

AI and Machine Learning on AWS

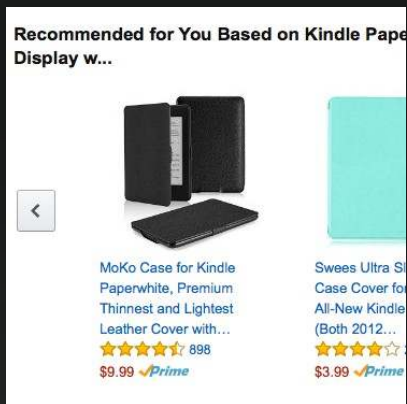
Julien Simon
Global Evangelist, AI & Machine Learning
Amazon Web Services

[@julsimon](#)

<https://medium.com/@julsimon>

<https://youtube.com/juliensimonfr>

AMAZON'S MACHINE LEARNING INNOVATION



Our mission at AWS

Put machine learning in the
hands of every developer

WHY AWS FOR ML?



Broadest and deepest set of AI and ML services

200+ new features and services launched this last year alone

Solutions for everyone from ML scientists to application developers

Support all major frameworks



Accelerate your adoption of ML with SageMaker

Single IDE for the entire ML workflow

At least 54% lower TCO

Up to 70% cost reduction in data-labeling

Up to 90% cost reduction with managed spot training



Built on the most comprehensive cloud platform

Highly secure, reliable, fully featured data store

The strongest set of compute, storage, security, database, and analytics capabilities to build upon

85% TensorFlow in the cloud runs on AWS











AI and Machine Learning on AWS

Tens of thousands of active customers – all sizes, all verticals




The AWS ML Stack

AI SERVICES

VISION	SPEECH	TEXT	SEARCH	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD	DEVELOPMENT	CONTACT CENTERS
									
Amazon Rekognition	Amazon Polly Amazon Transcribe <i>+Medical</i>	Amazon Comprehend <i>+Medical</i> Amazon Translate Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Amazon CodeGuru	Contact Lens <i>For Amazon Connect</i>

ML SERVICES

 Amazon SageMaker	Ground Truth	ML Marketplace	SageMaker Studio IDE								Neo	Augmented AI
			Built-in algorithms	Notebooks	Experiments	Processing & Model Evaluation	Model training & tuning	Debugger	Autopilot	Model hosting	Model Monitor	

ML FRAMEWORKS &

INFRASTRUCTURE

 TensorFlow  mxnet  GLUON  Keras  PYTORCH  Scikit-learn  Deep Graph Library	Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Inferentia	FPGA
--	------------------------------------	----------------	----------------------	------------	------

REDUCE LOCALIZATION COSTS AND IMPROVE ACCURACY

Website & document translation | Recorded call analysis | Video subtitling | Accessibility



Hotels.com



Lionbridge



POLLY



TRANSCRIBE



TRANSLATE



COMPREHEND

ACCURATELY FORECAST FUTURE BUSINESS OUTCOMES

Workforce planning | Product and advertising demand | Sales by store |
Web traffic projection | Inventory optimization | AWS usage



FORECAST

INCREASE EFFICIENCY WITH AUTOMATED DOCUMENT ANALYSIS

Fast archive search | Automated form processing | Systematic redaction



TEXTTRACT



COMPREHEND &
COMPREHEND
MEDICAL

Amazon SageMaker helps you build, train, and deploy models


Prepare Build

Train & Tune

Deploy & Manage


Web-based IDE for machine learning

Fully managed data processing jobs and data labeling workflows




Collect and prepare training data

One-click collaborative notebooks and built-in, high performance algorithms and models




Choose or build an ML algorithm

One-click training




Set up and manage environments for training

Debugging and optimization



Train, debug, and tune models


Visually track and compare experiments



Manage training runs


Automatically build and train models

One-click deployment and auto scaling




Deploy model in production

Automatically spot concept drift




Monitor models

Add human review of predictions



Validate predictions

Fully managed with auto-scaling for 75% less cost



Scale and manage the production environment

Modular service and APIs, from experimentation to production



Amazon SageMaker Studio

Fully integrated development environment (IDE) for machine learning



Collaboration at
scale

Share notebooks
without tracking code
dependencies



Easy experiment
management

Organize, track, and
compare thousands of
experiments



Automatic model
generation

Get accurate models
with full visibility & control
without writing code



Higher quality ML
models

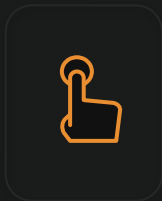
Automatically debug errors,
monitor models, & maintain
high quality



Increased
productivity

Code, build, train, deploy,
& monitor in a unified
visual interface

Modeling options



AWS Marketplace
for Machine
Learning



Training code



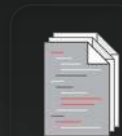
Amazon SageMaker
AutoPilot

Factorization Machines
Linear Learner
Principal Component
Analysis
K-Means Clustering
Etc.

Built-in Algorithms (17)
No ML coding required



Built-in Frameworks
Bring your own code
Open source containers



Bring Your Own
Full control, run your container
R, C++, etc.

Fully managed training, spot instances included

Training a TensorFlow model

```
from sagemaker.tensorflow import TensorFlow

tf_estimator = TensorFlow(
    entry_point='my_script.py',
    role=role,
    train_instance_count=1,
    train_instance_type='ml.p3.2xlarge',      # 1 GPU
    framework_version='2.2.0', py_version='py3',
    hyperparameters={'epochs': 10} )

tf_estimator.fit('s3://bucket/path_to_training_data')
```

Training a TensorFlow model at scale

```
from sagemaker.tensorflow import TensorFlow

tf_estimator = TensorFlow(
    entry_point='my_script.py',
    role=role,
    train_instance_count=8,
    train_instance_type='ml.p3.16xlarge',    # 64 GPUs total
    framework_version='2.2.0', py_version='py3',
    hyperparameters={'epochs': 100} )

tf_estimator.fit('s3://bucket/path_to_training_data')
```

Deployment options



Model in Amazon S3

Amazon SageMaker
real-time endpoint

1 line of code
Vanilla HTTPS
Post data, get
predictions
Any tool, any language
Auto Scaling available

Amazon SageMaker
batch transform

1 line of code
Predict data stored in S3
Read results in S3

Amazon
container services
(ECS, EKS, Fargate)

Use AWS Deep Learning
containers
Use your own container

Anywhere
you like

Grab the model
in S3 and run!

Fully managed deployment

Deploying a TensorFlow model to an HTTPS endpoint

```
from sagemaker.tensorflow import TensorFlow

tf_estimator = TensorFlow(entry_point='tf-train.py', ...)
tf_estimator.fit(inputs)

predictor = tf_estimator.deploy(initial_instance_count=1, instance_type='ml.c5.2xlarge')
```

```
from sagemaker.tensorflow import TensorFlowModel

tf_model = TensorFlowModel(model_data='s3://mybucket/model.tar.gz', ...,
                           entry_point='entry.py', name='model_name')

predictor = tf_model.deploy(initial_instance_count=1, instance_type='ml.c5.x2large')
```


https://ml.aws

Julien Simon
Global Evangelist, AI & Machine Learning
Amazon Web Services

[@julsimon](#)

<https://medium.com/@julsimon>

<https://youtube.com/juliensimonfr>