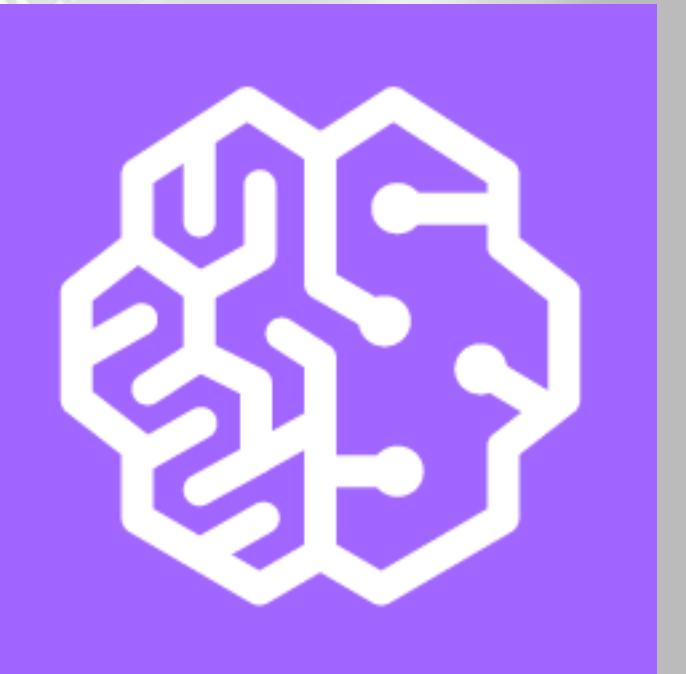
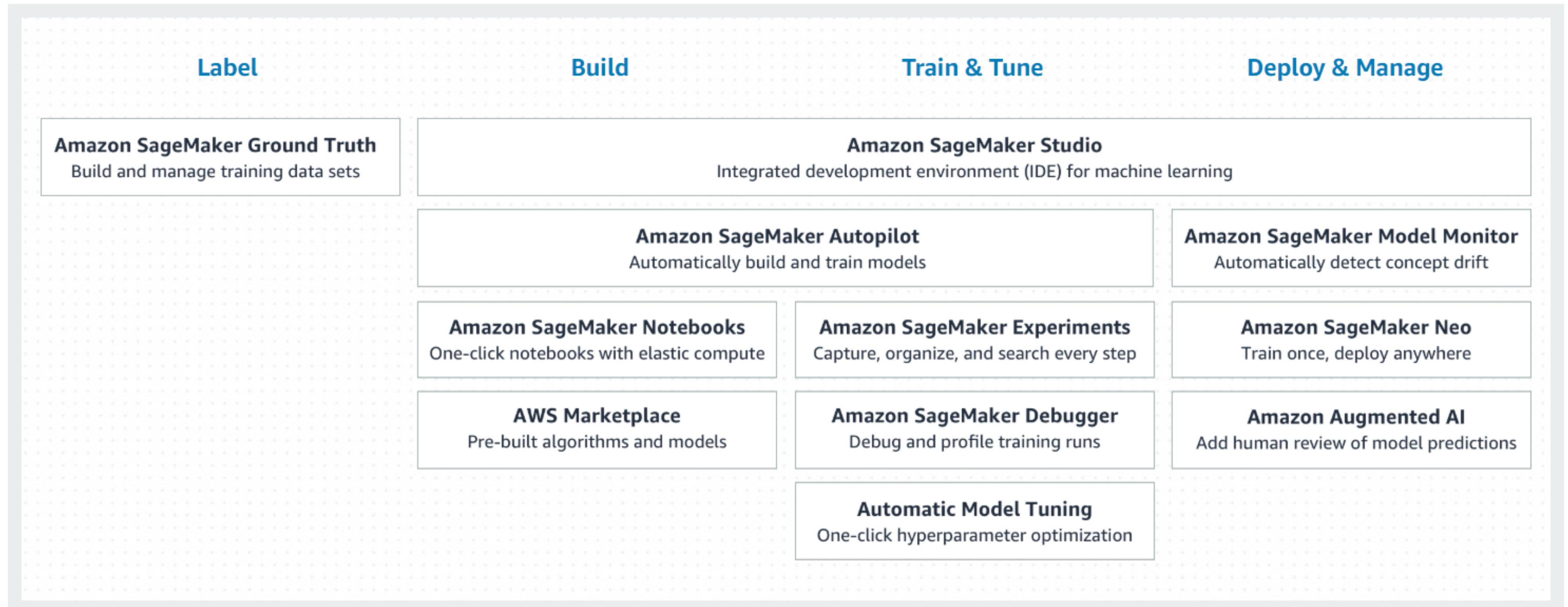


Deploying Models using AWS Sagemaker



What is Sagemaker

Amazon SageMaker is a fully managed service that provides every developer and data scientist with the ability to build, train, and deploy machine learning (ML) models quickly. SageMaker removes the heavy lifting from each step of the machine learning process to make it easier to develop high quality models.



Features of Sagemaker

0. Can be run with jupyter
1. Data labelling service
2. Training a Model
 - a. Prebuilt Algorithms (XGBoost, LinearLearner)
 - b. SDKs for frameworks like Pytorch, Tensorflow, MXNet, Sci-kit Learn
3. Creating/Updating Endpoints (A/B Testing, updating models etc.)
4. Automatic Hyperparameter Tuning
5. Monitoring model performance
6. Various docker containers available

AWS Lambda

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume.

With Lambda, you can run code for virtually any type of application or backend service - all with zero administration. Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically trigger from other AWS services or call it directly from any web or mobile app.

AWS API Gateway

Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. APIs act as the "front door" for applications to access data, business logic, or functionality from your backend services. Using API Gateway, you can create RESTful APIs and WebSocket APIs that enable real-time two-way communication applications. API Gateway supports containerized and serverless workloads, as well as web applications.

Today's Agenda

Things we will be doing today

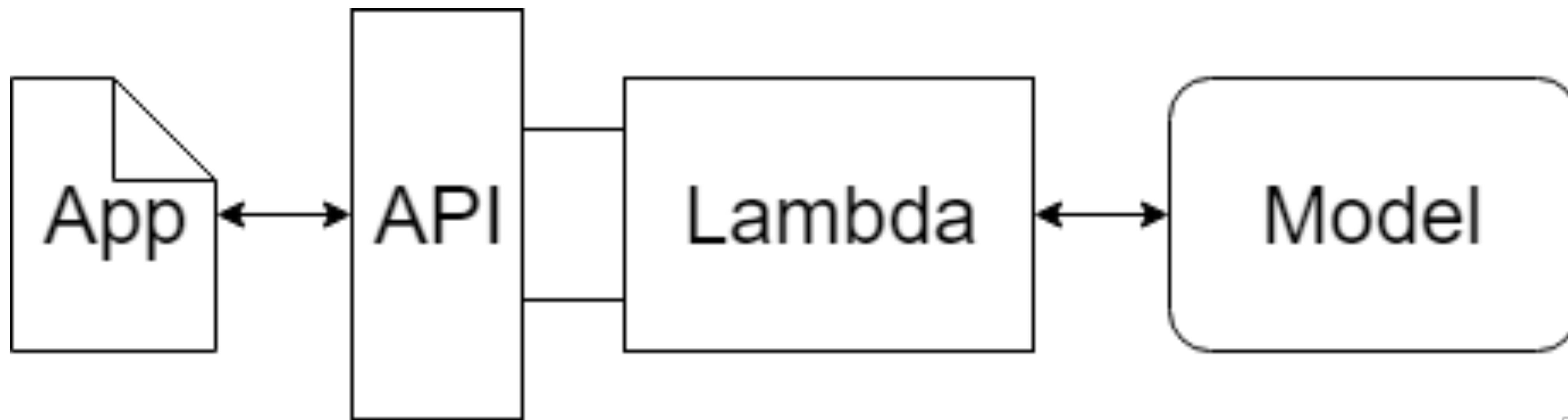
0. Sagemaker, Lambda and API Gateway
1. Take a Model trained with Tensorflow (tf.keras)
2. Convert the keras model to protobuf
3. Save the model artifacts in a sagemaker readable format
4. Load and Create the model in Sagemaker
5. Deploy the Model as an Endpoint
6. Invoking the endpoint using AWS Lambda
7. Communicating to Lambda from our App via API Gateway

Things we will NOT be doing today

1. Training a model in Sagemaker
2. Hyperparameter Tuning
3. Updating and monitoring models
4. Making Frontends

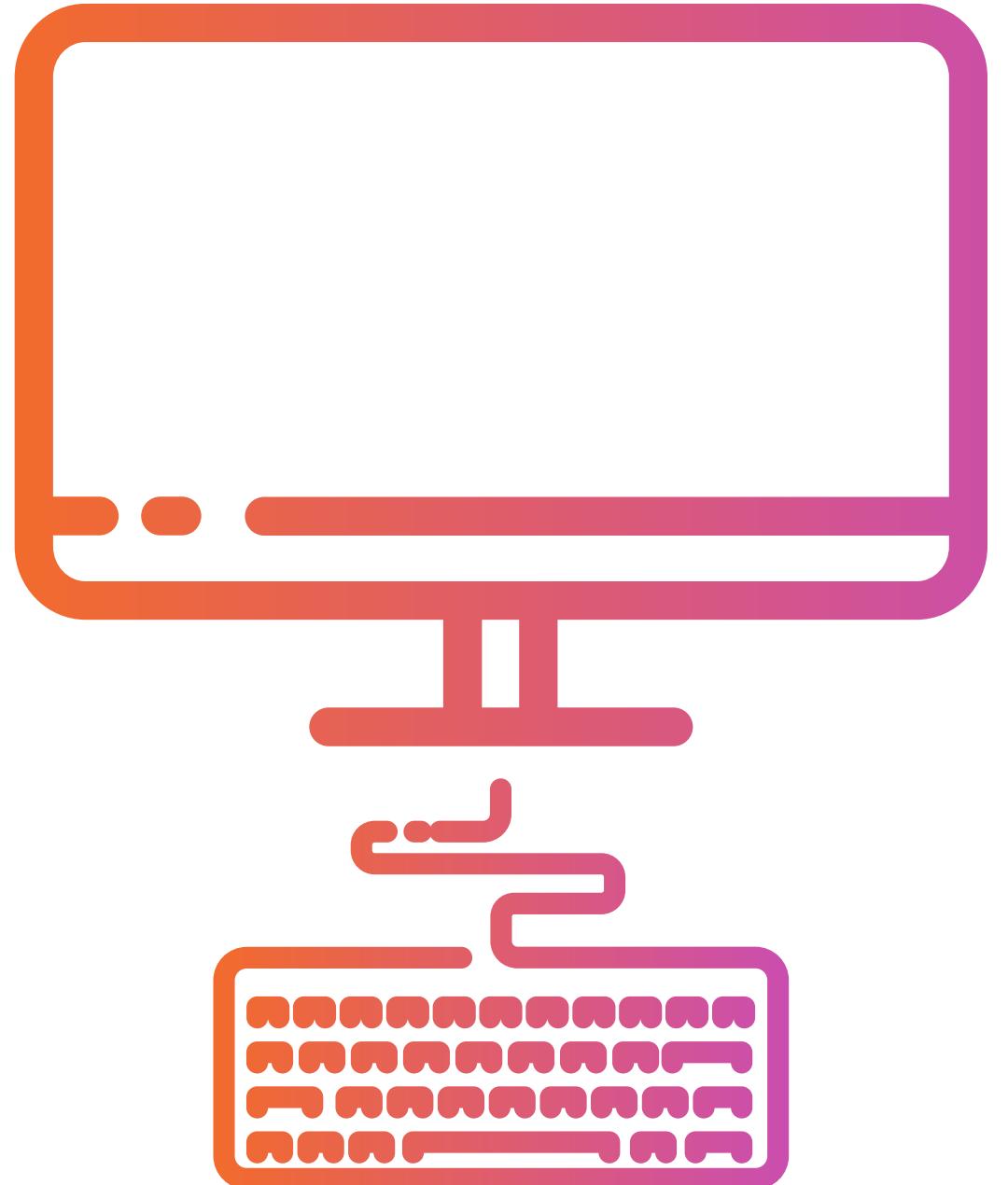
The above topics will be covered in
my Full-Stack AI Webinar Series

Basic Workflow



Thank You

Presented by:
Bhavesh Laddagiri



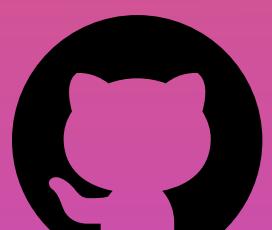
Contact Me at



bhavesh.laddagiri1@gmail.com



Bhavesh Laddagiri



@theneuralbeing

Sources:
<https://github.com/awslabs/amazon-sagemaker-examples/>
<https://sagemaker.readthedocs.io/>

<https://github.com/awslabs/amazon-sagemaker-examples/>
<https://sagemaker.readthedocs.io/>