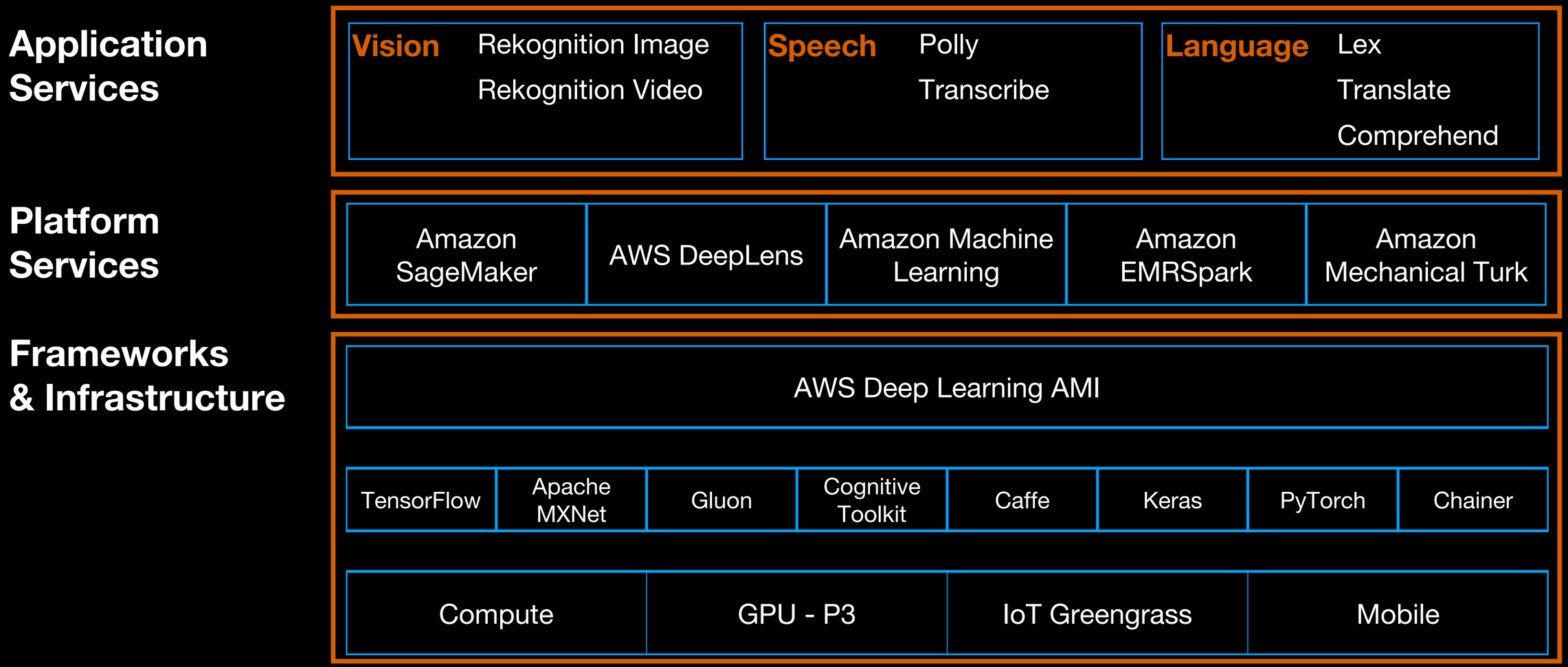
Amazon SageMaker

End-to-End Managed Machine Learning

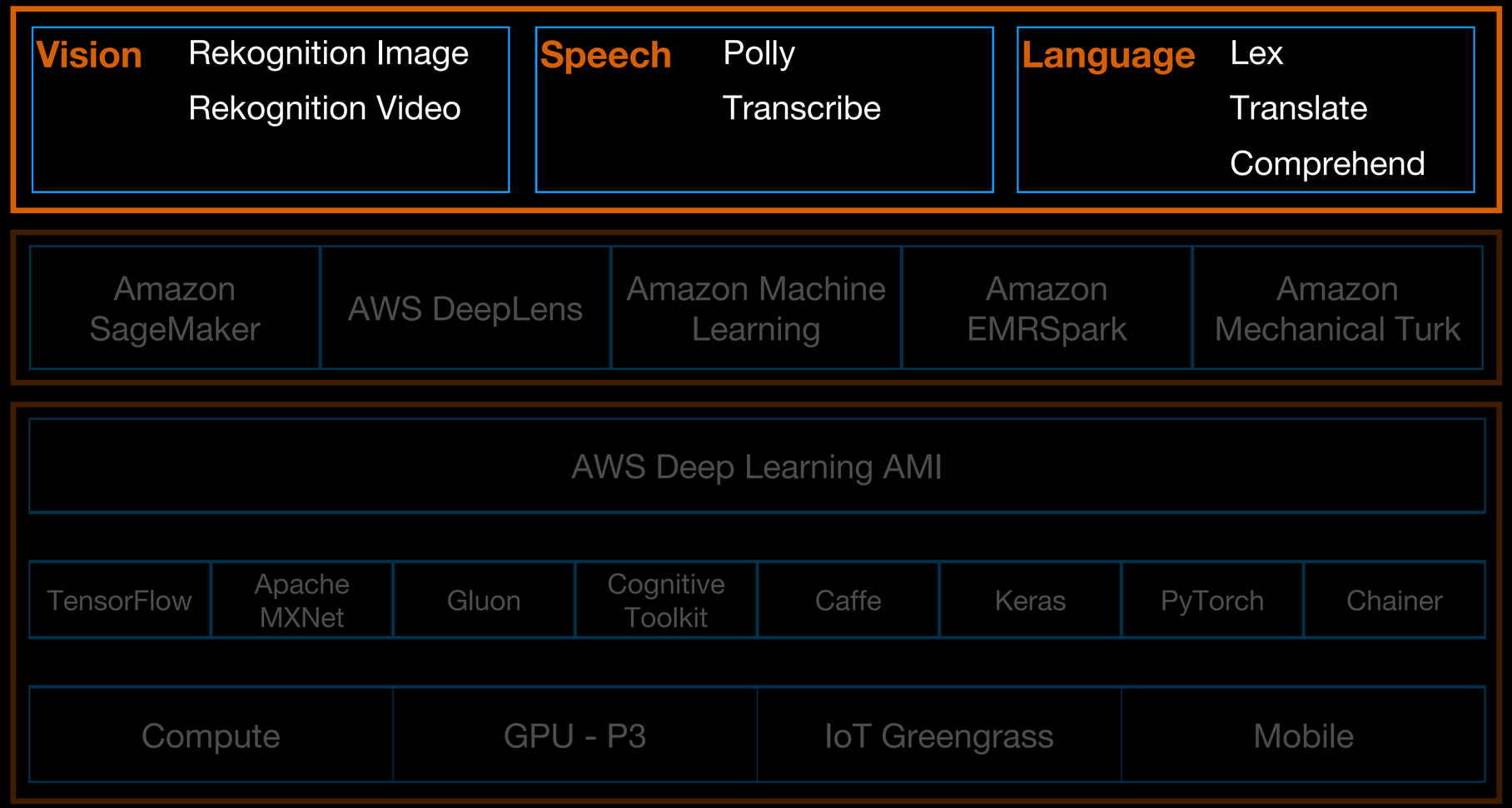
Our Mission

Put machine learning in the hands of every developer, data scientist and architect.

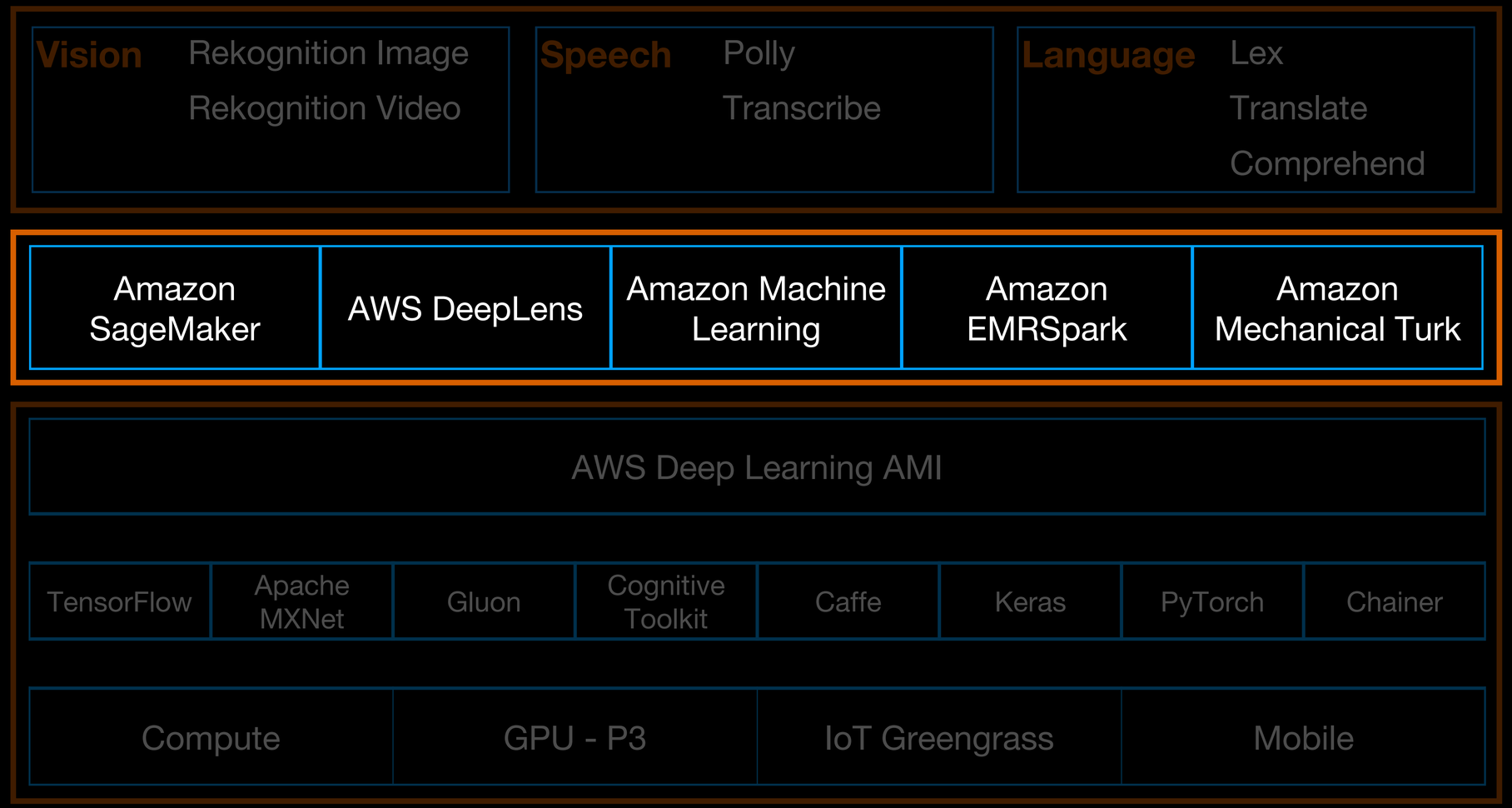
The AWS Machine Learning Stack



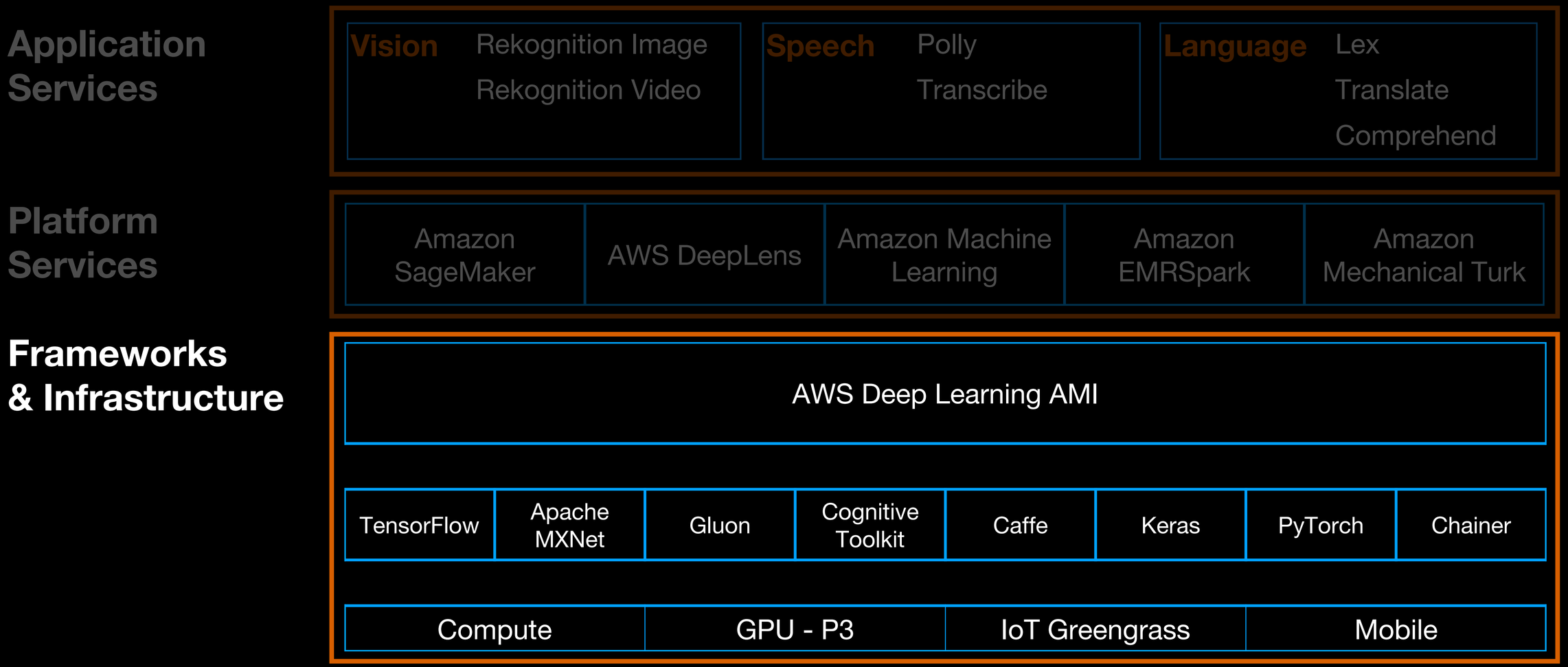
The AWS Machine Learning Stack



The AWS Machine Learning Stack



The AWS Machine Learning Stack



Amazon EC2 P3 Instances

The fastest, most powerful GPU instances in the cloud

- Up to 8 NVIDIA Tesla V100 GPUs
  - 16GB GPU memory with 900 GB/sec peak bandwidth
- 1 PetaFLOPs of computational performance
  - 14x better than P2
- 300 GB/s GPU-to-GPU communication (NVLink)
  - 9X better than P2

NEW!

Amazon ML Lab

Lots of companies doing Machine Learning

Lack ML expertise

Unable to unlock business potential

Amazon ML Lab provides the missing ML expertise

Leverage Amazon experts with decades of ML experience with technologies like Amazon Echo, Amazon Alexa, Prime Air and Amazon Go

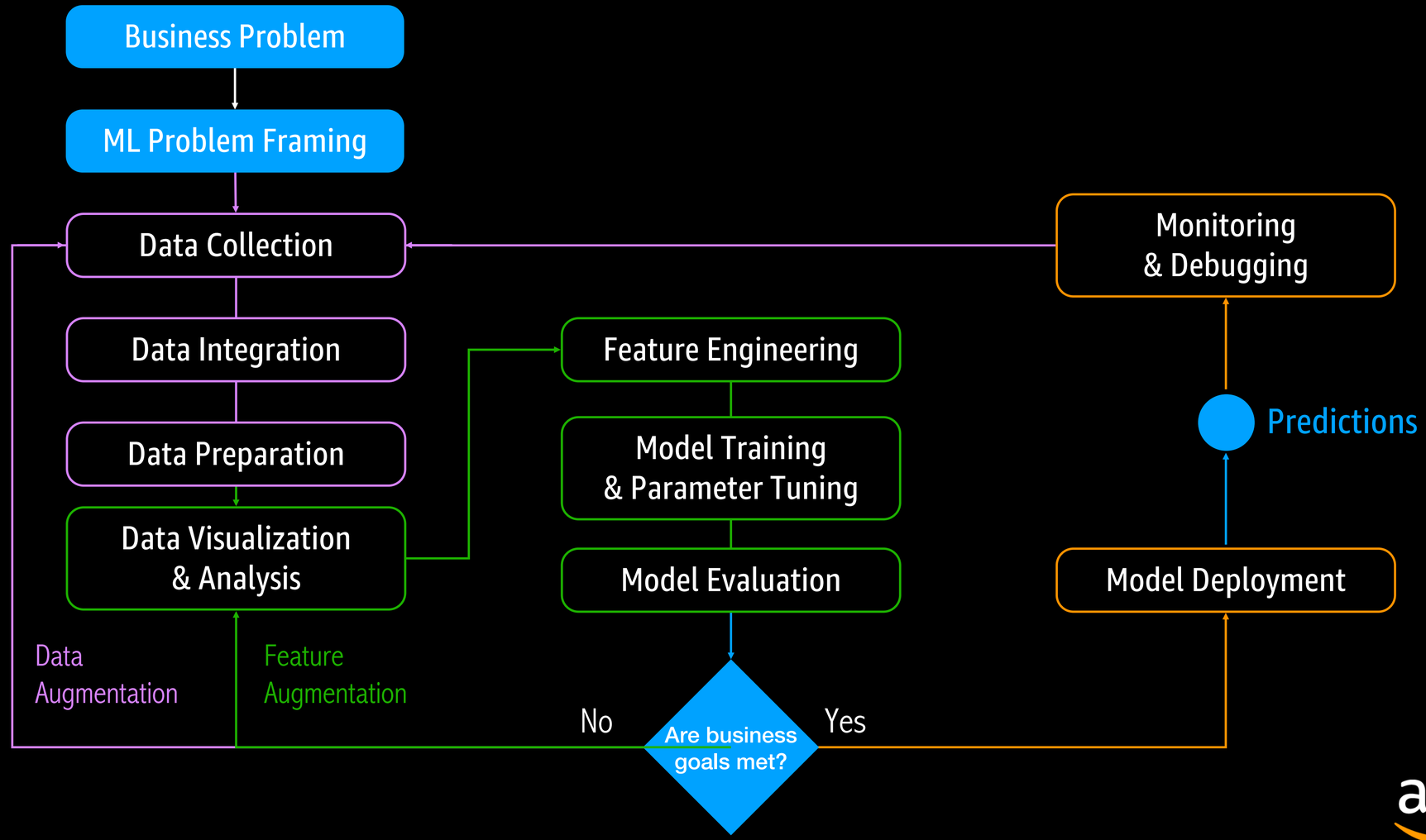
Brainstorming

Modeling

Teaching

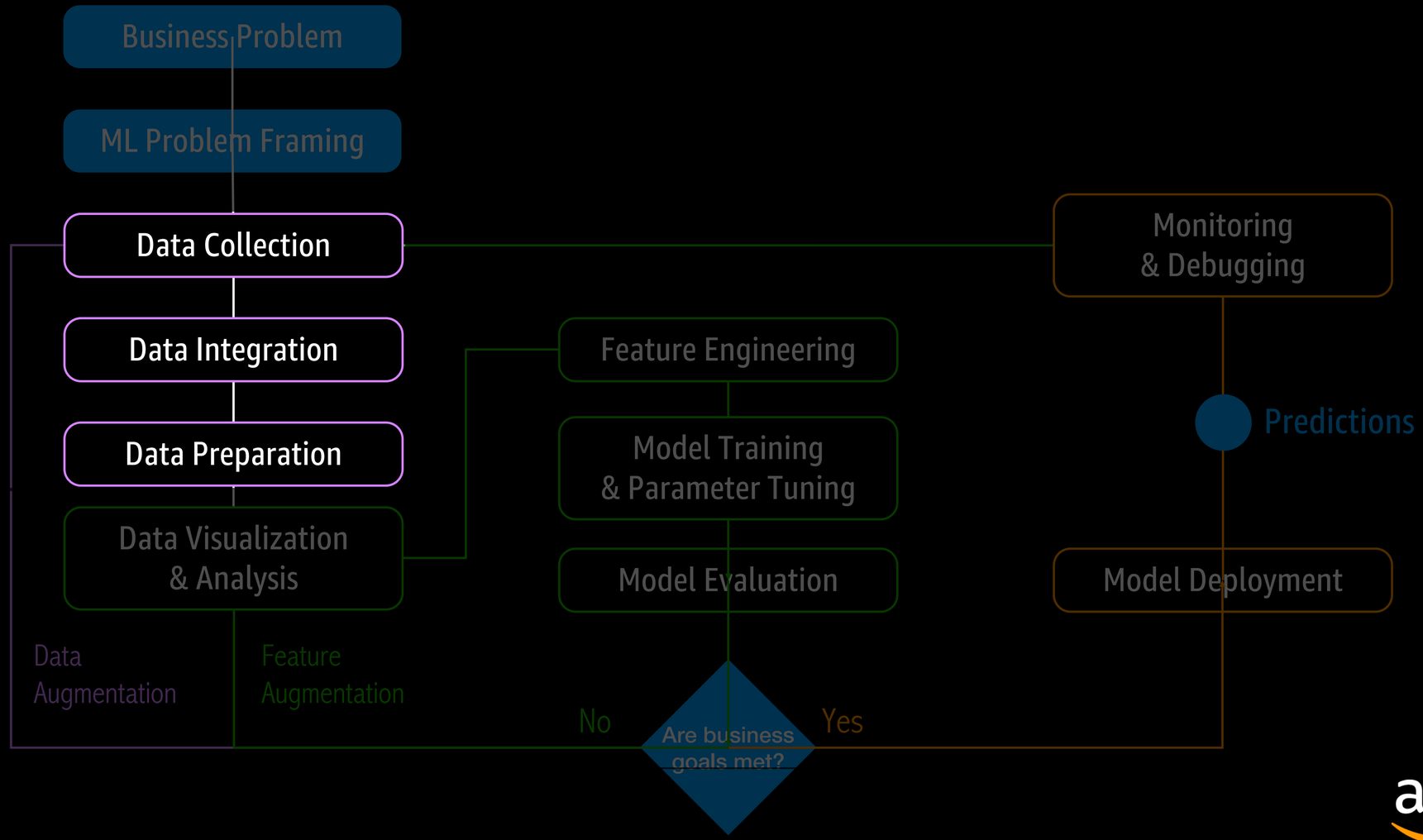
Let’s Review the ML Process

The Machine Learning Process



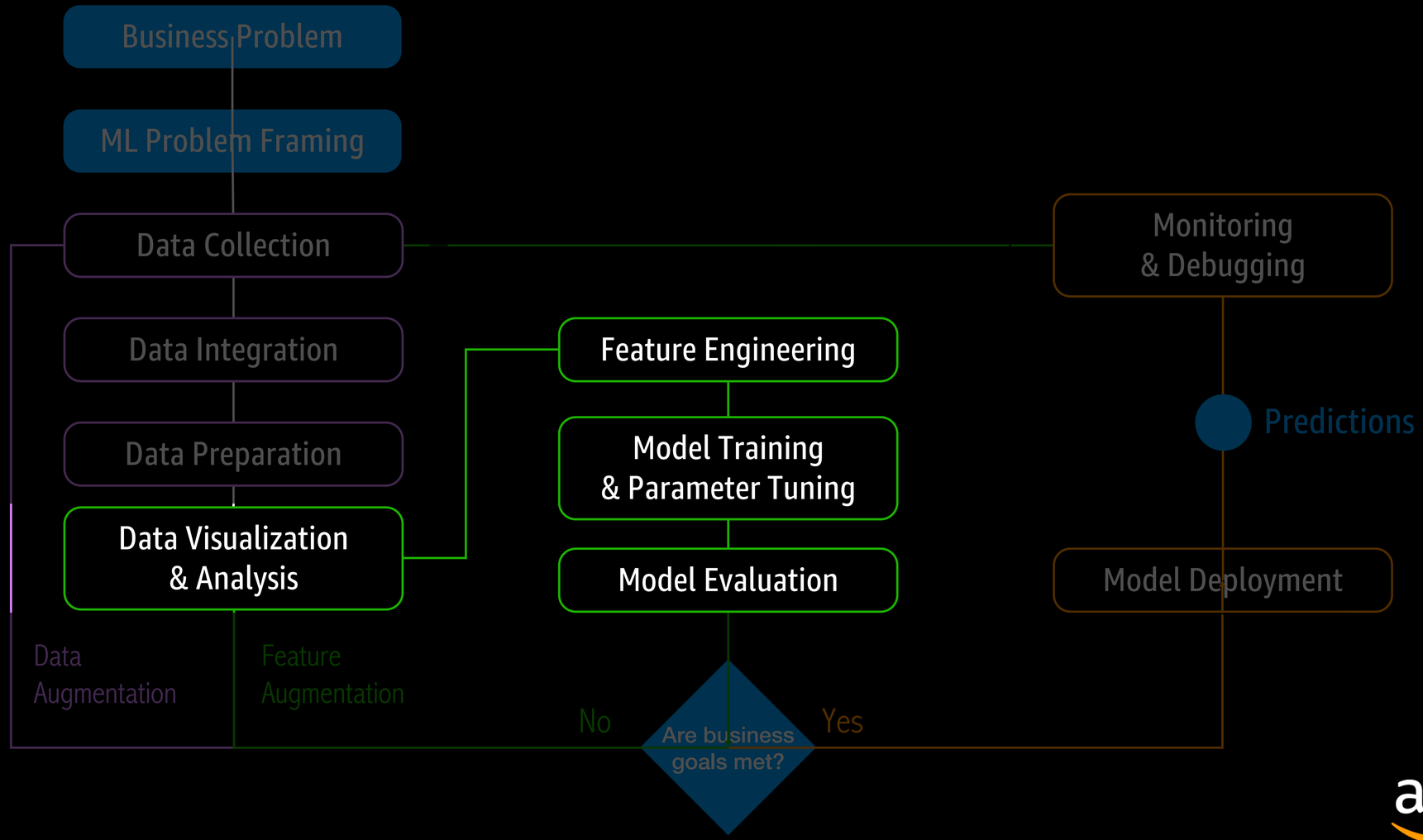
The Machine Learning Process
Integration: The Data Architecture

- Build the data platform:
  - Amazon S3
  - Amazon Athena
  - Amazon EMR
  - Amazon Redshift Spectrum
  - AWS Glue



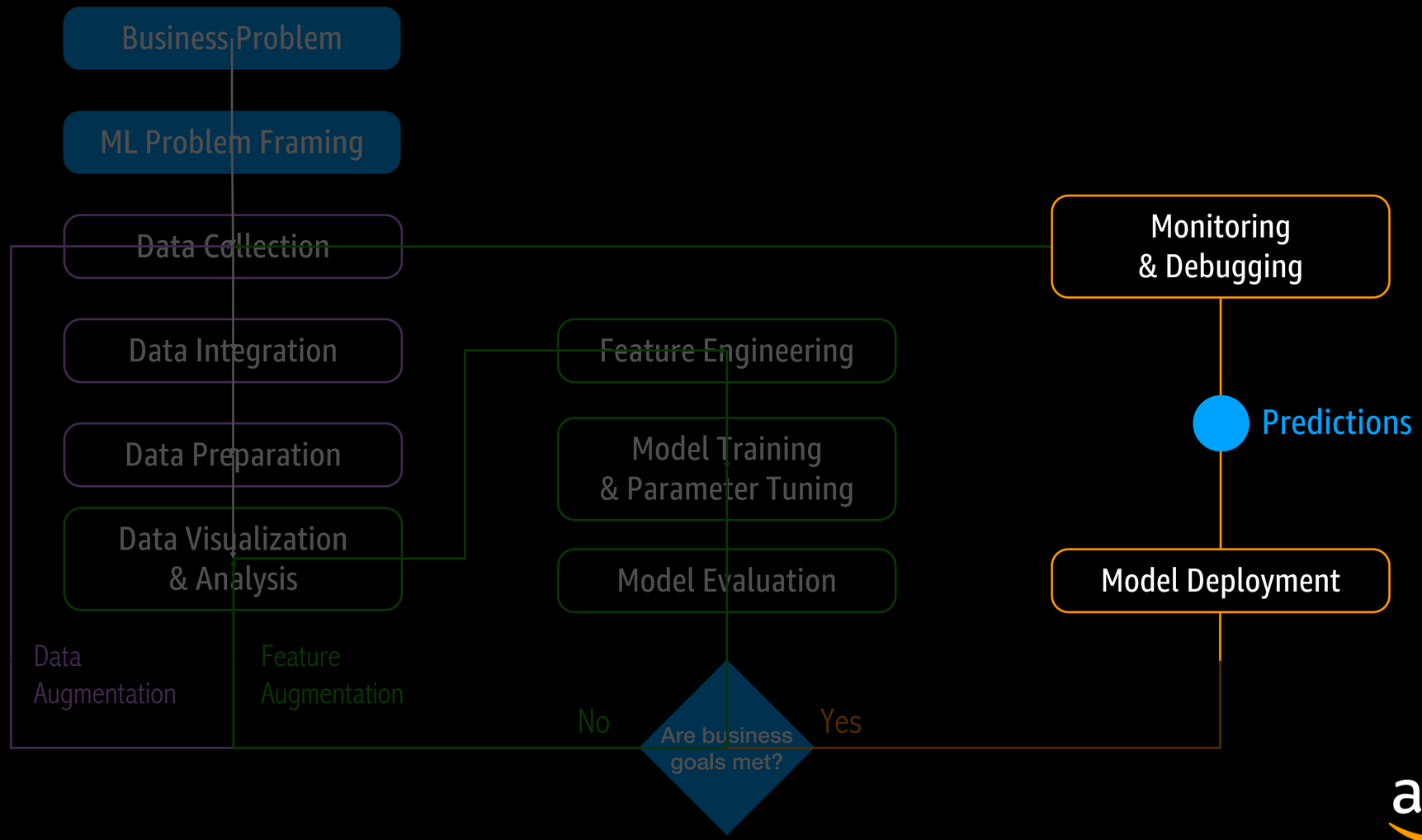
The Machine Learning Process
The Model Training: Undifferentiated Heavy Lifting

- Setup and manage
  - Notebook Environments
  - Training Clusters
- Write Data Connectors
- Scale ML algorithms to large datasets
- Distribute ML training algorithm to multiple machines
- Secure Model artifacts

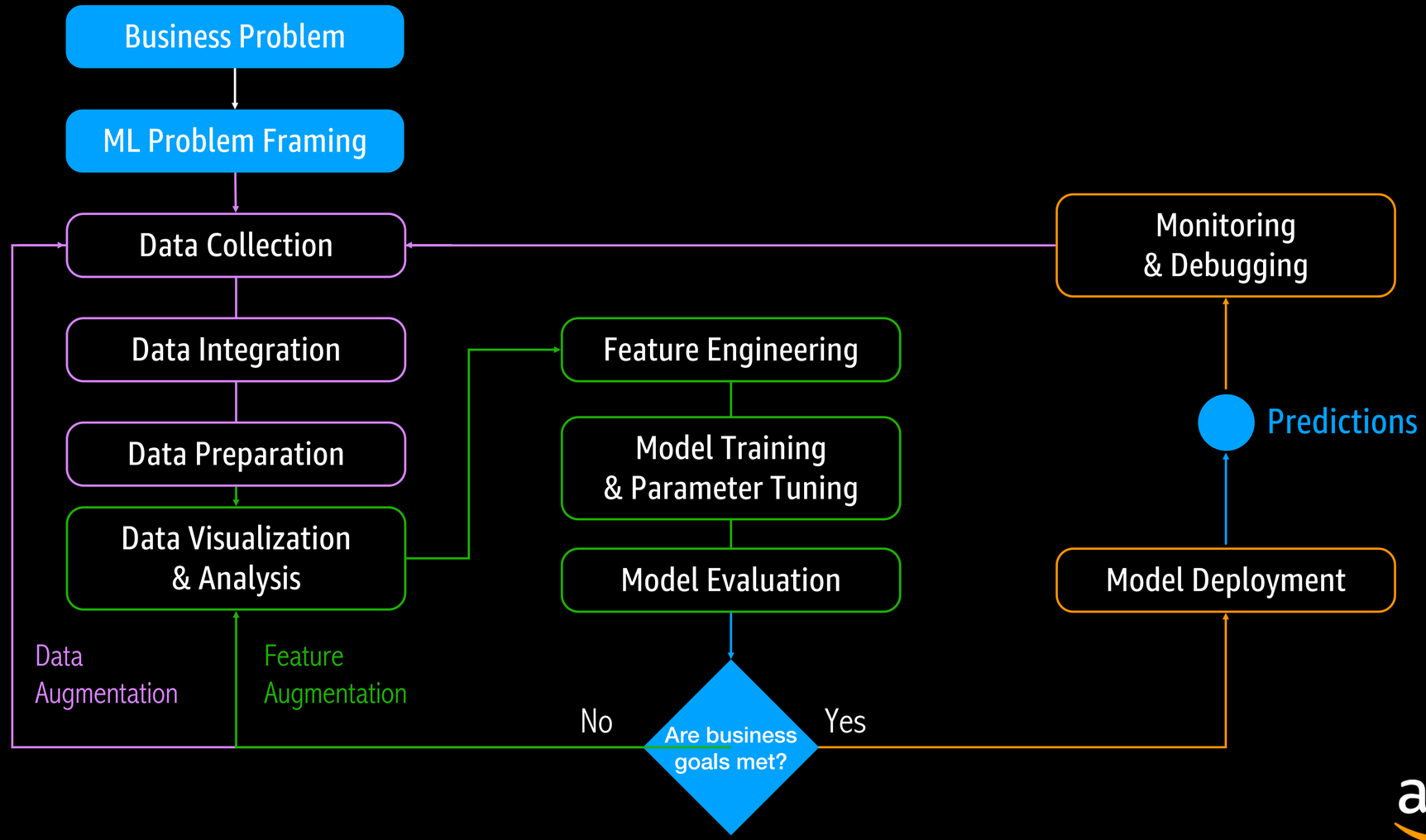


The Machine Learning Process
DevOps: Undifferentiated Heavy Lifting

- Setup and manage Model Inference Clusters
- Manage and Scale Model Inference APIs
- Monitor and Debug Model Predictions
- Models versioning and performance tracking
- Automate New Model version promotion to production (A/B testing)



The Machine Learning Process
Why We Built SageMaker



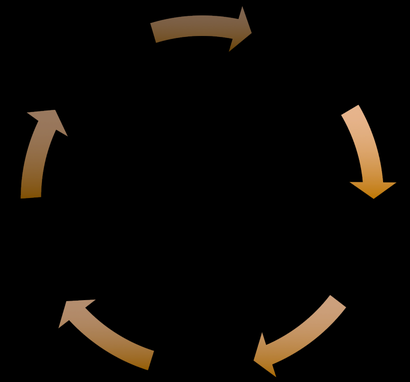
Amazon SageMaker

A fully-managed service that enables data scientists and developers to quickly and easily build machine-learning based models into smart production applications.

aws

Amazon SageMaker
Dramatically Accelerates the Machine Learning Process

Build **Train** **Deploy**



Easier training with hyperparameter optimization

Highly-optimized machine learning algorithms

One-click training for ML, DL, and custom algorithms

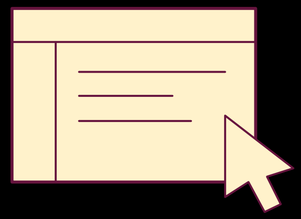
Easier training with hyperparameter optimization

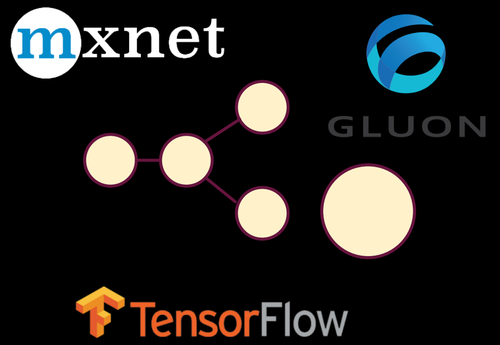
Fully-managed hosting at scale

Deployment without engineering effort

aws

Amazon SageMaker
Dramatically Accelerates the Machine Learning Process





End-to-End Machine Learning Platform

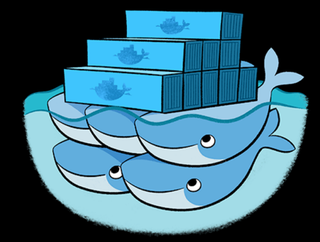
Zero setup

Flexible Model Training

Pay by the second

aws

A Fully-Dockerized Lifecycle
From discovery to development and deployment



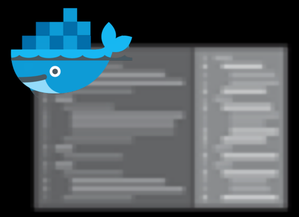
Data Scientists

Amazon SageMaker

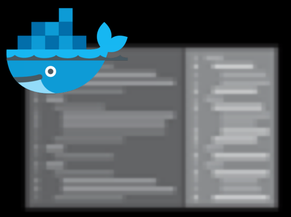
Model Artifacts

Amazon S3

Training Data



Training Algorithm

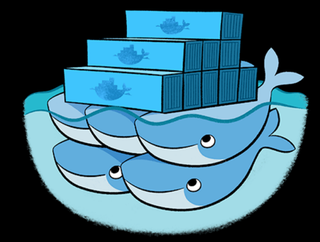


Inference Algorithm

Amazon ECR

aws

A Fully-Dockerized Lifecycle
From discovery to development and deployment



Developers and Operations

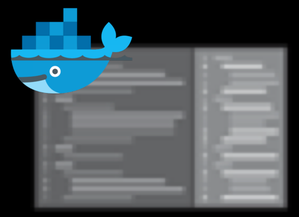
Amazon SageMaker

EndPoint

Amazon S3

Training Data

Model Artifacts



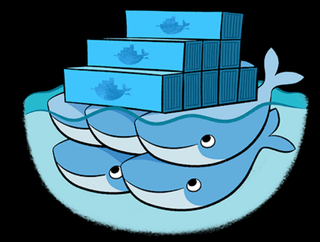
Training Algorithm

Inference Algorithm

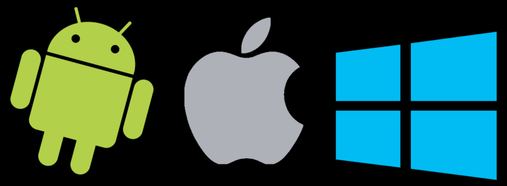
Amazon ECR

aws

A Fully-Dockerized Lifecycle
From discovery to development and deployment

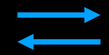


Delighted Customers



API Gateway

Identification /
Authorization /
Logging / Analytics

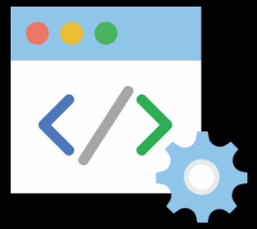


Predictive Model

aws

Amazon SageMaker

1

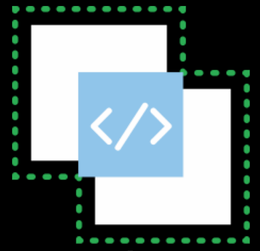


Notebook Instances

2

Algorithms

3



ML Training Service

4

ML Hosting Service

aws

Amazon SageMaker
Reference Architecture