

# AWS High Level ML Services

Tuesday, 26 July 2022 18:12

## Amazon Comprehend - Text Analytics, NLP, Extract: Entities, Key phrase, Language, Sentiment, Syntax

- Natural Language Processing and Text Analytics
- Input social media, emails, web pages, documents, transcripts, medical records (Comprehend Medical)
- Extract key phrases, entities, sentiment, language, syntax, topics, and document classifications
- Can train on your own data

### Key Phrases

Analyzed text

Amazon.com, Inc. is located in Seattle, WA and was founded July 5th, 1994 by Jeff Bezos, allowing customers to buy everything from books to blenders. Seattle is north of Portland and south of Vancouver, BC. Other notable Seattle-based companies are Starbucks and Boeing.

▼ Results

Q Search

Key phrases	Confidence
Amazon.com	0.86
Seattle, WA	0.95
July 5th, 1994	0.91
Jeff Bezos	0.99+
customers	0.99+
books	0.99+

### Syntax

▼ Results

Q Search

Word	Part of speech	Confidence
Amazon.com	Proper noun	0.94
,	Punctuation	0.99+
Inc.	Proper noun	0.99+
is	Auxiliary verb	0.98
located	Verb	0.99+
in	Adposition	0.99+
Seattle	Proper noun	0.99+

## Amazon Translate- Translate Text, Deep Learning, Custom terminology in CSV or TMX

- Auto detect Language and Translate
- Uses deep learning for translation
- Supports custom terminology
  - In CSV or TMX format
  - Appropriate for proper names, brand names, etc.

## Amazon Polly- Text to Speech, Pronunciation with Lexicons, Emphasis with SSML, Encode sentence/word ends for lip syncing with SpeechMarks

- Neural Text-To-Speech, many voices & languages
- Lexicons
  - Customize pronunciation of specific words & phrases
  - Example: "World Wide Web Consortium" instead of "W3C"

## Amazon Comprehend

### Extract Entities

- Extract Entities and categorize them

Analyzed text

Amazon.com, Inc. is located in Seattle, WA and was founded July 5th, 1994 by Jeff Bezos, allowing customers to buy everything from books to blenders. Seattle is north of Portland and south of Vancouver, BC. Other notable Seattle-based companies are Starbucks and Boeing.

▼ Results

Q Search

Entity	Category	Confidence
Amazon.com, Inc.	Organization	0.90
Seattle, WA	Location	0.89
July 5th, 1994	Date	0.99+
Jeff Bezos	Person	0.99+
customers	Person	0.99+
books	Person	0.99+

### Language

▼ Results

Language

English, en  
0.99 confidence

### Sentiment

- Positive, Negative, Neutral, Mixed

▼ Results

Sentiment

Neutral 0.99 confidence    Positive 0.00 confidence    Negative 0.00 confidence    Mixed 0.00 confidence

## Amazon Transcribe- Speech to Text, Custom Vocabulary, Identify Speaker, Auto-Lang detect, Identify Channel

- Speech to text
  - Input in FLAC, MP3, MP4, or WAV, in a specified language
  - Streaming audio supported (HTTP/2 or WebSocket)
    - French, English, Spanish only
- Speaker Identification
  - Specify number of speakers
- Channel Identification
  - i.e., two callers could be transcribed separately
  - Merging based on timing of "utterances"
- Automatic Language Identification
  - You don't have to specify a language; it can detect the dominant one spoken.
- Custom Vocabularies
  - Vocabulary Lists (just a list of special words – names, acronyms)
  - Vocabulary Tables (can include "SoundsLike", "IPA", and "DisplayAs")

- Lexicons
  - Customize pronunciation of specific words & phrases
  - Example: "World Wide Web Consortium" instead of "W3C"
- SSML
  - Alternative to plain text
  - Speech Synthesis Markup Language
  - Gives control over emphasis, pronunciation, breathing, whispering, speech rate, pitch, pauses.
- Speech Marks
  - Can encode when sentence / word starts and ends in the audio stream
  - Useful for lip-synching animation

#### Amazon Rekognition- Computer Vision, Object & Scene Detection, Facial Recg, Text in Image, Video Analysis

- Computer vision
- Object and scene detection
  - Can use your own face collection
- Image moderation
- Facial analysis
- Celebrity recognition
- Face comparison
- Text in image
- Video analysis
  - Objects / people / celebrities marked on timeline
  - People Pathing



#### Amazon Rekognition- Custom Labels for Unique items

- Train with a small set of labeled images
- Use your own labels for unique items
- Example: the NFL (National Football League in the US) uses custom labels to identify team logos, pylons, and foam fingers in images.

#### Extra Info to keep in mind

- Images come from S3, or provide image bytes as part of request
  - S3 will be faster if the image is already there
- Facial recognition depends on good lighting, angle, visibility of eyes, resolution
- Video must come from Kinesis Video Streams
  - H.264 encoded
  - 5-30 FPS
  - Favor resolution over framerate
- Can use with Lambda to trigger image analysis upon upload

#### Amazon Forecast- Forecast Time Series, AutoML

- Fully-managed service to deliver highly accurate forecasts with ML
- "AutoML" chooses best model for your time series data
  - ARIMA, DeepAR, ETS, NPTS, Prophet
- Works with any time series
  - Price, promotions, economic performance, etc.
  - Can combine with associated data to find relationships
- Inventory planning, financial planning, resource planning
- Based on "dataset groups," "predictors," and "forecasts."

#### Amazon Lex- Chatbot, Intents, Utterances, Slots, Lambda

- Billed as the inner workings of Alexa
- Natural-language chatbot engine
- A Bot is built around Intents
  - Utterances invoke intents ("I want to order a pizza")
  - Lambda functions are invoked to fulfill the intent
  - Slots specify extra information needed by the intent
    - Pizza size, toppings, crust type, when to deliver, etc.
- Can deploy to AWS Mobile SDK, Facebook Messenger, Slack, and Twilio

#### Amazon Lex- Automated Chatbot Designer from Transcripts (2022)

- You provide existing conversation transcripts
- Lex applies NLP & deep learning, removing overlaps & ambiguity
- Intents, user requests, phrases, values for slots are extracted
- Ensures intents are well defined and separated
- Integrates with Amazon Connect transcripts



#### Amazon Personalize - Recommendation Engine: recommend, personalized ranking from events.

- Fully-managed recommender engine
  - Same one Amazon uses
- API access
  - Feed in data (purchases, ratings, impressions, cart adds, catalog, user)

#### Amazon Personalize - Features: Real-Time or Batch, New users or new items, contextual rec, similar items, unstructured text input, user seg

- Real-time or batch recommendations
- Recommendations for new users and new items (the cold start problem)



- Same one Amazon uses
- API access
  - Feed in data (purchases, ratings, impressions, cart adds, catalog, user demographics etc.) via S3 or API integration
  - You provide an explicit schema in Avro format
  - Javascript or SDK
  - GetRecommendations
    - Recommended products, content, etc.
    - Similar items
  - GetPersonalizedRanking
    - Rank a list of items provided
    - Allows editorial control / curation

#### Amazon Personalize Terminology - Datasets, Recipes, Solutions, Campaigns

- Datasets
  - Users, Items, Interactions
- Recipes
  - USER\_PERSONALIZATION
  - PERSONALIZED\_RANKING
  - RELATED\_ITEMS
- Solutions
  - Trains the model
  - Optimizes for relevance as well as your additional objectives
    - Video length, price, etc. – must be numeric
  - Hyperparameter Optimization (HPO)
- Campaigns
  - Deploys your “solution version”
  - Deploys capacity for generating real-time recommendations

#### Maintaining Relevance - Updated Dataset, Real-Time user behaviour, retrain model increment 2 hours and full weekly

- Keep your datasets current
  - Incremental data import
- Use PutEvents operation to feed in real-time user behavior
- Retrain the model
  - They call this a new *solution version*
  - Updates every 2 hours by default
  - Should do a full retrain (trainingMode=FULL) weekly

#### Amazon Personalize Pricing - Data Ingestion, Training, Inference (Transaction per sec), Batch recomm / user / item

- Data ingestion: per-GB
- Training: per training-hour
- Inference: per TPS-hour
- Batch recommendations: per user or per item

#### Other ML Services

##### Textract (OCR, Forms), DeepRacer (Reinforcement), DeepLens (AI powered camera)

- Amazon Textract
  - OCR with forms, fields, tables support
- AWS DeepRacer
  - Reinforcement learning powered 1/18-scale race car
- DeepLens
  - Deep learning-enabled video camera
  - Integrated with Rekognition, SageMaker, Polly, Tensorflow, MXNet, Caffe

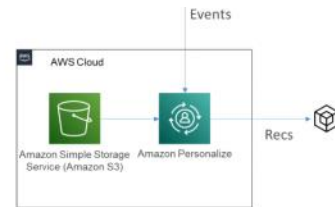
#### Amazon TorchServe - Model serving PyTorch Framework

- TorchServe
  - Model serving framework for PyTorch
  - Part of the PyTorch open source project from Facebook (Meta?)

#### AWS Neuron - SDK for ML Inference on EC2 Inf1 instance

#### recommendations

- Recommendations for new users and new items (the cold start problem)
- Contextual recommendations
  - Device type, time, etc.
- Similar items
- Unstructured text input
- Intelligent user segmentation
  - For marketing campaigns



#### Amazon Personalize Hyperparameters - User-Personalize, Personalized-ranking

- User-Personalization, Personalized-Ranking
  - hidden\_dimension (HPO)
  - bptt (back-propagation through time - RNN)
  - recency\_mask (weights recent events)
  - min/max\_user\_history\_length\_percentile (filter out robots)
  - exploration\_weight 0-1, controls relevance
  - exploration\_item\_age\_cut\_off – how far back in time you go
- Similar-items
  - item\_id\_hidden\_dimension (HPO)
  - item\_metadata\_hidden\_dimension (HPO with min & max range specified)

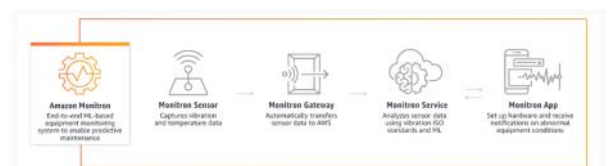
#### Amazon Personalize Security - Restrict Data, Encrypt in Rest and Transit, Access control IAM, S3 bucket policy, Monitor, Log and Audit

- Data not shared across accounts
- Data may be encrypted with KMS
- Data may be encrypted at rest in your region (SSE-S3)
- Data in transit between your account and Amazon's internal systems encrypted with TLS 1.2
- Access control via IAM
- Data in S3 must have appropriate bucket policy for Amazon Personalize to process it
- Monitoring & logging via CloudWatch and CloudTrail

#### Other ML Services - Industrial Applications

##### Amazon Lookout (Sensor Data, Computer Vision, Metrics), Amazon Monitron (Sensor device to monitor equipment)

- Amazon Lookout
  - Equipment, metrics, vision
  - Detects abnormalities from sensor data automatically to detect equipment issues
  - Monitors metrics from S3, RDS, Redshift, 3rd party SaaS apps
  - Vision uses computer vision to detect defects in silicon wafers, circuit boards, etc.
- Amazon Monitron
  - End to end system for monitoring industrial equipment & predictive maintenance



#### AWS Panorama - Appliance for Computer vision at the edge for existing IP cameras

- Computer Vision at the edge
- Brings computer vision to your existing IP cameras



## AWS Neuron - SDK for ML Inference on EC2 Inf1 instance

- AWS Neuron
  - SDK for ML inference specifically on AWS Inferentia chips
  - EC2 Inf1 instance type
  - Integrated with SageMaker or whatever else you want (deep learning AMI's, containers, Tensorflow, PyTorch, MXNet)

## AWS DeepComposer - AI-powered music piano keyboard

- AI-powered keyboard
- Composes a melody into an entire song
- For educational purposes

## Amazon CodeGuru - Automated code reviews: Java, Python

- Automated code reviews!
- Finds lines of code that hurt performance
- Resource leaks, race conditions
- Offers specific recommendations
- Powered by ML
- Supports Java and Python

## ContactLens for Amazon Connect - Search support calls and chat, Sentiment analysis, find utterances, categorize calls, talk speed, Theme detection

- For customer support call centers
- Ingests audio data from recorded calls
- Allows search on calls / chats
- Sentiment analysis
- Find "utterances" that correlate with successful calls
- Categorize calls automatically
- Measure talk speed and interruptions
- Theme detection: discovers emerging issues

## Amazon Augmented AI (A2I)- Workflows for Human review of low-confidence ML predictions. Integrated with SageMaker, Textract, Rekognition

- Human review of ML predictions
- Builds workflows for reviewing low-confidence predictions
- Access the Mechanical Turk workforce or vendors
- Integrated into Amazon Textract and Rekognition
- Integrates with SageMaker
- Very similar to Ground Truth

## Summary

- **Amazon Comprehend** - Text Analytics, NLP, Extract: Entities, Key phrase, Language, Sentiment, Syntax
- **Amazon Translate**- Translate Text, Deep Learning, Custom terminology in CSV or TMX
- **Amazon Transcribe**- Speech to Text, Custom Vocabulary, Identify Speaker, Auto-Lang detect, Identify Channel
- **Amazon Polly**- Text to Speech, Pronunciation with Lexicons, Emphasis with SSML, Encode sentence/word ends for lip syncing with SpeechMarks
- **Amazon Rekognition**- Computer Vision, Object & Scene Detection, Facial Recg, Text in Image, Video Analysis, Custom Labels for Unique items
- **Amazon Forecast**- Forecast Time Series, AutoML
- **Amazon Lex**- Chatbot, Intents for users, Utterances for words, Slots & Slot Value for input choices, Lambda for execution. Automated Chatbot Designer from Transcripts (preview 2022)
- **Amazon Personalize** - Recommendation Engine: recommend, personalized ranking from events.
  - a. Features: Real-Time or Batch, New users or new items, contextual recomm, similar items, unstructured text input, user segmentation

## existing IP cameras



## Amazon Fraud Detector - Build Model from historical fraud data and expose API to use in app

- Upload your own historical fraud data
- Builds custom models from a template you choose
- Exposes an API for your online application
- Assess risk from:
  - New accounts
  - Guest checkout
  - "Try before you buy" abuse
  - Online payments



## Amazon Kendra- Enterprise search with natural language from files, sharePoint, intranet

- Enterprise search with natural language
- For example, "Where is the IT support desk?" "How do I connect to my VPN?"
- Combines data from file systems, SharePoint, intranet, sharing services (JDBC, S3) into one searchable repository
- ML-powered (of course) – uses thumbs up / down feedback
- Relevance tuning – boost strength of document freshness, view counts, etc.
- Alexa's sister? I don't know, but that's one way to remember it ☺

## Summary

- **Textract** (OCR, Forms)
- **DeepRacer** (Reinforcement)
- **DeepLens** (AI powered camera)
- **Amazon Lookout - Detect anomalies** in Industrial Equipment (Sensor Data, Computer Vision, Metrics)
- **Amazon Monitron (Sensor Device to monitor equipment)** - End to end system for monitoring industrial equipment and predictive maintenance
- **Amazon TorchServe** - Model serving PyTorch Framework
- **AWS Panorama** - Appliance for Computer vision at the edge for existing IP cameras
- **AWS Neuron** - SDK for ML Inference on EC2 Inf1 instance
- **AWS DeepComposer** - AI-powered music piano keyboard
- **Amazon Fraud Detector** - Build Model from historical fraud data and expose API to use in app
- **Amazon CodeGuru** - Automated code reviews: Java, Python
- **ContactLens for Amazon Connect** - Search support calls and chat, Sentiment analysis, find utterances, categorize calls, talk speed, Theme detection
- **Amazon Kendra**- Enterprise search with natural language from files, sharePoint, intranet. Central search repository
- **Amazon Augmented AI (A2I)**- Workflow for Human review of low confidence ML predictions

2022)

- **Amazon Personalize** - Recommendation Engine: recommend, personalized ranking from events.
  - a. Features: Real-Time or Batch, New users or new items, contextual recomm, similar items, unstructured text input, user segmentation
  - b. Terminology: Datasets for input, Recipes pre-built, Solutions to train model with relevance, Campaigns to deploy solution version
  - c. Maintaining Relevance: Updated Dataset, Real-Time user behaviour, retrain model every 2 hours incremented and full weekly
  - d. Security: Restrict Data across accounts, Encrypt in Rest and Transit, Access control IAM, S3 bucket policy, Monitor, Log and Audit
  - e. Pricing : Data Ingestion, Training, Inference (Transaction per sec), Batch recomm / user / item
- **ContactLens for Amazon Connect** - Search support calls and chat, Sentiment analysis, find utterances, categorize calls, talk speed, Theme detection
- **Amazon Kendra**- Enterprise search with natural language from files, sharePoint, intranet. Central search repository
- **Amazon Augmented AI (A2I)**- Workflows for Human review of low-confidence ML predictions. Integrated with SageMaker, Textract, Rekognition