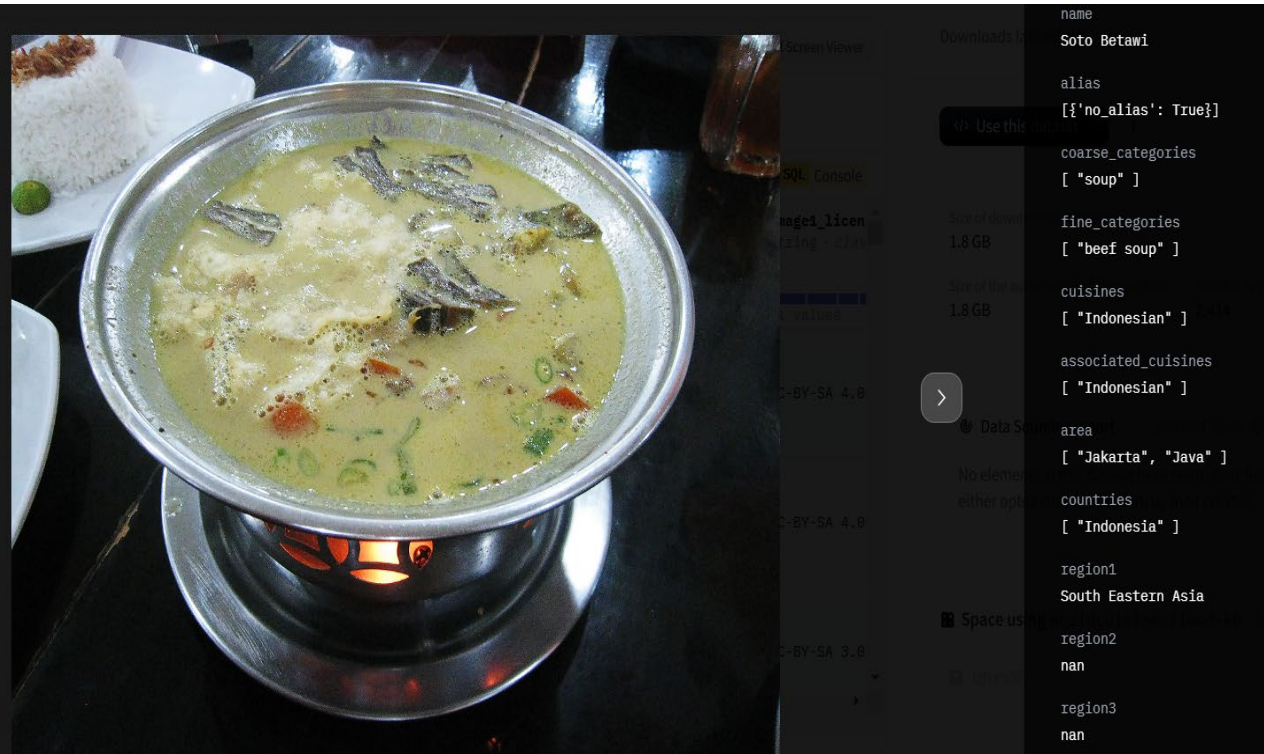
Moreover, even with significant investment in these processes, there is no guarantee of achieving strong model performance

Multi-modal LLMs such as Google Gemini Pro significantly reduced the time and improved the confidence in solving such business problems, compared with building Deep Neural Nets such as the CNN (Convolutional Neural Nets) Sequential model using TensorFlow framework

## Business Problem

IATA estimates that passenger flights generate approximately six million tonnes of waste per year. Some 20% of this is untouched food and drink, which the association estimates to carry a value of \$4 billion

Industry is exploring innovations for improving what's served up on the tray to reduce wastage and improve passenger experience



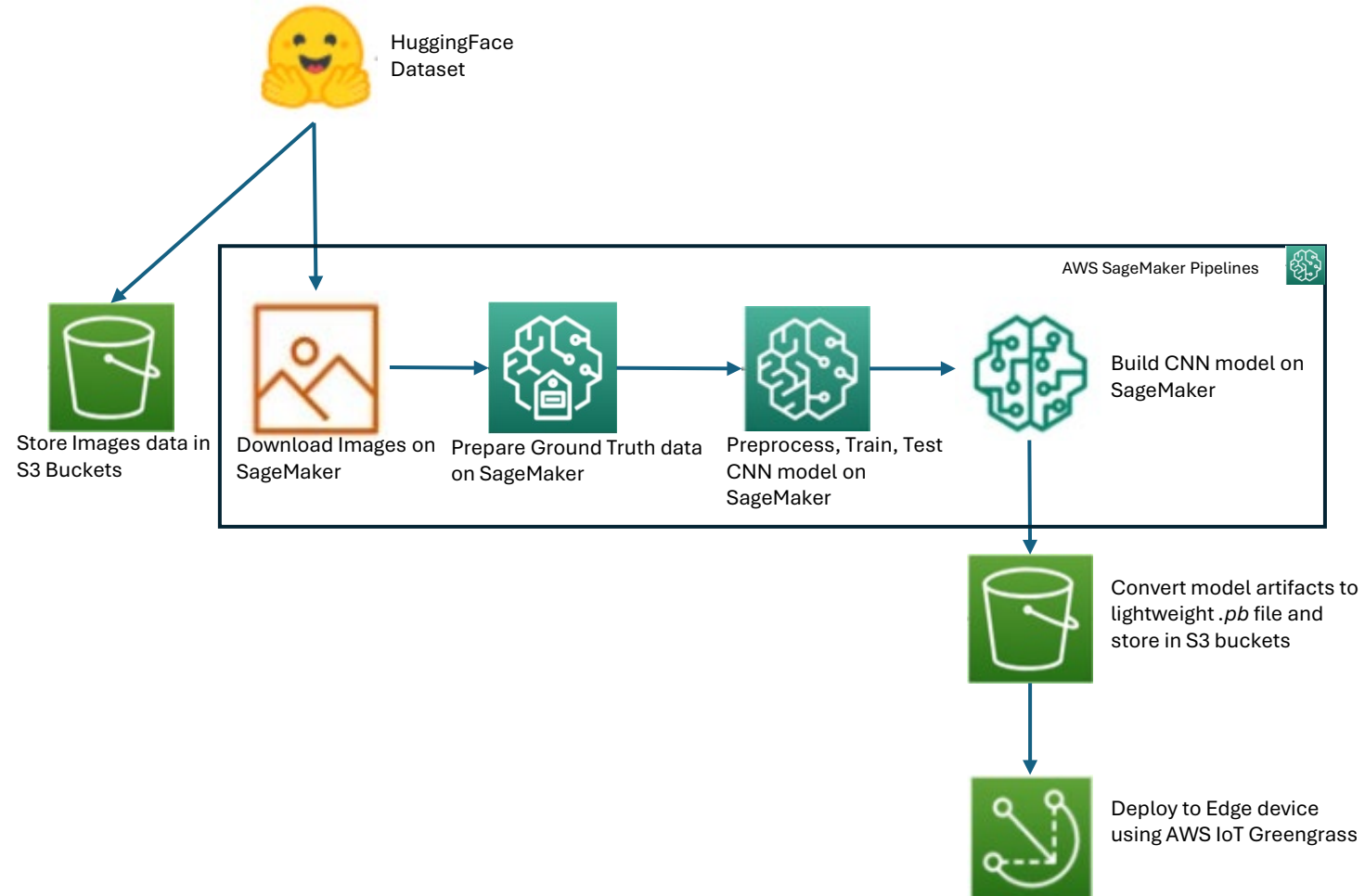
## Hugging Face Dataset

- 2400 different dishes from different cuisines
- Images and different dishes are identified but images are not graded
- worldcuisines/food-kb

# Two different AI solutions based on accuracy, latency and go-to-market speed

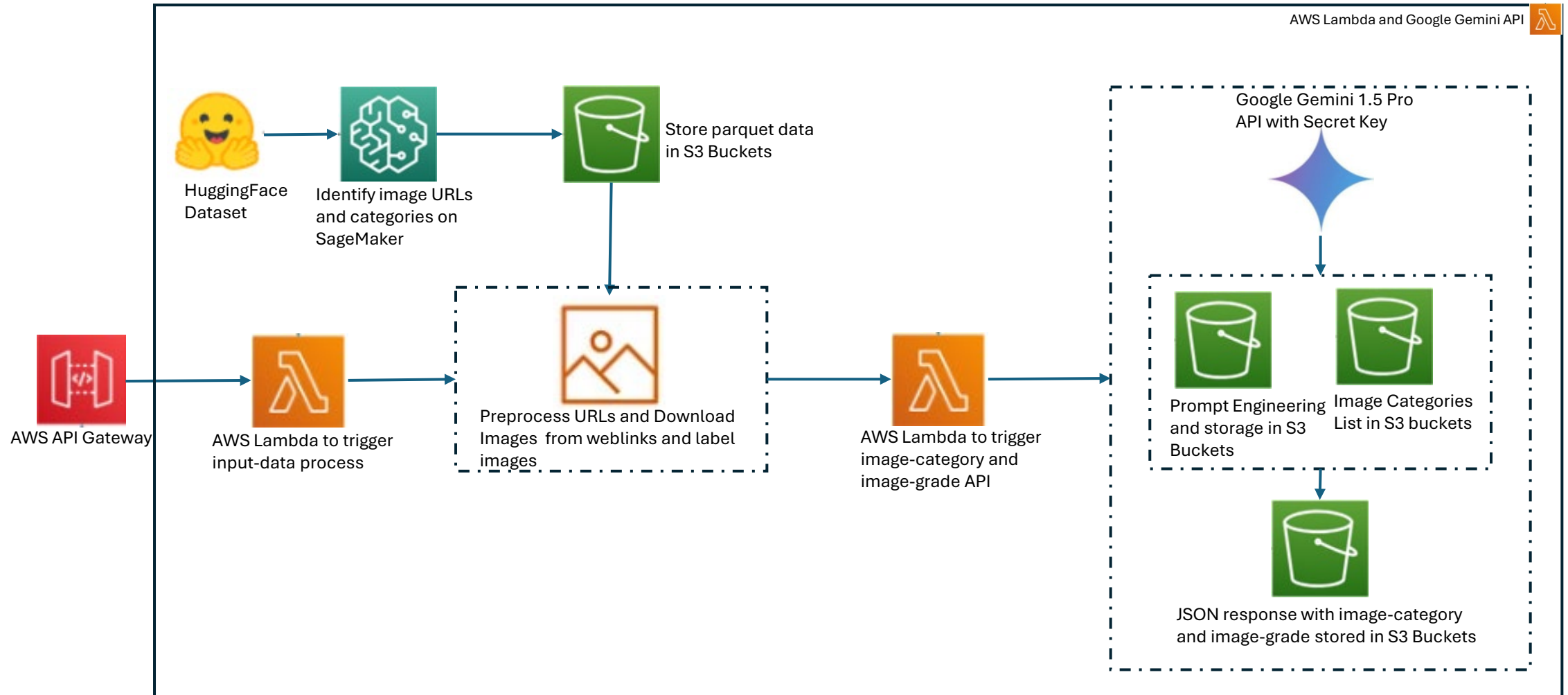
	Identify food in a dish	Grade food quantity
Approach-1	<p>Train and validate Image-classification model</p> <ul style="list-style-type: none"><li>-Convolutud Neural Networks</li><li>-Ground Truth data for model training</li></ul>	<p>Train and validate Image-grading model</p> <ul style="list-style-type: none"><li>-Convolutud Neural Networks</li><li>-Ground Truth data for model training</li></ul>
Approach-2	<p>Inference Google Gemini Pro LLM</p> <ul style="list-style-type: none"><li>-Prompt Engineering of Multi-modal Image-to-Text LLMs</li><li>-API Wrapper</li></ul>	<p>Inference Google Gemini Pro LLM</p> <ul style="list-style-type: none"><li>-Prompt Engineering of Multi-modal Image-to-Text LLMs</li><li>-API Wrapper</li></ul>

# Convolutud Neural Network





# End-to-End AI solution with Google Gemini LLM



# Performance Metrics

## Convolutional Neural Network Model:

- Classification Accuracy of CNN Model trained on 2000 images: 13.6%
- Limited image count within each category classification
- The performance is highly dependent on the Ground Truth data labeling and size

## Google Gemini LLM API Wrapper:

- Classification Accuracy: 95%
- Grading Accuracy: 60%
- Significantly low OpEx
- The performance is highly dependent on the Prompt Engineering