

AI: State of the Union

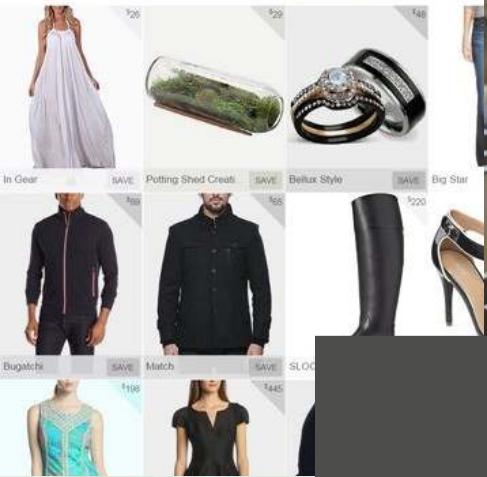
Julien Simon – Principal Evangelist, AI / ML



@julsimon

Beautiful things, up

ALL WOMEN MEN



Amazon AI



Agenda

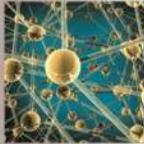
- An Introduction to AWS
- A Quick Look at Big Data on AWS
- An Overview of Amazon AI services
- End-to-end Machine Learning with Amazon SageMaker
- AWS DeepLens

All questions welcome at anytime 😊





amazon
fulfillment



Two Decades of Recommender Systems at Amazon.com

Amazon is well-known for personalization and recommendations, which help customers discover items they might otherwise not have found. In this update to our original article, we discuss some of the changes as Amazon has grown.

Brent Smith
Amazon.com

Greg Linden
Microsoft

For two decades now,¹ Amazon.com has been building a store for every customer. Each person who comes to Amazon.com sees it differently, because it's individually personalized based on their interests. It's as if you walked into a store and the shelves started rearranging themselves, with what you might want moving to the front, and what you're unlikely to be interested in shuffling further away.

From a catalog of hundreds of millions of items, Amazon.com's recommendations pick a small number of items you might enjoy based on your current context and your past behavior. The algorithms aren't magic; they simply share with you what other people have already discovered. The algorithm does all the work. It's computers helping people help other people, implicitly and anonymously.

Amazon.com launched item-based collaborative filtering in 1998, enabling recommendations at a previously unseen scale for millions of customers and a catalog of millions of items. Since we wrote about the algorithm in *IEEE Internet Computing* in 2003,² it has seen widespread use across the Web, including YouTube, Netflix, and many others. The algorithm's success has been from its simplicity, scalability, and often surprising and useful

recommendations, as well as desirable properties such as updating immediately based on new information about a customer and being able to explain why it recommended something in a way that's easily understandable.

What was described in our 2003 *IEEE Internet Computing* article has faced many challenges and seen much development over the years. Here, we describe some of the updates, improvements, and adaptations for item-based collaborative filtering, and offer our view on what the future holds for collaborative filtering, recommender systems, and personalization.

The Algorithm

As we described it in 2003, the item-based collaborative filtering algorithm is straightforward. In the mid-1990s, collaborative filtering was generally user-based, meaning the first step of the algorithm was to search across other users to find people with similar interests (such as similar purchase patterns), then look at what items those similar users found that you haven't found yet. Instead, our algorithm begins by finding related items for each item in the catalog. The term "related" could have several meanings here, but at this point,





amazon



JUST
WALK
OUT
SHOPPING

amazon

ML @ AWS: Our mission

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

AWS ML Stack

Application Services

API-driven services: Vision & Language Services, Conversational Chatbots

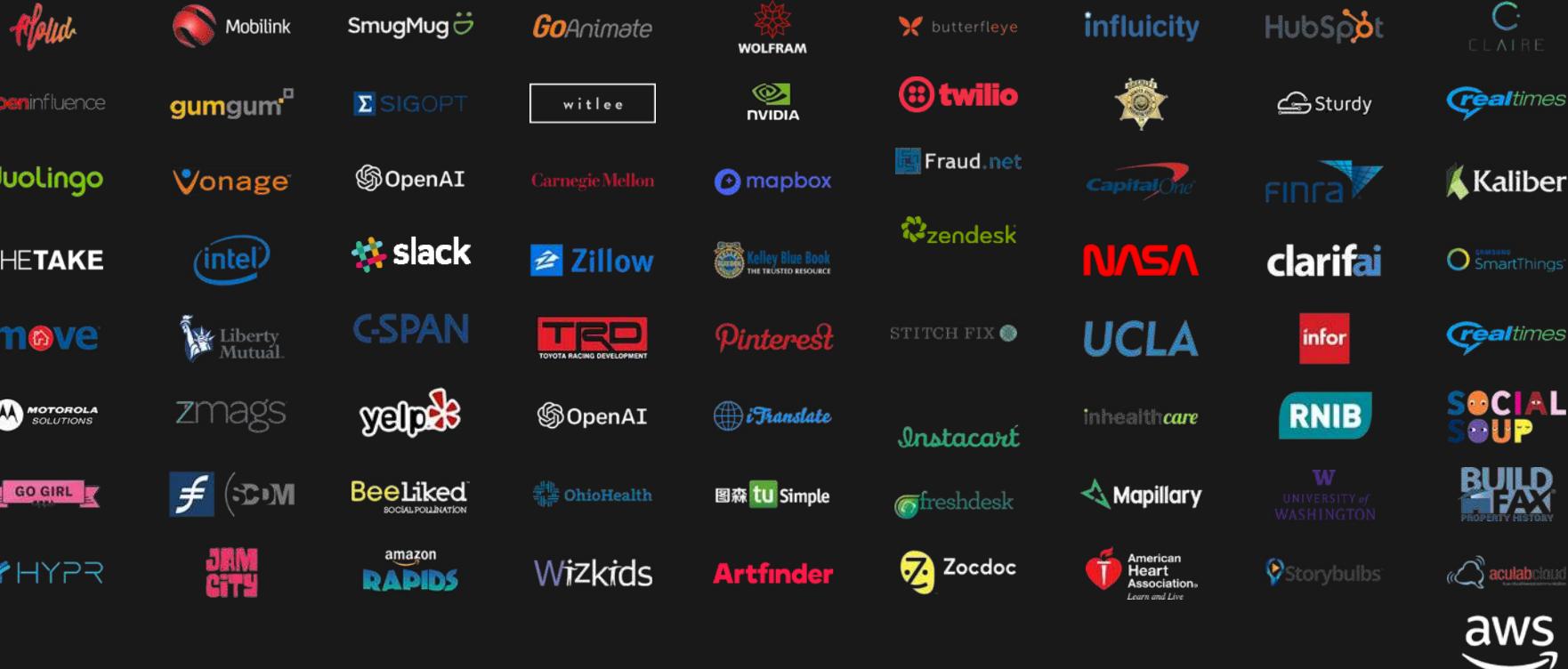
Platform Services

Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.

Frameworks & Infrastructure

Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run inference on trained models.

Customers Running ML on AWS Today



AWS ML Stack

Application Services

API-driven services: Vision & Language Services, Conversational Chatbots

Platform Services

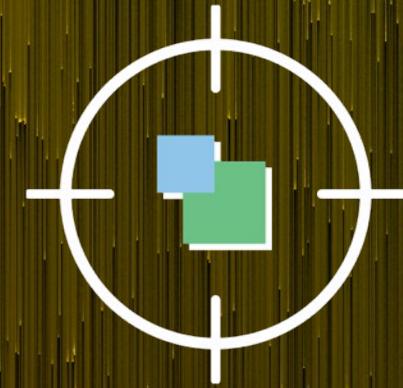
Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.

Frameworks & Infrastructure

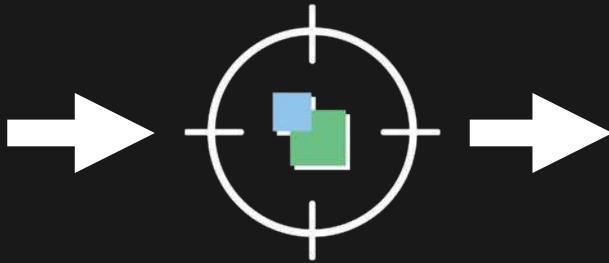
Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run inference on trained models.

Amazon Rekognition

Deep learning-based visual analysis service



Deep learning-based visual analysis service



Object and scene detection
Facial analysis
Face comparison
Celebrity recognition
Image moderation

Object & Scene Detection



City	91.4%
Downtown	91.4%
Metropolis	91.4%
Urban	91.4%
Building	56.2%
High Rise	56.2%
Dock	55.6%
Pier	55.6%
Dawn	52.4%
Dusk	52.4%

Facial Analysis



looks like a face	99.8%
appears to be female	100%
age range	23 - 38 years old
smiling	99.4%
appears to be happy	93.2%
wearing eyeglasses	99.9%
wearing sunglasses	97.6%

Crowd-Mode Face Detection

NEW



Facial Search

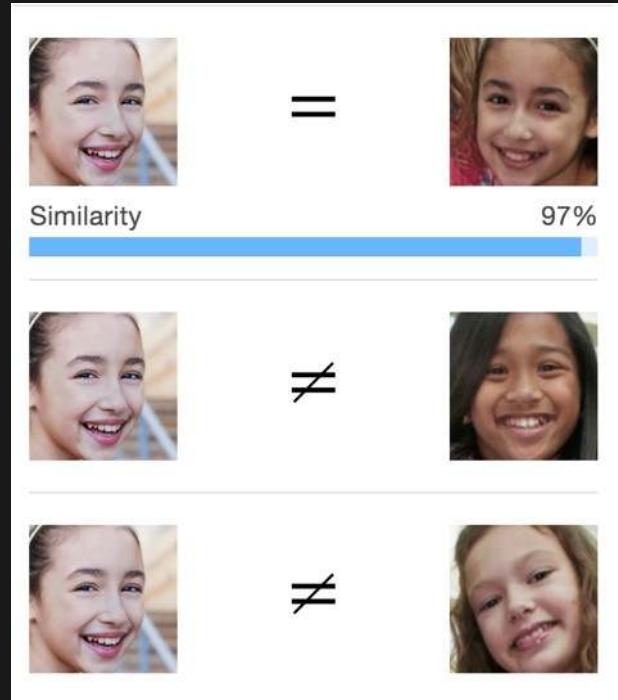


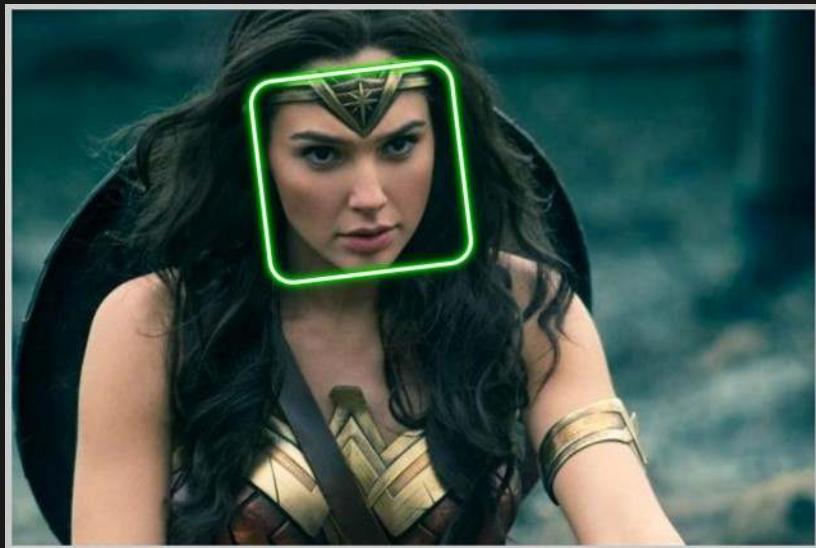
Image Moderation



▼ Results	
Suggestive	83.5%
Female Swimwear Or Underwear	83.5%

Explicit Nudity
Nudity
Graphic Male Nudity
Graphic Female Nudity
Sexual Activity
Partial Nudity
Suggestive
Female Swimwear or Underwear
Male Swimwear or Underwear
Revealing Clothes

Celebrity Recognition



▼ Results



Gal Gadot
[Learn More](#)

Match confidence 100%

Text in Image

NEW



▼ Results US English only

| 5T6E652 |

► Request

▼ Response

```
{ "TextDetections": [ { "Confidence": 96.38984680175781, "DetectedText": "5T6E652", "Geometry": { "BoundingBox": { "Height": 0.05606506019830704, "Left": 0.38905566930770874, "Top": 0.7383729219436646, "Width": 0.10844030976295471 }, "Polygon": [ { "X": 0.38905566930770874, "Y": 0.7383729219436646 }, { "X": 0.49749597907066345, "Y": 0.7726403474807739 }, { "X": 0.4896218776702881, "Y": 0.8287054300308228 } ] } } ] }
```

Rekognition API example

```
aws rekognition detect-labels
--image '{"S3Object":{"Bucket":"adhorn-reko","Name":"horse.jpg"}}'

{
  "Labels": [
    {
      "Confidence": 99.29136657714844,
      "Name": "Human"
    },
    {
      "Confidence": 99.29136657714844,
      "Name": "People"
    },
    {
      "Confidence": 99.29136657714844,
      "Name": "Person"
    },
    ....
  ]
}
```



Rekognition API example

```
aws rekognition detect-faces
  --image '{"S3Object":{"Bucket":"adhorn-reko","Name":"horse.jpg"}}'
  --attributes "ALL"

{

  "FaceDetails": [
    {
      "BoundingBox": {
        "Width": 0.05462963134050369,
        "Top": 0.2880098819732666,
        "Left": 0.4722222089767456,
        "Height": 0.07292954623699188
      },
      "Landmarks": [
        {
          "Y": 0.31606796383857727,
          "X": 0.48852023482322693,
          "Type": "eyeLeft"
        }
      ]
    }
  ]
}
```

Amazon Rekognition Helps Marinus Analytics Fight Human Trafficking

Marinus Analytics provides law enforcement with tools, founded in artificial intelligence, to turn big data into actionable intelligence. The Marinus flagship software, Traffic Jam, is a suite of tools for use by law enforcement agencies on sex trafficking investigations.

Traffic Jam began as research by CEO Emily Kennedy at the Carnegie Mellon University Robotics Institute in 2011. In 2014, Ms. Kennedy and her team founded the social innovation company, Marinus Analytics, to bring research and technology development solutions to make big data useful to law enforcement over the long term. Traffic Jam makes use of the huge amount of data online to proactively find victims, target criminals, and identify transnational criminal networks. These are a complex and evolving issues: "I am inspired by the impact we've seen thus far in the human trafficking space. Daily, we empower local, state, and federal law enforcement across the United States and Canada to rescue hundreds of victims and apprehend those who exploit them," Kennedy says.



Recently, Marinus Analytics added facial recognition to the Traffic Jam platform using the Amazon Rekognition AI service. Marinus is the first group to effectively turn facial recognition technology against the vast, clandestine networks of human traffickers that prey upon our most vulnerable citizens. In the first few days

“

Now with Traffic Jam's FaceSearch, powered by Amazon Rekognition, investigators are able to take effective action by searching through millions of records in seconds to find victims.

following the release of Traffic Jam's FaceSearch, law enforcement agents made multiple successful identifications of minors being sold online for sex. In one case, detectives identified the missing victim by searching her social media photos in FaceSearch. A positive match was quickly made and law enforcement was provided with the critical information to locate and rescue the victim.

NEW

Amazon Rekognition Video

Deep learning-based visual analysis service

(GA)



Amazon Rekognition Video



Video in. People, activities, and details out.

Objects, scenes, and activities

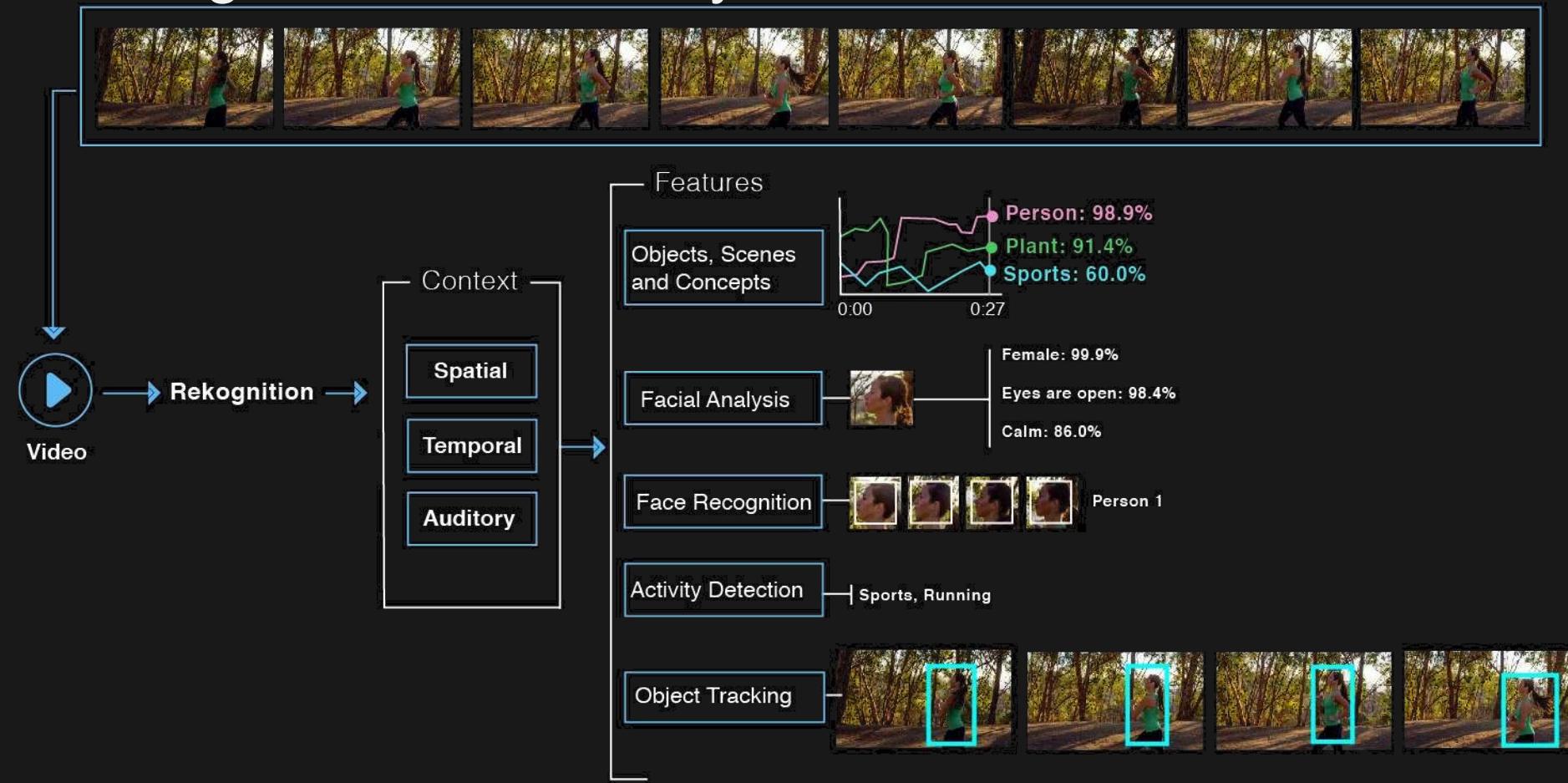
Person detection and recognition

Inappropriate content detection

Celebrity recognition

Person tracking

Rekognition Video Analysis Service



Results

Here are the results we found sorted by category. The marks on the timeline show where we found results that are selected.

57

All results

1

People

0

Celebrities

56

Objects and activities

0

Moderated labels

 Search

Select all

▼ Objects and activities

Sunset Nature Outdoors Mountain Sky

Mountain Range Crest Sunrise Dawn Dusk Red Sky

Water Promontory Ocean Sea Sea Waves Fire

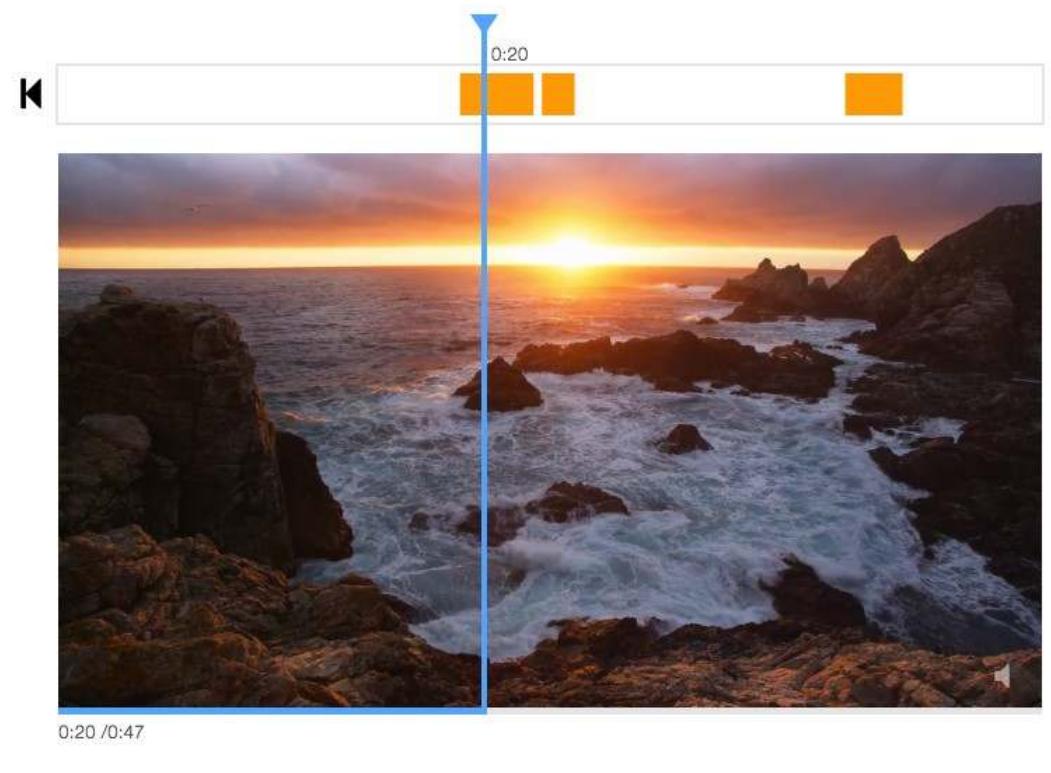
Campfire Flame Bonfire Camping Leisure Activities

Rock Person

Show more

▼ Download response

[Download response as JSON file](#)



Results

Here are the results we found sorted by category. The marks on the timeline show where we found results that are selected.

57

All results

1

People

0

Celebrities

56

Objects and activities

0

Moderated labels

Search

Select all

▼ People



Person: 7

▼ Download response

[Download response as JSON file](#)

0:44 / 0:47

TIMESCAPES

Rekognition Video API example

```
aws rekognition start-label-detection
  --video '{"S3Object":{"Bucket":"adhorn-reko","Name":"bourne.mp4"} }'

{
  "JobId": "a89eeae89ec38d8579a3a0bfc2bbf522ea5a939cdf751df4b3872d04e8394496"
}
```

```
awsrekognition get-label-detection
  --jobId "a89eeae89ec38d8579a3a0bfc2bbf522ea5a939cdf751df4b3872d04e8394496"
```



Launch customers

<https://aws.amazon.com/rekognition/customers/>



Amazon Polly

Deep learning-based Text-to-Speech service

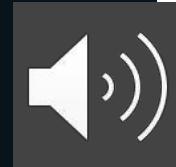


Amazon Polly: Text In, Life-like Speech Out

“Hejsan! Jag heter Astrid och läser upp det som skrivs här.”



Amazon Polly



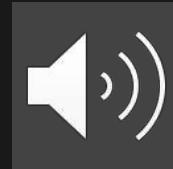
A Focus On Voice Quality & Pronunciation

Support for Speech Synthesis Markup Language (SSML) Version 1.0
<https://www.w3.org/TR/speech-synthesis>

```
<speak xml:lang="en-US">
```

The price of this book is <prosody rate="60%">€45</prosody>

```
</speak>
```



Polly API example

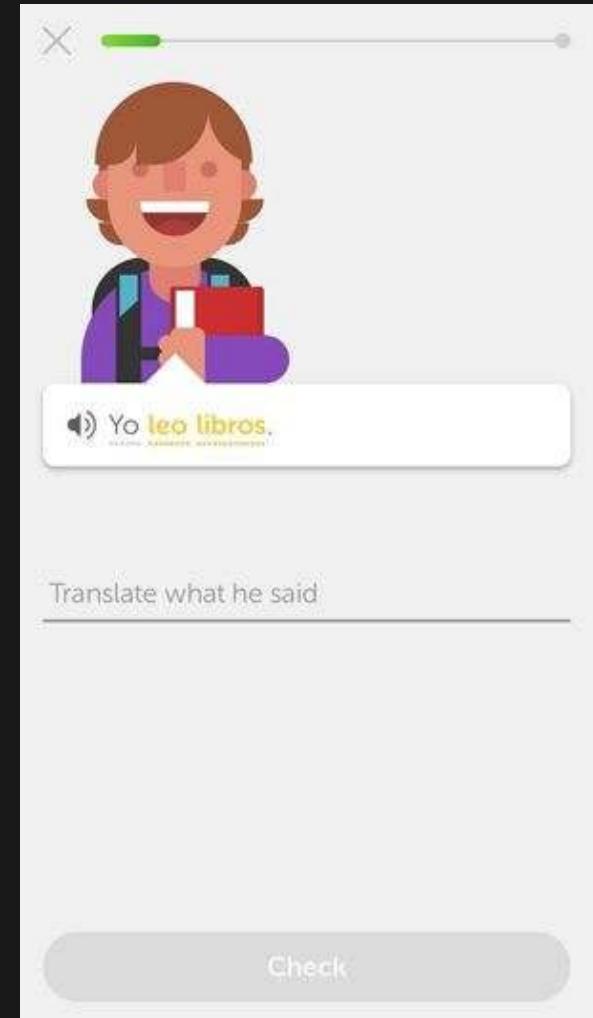
```
aws polly synthesize-speech  
  --text "It was nice to live such a wonderful live show"  
  --output-format mp3  
  --voice-id Joanna  
  --text-type text johanna.mp3
```

```
aws polly synthesize-speech  
  --text-type ssml  
  --text file://ssml_polly  
  --output-format mp3  
  --voice-id Joanna speech.mp3
```

duolingo

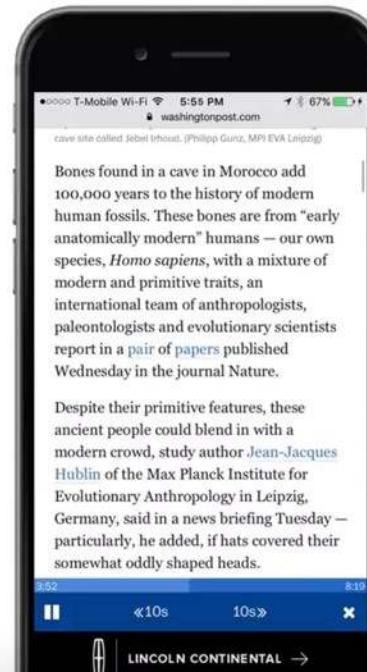
“With Amazon Polly our users benefit from the most lifelike Text-to-Speech voices available on the market.”

Severin Hacker
CTO, Duolingo



The Washington Post to start experimenting with audio articles using Amazon Polly

By WashPostPR June 9



NEW

Amazon Translate

Neural Machine Translation Service

(Preview Today)



Amazon Translate

Natural and fluent language translation

"Hello, what's up? Do you want to go see a movie tonight?"



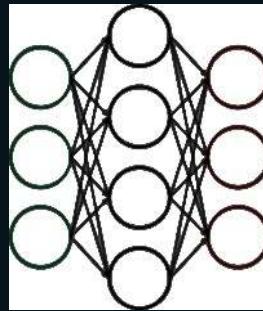
"Bonjour, quoi de neuf ? Tu veux aller voir un film ce soir ?"



Automatically translates text between languages



Real-time translation



Powered by deep learning



12 Language pairs
(more to come)



Language detection



Translate API example

```
aws translate translate-text
  --endpoint-url https://translate.us-east-1.amazonaws.com
  --region us-east-1
  --text "hello, what's up? Do you want to go see a movie tonight?"
  --source-language-code "en"
  --target-language-code "fr"

{
  "TargetLanguageCode": "fr",
  "TranslatedText": "Bonjour, quoi de neuf ? Tu veux aller voir un film ce soir ?",
  "SourceLanguageCode": "en"
}
```



Translate API example

```
aws translate translate-text  
  --endpoint-url https://translate.us-east-1.amazonaws.com  
  --region us-east-1  
  --text "hello, what's up? Do you want to go see a movie tonight?"  
  --source-language-code "en"  
  --target-language-code "fr"  
  
{  
  "TargetLanguageCode": "fr",  
  "TranslatedText": "Bonjour, quoi de neuf ? Tu veux aller voir un film ce soir ?",  
  "SourceLanguageCode": "en"  
}
```

Context Awareness



DEMO – Translation service

Amazon Translate Demo

Source Language Code: en

Target Language Code: es

hello, what's up?

Hola, ¿qué pasa?



Launch customers

<https://aws.amazon.com/translate/customers/>

Hotels.com®

 isentia

NEW

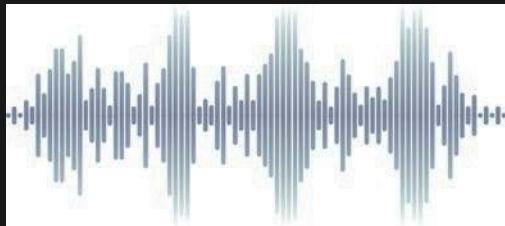
Amazon Transcribe

Automatic speech recognition service

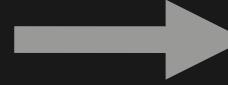
(Preview Today)



Automatic speech recognition service



Amazon
Transcribe



“Hello, this is Allan
speaking”

Automatic speech recognition service



**Available
in preview
today**



Multiple
languages



Intelligent
punctuation and
formatting



Timestamp
generation



Support for
telephony audio



Recognize
multiple
speakers



Custom
vocabulary

Launch customers

<https://aws.amazon.com/transcribe/customers/>



End-to-end
communications platform
for sales teams.



Analyze and monitor the
media coverage for
brands.

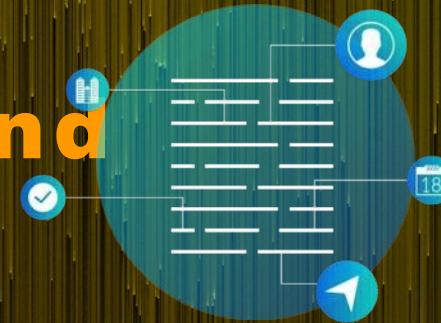


NEW

Amazon Comprehend

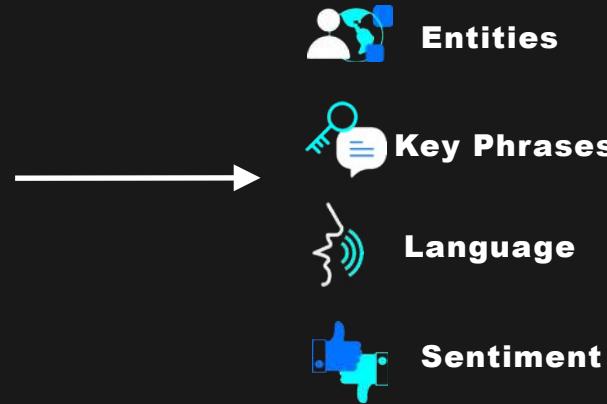
Natural Language Processing

(GA)

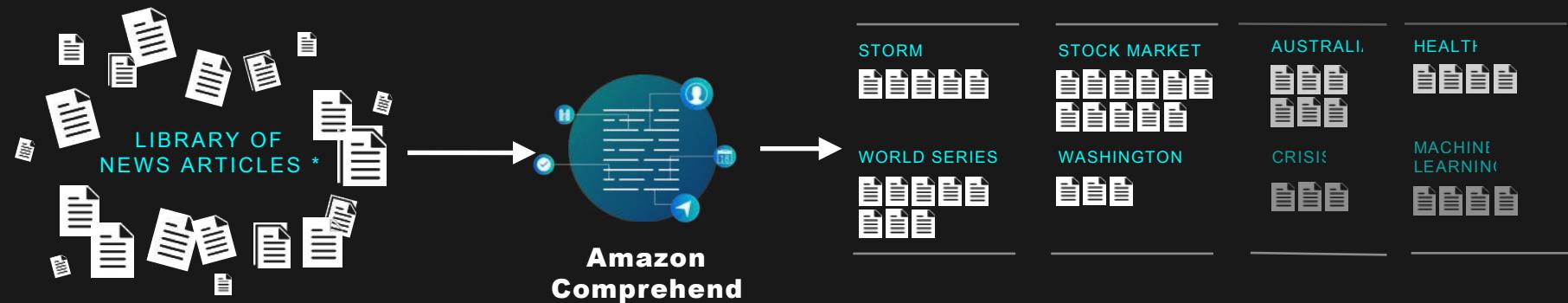


Fully managed natural language processing

Discover valuable insights from text



Support for large data sets and topic modeling



* Integrated with Amazon S3 and AWS Glue

API explorer

Paste the text that you would like to analyze with natural language processing.

[Clear text](#)

The accompanying shots from the **Newcastle Herald**'s award-winning team of photographers are worth more than 1000 words apiece. But that's the maximum number of words you will require to have a shot at winning the **Herald**'s sixth annual short story competition. It's time to let your inner author off the leash and allow your imagination to run wild. Stories can be of any genre but must be your own work. The plot must draw inspiration in some way from one of these four images. Herald photographers **Simone De Peak**, **Marina Neil**, **Jonathan Carroll** and **Max Mason-Hubers** captured the evocative pictures. Stories submitted for the competition must be between 900 and 1000 words. Judges include **Newcastle Writers Festival** director **Rosemarie Milsom**, **Herald** deputy editor **Matt Carr**, **Hunter Writers Centre** director **Karen Crofts** and president **Megan Buxton**, and **Fairfax Media**'s **Newcastle-Hunter** group managing editor **Chad Watson**.

4

916 of 1000 characters used

Language

English

Analyze

Entity

This API returns the named entities ("Person", "Organization", "Locations", etc.) within the text you analyzed.

Entity	Category	Count	Confidence
Newcastle Herald	Organization	1	0.89
more than 1000 words	Quantity	1	0.91
Herald	Organization	2	0.46
sixth annual short story competition	Quantity	1	0.79
one	Quantity	1	0.4
four images	Quantity	1	0.96
Simone De Peak	Person	1	0.99
Marina Neil	Person	1	0.99+
Jonathan Carroll	Person	1	0.99+
Max Mason-Hubers	Person	1	0.99
900	Quantity	1	0.99

Show all

Topic modeling

Amazon Comprehend will manage the lifetime on your completed topic detection jobs. Completed jobs are cleaned up periodically. The results from successfully completed jobs are persisted indefinitely, until you choose to remove them.

Create



viewing 1 to 1 items				
Job name	Job ID	Start time	End time	Status
TopicModelJob	346f527bbd09fb370be7e9f46ad7d249	11/12/2017, 22:34:29	11/12/2017, 22:42:38	Completed

Job info

Job name TopicModelJob

Job ID 346f527bbd09fb370be7e9f46ad7d249

Analysis type Topic modeling

Number of topics 10

Start time 11 Dec 2017 20:34:29 GMT

End time 11 Dec 2017 20:42:38 GMT

Status COMPLETED

Job information --

Actions

Input data location s3://public-sample-us-east-1

Output data location s3://adhorn-topicmodeling-output/322549714802-346f527bbd09fb370be7e9f46ad7d249-1513024469818/output/output.tar.gz

Comprehend API example

```
aws comprehend detect-sentiment
  --text "I love you"
  --language-code "en"

{

  "SentimentScore": [
    {
      "Mixed": 0.005664939060807228,
      "Positive": 0.9262985587120056,
      "Neutral": 0.06511948257684708,
      "Negative": 0.0029170133639127016
    },
    "Sentiment": "POSITIVE"
  }
}
```



Launch customers

<https://aws.amazon.com/comprehend/customers/>

Hotels.com®



The Washington Post

 isentia

The isentia logo features a blue, stylized 'i' shape composed of two curved segments, followed by the word "isentia" in a light gray, sans-serif font.

Amazon Lex

Conversational Interfaces





Intents

A particular goal that the user wants to achieve



Utterances

Spoken or typed phrases that invoke your intent

I'd like to book a hotel.

Sure, which city?

New York City

What date do you check in?

...

November 30th.



Slots

Data the user must provide to fulfill the intent

Prompts

Questions that ask the user to input data

Fulfillment

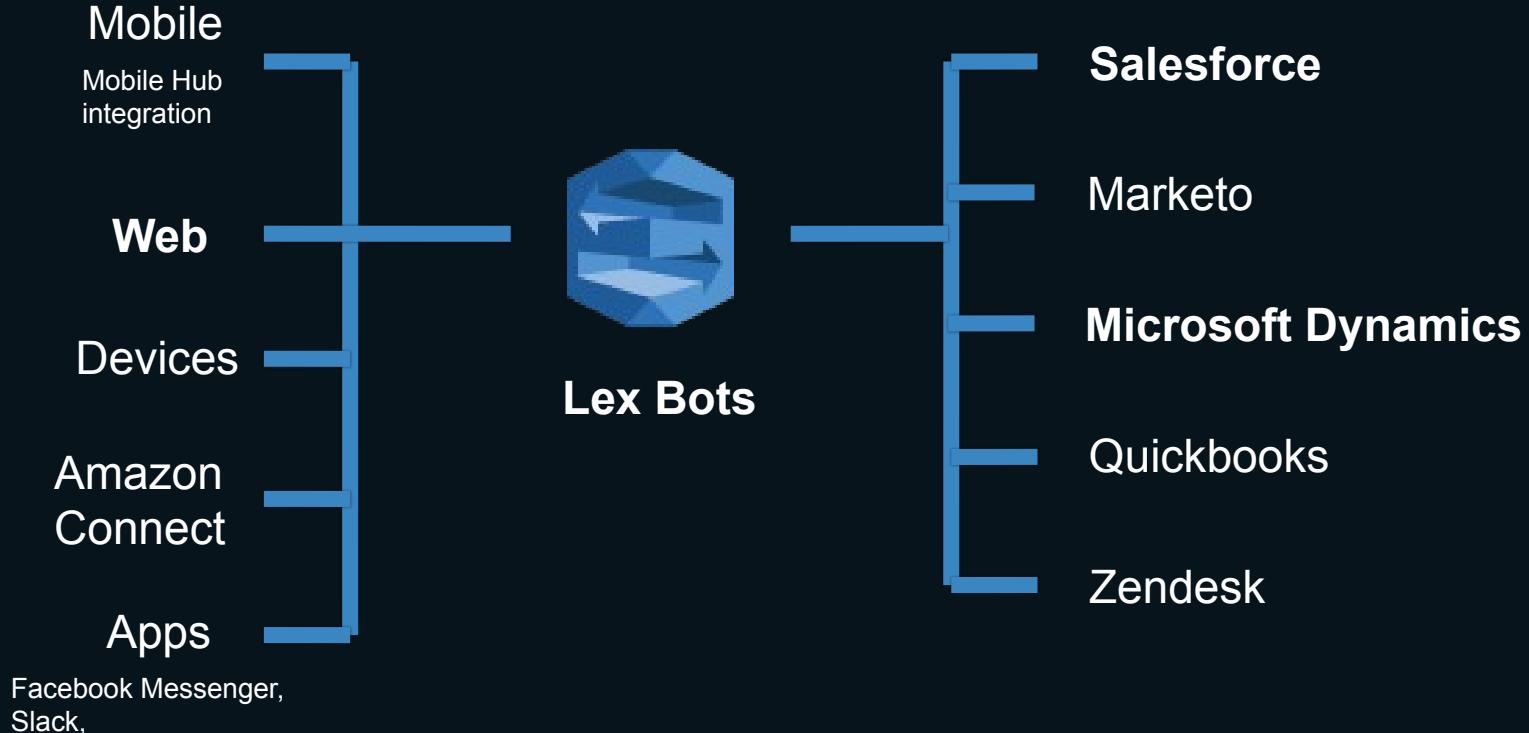
The business logic required to fulfill the user's intent

Are you sure you want to book the hotel in New York City?

Yes.

Thank you. The reservation went through successfully.

Amazon Lex: Conversational Chatbots



AWS ML Stack

Application Services

API-driven services: Vision & Language Services, Conversational Chatbots

Platform Services

Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.

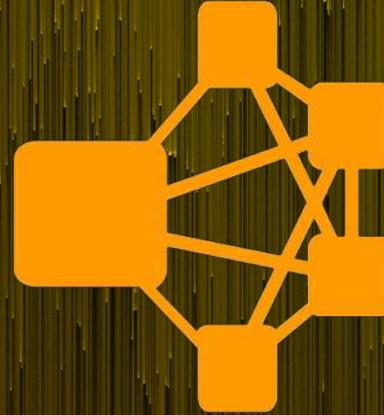
Frameworks & Infrastructure

Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run inference on trained models.



Amazon EMR

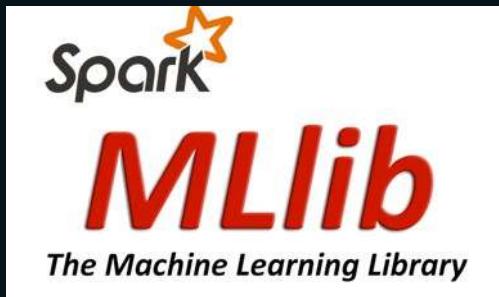
Easily Run and Scale Apache Hadoop,
Spark, HBase, Presto, Hive, and other
Big Data Frameworks



ML Applications on Amazon EMR



Amazon EMR (Elastic MapReduce)



Spark 1.2.2 Overview Programming Guides ▾ API Docs ▾ Deploying ▾ More ▾

Machine Learning Library (MLlib) Programming Guide

MLlib is Spark's scalable machine learning library consisting of common learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as underlying optimization primitives, as outlined below:

- [Data types](#)
- [Basic statistics](#)
 - summary statistics
 - correlations
 - stratified sampling
 - hypothesis testing
 - random data generation
- [Classification and regression](#)
 - linear models (SVMs, logistic regression, linear regression)
 - naive Bayes
 - decision trees
 - ensembles of trees (Random Forests and Gradient-Boosted Trees)
- [Collaborative filtering](#)
 - alternating least squares (ALS)
- [Clustering](#)
 - k-means
- [Dimensionality reduction](#)
 - singular value decomposition (SVD)
 - principal component analysis (PCA)
- [Feature extraction and transformation](#)
- [Optimization \(developer\)](#)
 - stochastic gradient descent
 - limited-memory BFGS (L-BFGS)

MLlib is under active development. The APIs marked Experimental/DeveloperApi may change in future releases, and the migration guide below will explain all changes between releases.





Buy

Rent

Sell

Mortgages

Agent finder

Home design

More

Advertise

Sign in or Join

Find your way home™

[Buy](#)[Rent](#)[Sell](#)[Zestimate](#) Enter an address, neighborhood, city or ZIP code[Search](#)

House obsessed? Watch our new show for the open house enthusiast.

[Watch now](#)

Open houses

75 upcoming open houses in Ashburn

5 photos

NEW CONSTRUCTION
\$689,094 3 bds • 2 ba • 2,436 sqft
Beazer Homes • 43460 Robey Sq, Ashburn, VA

16 photos

NEW CONSTRUCTION
\$629,990+ 3 bds • 3 ba • 2,870+ sqft
Winchester Homes • Landon Plan, Vistas at Lansdowne

10 photos

NEW CONSTRUCTION
\$854,990+ 3 bds • 4 ba • 3,725+ sqft
Miller and Smith • IL Next Level Home Plan, Upper Wes...

[See more listings](#)

New to market

75 homes in Ashburn listed in the past week

25 photos

COMING SOON
\$539,900 4 bds • 4 ba • 1,540 sqft
On market Jun 13 • 20686 Mandalay Ct, Ashburn, VA

2 photos

COMING SOON
\$315,000 2 bds • 2 ba • 1,361 sqft
On market Jun 15 • 20404 Alderleaf Ter, Ashburn, VA

4 photos

COMING SOON
\$585,000 3 bds • 2 ba • 2,618 sqft
On market Jun 8 • 20857 Ashburn Rd, Ashburn, VA

[See more listings](#)

NEW

Amazon SageMaker

A fully managed service to quickly and easily
build machine-learning based models

(GA)



Amazon SageMaker

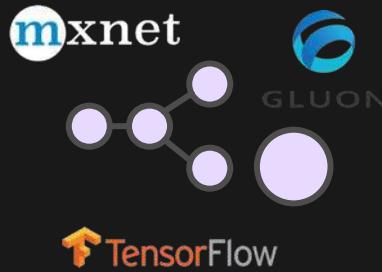
Build, train, and deploy machine learning models at scale



End-to-End
Machine Learning
Platform



Zero setup



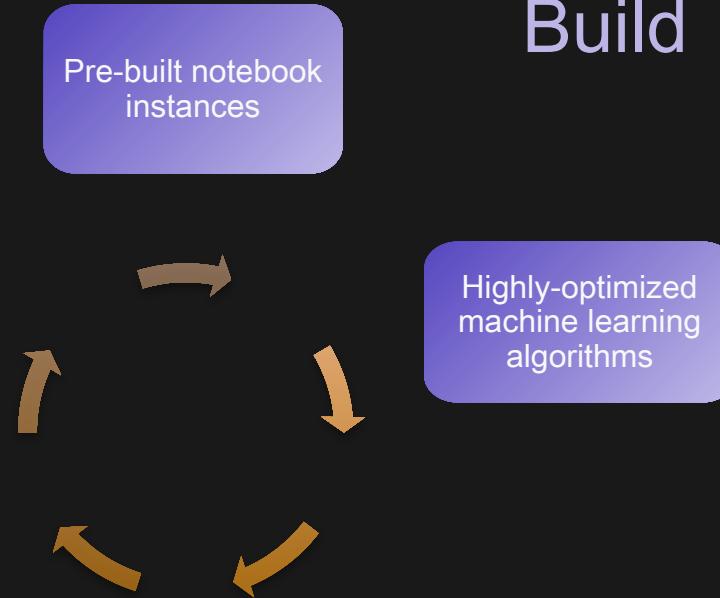
Flexible Model
Training



Pay by the second

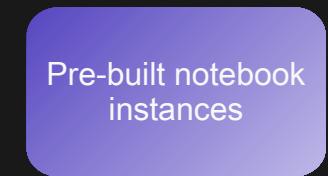
Amazon SageMaker

Build



Amazon SageMaker

Build



Highly-optimized
machine learning
algorithms



Train

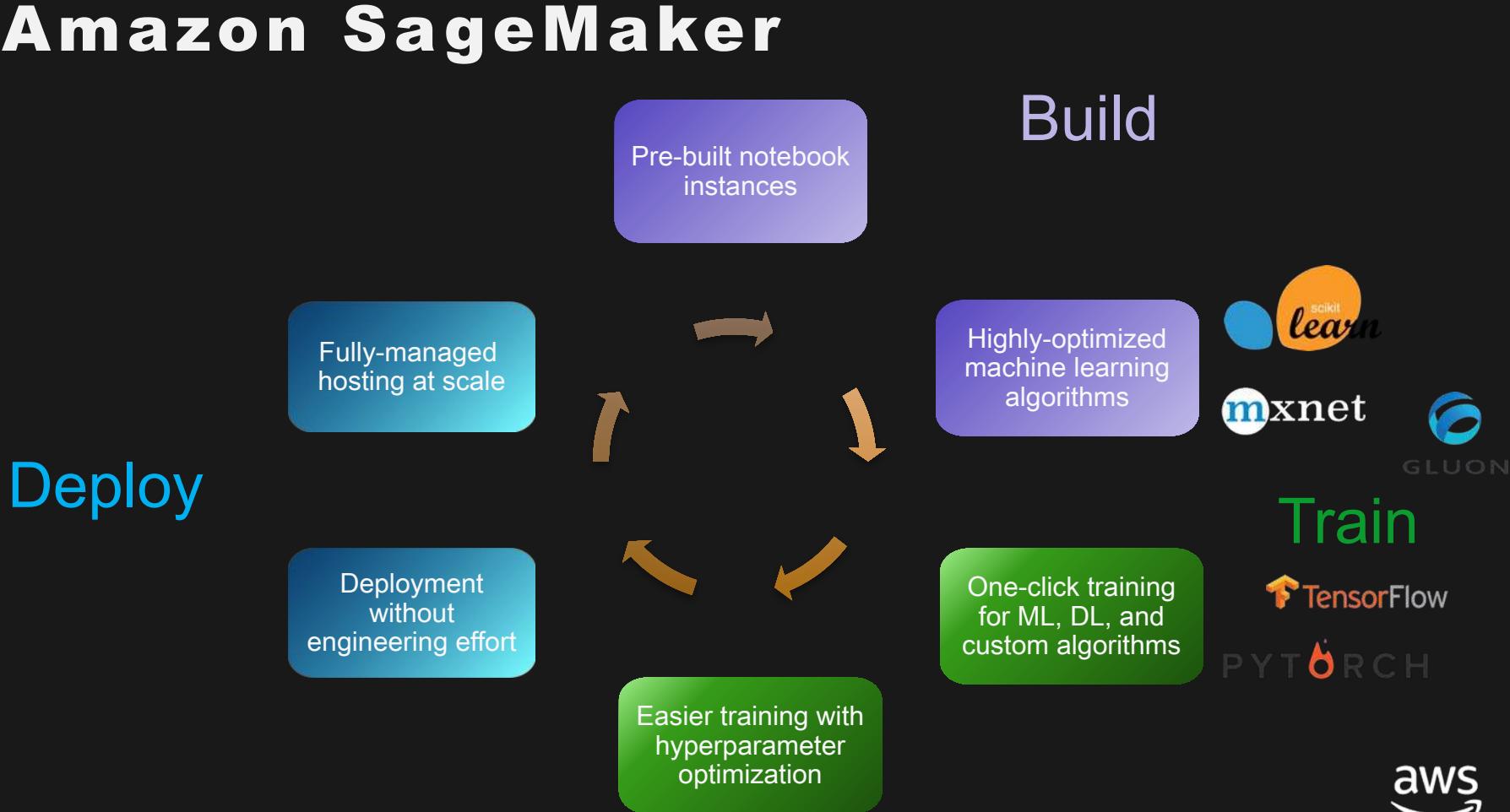
One-click training
for ML, DL, and
custom algorithms



Easier training with
hyperparameter
optimization



Amazon SageMaker



Launch Customers

Intuit®



ZipRecruiter®

Hotels.com

THOMSON REUTERS®

aws

AWS ML Stack

Application Services

API-driven services: Vision & Language Services, Conversational Chatbots

Platform Services

Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.

Frameworks & Infrastructure

Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run inference on trained models.

Amazon EC2 P3 Instances (October 2017)

NEW

The fastest, most powerful GPU instances in the cloud

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



AWS Deep Learning AMI

- Easy-to-launch tutorials
- Hassle-free setup and configuration
- Pay only for what you use
- Accelerate your model training and deployment
- Support for popular deep learning frameworks



NEW

AWS DeepLens

Deep learning enabled video camera for
developers

(Pre-order Today)



AWS DeepLens

World's first deep learning enabled video camera for developers



A new way to learn

Custom built for deep learning

Broad Framework Support

Deploy models from Amazon SageMaker

Integrated with AWS

Full programmable with AWS Lambda



10 minutes to your first deep learning project

1

Choose your deep learning model from the AWS DeepLens pre-trained model library, or your own models trained with Amazon SageMaker.

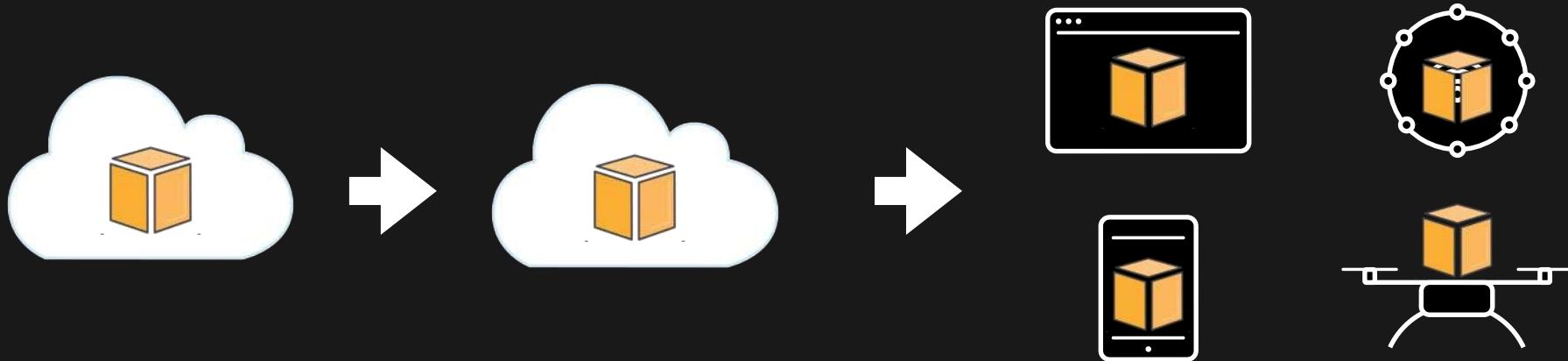
2

Deploy your model to the device with a single click.

3

Watch the results in real time in the AWS Management Console .

Infrastructure to support model build and deploy



Model
Training

Inference
in the Cloud

Inference
at the Edge

NEW

Amazon ML Lab

Provides the missing ML expertise



Amazon ML Lab



Lots of companies
doing Machine
Learning

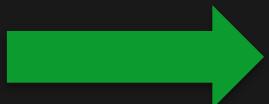


Lack ML
expertise



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



Amazon ML Lab Customers

Johnson & Johnson



The Washington Post

Summary

Democratization of AI

APPLICATION SERVICES



Amazon
Rekognition



Amazon
Rekognition
Video



Amazon
Polly



Amazon
Transcribe



Amazon
Lex



Amazon
Translate

Amazon
Comprehend

PLATFORM SERVICES

Amazon SageMaker

AWS DeepLens

Amazon EMR

FRAMEWORKS AND INTERFACES

Deep
Learning AMI

Apache MXNet

Caffe2

CNTK

PyTorch

TensorFlow

Theano

Torch

Keras

Gluon

The background features a dark teal or blue gradient with a subtle texture. Overlaid on this are numerous small, semi-transparent light gray dots of varying sizes, some connected by thin white lines, creating a network-like pattern.

GO BUILD