

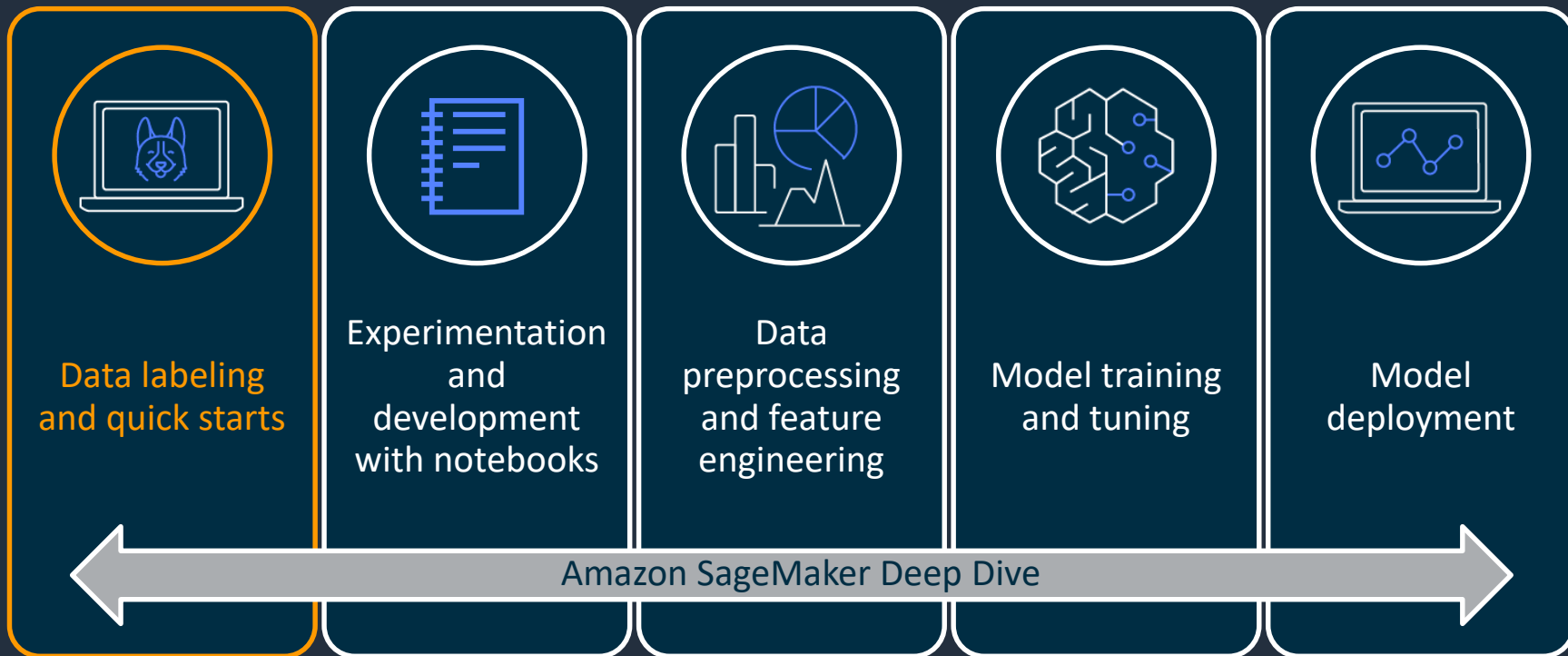


# Data labeling and quick starts

Amazon SageMaker Deep Dive Series



# Amazon SageMaker Deep Dive Series



# Amazon SageMaker key benefits

Most complete,  
end-to-end ML service



## Accelerate ML development

20+ tools covering the entire ML development lifecycle



## Boost data scientist productivity

The world's first integrated development environment (IDE)



## Reduce cost

Eliminate costs of writing custom integration code with integrated functionality optimized for ML

# Amazon SageMaker overview

## Amazon SageMaker

### PREPARE

#### SageMaker Ground Truth

Label training data for machine learning

#### SageMaker Data Wrangler

Aggregate and prepare data for machine learning

#### SageMaker Processing

Built-in Python, BYO R/Spark

#### SageMaker Feature Store

Store, update, retrieve, and share features

#### SageMaker Clarify

Detect bias and understand model predictions

### BUILD

#### SageMaker Studio Notebooks

Jupyter notebooks with elastic compute and sharing

#### Built-in and Bring your-own Algorithms

Dozens of optimized algorithms or bring your own

#### SageMaker Autopilot

Automatically create machine learning models with full visibility

#### SageMaker JumpStart

Pre-built solutions for common use cases

#### SageMaker Canvas

Generate accurate machine learning predictions—no code required

#### SageMaker Studio Lab

Learn and experiment with ML using a no-setup, free development environment

#### RStudio

Fully integrated development environment for machine learning

### TRAIN & TUNE

#### Managed Training

Distributed infrastructure management

#### SageMaker Experiments

Capture, organize, and compare every step

#### Automatic Model Tuning

Hyperparameter optimization

#### Distributed Training Libraries

Training for large datasets and models

#### SageMaker Debugger

Debug and profile training runs

#### Managed Spot Training

Reduce training cost by 90%

#### Managed Training Compiler

Accelerate training of deep learning models by up to 50%

### DEPLOY & MANAGE

#### Managed Deployment

Fully managed, ultra low latency, high throughput

#### Kubernetes & KubeFlow Integration

Simplify Kubernetes-based machine learning

#### Multi-Model Endpoints

Reduce cost by hosting multiple models per instance

#### SageMaker Model Monitor

Maintain accuracy of deployed models

#### SageMaker Edge Manager

Manage and monitor models on edge devices

#### SageMaker Pipelines

Workflow orchestration and automation

#### SageMaker Inference Recommender

Automate load testing and optimize model performance across ML instances

### SageMaker Studio

Integrated development environment (IDE) for ML

# Features covered in this session

## Amazon SageMaker

### PREPARE

#### SageMaker Ground Truth

Label training data for machine learning

#### SageMaker Data Wrangler

Aggregate and prepare data for machine learning

#### SageMaker Processing

Built-in Python, BYO R/Spark

#### SageMaker Feature Store

Store, update, retrieve, and share features

#### SageMaker Clarify

Detect bias and understand model predictions

### BUILD

#### SageMaker Studio Notebooks

Jupyter notebooks with elastic compute and sharing

#### Built-in and Bring your-own Algorithms

Dozens of optimized algorithms or bring your own

#### SageMaker Autopilot

Automatically create machine learning models with full visibility

#### SageMaker JumpStart

Pre-built solutions for common use cases

#### SageMaker Canvas

Generate accurate machine learning predictions—no code required

#### SageMaker Studio Lab

Learn and experiment with ML using a no-setup, free development environment

#### RStudio

Fully integrated development environment for machine learning

### TRAIN & TUNE

#### Managed Training

Distributed infrastructure management

#### SageMaker Experiments

Capture, organize, and compare every step

#### Automatic

#### Model Tuning

Hyperparameter optimization

#### Distributed Training

#### Libraries

Training for large datasets and models

#### SageMaker Debugger

Debug and profile training runs

#### Managed Spot Training

Reduce training cost by 90%

#### Managed Training Compiler

Accelerate training of deep learning models by up to 50%

### DEPLOY & MANAGE

#### Managed Deployment

Fully managed, ultra low latency, high throughput

#### Kubernetes & Kubeflow Integration

Simplify Kubernetes-based machine learning

#### Multi-Model Endpoints

Reduce cost by hosting multiple models per instance

#### SageMaker Model Monitor

Maintain accuracy of deployed models

#### SageMaker Edge Manager

Manage and monitor models on edge devices

#### SageMaker Pipelines

Workflow orchestration and automation

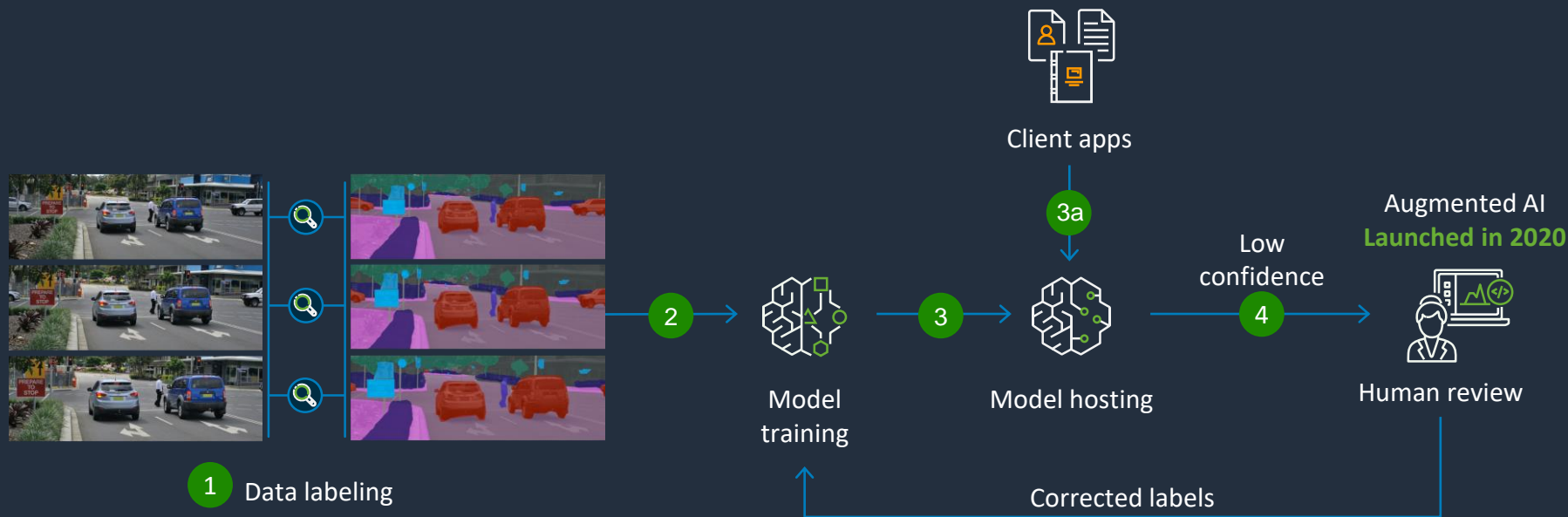
#### SageMaker Inference Recommender

Automate load testing and optimize model performance across ML instances

### SageMaker Studio

Integrated development environment (IDE) for ML

# Data labeling in machine learning

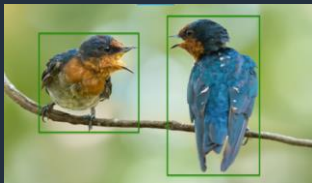


SageMaker Ground Truth  
Launched in 2018

# 30+ data labeling workflows



Image classification



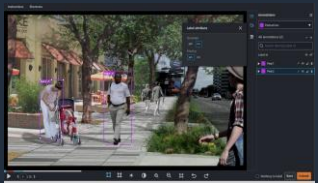
Bounding boxes



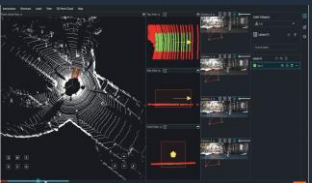
Semantic segmentation



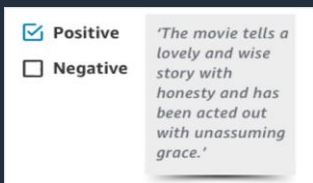
Label Verification



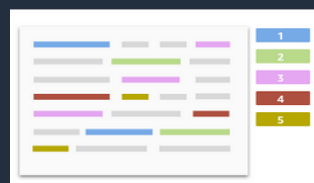
Video



LIDAR 3D Point Cloud



Text classification



Named entity recognition

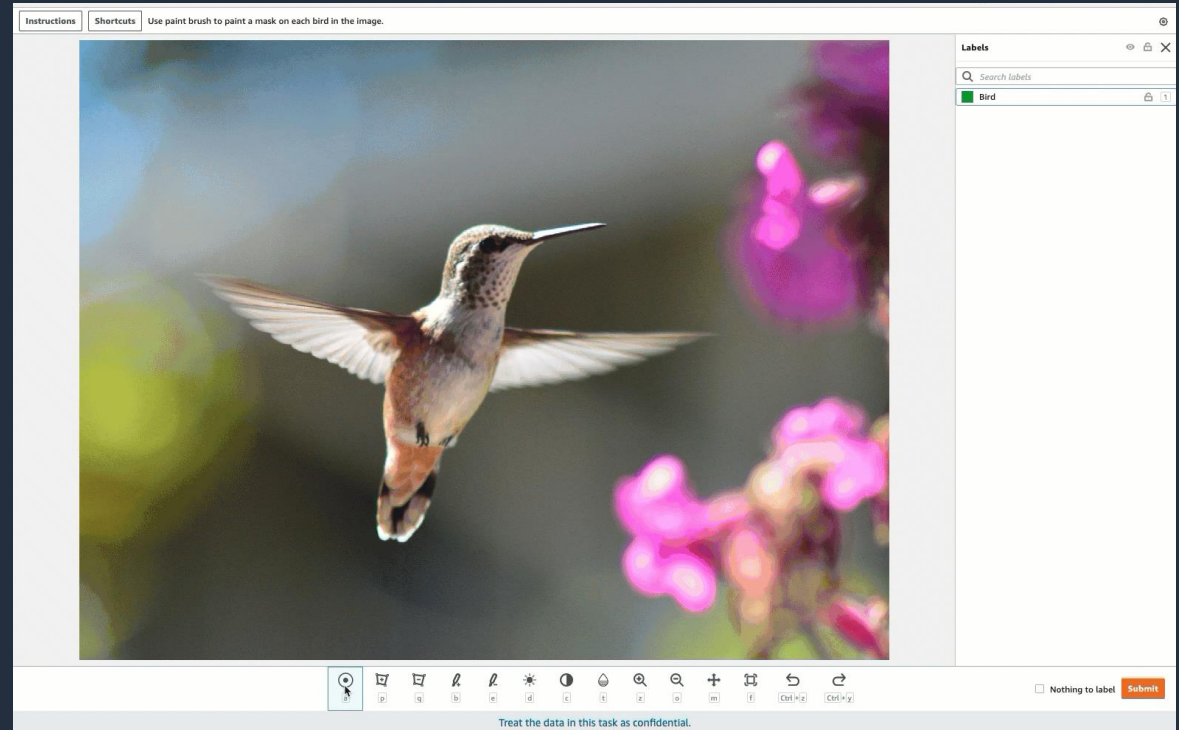


Custom (25 Templates)

- Key point
- Line
- Instance Segmentation

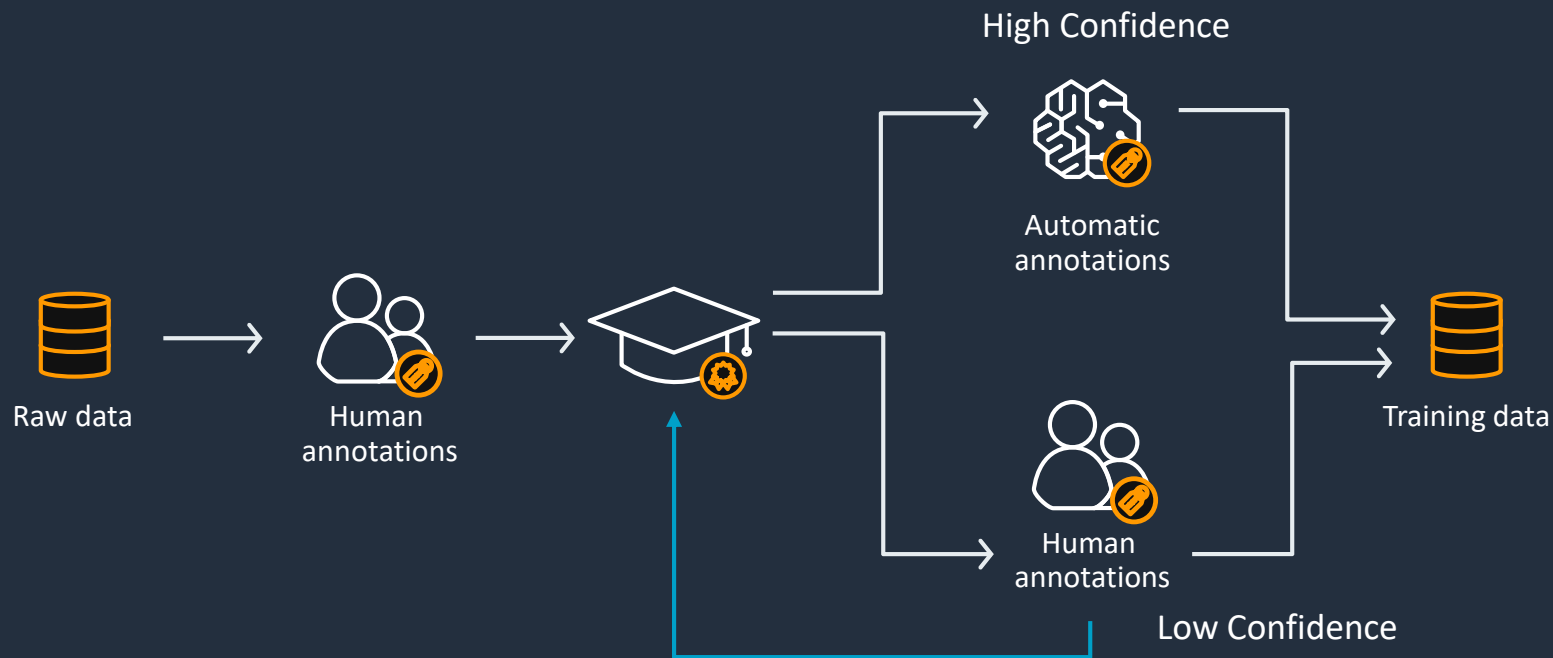
# Assistive tooling to improve labeling efficiency

- Auto-Segmentation
- Predict Next Frame
- Interpolation
- Ground Detection
- Cuboid Snapping





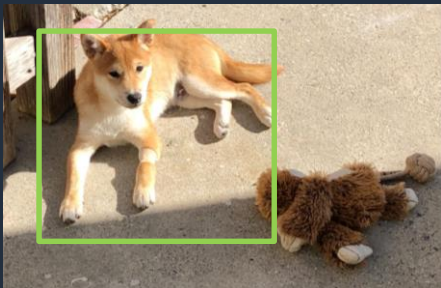
# Auto-labeling to reduce labeling cost by up to 70%



# Annotation consolidations to improve label quality by 20%



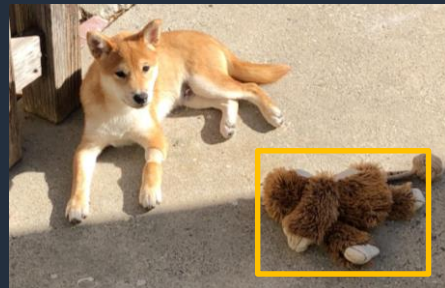
Bob



Shirley



Dan



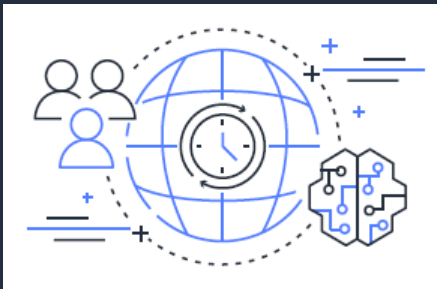
*Amazon SageMaker Ground Truth uses  
annotation consolidation to output a  
single high-accuracy label*

# Manifest files

```
{ "source-ref": " s3://AWSDOC-EXAMPLE-BUCKET/image_example.png " }  
{ "source-ref": " s3://AWSDOC-EXAMPLE-BUCKET/image_example2.png " }  
...  
{ "source-ref": " s3://AWSDOC-EXAMPLE-BUCKET/image_examp1en.png " }
```

```
{  
  "source-ref": "s3://AWSDOC-EXAMPLE-BUCKET/image_example.png",  
  "sport": 0,  
  "sport-metadata":  
  {  
    "class-name": "football",  
    "confidence": 0.00,  
    "type": "groundtruth/image-classification",  
    "job-name": "identify-sport",  
    "human-annotated": "yes",  
    "creation-date": "2018-10-18T22:18:13.527256"  
  }  
}
```

# Your choice of workforce options



## Mechanical Turk

An on-demand 24x7 workforce of over 500,000 independent contractors worldwide, powered by Amazon Mechanical Turk



## Private

A team of workers that you have sourced yourself, including your own employees or contractors, for handling data that needs to stay within your organization



## Vendor

A curated list of third-party vendors that specialize in providing data labeling services, available via the AWS Marketplace

# SageMaker Ground Truth Plus



Turnkey  
solution



Guaranteed  
throughput



Higher quality and  
feedback loop



Predictable and  
lower cost



Tracking &  
visibility



Data security and operational excellence

# A turnkey data labeling service



# Demo

SageMaker Ground Truth

# SageMaker Canvas

Build ML models and generate accurate predictions—no code required



Quickly access and prepare data for Machine Learning



Built-in AutoML to build models and generate accurate predictions



Share ML models and collaborate with data science teams



Usage-based pricing to avoid licensing fees and reduce TCO



# SageMaker Canvas use cases



## Sales and Marketing

1. Sales conversion
2. Sales forecasting
3. Propensity to churn
4. Customer lifetime value prediction
5. Marketing mix modeling



## Finance and Accounting

1. Credit risk scoring
2. Delayed payments prediction
3. Fraud detection
4. Portfolio optimization
5. Account payables automation



## Operations and Logistics

1. Demand forecasting
2. Inventory planning and scheduling
3. Delivery time forecasting
4. Predictive Maintenance

*and many  
more...*

# Demo

SageMaker Canvas

# SageMaker Autopilot

Automatic model creation for tabular data with full visibility and control



## Quick to start

Provide your data in a  
tabular form and specify  
target prediction



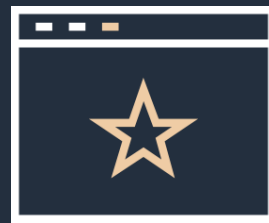
## Automatic model creation

Get ML models with  
feature engineering &  
model tuning automatically  
done



## Visibility and control

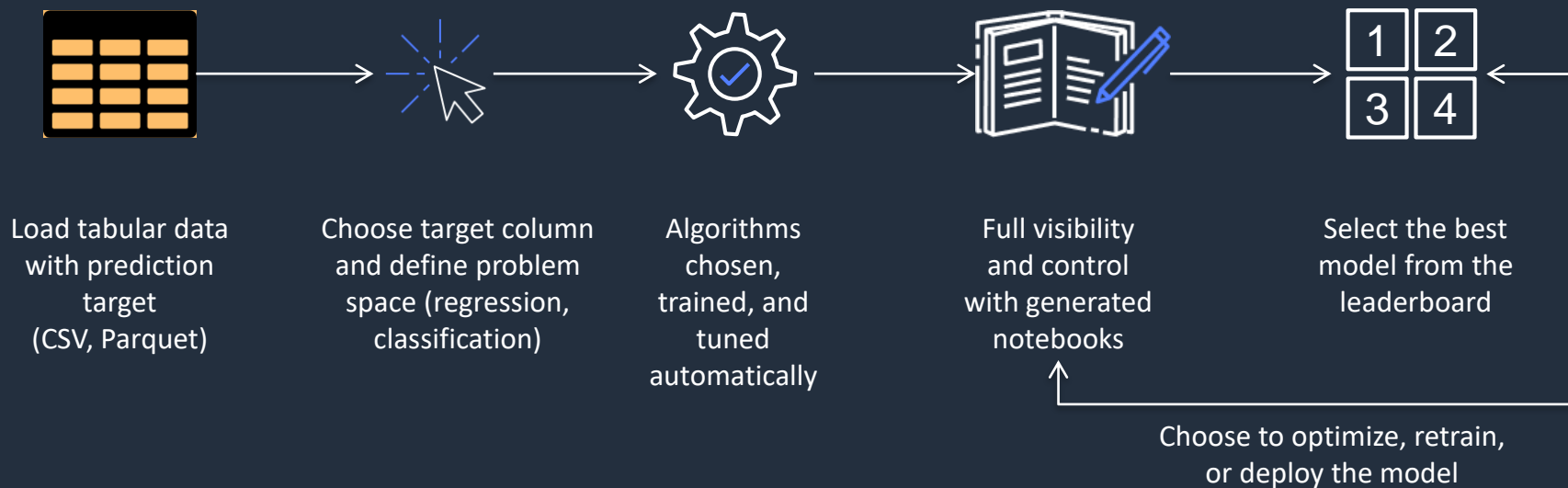
Get notebooks for your  
models with source code



## Recommendations and optimization

Get a leaderboard &  
continue to improve your  
model

# How Autopilot works



# Demo

SageMaker Autopilot

# SageMaker JumpStart

Easily and quickly bring  
machine learning  
applications to market



16 pre-built solutions for common ML use cases



Accelerate time to deploy over 300 open source models



Get started with just a few clicks

# SageMaker JumpStart Solutions

## Predictive Maintenance

*Detecting potential  
equipment failure for  
manufacturing*

*Predictive maintenance  
for vehicle fleets*

## Demand Forecasting

*Deep Demand  
Forecasting*

## Fraud Detection

*Detect Malicious Users  
and Transactions*

*Fraud Detection in  
Financial Transactions*

## Credit Risk Prediction

*Corporate Credit  
Rating Prediction*

*Explain Credit  
Decisions*

## Extract & Analyze Data from Documents

*Document  
Understanding*

*Handwriting  
Recognition*

*Intelligently fill in  
missing form data*

## Computer Vision

*Product Defect  
Detection*

## Autonomous Driving

*Visual Perception with  
Active Learning*

*Reinforcement  
Learning for  
Battlesnake AI*

## Personalized Recommendations

*Entity resolution for  
smart advertising*



*Purchase Modeling*

## Churn Prediction

*Churn Prediction*

# SageMaker JumpStart open-source models

300+ pre-trained, state-of-the-art, open source models from PyTorch Hub, TensorFlow Hub, Hugging Face ...

	TASKS	MODELS
 <b>VISION</b>	Image Classification Image Embedding Object Detection Semantic Segmentation	ResNet, Inception, MobileNet, SSD, Faster RCNN, YOLO, & More
 <b>TEXT</b>	Sentence Classification Text Classification Question Answering Summarization Text Generation, Translation, Named Entity Recognition	BERT, RoBERTa, DistilBERT, Distillbart xsum, GPT2, ELECTRA, & More
 <b>TABULAR</b>	Classification, Regression	LightGBM, CatBoost, XGBoost



# Demo

SageMaker JumpStart

# Features covered in this session

## Amazon SageMaker

### PREPARE

#### SageMaker Ground Truth

Label training data for machine learning

#### SageMaker Data Wrangler

Aggregate and prepare data for machine learning

#### SageMaker Processing

Built-in Python, BYO R/Spark

#### SageMaker Feature Store

Store, update, retrieve, and share features

#### SageMaker Clarify

Detect bias and understand model predictions

### BUILD

#### SageMaker Studio Notebooks

Jupyter notebooks with elastic compute and sharing

#### Built-in and Bring your-own Algorithms

Dozens of optimized algorithms or bring your own

#### SageMaker Autopilot

Automatically create machine learning models with full visibility

#### SageMaker JumpStart

Pre-built solutions for common use cases

#### SageMaker Canvas

Generate accurate machine learning predictions—no code required

#### SageMaker Studio Lab

Learn and experiment with ML using a no-setup, free development environment

#### RStudio

Fully integrated development environment for machine learning

### TRAIN & TUNE

#### Managed Training

Distributed infrastructure management

#### SageMaker Experiments

Capture, organize, and compare every step

#### Automatic

#### Model Tuning

Hyperparameter optimization

#### Distributed Training

#### Libraries

Training for large datasets and models

#### SageMaker Debugger

Debug and profile training runs

#### Managed Spot Training

Reduce training cost by 90%

#### Managed Training Compiler

Accelerate training of deep learning models by up to 50%

### DEPLOY & MANAGE

#### Managed Deployment

Fully managed, ultra low latency, high throughput

#### Kubernetes & Kubeflow Integration

Simplify Kubernetes-based machine learning

#### Multi-Model Endpoints

Reduce cost by hosting multiple models per instance

#### SageMaker Model Monitor

Maintain accuracy of deployed models

#### SageMaker Edge Manager

Manage and monitor models on edge devices

#### SageMaker Pipelines

Workflow orchestration and automation

#### SageMaker Inference Recommender

Automate load testing and optimize model performance across ML instances

### SageMaker Studio

Integrated development environment (IDE) for ML

# Resources

[Documentation] [Amazon SageMaker Ground Truth and Ground Truth Plus](#)

[Code Sample] [Amazon SageMaker Ground Truth sample task Uis](#)

[Tutorials] [Amazon SageMaker Ground Truth](#)

[Workshop] [Easily label training data for machine learning at scale](#)

[Documentation] [Amazon SageMaker Canvas](#)

[Workshop] [Amazon SageMaker Canvas Immersion Day](#)

[Documentation] [Amazon SageMaker Autopilot](#)

[Code Samples] [Amazon SageMaker Autopilot](#)

[Tutorials] [Amazon SageMaker Autopilot](#)

[Workshop] [Amazon SageMaker Autopilot](#)

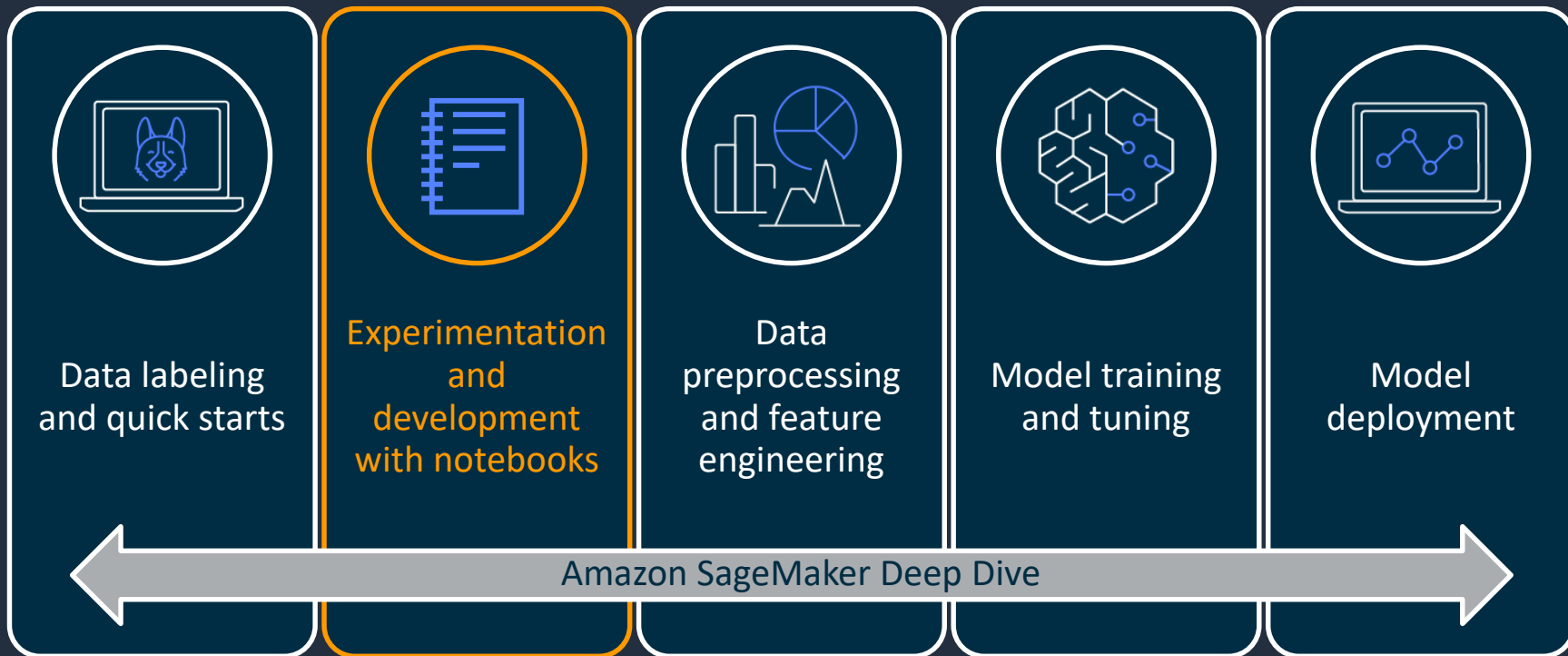
[Documentation] [Amazon SageMaker JumpStart](#)

[Code Samples] [Amazon SageMaker JumpStart](#)

[Tutorials] [Amazon SageMaker JumpStart Industry](#)

[Tutorials] [Use Prebuilt Models with SageMaker JumpStart](#)

# Join us in the next session





# Thank you!

## Australia & New Zealand

Brad Ryan

AI/ML Partner Development Specialist

[brdryn@amazon.com](mailto:brdryn@amazon.com)

Sara van de Moosdijk

AI/ML Partner Solutions Architect

[sarmoosd@amazon.com](mailto:sarmoosd@amazon.com)

## ASEAN

Ling Chang

AI/ML Partner Development Specialist

[linglych@amazon.com](mailto:linglych@amazon.com)

Vasileios Vonikakis

AI/ML Partner Solutions Architect

[vonikakv@amazon.com](mailto:vonikakv@amazon.com)

## India

Ankit Kandoi

Data Analytics & ML Partner  
Development Specialist

[akkandoi@amazon.com](mailto:akkandoi@amazon.com)