AI & Machine Learning at Amazon

Ian Massingham  
Developer Technology Evangelism, Amazon Web Services  
ianm@amazon.com
At Amazon, we’ve been making investments in ML for the last 20 years...
At Amazon, we’ve been making investments in ML for the last 20 years...
At Amazon, we’ve been making investments in ML for the last 20 years…
Dany's "sun-and-stars" - Daenerys' deceased husband; A great khal of the Dothraki; Was never defeated in battle; Died of a wound that festered and was helped along by Mirri Maz Duur

From Shelfari.com

Chapter 2

They had been given to her when she wed Khal Drogo. It was Drogo who had given her the pelt she wore, the head and hide of a krakkar, the white lion of the Dothraki sea.

Page 31 · Loc 648

Chapter 5

"That did not stop you selling her to Khal Drogo ..."

Drogo neither buy nor sell. Say rather that her brother Viserys gave her to Drogo to win the khal's friendship.

Page 73 · Loc 1470

Chapter 23

*Bring the grey linen gown with the pearls on the..."
At Amazon, we’ve been making investments in ML for the last 20 years...
ML @ AWS: OUR MISSION

Put machine learning in the hands of every developer and data scientist
The Amazon ML Stack

Application Services
- Rekognition
- Transcribe
- Translate
- Polly
- Comprehend
- Lex

Platform Services
- Amazon SageMaker
- AWS DeepLens

Frameworks & Interfaces
- AWS Deep Learning AMIs
- Caffe2
- CNTK
- Apache MXNet
- PyTorch
- TensorFlow
- Torch
- Keras
- Gluon
AWS Machine Learning
Language Services
<table>
<thead>
<tr>
<th>Service</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Transcribe</td>
<td>Automatic conversion of speech into accurate, grammatically correct text</td>
</tr>
<tr>
<td>Amazon Translate</td>
<td>Natural and fluent language translation</td>
</tr>
<tr>
<td>Amazon Polly</td>
<td>Turn text into lifelike speech using deep learning</td>
</tr>
<tr>
<td>Amazon Comprehend</td>
<td>Discover insights and relationships in text</td>
</tr>
<tr>
<td>Amazon Lex</td>
<td>Conversational interfaces for text-based and voice-based applications</td>
</tr>
</tbody>
</table>
Benefits of ML Language Services

• **Easily add intelligence to apps**
  with pre-trained APIs for speech, transcription, translation, language analysis, and chatbot functionality

• **Connect to comprehensive analytics**
  including data warehousing, business intelligence, batch processing, stream processing, and workflow orchestration

• **Integrate with the most complete big data platform**
  including the data lake and database tools to run machine learning workloads
Common Language Use Cases

- Voice of Customer Applications
- Customer Service/Call Centers
- Enterprise Digital Assistant
- Personalization
- Knowledge Management
- Semantic Search
- Captioning Workflows
- Localization
- Education
- Accessibility
- Information Bots
Deep dive/Components
Amazon Comprehend

Discover insights and relationships in text using deep learning
Amazon Comprehend

Discover insights and relationships in text using deep learning

Library of news articles

Amazon Comprehend

Storm

World Series

Australia

Crisis

Stock market

Washington

Health

Machine learning
Amazon Transcribe

Automatic conversion of speech into accurate, grammatically correct text

- Multiple languages
- Support for telephony audio
- Intelligent punctuation and formatting
- Recognize multiple speakers
- Timestamp generation
- Custom vocabulary
Amazon Translate
Natural and fluent language translation

- Real-time translation
- Batch analysis
- Automatic language recognition
- Low cost
Amazon Polly

Turn text into lifelike speech using deep learning

- Wide selection of voices and languages
- Synchronize speech
- Fine-grained control
- Unlimited replay
Amazon Lex

Building conversational interfaces into any application using voice and text

- Integrated development in the AWS console
- Trigger Lambda functions
- Multi-step conversations
- One-click deployment
- Enterprise connectors
- Fully managed
CUSTOMERS RUNNING MACHINE LEARNING ON AWS TODAY
**AWS ML Language Services Customers**

“Building intelligent applications to help customers drive their businesses is our entire focus. *Amazon Comprehend* allows us to analyze unstructured text within search, chat, and documents to understand intent and sentiment. This capability enables us to train our Coleman AI skillset, and also provide a truly focused and tailored search experience for our customers.”

Manjunath Ganimasty, V.P. Software Development with Infor

“Through *Amazon Lex*, we’re adding sophisticated natural language processing capabilities that help GrowthBot provide a more intuitive UI for our users. Amazon Lex lets us take advantage of advanced AI and machine learning without having to code the algorithms ourselves”

Dharmesh Shah, HubSpot CTO and Founder

“I can’t think of many use cases where accurate pronunciation is more important than when you’re learning a new language. We have found that the *Amazon Polly* voices are not just high in quality, but are as good as natural human speech for teaching a language.”

Severin Hacker, CTO, Duolingo

“At Isentia, we built our media intelligence software in a single language. Having tried multiple Machine Translation services in the past, we are impressed with how easy it is to integrate *Amazon Translate* into our pipeline and its ability to scale to handle any volume we throw at it. The translations also came out more accurate and nuanced and met our high standards for clients.”

Andrea Walsh - CIO, Isentia

“RingDNA is an end-to-end communications platform for sales teams. A critical component of RingDNA’s Conversation AI requires best of breed speech-to-text to deliver transcriptions of every phone call. RingDNA is excited about *Amazon Transcribe* since it provides high-quality speech recognition at scale, helping us to better transcribe every call to text.”

Howard Brown – CEO & Founder, RingDNA
Lex Use Case: Digital Assistant to Book a Hotel

“Book a hotel in NYC”

Automatic speech recognition

Hotel booking

Natural language understanding

<table>
<thead>
<tr>
<th>Utterances</th>
<th>Hotel booking</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>New York City</td>
</tr>
<tr>
<td>Check in</td>
<td>November 30</td>
</tr>
<tr>
<td>Check out</td>
<td>December 2</td>
</tr>
</tbody>
</table>

“Can I go ahead with the booking?”

Amazon Polly

Confirmation: “Your hotel is booked for November 30.”
Amazon Rekognition
Amazon Rekognition

Easy-to-use deep learning-based computer vision analysis

- Fully managed service
- Image analysis
- Video analysis
- Easy-to-use API
- Real-time and batch analysis
- Low cost
Amazon Rekognition Image

Detect objects, scenes, and faces, extract text, recognize celebrities, and identify unsafe content in images.

- Object and Scene Detection
- Facial Analysis
- Face Recognition
- Unsafe Image Detection
- Celebrity Recognition
- Text in Image
- Face Comparison
Video Analysis
Limitations of traditional solutions

- Temporal information lost
- Motion context lost
- Manually intensive
- Slow and error-prone
- Expensive
Amazon Rekognition Video
Analyze activity, recognize, and track people in stored and live video

- Object and activity detection
- Person tracking
- Real-time live stream
- Celebrity recognition
- Unsafe video detection
- Face recognition
- Facial Analysis
Deep Dive/Components
Amazon Rekognition Video
Media and entertainment

Create metadata for celebrities, emotions, key topics in video with time segments for recommendation engines and ad placement.

Automatically detect unsafe content, based on market requirements.

Extract data in streaming mode to enhance user engagement.
1. Video is uploaded and stored to S3
2. Amazon Rekognition Video creates metadata for celebrities, emotions, key topics in video with time segments for search
3. The output is persisted as metadata into DynamoDB to ensure durability
4. Lambda also pushes the metadata and confidence scores into Elasticsearch
Face Search—Media and Entertainment

C-SPAN

Automating Footage Tagging with Amazon Rekognition

Indexed 99,000 people

Saves ~9,000 hours a year in labor
Amazon Rekognition Customers

- witlee
- SOCIAL SOUP
- realtime
- butterfleye
- influicity
- THETAKE
- C-SPAN
- Artfinder
- GO GIRL
- HYPR
- Openinfluence
- The Washington Post
- Sturdy
- SmugMug
- MARINUS ANALYTICS
- zmags
Amazon SageMaker
Bottom Layer: Frameworks & Interfaces

- **P3**
  - Tesla V100 GPUs
  - 5,120 tensor cores
  - 128 GB of memory
  - 1 petaflop of compute
  - NVLink 2.0
  - ~14X faster than P2

- **C5**
  - 72 vCPUs
  - 3.0 GHz Intel Xeon (Skylake) processors
  - Advanced Vector Extension (AVX) 512
  - 144 GB of memory
  - Nitro Hypervisor
  - 25% improvement in price/perf. than C4

- **AWS Deep Learning AMI**
  - Caffe2
  - CNTK
  - MXNet
  - PyTorch
  - Keras
  - Gluon

*Note: The diagram includes logos of various software frameworks and interfaces.*
ML is still too complicated for everyday developers

Collect and prepare training data
Choose and optimize your ML algorithm
Set up and manage environments for training
Train and tune model (trial and error)
Deploy model in production
Scale and manage the production environment
Amazon SageMaker

Easily build, train, and deploy machine learning models

1. Collect and prepare training data
2. Choose and optimize your ML algorithm
3. Set up and manage environments for training
4. Train and tune model (trial and error)
5. Deploy model in production
6. Scale and manage the production environment
Amazon SageMaker

Pre-built notebooks for common problems

Built-in, high-performance algorithms

ALGORITHMS
- K-Means Clustering
- Principal Component Analysis
- Neural Topic Modeling
- Factorization Machines
- Linear Learner - Regression
- XGBoost
- Latent Dirichlet Allocation
- Image Classification
- Seq2Seq
- Linear Learner - Classification

FRAMEWORKS
- Apache MXNet
- TensorFlow
- Caffe2, CNTK, PyTorch, Torch

Build
Amazon SageMaker

Pre-built notebooks for common problems
Built-in, high-performance algorithms
One-click training
Hyperparameter optimization

Build
Train

Deploy model in production
Scale and manage the production environment
Amazon SageMaker

- Pre-built notebooks for common problems
- Built-in, high-performance algorithms
- One-click training
- Hyperparameter optimization
- One-click deployment
- Fully managed hosting with auto-scaling

Build | Train | Deploy
AWS DeepLens

The world's first deep learning-enabled video camera for developers

HD video camera

Custom-designed deep learning
inference engine

HD video camera
with on-board
compute optimized
for deep learning

From unboxing
to first inference
in <10 minutes

Integrates with
Amazon SageMaker
and AWS Lambda

Tutorials, examples,
demos, and pre-
built models

Micro-SD
Mini-HDMI
USB
USB
Reset
Audio out
Power
Get Started with Deep Learning

It takes less than 10 minutes with AWS DeepLens

Build custom deep-learning models in the cloud using Amazon SageMaker, or use the collection of pre-trained models included with AWS DeepLens
Thank You

Ian Massingham  IanMmm
Developer Technology Evangelism, Amazon Web Services
ianm@amazon.com