AWS INNOVATE

ONLINE CONFERENCE 2019

MACHINE LEARNING & AI EDITION

Opening Keynote

Swami Sivasubramanian, VP AI & Machine Learning, AWS

Olivier Klein, Head of Emerging Technologies, AWS

aws

Self-driving cars or assisted driving

Voice assistants, IVR and chatbots

Recommendation engines

Fraud detection

Automatic Fulfillment Centers

Computer assisted health diagnosis

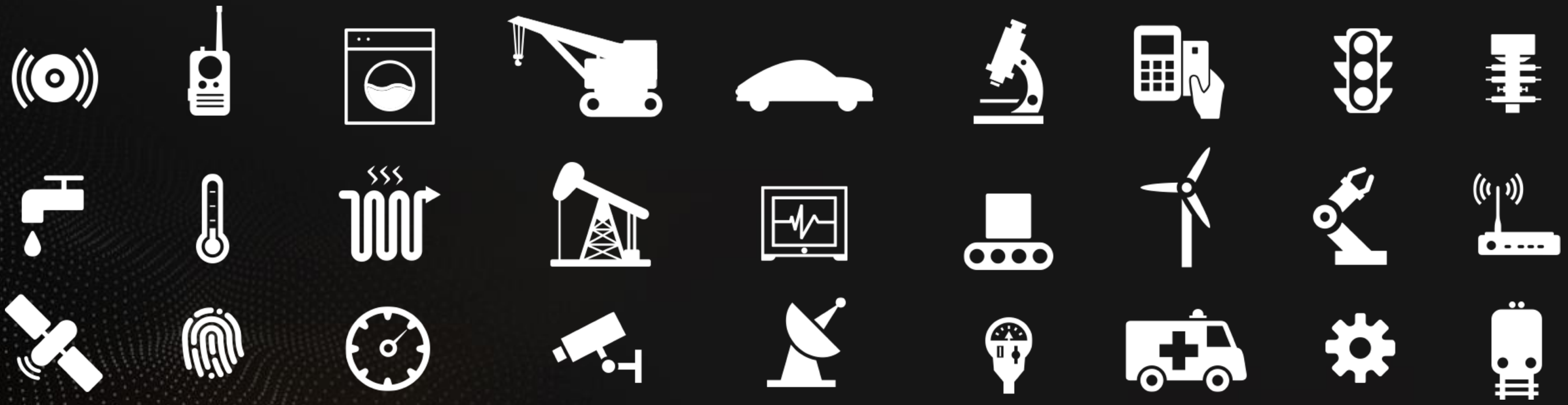
Anomaly detection in manufacturing

Demand forecasting in F&B

Spam E-Mail filtering

Facial recognition airport check-in

HUMAN CENTRIC



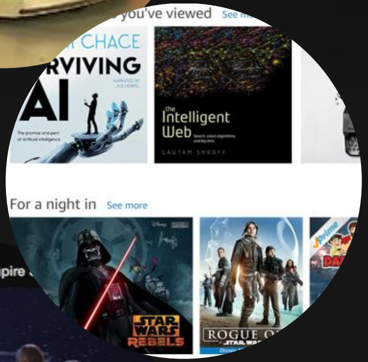
AUTOMATION
& ROBOTICS

NATURAL LANGUAGE
UNDERSTANDING
& VOICE PROCESSING

COMPUTER VISION

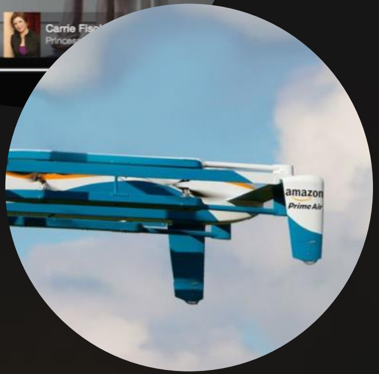
At Amazon,
we've been making investments in ML for the last 20 years…

Fulfillment & logistics



Search & discovery

Existing products



New products

AWS INNOVATE

ONLINE CONFERENCE 2019

MACHINE LEARNING & AI EDITION

KEYNOTE

- AI & ML ON AWS

TRACKS

- APPLCATION OF AI/ML
- BUILD AND DEPLOY ML MODELS
- FULLY MANAGED AI/ML SERVICES
- END TO END ML
- LIVE CODING

LIVE DEMO

- ROBOTICS AND AUTOMATION

AWS INNOVATE

ONLINE CONFERENCE 2019

MACHINE LEARNING & AI EDITION

Swami Sivasubramanian
VP AI & Machine Learning, AWS

Our Mission at AWS

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

Centerpiece for Digital Transformation

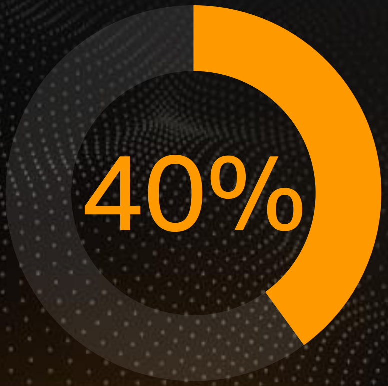
Customer experience

Business operations

Decision making

Innovation

Competitive advantage



OF DIGITAL TRANSFORMATION INITIATIVES SUPPORTED BY AI IN 2019

The AWS Machine Learning Stack

AI Services

(App Developers with no knowledge of ML)

VISION
- REKOGNITION IMAGE
- REKOGNITION VIDEO

SPEECH
- POLLY
- TRANSCRIBE

LANGUAGE
- TRANSLATE
- COMPREHEND

CHATBOTS & CONTACT CENTERS
- LEX

ML Services

AMAZON SAGEMAKER

BUILD | TRAIN | DEPLOY

ML Frameworks + Infrastructure

FRAMEWORKS
- TensorFlow
- mxnet
- PYTORCH
- Chainer
- HOROVOD

INTERFACES
- GLUON
- Keras

INFRASTRUCTURE
- EC2 P3
- EC2 C5
- FPGAs
- GREENGRASS

aws

More ML Happens on AWS than Anywhere Else

Tens of thousands
of active developers running ML on AWS

250%
growth YoY

2x
customer references

8 out of 10
of all TensorFlow workloads in the cloud run on AWS

Nucleus Research, November 2018

More ML Happens on AWS than Anywhere Else

Our Unique Approach

Customer-focused
90%+ of our ML roadmap is defined by customers

Pace of innovation
200+ new ML launches in the last year

Breadth and depth
More AI and ML services in production than any other provider

Multi-framework
Support for the most popular frameworks

Security and analytics
Deepest set of security and encryption features, with robust analytics capabilities

Embedded R&D
Customer centric approach to advancing the state of the art

Driving Better Healthcare Outcomes

Using Amazon SageMaker, GE Healthcare developed an ML model that can learn from thousands of medical scans to detect anomalies more accurately and efficiently, allowing radiologists to prioritize patients needing immediate attention

aws

Enhancing the Fan Experience

One week of NFL games now creates 3TB of data

NFL uses Amazon SageMaker to analyze telemetry data to predict plays

Computations that could take months to refine now take only weeks or days

aws

Accelerating Financial Analysis

Using TensorFlow on Amazon SageMaker, Siemens Financial Services developed an NLP model to extract critical information to accelerate investment due diligence, reducing time to summarize diligence documents from 12 hours down to 30 seconds

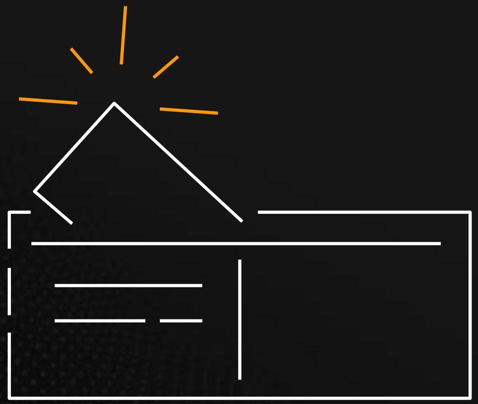
Meaningful Customer Interactions

Vonage uses Amazon Lex and Amazon Polly to develop natural language-driven conversational apps to respond to customer requests with real-time and contextual intelligence, improving response time and quality of service

aws

We’re focused on solving the toughest challenges that hold back success with machine learning

Three of the Biggest

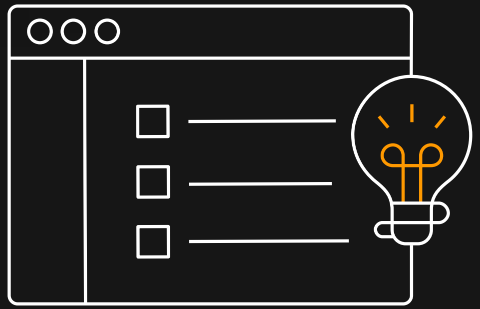


Cost

Making predictions at scale can still be very expensive

Data

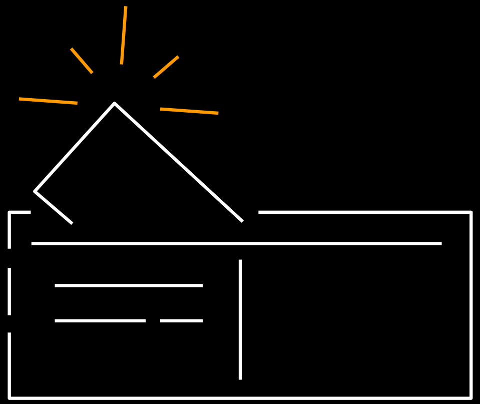
Preparing data for machine learning can be a major roadblock



Ease of use

There are many pitfalls and speedbumps that still exist

Three of the Biggest

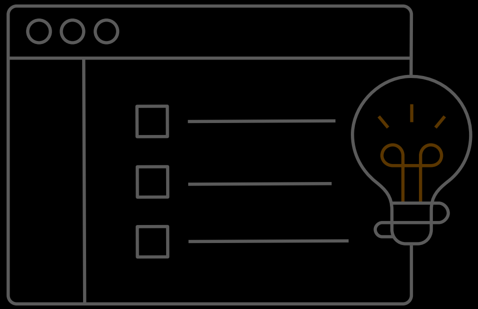


Cost

Making predictions at scale can still be very expensive

Data

Preparing data for machine learning can be a major roadblock



Ease of use

There are many pitfalls and speedbumps that still exist

NEW

Amazon EC2 P3dn Instance

The largest P3 instance,
optimized for distributed training

Amazon EC2 P3dn Instance

NEW

The largest P3 instance, optimized for distributed training

Reduce machine learning training time

Better GPU utilization

Support larger, more complex models

KEY FEATURES

100Gbps of networking bandwidth (4x more than P3)

8 NVIDIA Tesla V100 GPUs

32GB of memory per GPU (2x more P3)

96 Intel Skylake vCPUs (50% more than P3) with AVX-512

The Best Place to Run TensorFlow

Stock
TensorFlow

65%

scaling efficiency
with 256 GPUs

The best place to run TensorFlow

Stock TensorFlow

65%

scaling efficiency with 256 GPUs

NEW

AWS-Optimized TensorFlow

90%

scaling efficiency with 256 GPUs

Available with Amazon SageMaker and the AWS Deep Learning AMIs

The best place to run TensorFlow

NEW

Stock TensorFlow		AWS-Optimized TensorFlow
65% scaling efficiency with 256 GPUs	→	90% scaling efficiency with 256 GPUs
30m training time	→	14m training time

Fastest time for TensorFlow

Available with Amazon SageMaker and the AWS Deep Learning AMIs

NEW

Amazon Elastic Inference

Reduce deep learning inference costs up to 75%

Lower inference costs

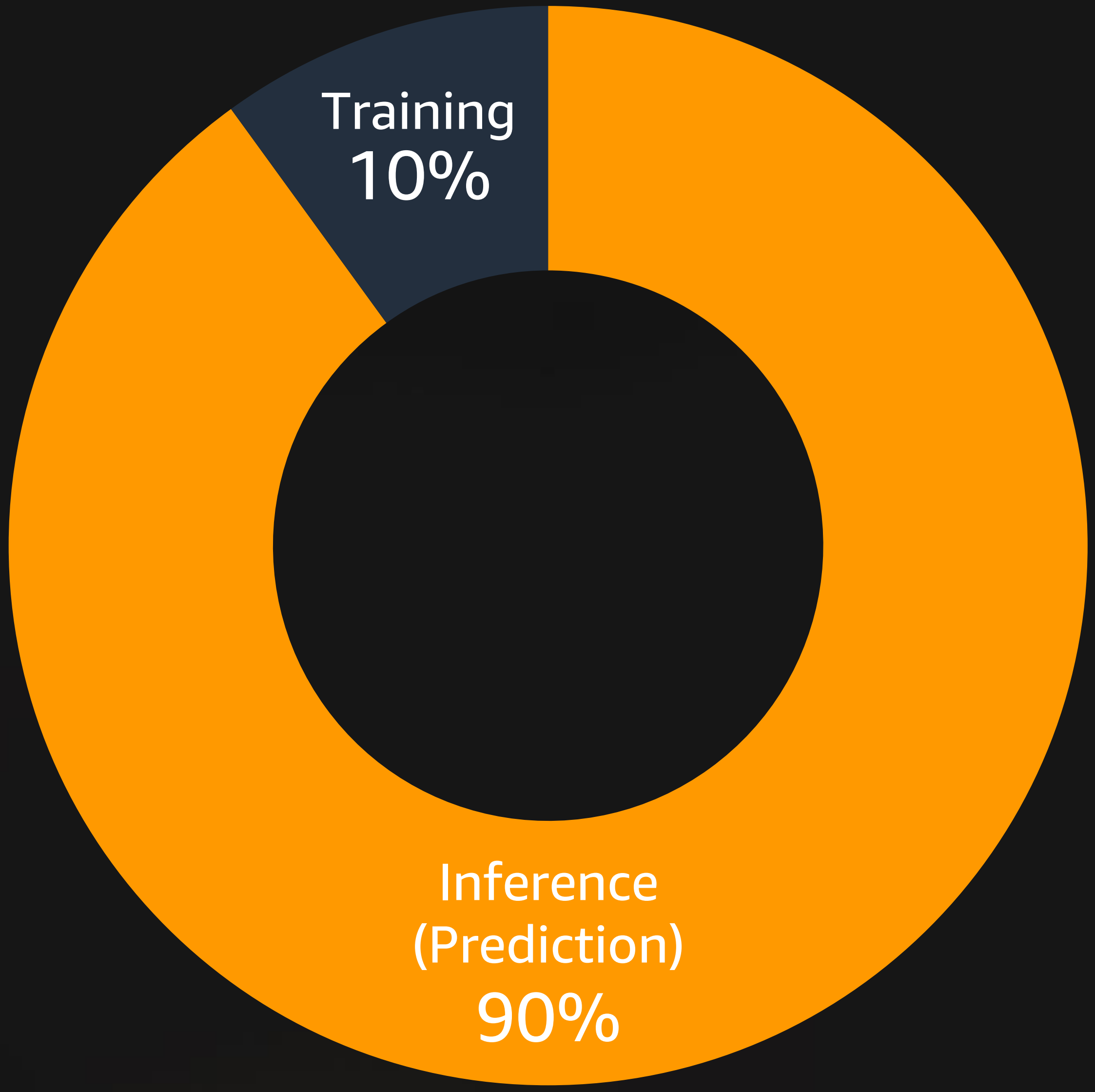
Match capacity to demand

Get exactly what you need, nothing more

AUTODESK. EAGLEVIEW TECHNOLOGIES gumgum

It’s never been easier, faster, or more cost effective to train machine learning models

But predictions drive the complexity and cost in production

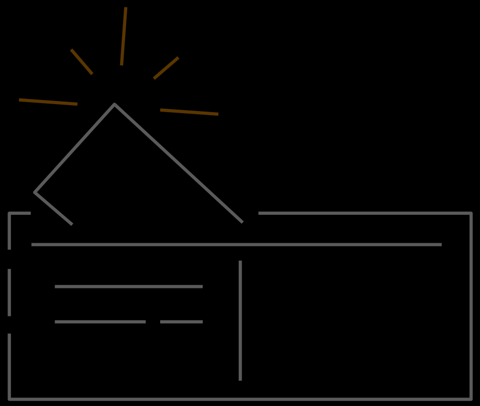


The Challenges of Prediction in Production

One size does not fit all

Elasticity is important

Three of the Biggest

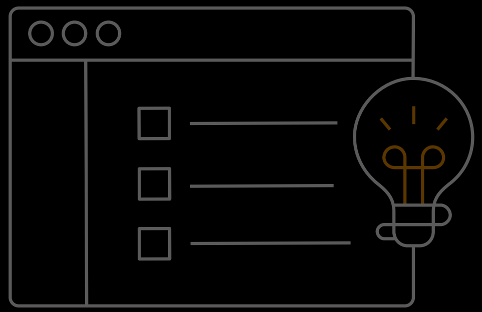


Cost

Making predictions at scale can still be very expensive

Data

Preparing data for machine learning can be a major roadblock



Ease of use

There are many pitfalls and speedbumps that still exist

Last Year, ML was Still Too Complicated

- Collect and prepare training data
- Choose and optimize your ML algorithm
- Set up and manage environments for training
- Train and tune model
- Deploy model in production
- Scale and manage the production environment

Amazon SageMaker: Build, Train, and Deploy ML

Pre-built notebooks for common problems

Collect and prepare training data

Choose and optimize your ML algorithm

Set up and manage environments for training

Train and tune model

Deploy model in production

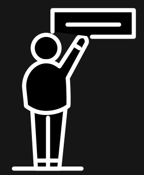
Scale and manage the production environment

Amazon SageMaker: Build, Train, and Deploy ML

Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment
Pre-built notebooks for common problems	Built-in, high performance algorithms				

Amazon SageMaker: Build, Train, and Deploy ML

Amazon EC2 P3dn Instances

Pre-built notebooks for common problems	Built-in, high performance algorithms	One-click training on the highest performing infrastructure			
Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment

Amazon SageMaker: Build, Train, and Deploy ML

Amazon EC2 P3dn Instances

Pre-built notebooks for common problems	Built-in, high performance algorithms	One-click training on the highest performing infrastructure	Model optimization		
Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment

Amazon SageMaker: Build, Train, and Deploy ML

Amazon EC2 P3dn Instances

Pre-built notebooks for common problems	Built-in, high performance algorithms	One-click training on the highest performing infrastructure	Model optimization	One-click deployment	
Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment

Amazon SageMaker: Build, Train, and Deploy ML

	Amazon EC2 P3dn Instances				Amazon Elastic Inference
Pre-built notebooks for common problems	Built-in, high performance algorithms	One-click training on the highest performing infrastructure	Model optimization	One-click deployment	Fully managed with auto-scaling for 75% less
Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment

It All Starts With Data

Successful Models Require High-Quality Data

Successful Models Require High-Quality Data

NEW

Amazon SageMaker Ground Truth

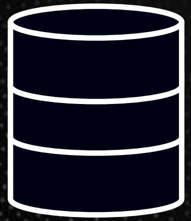
Label machine learning training data easily and accurately

Quickly label training data

Easily integrate human labelers

Get accurate results with low cost

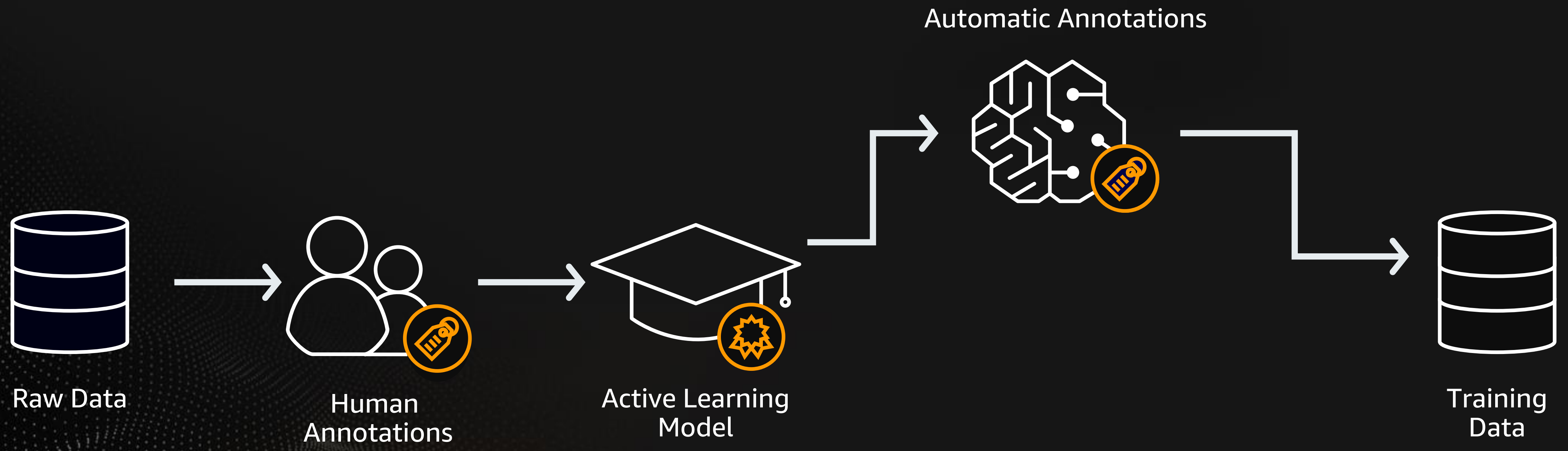
Using Machine Learning to Improve Labeling



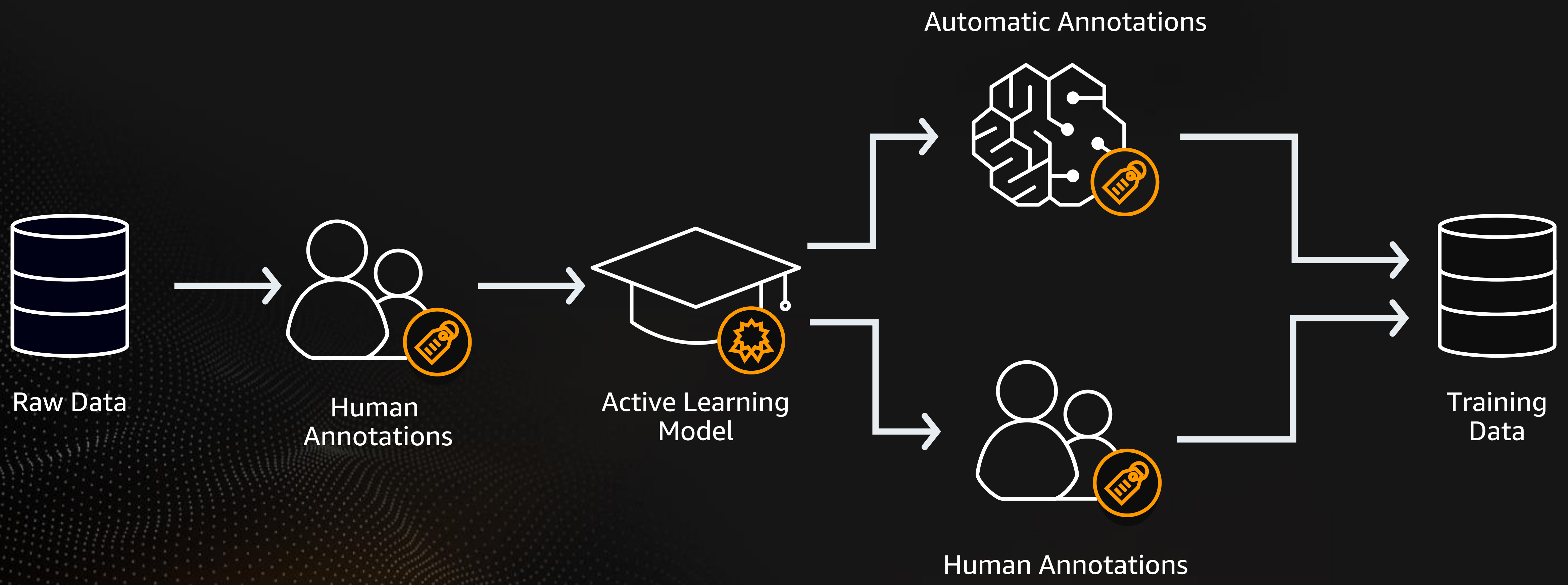
Raw Data

Using Machine Learning to Improve Labeling

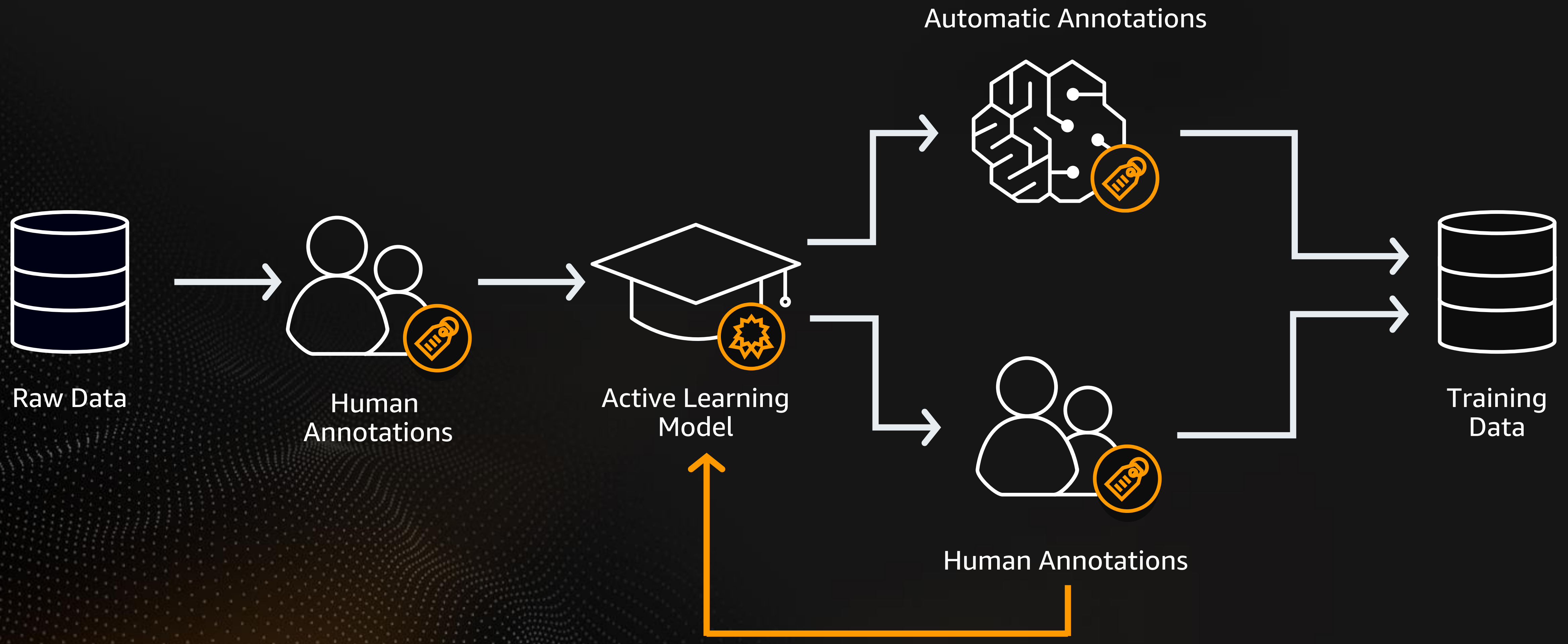
Using Machine Learning to Improve Labeling



Using Machine Learning to Improve Labeling



Using Machine Learning to Improve Labeling



Creating Training Data

Mechanical turk workers

Private labeling workforce

Third-party vendors

NEW

Amazon SageMaker Ground Truth

Label machine learning training data easily and accurately

Quickly label training data

Easily integrate human labelers

Get accurate results with low cost

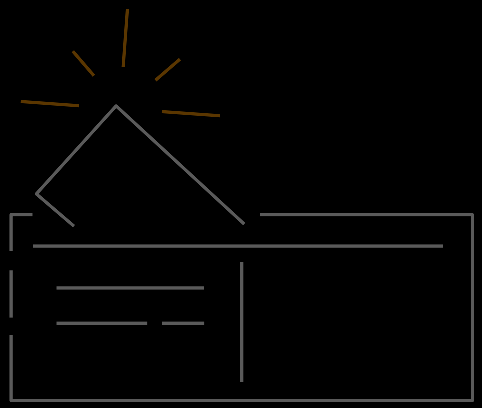
Tyson

CHANGE HEALTHCARE

Automagi

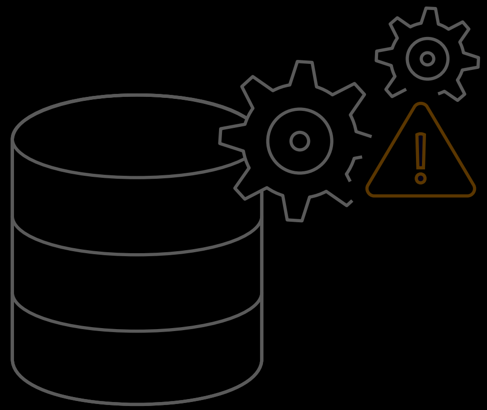
gumgum

Three of the Biggest



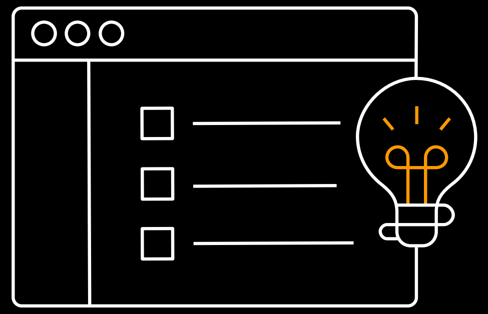
Cost

Making predictions at scale can still be very expensive



Data

Preparing data for machine learning can be a major roadblock



Ease of use

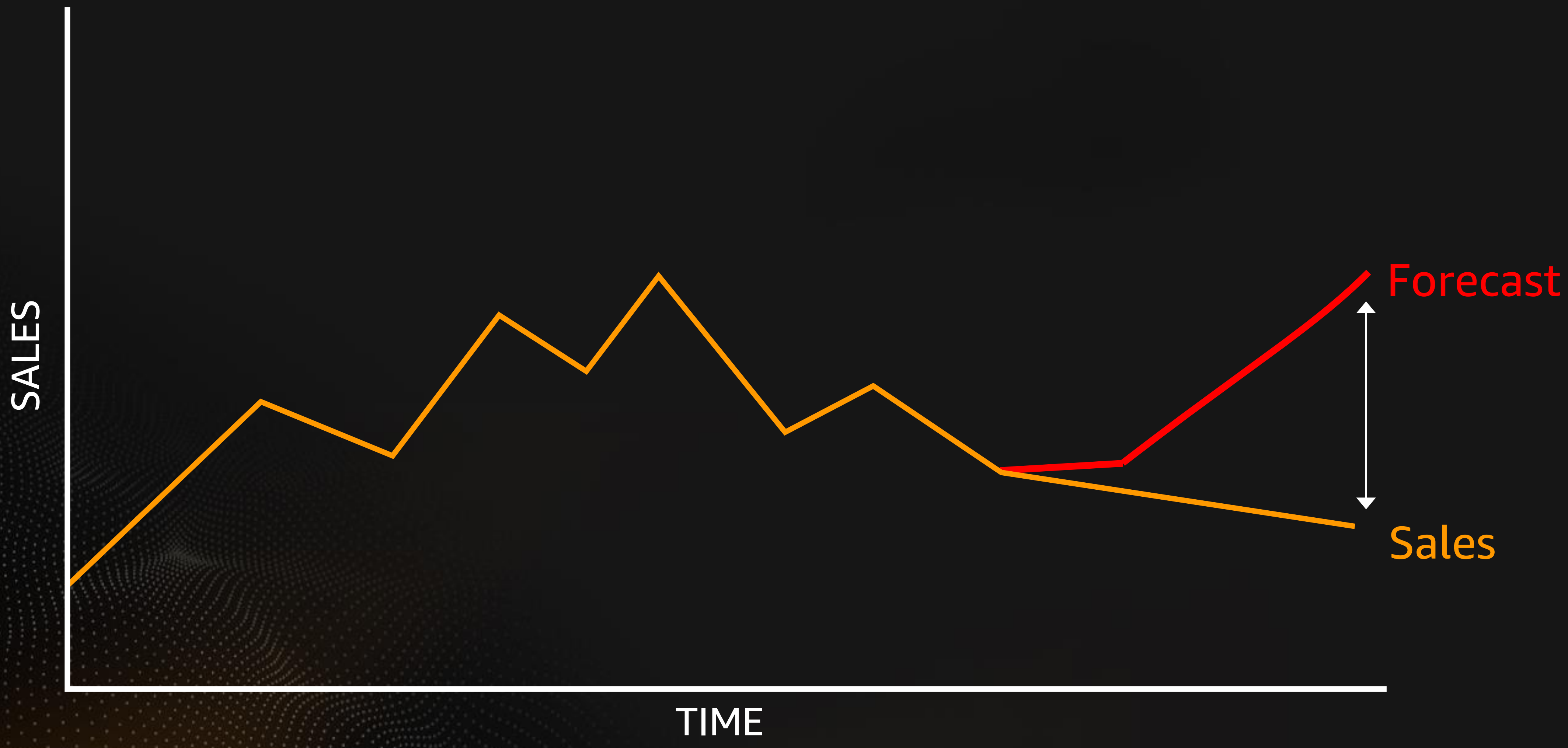
There are many pitfalls and speedbumps that still exist

Recap: Updates to Amazon SageMaker

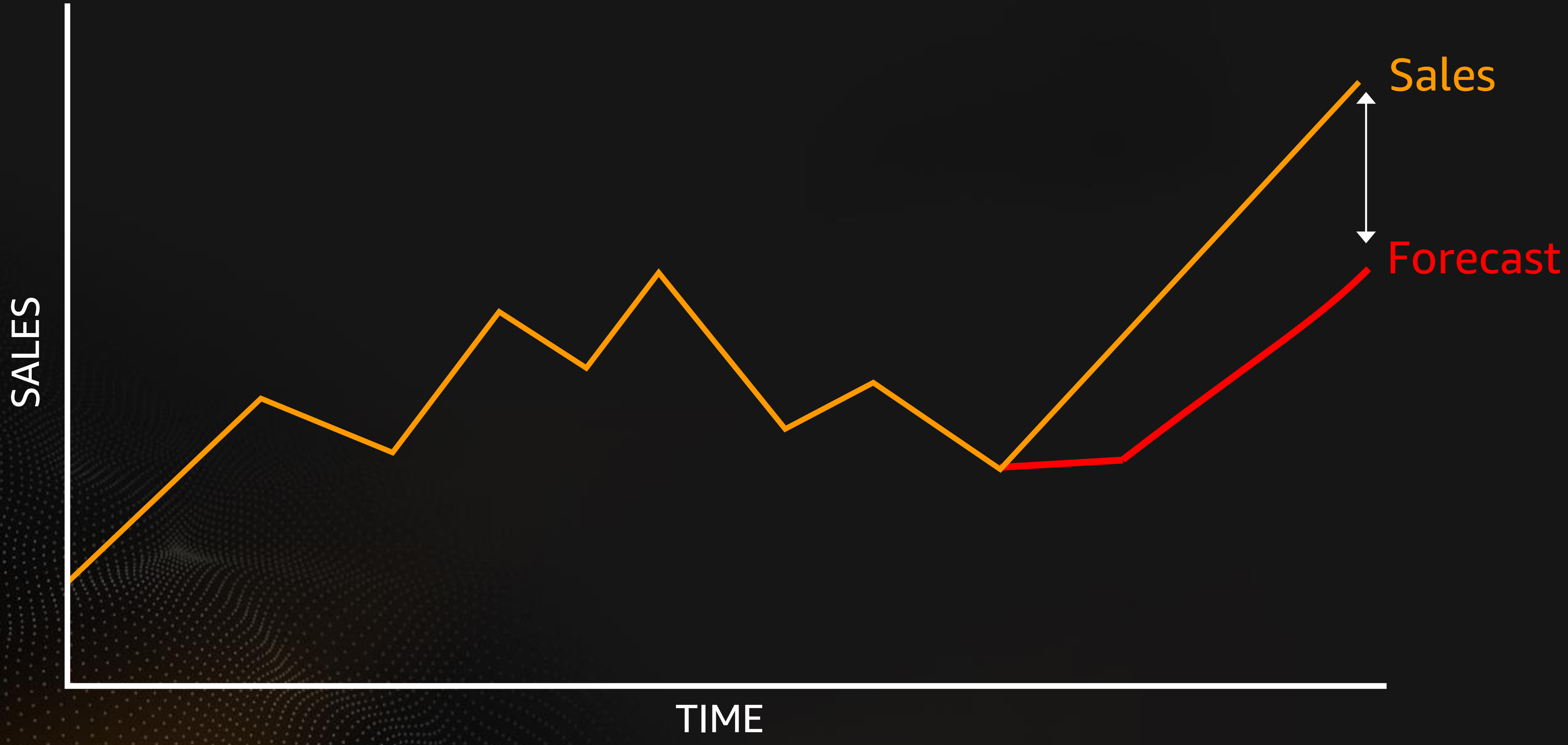
Amazon SageMaker Ground Truth	AWS Marketplace for Machine Learning	Amazon EC2 P3dn Instances	Amazon SageMaker Neo		Amazon Elastic Inference
Data labeling & Pre-built notebooks for common problems	Model and algorithm marketplace & Built-in, high performance algorithms	One-click training on the highest performing infrastructure	Train once, run anywhere & Model optimization	One-click deployment	Train once, run anywhere & Model optimization
Choose your ML algorithm	Optimize your ML algorithm	Set up and manage environments for training	Train and tune model	Deploy model in production	Scale and manage the production environment

Customers often ask…
"How can we tap into Amazon's experience in machine learning?"

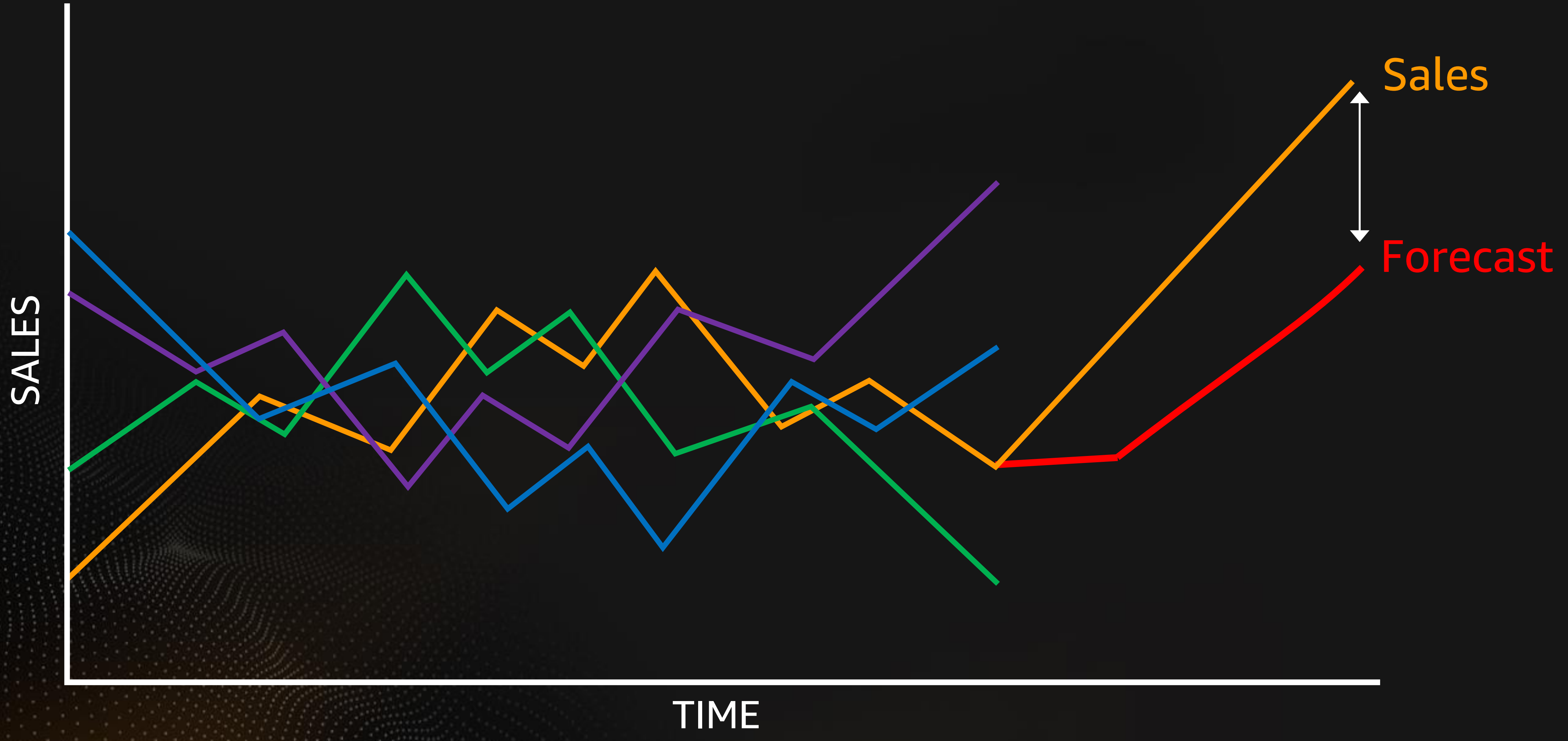
Challenges in Forecasting



Challenges in Forecasting



Challenges in Forecasting



NEW

Amazon Forecast

Improve forecasting accuracy by up to 50% at 1/10th the cost

Accurate forecasts

Get to results quickly

Works with any historical time-series

Recommendations at Amazon.com

Music	Film	Products	Content
Tracks	Actors	Pricing	Themes
Artists	Directors	Category	Demographics
Albums	Genres	Promotions	Breaking News

NEW

Amazon Personalize

Improve customer experiences with personalization and recommendations

Deliver high quality recommendations

Real-time

Get results quickly

Works with any product or content

Dealing with Documents is Demanding

How can we make it easy?

Traditional OCR Systems

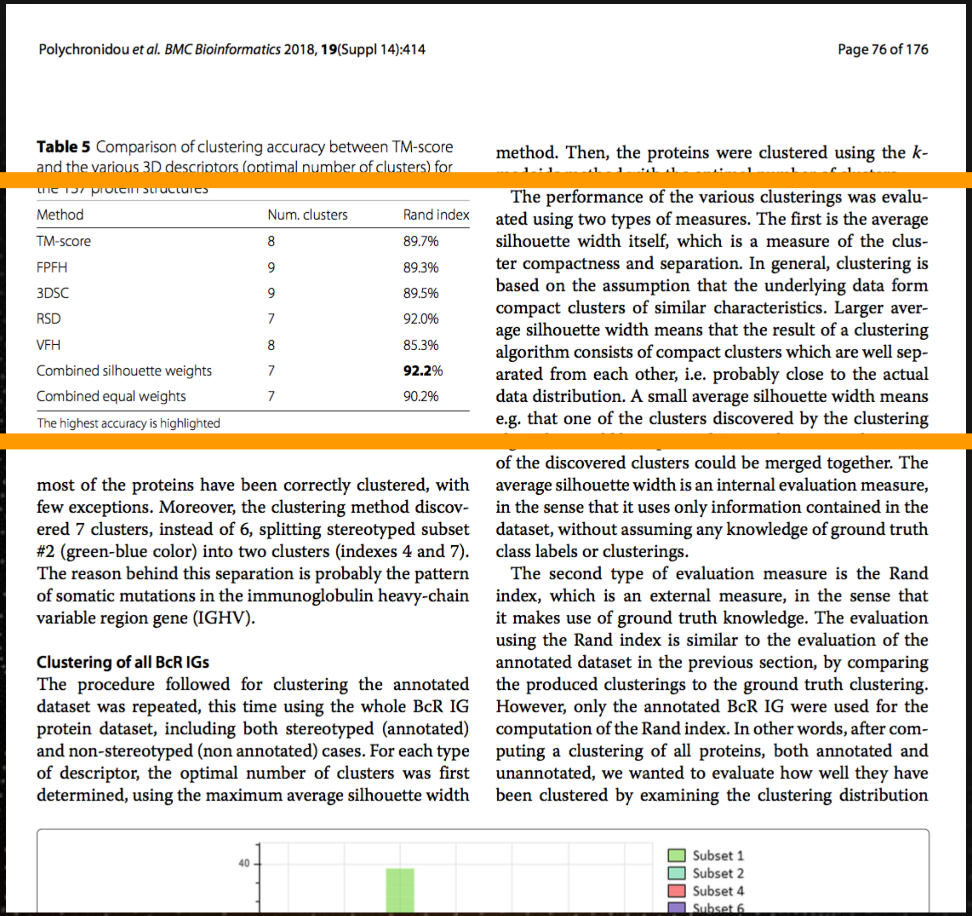


Table 5 Comparison of clustering accuracy between TM-score and the various 3D descriptors (optimal number of clusters) for the 137 protein structures

Method	Num. clusters	Rand index
TM-score	8	89.7%
FPFH	9	89.3%
3DSC	9	89.5%
RSD	7	92.0%
VFH	8	85.3%
Combined silhouette weights	7	**92.2%**
Combined equal weights	7	90.2%

The highest accuracy is highlighted

most of the proteins have been correctly clustered, with few exceptions. Moreover, the clustering method discovered 7 clusters, instead of 6, splitting stereotyped subset #2 (green-blue color) into two clusters (indexes 4 and 7). The reason behind this separation is probably the pattern of somatic mutations in the immunoglobulin heavy-chain variable region gene (IGHV).

Clustering of all BcR IGs

The procedure followed for clustering the annotated dataset was repeated, this time using the whole BcR IG protein dataset, including both stereotyped (annotated) and non-stereotyped (non annotated) cases. For each type of descriptor, the optimal number of clusters was first determined, using the maximum average silhouette width method. Then, the proteins were clustered using the *k*-medoids method with the optimal number of clusters.

The performance of the various clusterings was evaluated using two types of measures. The first is the average silhouette width itself, which is a measure of the cluster compactness and separation. In general, clustering is based on the assumption that the underlying data form compact clusters of similar characteristics. Larger average silhouette width means that the result of a clustering algorithm consists of compact clusters which are well separated from each other, i.e. probably close to the actual data distribution. A small average silhouette width means e.g. that one of the clusters discovered by the clustering

of the discovered clusters could be merged together. The average silhouette width is an internal evaluation measure, in the sense that it uses only information contained in the dataset, without assuming any knowledge of ground truth class labels or clusterings.

The second type of evaluation measure is the Rand index, which is an external measure, in the sense that it makes use of ground truth knowledge. The evaluation using the Rand index is similar to the evaluation of the annotated dataset in the previous section, by comparing the produced clusterings to the ground truth clustering. However, only the annotated BcR IG were used for the computation of the Rand index. In other words, after computing a clustering of all proteins, both annotated and unannotated, we wanted to evaluate how well they have been clustered by examining the clustering distribution

40
Subset 1
Subset 2
Subset 4
Subset 6

MethOd Num' C'USte'S Rand mdex ated using two types of measures. The first is the average

TM~score 8 89.7% silhouette width itself, which is a measure of the clus-

ppm 9 39,396 ter compactness and separation. In general, clustering is

305C 9 895% based on the assumption that the underlying data form

compact clusters of similar characteristics. Larger aver-

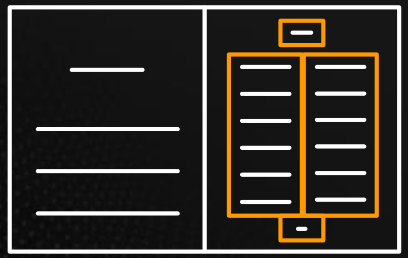
R50 7 92.096

age Silhouette Width means that the result of a clustering

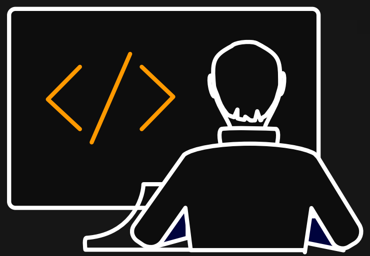
NEW

Amazon Textract

Extract text and data from virtually any document



Extract data quickly and accurately



Eliminate manual effort

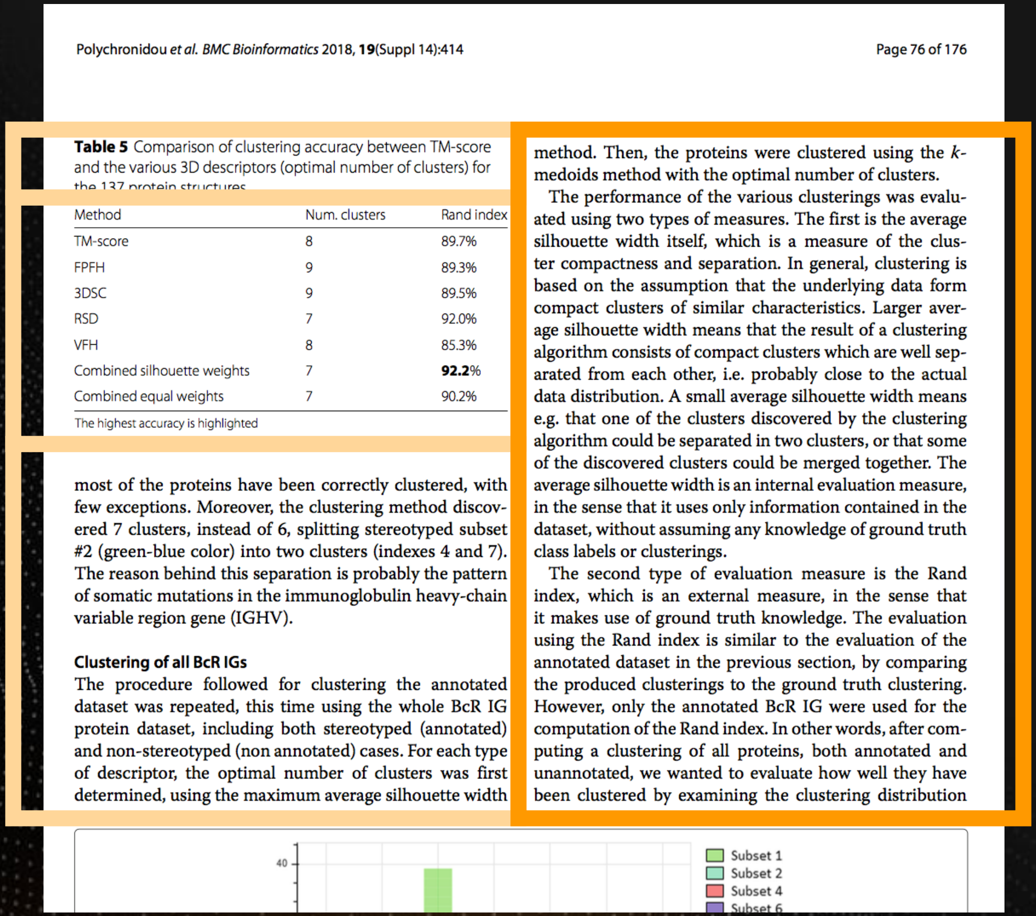
Lower document processing costs

Alfresco®

Cox Automotive

Roche

Amazon Textract OCR



Polychronidou *et al. BMC Bioinformatics* 2018, **19**(Suppl 14):414 Page 76 of 176

Table 5 Comparison of clustering accuracy between TM-score and the various 3D descriptors (optimal number of clusters) for the 137 protein structures

Method	Num. clusters	Rand index
TM-score	8	89.7%
FPFH	9	89.3%
3DSC	9	89.5%
RSD	7	92.0%
VFH	8	85.3%
Combined silhouette weights	7	**92.2%**
Combined equal weights	7	90.2%

The highest accuracy is highlighted

most of the proteins have been correctly clustered, with few exceptions. Moreover, the clustering method discovered 7 clusters, instead of 6, splitting stereotyped subset #2 (green-blue color) into two clusters (indexes 4 and 7). The reason behind this separation is probably the pattern of somatic mutations in the immunoglobulin heavy-chain variable region gene (IGHV).

Clustering of all BcR IGs

The procedure followed for clustering the annotated dataset was repeated, this time using the whole BcR IG protein dataset, including both stereotyped (annotated) and non-stereotyped (non annotated) cases. For each type of descriptor, the optimal number of clusters was first determined, using the maximum average silhouette width method. Then, the proteins were clustered using the *k*-medoids method with the optimal number of clusters.

The performance of the various clusterings was evaluated using two types of measures. The first is the average silhouette width itself, which is a measure of the cluster compactness and separation. In general, clustering is based on the assumption that the underlying data form compact clusters of similar characteristics. Larger average silhouette width means that the result of a clustering algorithm consists of compact clusters which are well separated from each other, i.e. probably close to the actual data distribution. A small average silhouette width means e.g. that one of the clusters discovered by the clustering algorithm could be separated in two clusters, or that some of the discovered clusters could be merged together. The average silhouette width is an internal evaluation measure, in the sense that it uses only information contained in the dataset, without assuming any knowledge of ground truth class labels or clusterings.

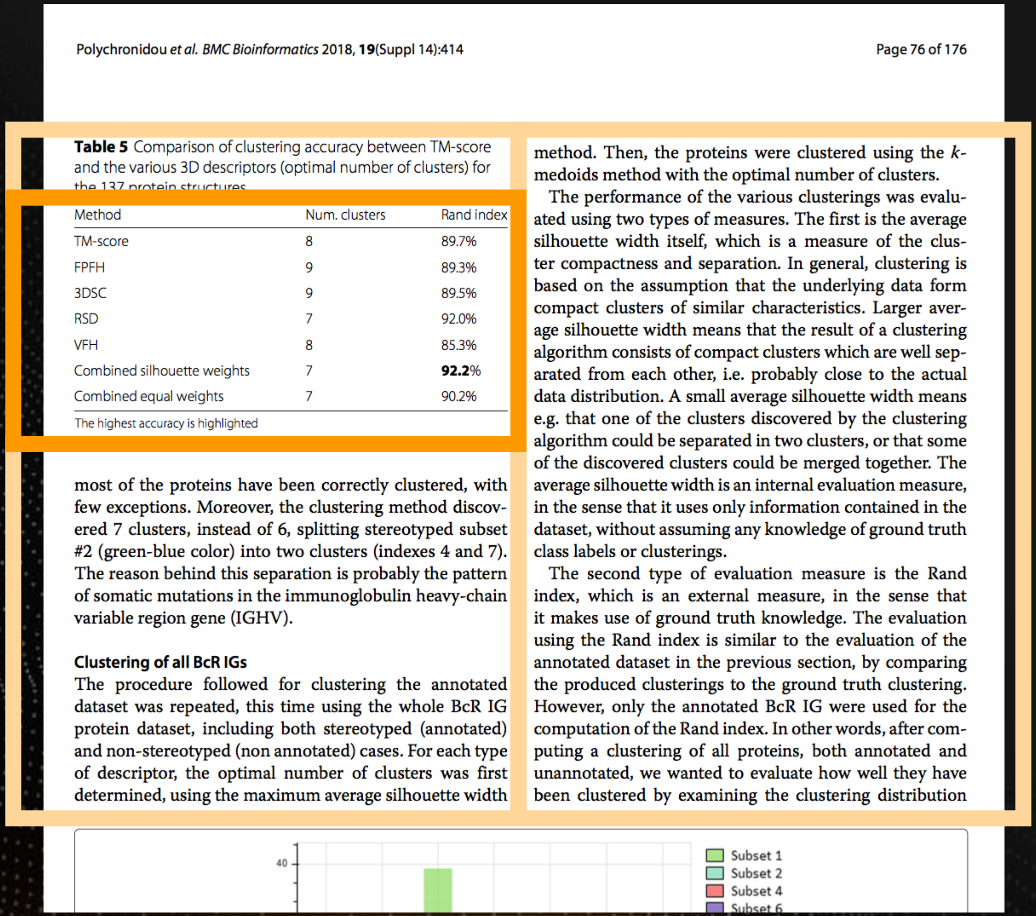
The second type of evaluation measure is the Rand index, which is an external measure, in the sense that it makes use of ground truth knowledge. The evaluation using the Rand index is similar to the evaluation of the annotated dataset in the previous section, by comparing the produced clusterings to the ground truth clustering. However, only the annotated BcR IG were used for the computation of the Rand index. In other words, after computing a clustering of all proteins, both annotated and unannotated, we wanted to evaluate how well they have been clustered by examining the clustering distribution

TEXT

method. Then, the proteins were clustered using the k- medoids method with the optimal number of clusters.

The performance of the various clusterings was evalu- ated using two types of measures. The first is the average silhouette width itself, which is a measure of the clus- ter compactness and separation. In general, clustering is based on the assumption that the underlying data form compact clusters of similar characteristics. Larger aver- age silhouette width means that the result of a clustering algorithm consists of compact clusters which are well sep- arated from each other, i.e. probably close to the actual data distribution. A small average silhouette width means e.g. that one of the clusters …

Amazon Textract Table Detection



Polychronidou *et al. BMC Bioinformatics* 2018, **19**(Suppl 14):414 Page 76 of 176

Table 5 Comparison of clustering accuracy between TM-score and the various 3D descriptors (optimal number of clusters) for the 137 protein structures

Method	Num. clusters	Rand index
TM-score	8	89.7%
FPFH	9	89.3%
3DSC	9	89.5%
RSD	7	92.0%
VFH	8	85.3%
Combined silhouette weights	7	**92.2%**
Combined equal weights	7	90.2%

The highest accuracy is highlighted

most of the proteins have been correctly clustered, with few exceptions. Moreover, the clustering method discovered 7 clusters, instead of 6, splitting stereotyped subset #2 (green-blue color) into two clusters (indexes 4 and 7). The reason behind this separation is probably the pattern of somatic mutations in the immunoglobulin heavy-chain variable region gene (IGHV).

Clustering of all BcR IGs

The procedure followed for clustering the annotated dataset was repeated, this time using the whole BcR IG protein dataset, including both stereotyped (annotated) and non-stereotyped (non annotated) cases. For each type of descriptor, the optimal number of clusters was first determined, using the maximum average silhouette width method. Then, the proteins were clustered using the *k*-medoids method with the optimal number of clusters.

The performance of the various clusterings was evaluated using two types of measures. The first is the average silhouette width itself, which is a measure of the cluster compactness and separation. In general, clustering is based on the assumption that the underlying data form compact clusters of similar characteristics. Larger average silhouette width means that the result of a clustering algorithm consists of compact clusters which are well separated from each other, i.e. probably close to the actual data distribution. A small average silhouette width means e.g. that one of the clusters discovered by the clustering algorithm could be separated in two clusters, or that some of the discovered clusters could be merged together. The average silhouette width is an internal evaluation measure, in the sense that it uses only information contained in the dataset, without assuming any knowledge of ground truth class labels or clusterings.

The second type of evaluation measure is the Rand index, which is an external measure, in the sense that it makes use of ground truth knowledge. The evaluation using the Rand index is similar to the evaluation of the annotated dataset in the previous section, by comparing the produced clusterings to the ground truth clustering. However, only the annotated BcR IG were used for the computation of the Rand index. In other words, after computing a clustering of all proteins, both annotated and unannotated, we wanted to evaluate how well they have been clustered by examining the clustering distribution

TABLE DATA

Method	Num. clusters	Rand index
TM-score	8	89.7%
FPFH	9	89.3%
3DSC	9	89.5%
RSD	7	92.0%
VFH	8	85.3%
Combined silhouette weights	7	92.2%
Combined equal weights	7	90.2%

The Challenge with Forms

22222

a Employee's social security number

OMB No. 1545-0008

b Employer identification number (EIN)

1 Wages, tips, other compensation

2 Federal income tax withheld

c Employer's name, address, and ZIP code

3 Social security wages

4 Social security tax withheld

5 Medicare wages and tips

6 Medicare tax withheld

7 Social security tips

8 Allocated tips

d Control number

9 Verification code

10 Dependent care benefits

e Employee's first name and initial Last name Suff.

11 Nonqualified plans

12a Code

13 Statutory employee Retirement plan Third-party sick pay

12b Code

14 Other

12c Code

12d Code

f Employee's address and ZIP code

15 State	Employer's state ID number	16 State wages, tips, etc.	17 State income tax	18 Local wages, tips, etc.	19 Local income tax	20 Locality name

Form **W-2** **Wage and Tax Statement** 2017

Department of the Treasury—Internal Revenue Service

Copy 1—For State, City, or Local Tax Department

aws

The Challenge with Forms

W2 e file 2017

Employee Reference Copy

d. control number	Dept	Corp.	Employer use only
438209	8840		A 439

c. Employer's name, address, and ZIP code
PARIS BAKERY
123 MAIN STREET
HANOVER, NH 03755

e/f Employee's name, address, and ZIP code
LILLIAN CRANE
1893 ORCHARD RD
WHITE RIVER JUNCTION, NH 03789

b Employer's FED ID number 32-3939201	a Employee's SSA number 111-22-3333
1 Wages, tips, other comp. $39,010.32	2 Federal income tax withheld $5,451.12
3 Social security wages $39,010.32	4 Social security tax withheld $2439.08
5 Medicare wages and tips $39,010.32	6 Medicare tax withheld $456.90
7 Social security tips	8 Allocated tips
9	10 Dependent care benefits
11 Nonqualified plans	12a See instructions for box 12 C $209
14 Other	12 b D $395.16
	12 c
	12 d
	13 Stat emp. / Ret. plan X / 3rd party sick pay
15 State NH / Employer's State ID no.	16 State, wages, tips, etc.
17 State Income Tax	18 Local wages, tips, etc.
19 Local Income Tax	20 Locality name

a Employee's SSA number 111-22-3333	1 Wages, tips, other comp. $39,010.32	2 Federal income tax withheld $5,451.12
	3 Social security wages $39,010.32	4 Social security tax withheld $2439.08
b Employer's FED ID number 32-3939201	5 Medicare wages and tips $39,010.32	6 Medicare tax withheld $456.90

c. Employer's name, address, and ZIP code
PARIS BAKERY
123 MAIN STREET
HANOVER, NH 03755

e. Employee's first name and initial	Last name	Suff.
LILLIAN	CRANE	

1893 ORCHARD RD
WHITE RIVER JUNCTION, NH 03789
f. Employee's address and ZIP code

d. control number 438209	7 Social security tips	8 Allocated tips
9	10 Dependent care benefits	11 Nonqualified plans
12a See instructions for box 12 C $209	14 Other	
12 b D $395.16		
12 c		
12 d		

13 Statutory employee [x] Retirement plan [ ] Third-party sick pay [ ]

15 State NH	Employer's State ID no.	16 State, wages, tips, etc.	17 State Income Tax
18 Local wages, tips, etc.		19 Local Income Tax	20 Locality name

W2 Wage and Tax Statement 2017 Copy B Employee Reference Copy

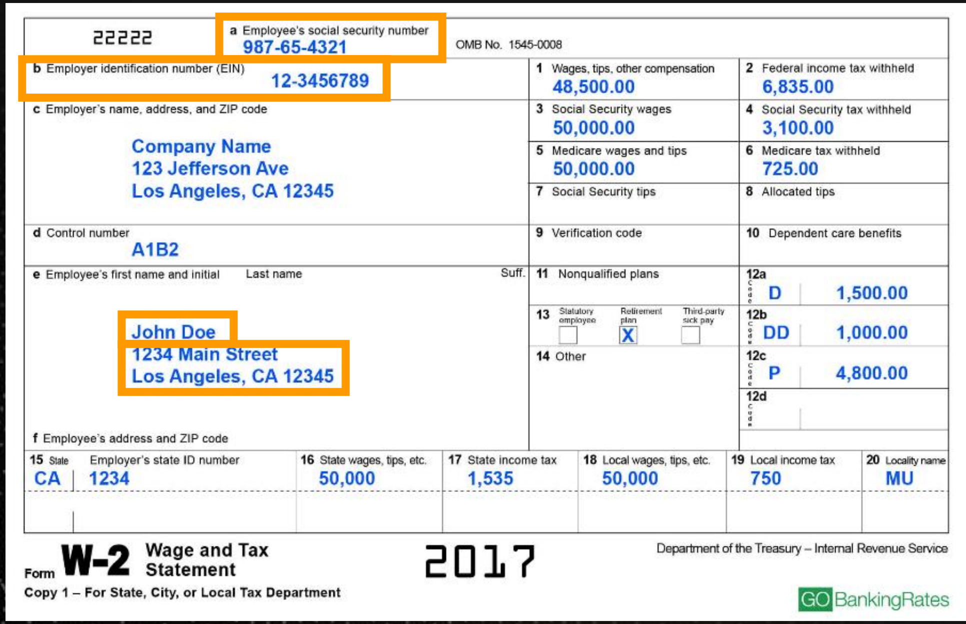
Form W2 Wage and Tax Statement 2017

Copy B Employee Reference Copy

d. control number 438209	Void	c. Employer's name, address, and ZIP code PARIS BAKERY 123 MAIN STREET HANOVER, NH 03755	1 Wages, tips, other comp. $39,010.32	2 Federal income tax withheld $5,451.12
b Employer's FED ID number 32-3939201	a Employee's SSA number 111-22-3333		3 Social security wages $39,010.32	4 Social security tax withheld $2439.08
13 Stat employee / Retirement plan X / 3rd party sick pay			5 Medicare wages and tips $39,010.32	6 Medicare tax withheld $456.90
12 See instrs. For Box 12 C $209 D $395.16	14 Other	e Employee's name, address, and ZIP code LILLIAN CRANE 1893 ORCHARD RD WHITE RIVER JUNCTION, NH 03789	7 Social security tips	8 Allocated tips
			10 Dependent care benefits	11 Nonqualified plans

15 State	Employer's State ID no.	16 State, wages, tips, etc.	17 State Income Tax	18 Local wages, tips, etc.	19 Local Income Tax	20 Locality name
NH						

Amazon Textract Forms


22222
a Employee's social security number
987-65-4321
OMB No. 1545-0008
b Employer identification number (EIN)
12-3456789
1 Wages, tips, other compensation
48,500.00
2 Federal income tax withheld
6,835.00
c Employer's name, address, and ZIP code
Company Name
123 Jefferson Ave
Los Angeles, CA 12345
3 Social Security wages
50,000.00
4 Social Security tax withheld
3,100.00
5 Medicare wages and tips
50,000.00
6 Medicare tax withheld
725.00
7 Social Security tips
8 Allocated tips
d Control number
A1B2
9 Verification code
10 Dependent care benefits
e Employee's first name and initial Last name Suff.
John Doe
1234 Main Street
Los Angeles, CA 12345
f Employee's address and ZIP code
11 Nonqualified plans
12a Code D 1,500.00
13 Statutory employee Retirement plan X Third-party sick pay
12b Code DD 1,000.00
14 Other
12c Code P 4,800.00
12d Code

15 State	Employer's state ID number	16 State wages, tips, etc.	17 State income tax	18 Local wages, tips, etc.	19 Local income tax	20 Locality name
CA	1234	50,000	1,535	50,000	750	MU

Form W-2 Wage and Tax Statement 2017
Department of the Treasury – Internal Revenue Service
Copy 1 – For State, City, or Local Tax Department
GOBankingRates

SSN 987-65-4321

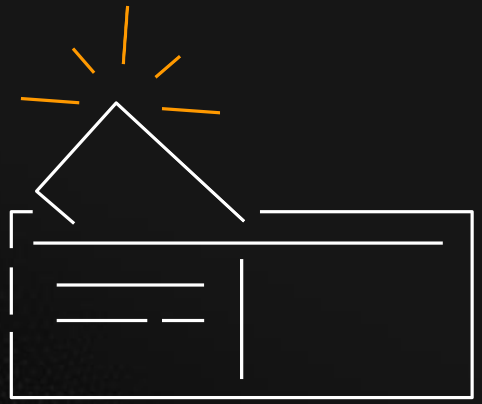
EIN 12-3456789

NAME John Doe

ADDRESS 1234 Main Street
Seattle, WA 98101

and so on…

Three of the Biggest



Cost

Making predictions at scale
can still be very expensive

Amazon Elastic Inference

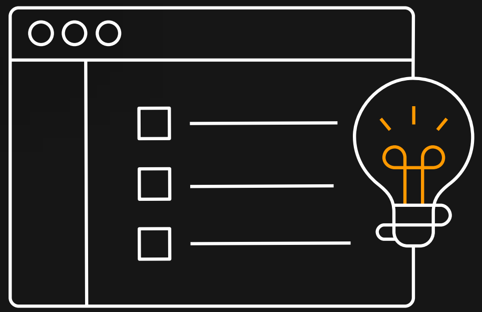
AWS Inferentia

Data

Preparing data for machine
learning can be a major roadblock

Amazon SageMaker Ground Truth

Amazon SageMaker RL



Ease of use

There are many pitfalls and
speedbumps that still exist

AWS Marketplace for Machine Learning
Amazon SageMaker Neo
Amazon Textract
Amazon Forecast
Amazon Personalize

The Amazon ML Stack

AI Services

VISION		SPEECH			LANGUAGE	CHATBOTS		FORECASTING	RECOMMENDATIONS
REKOGNITION IMAGE	REKOGNITION VIDEO	TEXTRACT	POLLY	TRANSCRIBE	TRANSLATE	COMPREHEND	LEX	FORECAST	PERSONALIZE

ML Services

AMAZON SAGEMAKER

GROUND TRUTH

NOTEBOOKS

ALGORITHMS

MARKETPLACE

SUPERVISED LEARNING

UNSUPERVISED LEARNING

REINFORCEMENT LEARNING

TRAINING

OPTIMIZATION (NEO)

DEPLOYMENT

HOSTING

ML Frameworks + Infrastructure

FRAMEWORKS

TensorFlow

mxnet

PYTORCH

Chainer

HOROVOD

INTERFACES

GLUON

Keras

INFRASTRUCTURE

EC2 P3 & P3DN

EC2 C5

FPGAs

GREENGRASS

ELASTIC INFERENCE

INFERENTIA

AWS INNOVATE

ONLINE CONFERENCE 2019

MACHINE LEARNING & AI EDITION

Olivier Klein
Head of Emerging Technologies, AWS

The Amazon ML Stack

AI Services

- **VISION**: AMAZON REKOGNITION; AMAZON TEXTRACT (NEW)
- **SPEECH**: AMAZON POLLY; AMAZON TRANSCRIBE
- **LANGUAGES**: AMAZON COMPREHEND; AMAZON TRANSLATE
- **CHATBOTS**: AMAZON LEX
- **VERTICAL**: AMAZON PERSONALIZE (NEW); AMAZON FORECAST (NEW)

ML Services

- AMAZON SAGEMAKER
- AMAZON SAGEMAKER RL (NEW)
- AMAZON SAGEMAKER GROUND TRUTH (NEW)
- AWS DEEPRACER (NEW)
- AWS DEEPLENS
- AWS MARKETPLACE FOR ML (NEW)
- AMAZON SAGEMAKER NEO (NEW)

ML Frameworks + Infrastructure

- **FRAMEWORKS**: mxnet, PYTORCH, TensorFlow
- **INTERFACES**: KERAS, GLUON
- **INFRASTRUCTURE**: P3, P3dn (NEW), C5, C5n (NEW); AMAZON ELASTIC INFERENCE (NEW); AWS INFERENTIA (NEW); AWS GREENGRASS

aws

AWS is Framework Agnostic

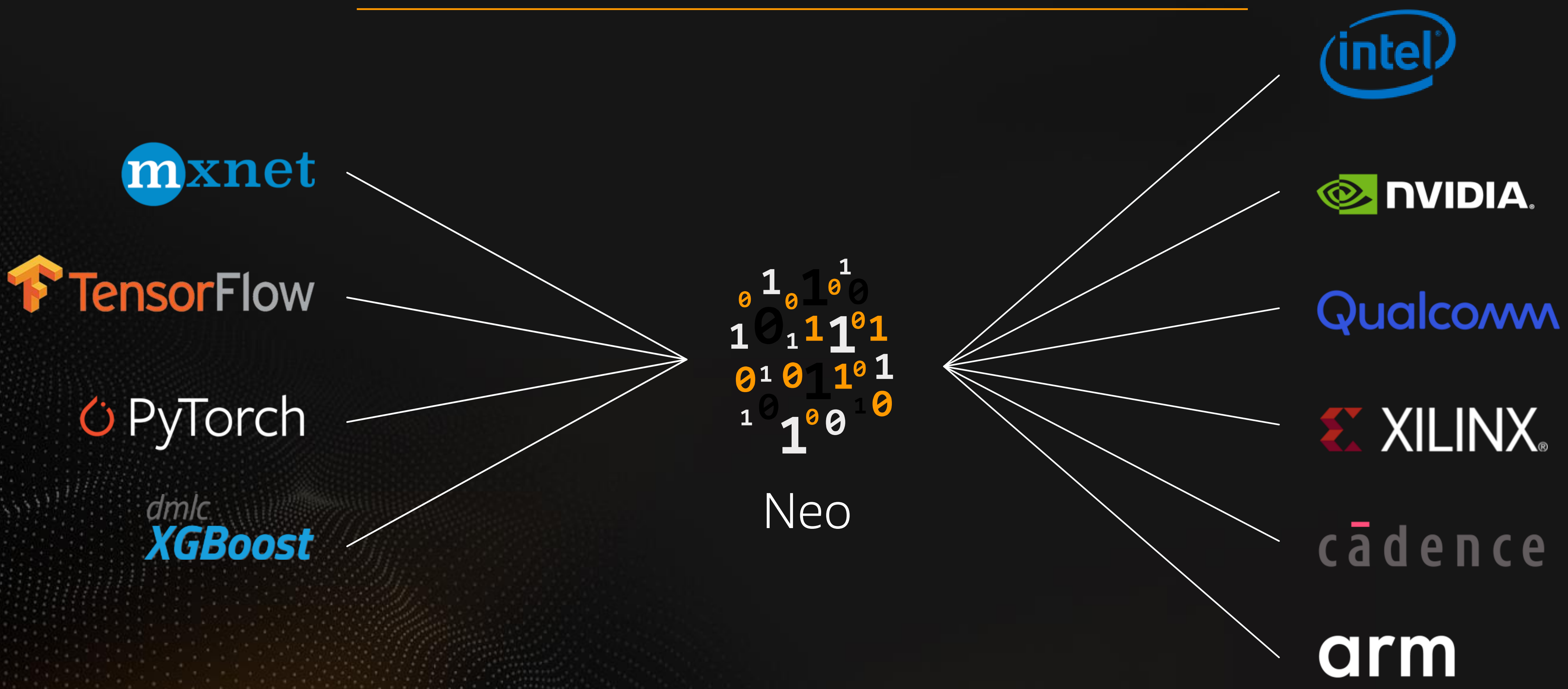
Choose from popular frameworks

TensorFlow mxnet PYTORCH Chainer Caffe2 ONNX Keras GLUON

Run them fully managed

Or run them yourself

Amazon Sagemaker Neo



aws

Active.Ai

Conversational services for FSI

capillary

AI-powered Retail Analytics

qure.ai

Deep Learning for Medical Imaging

TESSACT
MAKING MEDIA SMARTER

Intelligent Video Recognition

UNBXD

AI driven eCommerce Search Platform

haptik™

Conversational AI Platform

aws

KIA

KIA MOTORS

SAMSUNG

SAMSUNG HEAVY INDUSTRIES

aws

Access to AI and ML capabilities without any prior ML knowledge

Broadest and deepest set of machine learning and AI services for your business

Choice and flexibility with ML frameworks

Plenty of learning tools and resources

Learn Why Customers Choose AWS for Machine Learning

Demo Arena

Watch how machine learning is used in real life applications

Ask the Experts

Get your questions answered by AWS Experts

Machine Learning on AWS

https://aws.amazon.com/machine-learning/

Thank You for Attending AWS Innovate

We hope you found it interesting! A kind reminder to complete the survey. Let us know what you thought of today's event and how we can improve the event experience for you in the future.

- aws-apac-marketing@amazon.com
- twitter.com/AWSCloud
- facebook.com/AmazonWebServices
- youtube.com/user/AmazonWebServices
- slideshare.net/AmazonWebServices
- twitch.tv/aws