



Introduction to AI on AWS

Boaz Ziniman - Technical Evangelist, AWS

 @ziniman

GPU Technology Conference – Tel Aviv

October 2017





2006

EC2 & S3

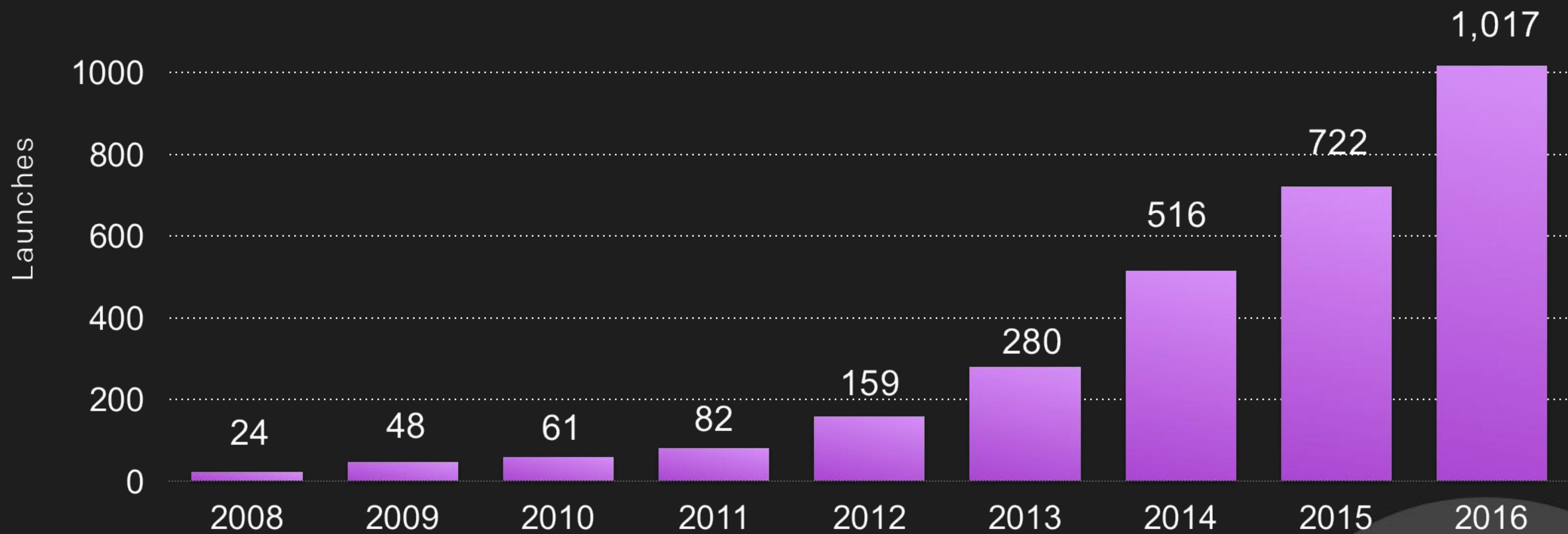
2017

90+ Managed Services

Kinesis Streaming Analytics, Mobile Testing,
Redshift Datawarehouse, Code Deploy/Build
Tools, Elastic Container Service, Application Load
Balancer, Lambda, API Gateway, DynamoDB,
Elastic Map, AI, Reduce
(Hadoop/Spark/Presto/etc), Elastic Beanstalk,
Elastic Transcoder, RDS, Elasticsearch, IoT,
more..



AWS Rapid Pace Of Innovation





Most Fully Featured Technology Platform

TECHNICAL & BUSINESS SUPPORT

Support Professional Services Optimization Guidance Partner Ecosystem Training & Certification Solutions Management Account Management Security & Billing Reports Personalized Dashboard

MARKETPLACE

Business Apps Business Intelligence DevOps Tools Security Networking Databases Storage

ANALYTICS	DEV/OPS	MOBILE SERVICES	IoT	AI	ENTERPRISE APPS	HYBRID ARCHITECTURE	.Migrations	
Data Warehousing Business Intelligence Hadoop /Spark Streaming Data Analysis Streaming Data Collection	Elasticsearch Data Pipelines Interactive SQL Queries ETL	One-click App Deployment Resource Templates Build and Test Application Lifecycle Management DevOps Resource Management	API Gateway Single Integrated Console Identity Sync Mobile Analytics	Rules Engine Device Shadows Device SDKs Device Gateway Registry	Machine Learning Image Recognition Speech to Text Conversation AI Deep Learning Frameworks	Virtual Desktops Sharing & Collaboration Corporate Email App Streaming Communications	Integrated Networking Direct Connect Identity Federation Integrated App Deployments Data Backups VMware Integration Integrated Resource Management	Schema Conversion Exabyte-Scale Data Migration Application Migration Database Migration Server Migration
APP SERVICES								
Queuing & Notifications Workflow Search	Email Transcoding	Triggers Containers Analyze and Debug	Mobile App Testing Targeted Push Notifications					

INFRASTRUCTURE

Regions Availability Zones Points of Presence

CORE SERVICES

Compute VMs, Auto-scaling, Load Balancing, Containers, Virtual Private Servers, Batch Computing, Cloud Functions, Elastic GPUs, Edge Computing

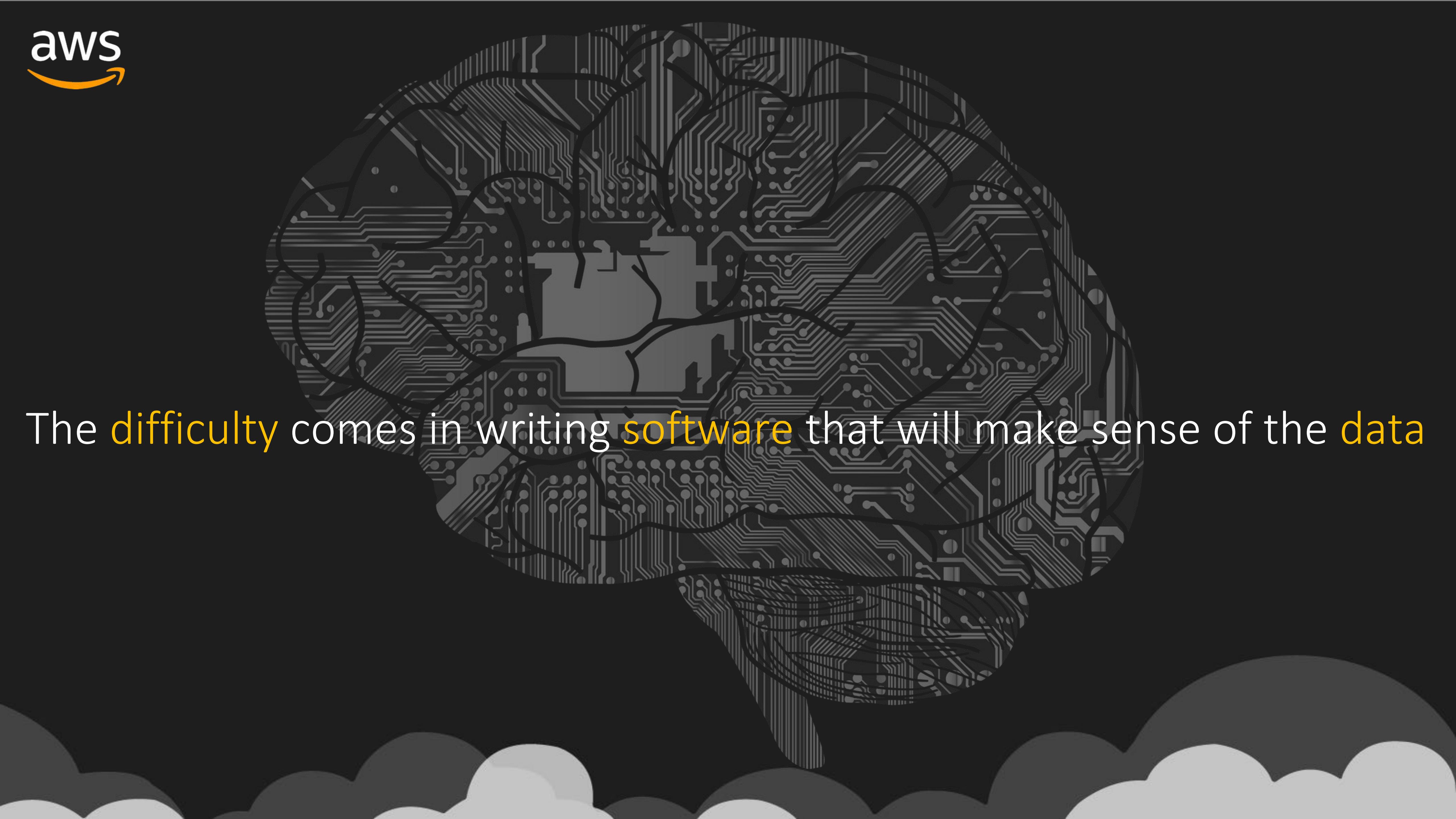
Storage Object, Blocks, File, Archives, Import/Export, Exabyte-scale data transfer

Networking VPC, DX, DNS

SECURITY & COMPLIANCE

Identity Management Access Control Monitoring & Logs Assessment and Reporting Web Application Firewall

Configuration Compliance Key Management & Storage Account Grouping Resource & Usage Auditing DDOS Protection



The difficulty comes in writing software that will make sense of the data



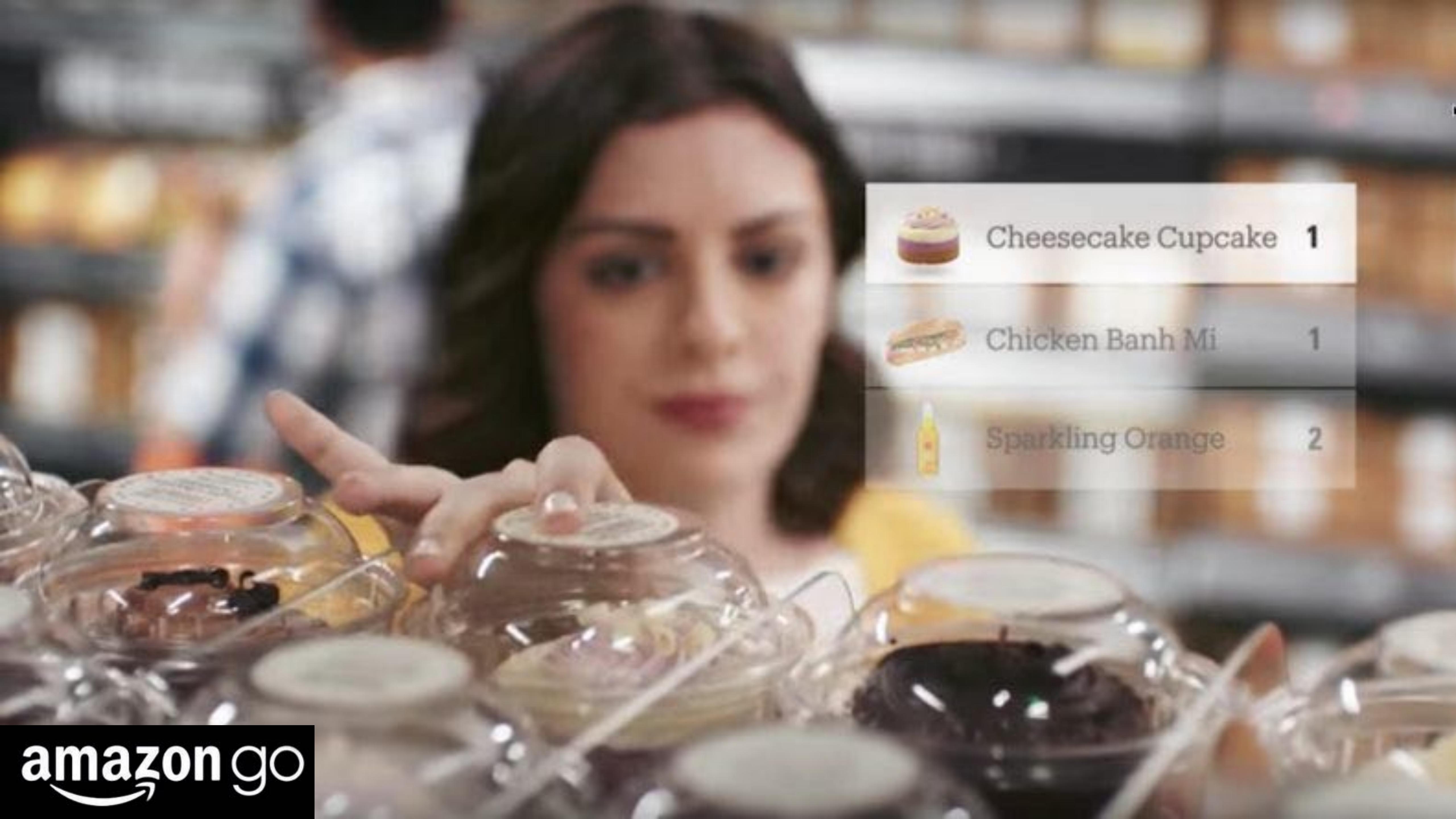
Artificial Intelligence

A system or service which can perform tasks that usually require human intelligence





amazon



amazon go

Cheesecake Cupcake 1

Chicken Banh Mi 1

Sparkling Orange 2

A black cylindrical Amazon Echo smart speaker is positioned on the left side of the frame. It has a vertical band of small circular holes near the bottom half. The word "amazon" is printed in white at the very bottom of the device.

amazon
echo

25,000 skills



And a few more examples...

Fraud detection

Detecting fraudulent transactions, filtering spam emails, flagging suspicious reviews, ...

Personalization

Recommending content, predictive content loading, improving user experience, ...

Targeted marketing

Matching customers and offers, choosing marketing campaigns, cross-selling and up-selling, ...

Content classification

Categorizing documents, matching hiring managers and resumes, ...

Churn prediction

Finding customers who are likely to stop using the service, free-tier upgrade targeting, ...

Customer support

Predictive routing of customer emails, social media listening, ...



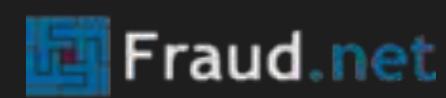
Machine Learning On AWS Today



realnetworks.



The Washington Post



infor



Carnegie Mellon



NETFLIX



Howard Hughes.



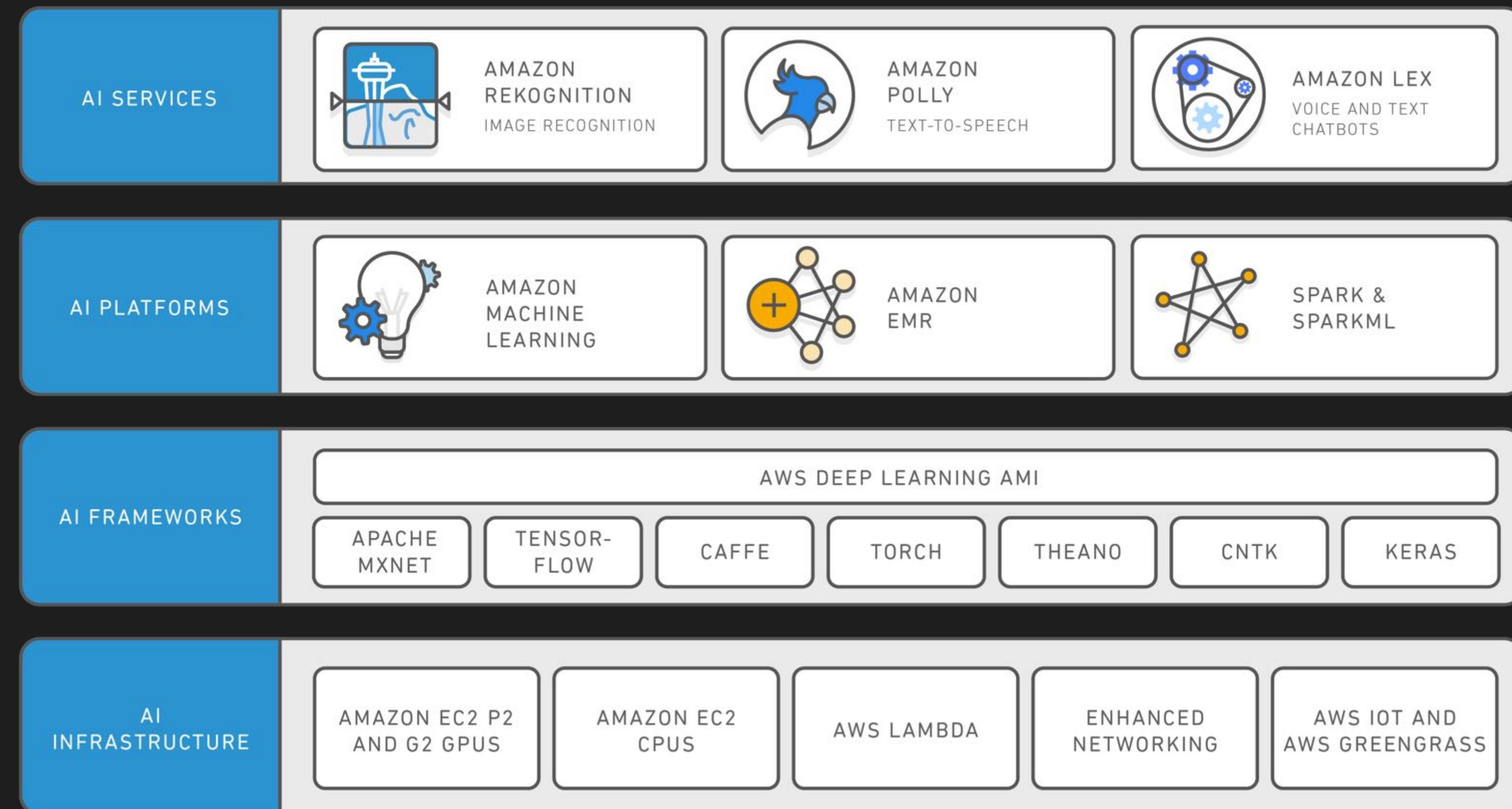
Artfinder



Openinfluence

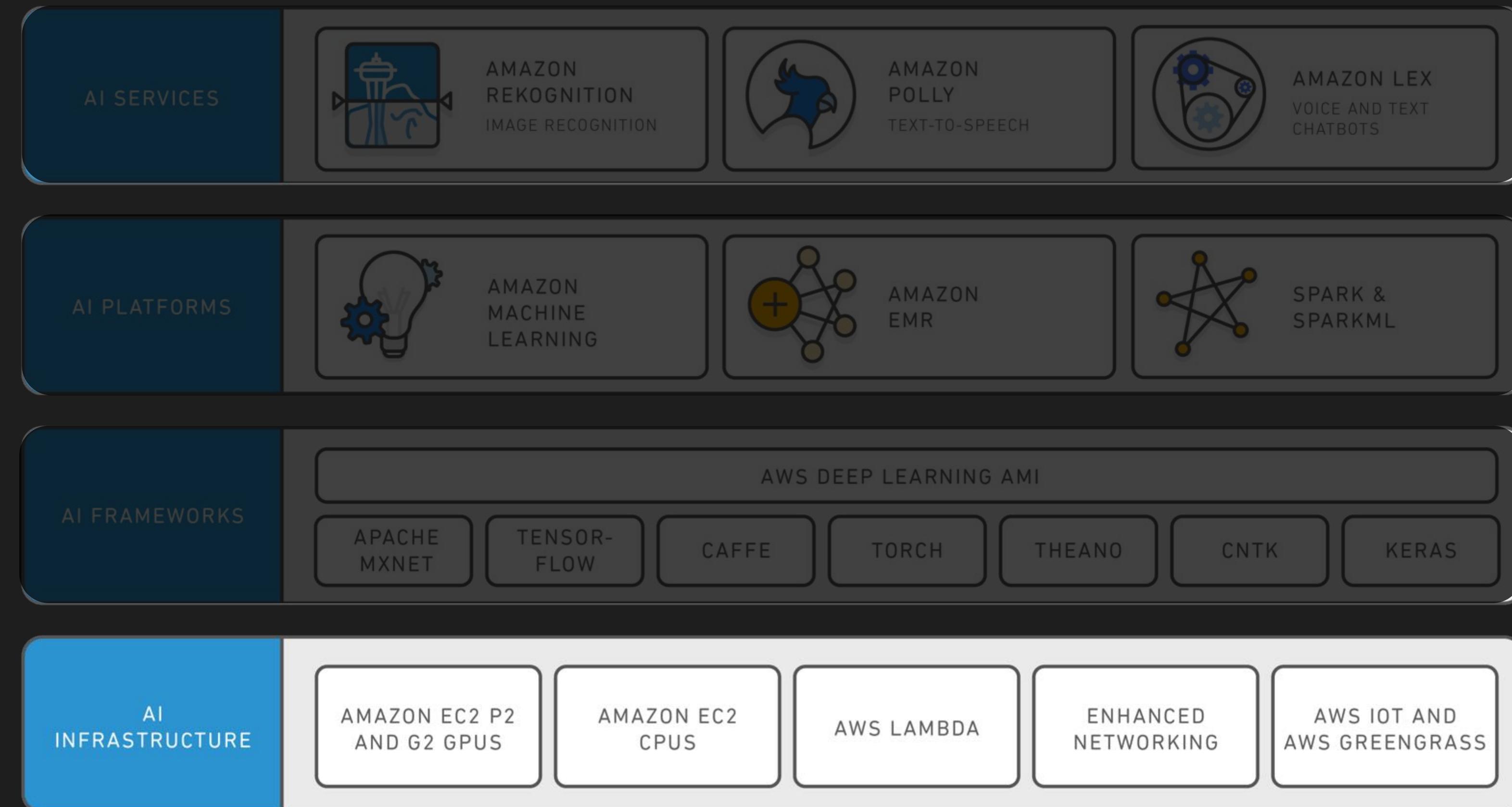


Amazon AI Ecosystem

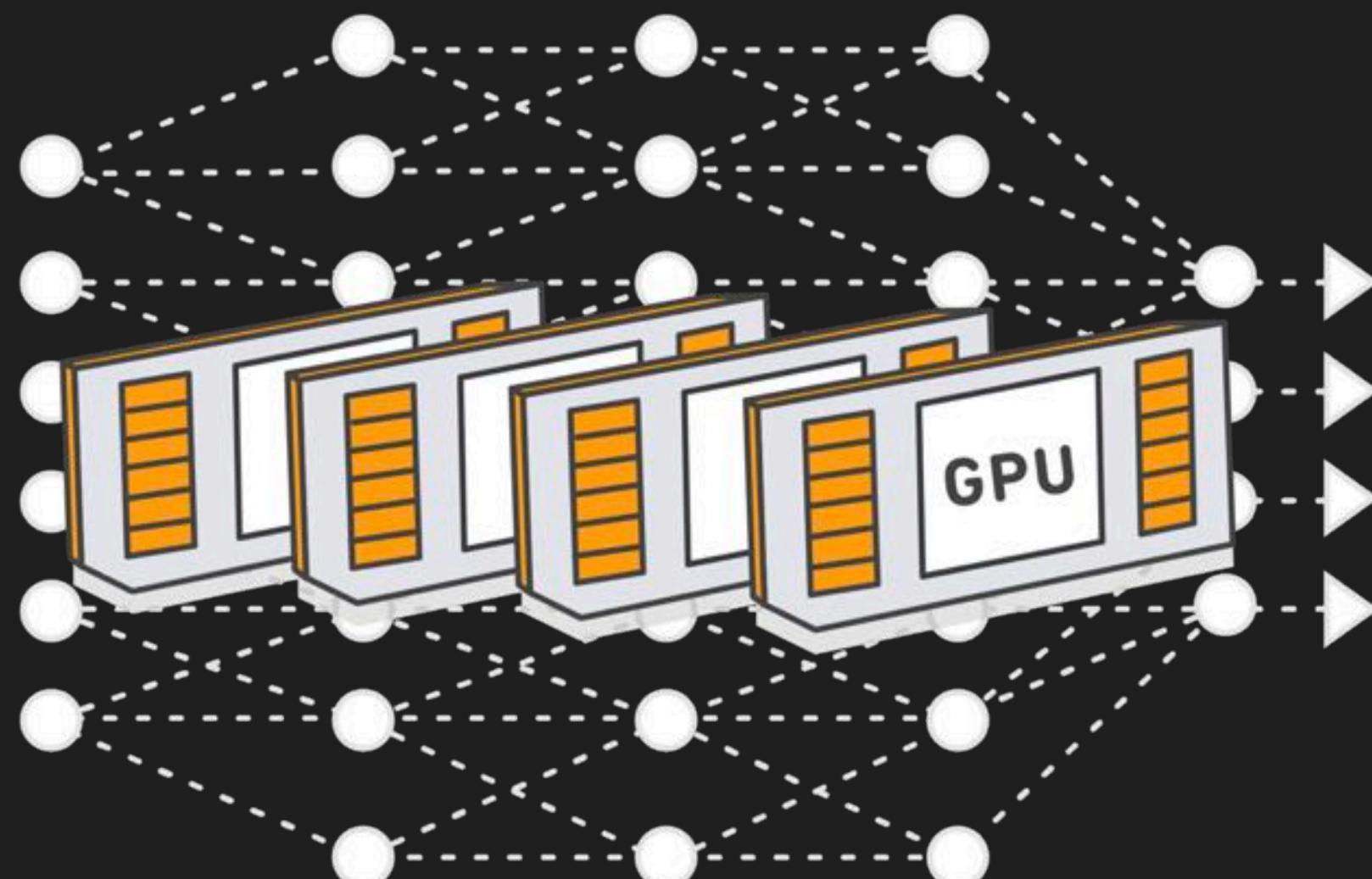




Amazon AI Ecosystem



Amazon EC2 P2 Instances



General-purpose GPU compute applications.

Features:

- High Frequency Intel Xeon E5-2686v4 (Broadwell)
- High-performance NVIDIA K80 GPUs, each with 2,496 parallel processing cores and 12GiB of GPU memory
- Supports GPUDirect™
- Enhanced Networking
- EBS-optimized by default at no additional cost

Instance Size	GPUs	GPU Peer to Peer	vCPUs	Memory (GiB)	Network Bandwidth*
p2.xlarge	1	-	4	61	1.25Gbps
p2.8xlarge	8	Y	32	488	10Gbps
p2.16xlarge	16	Y	64	732	20Gbps



Expedia - Ranking Hotel Images

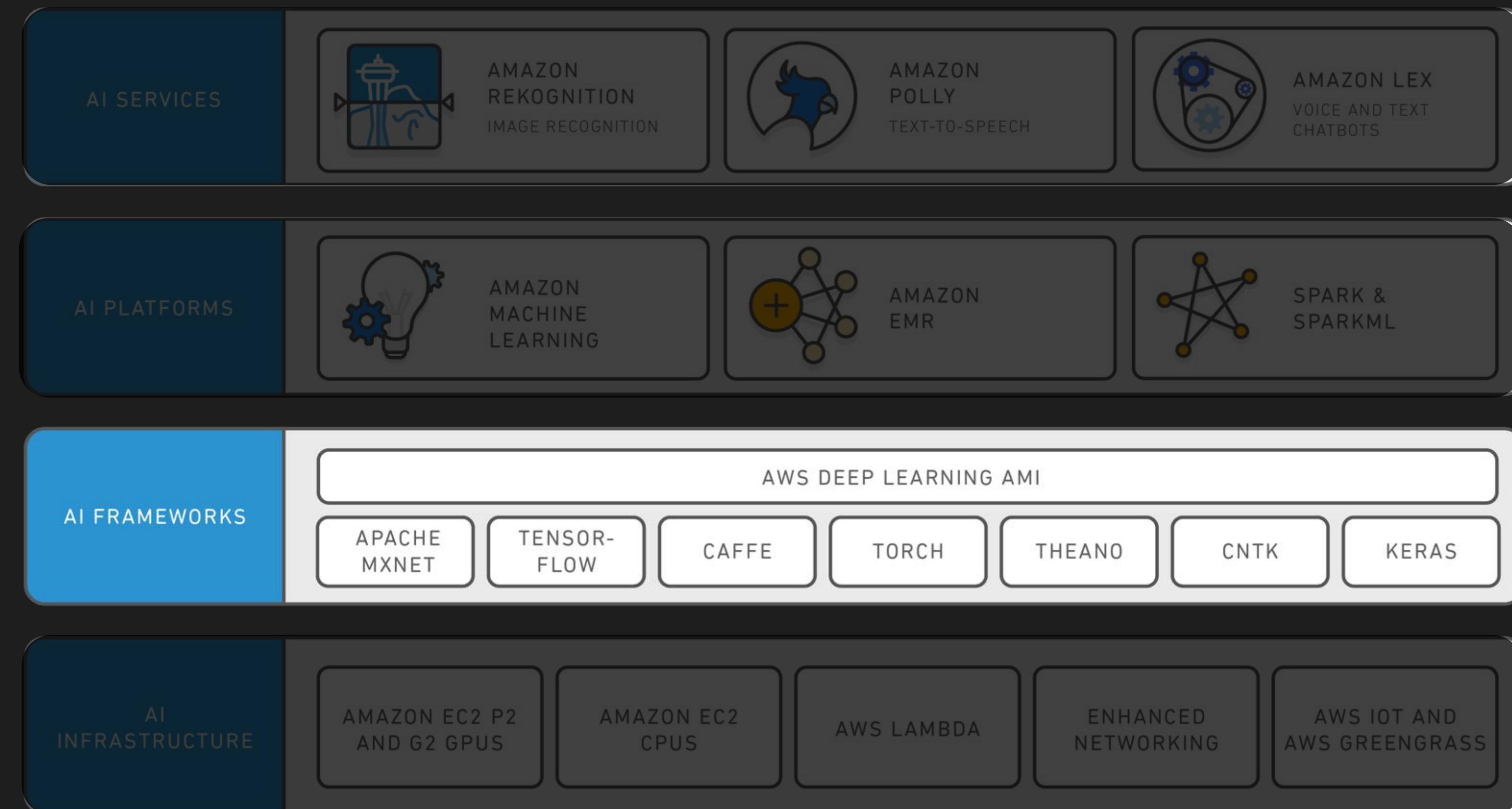
- Leading online travel company
- How to improve hotels listing images?
 - Use GPUs and deep learning to rank hotel images
 - Build a model based on 100K images
- How to rank 10M images?
 - Use data parallelization across multiple GPUs on AWS
 - Improve ranking time from more than a week to 1 day

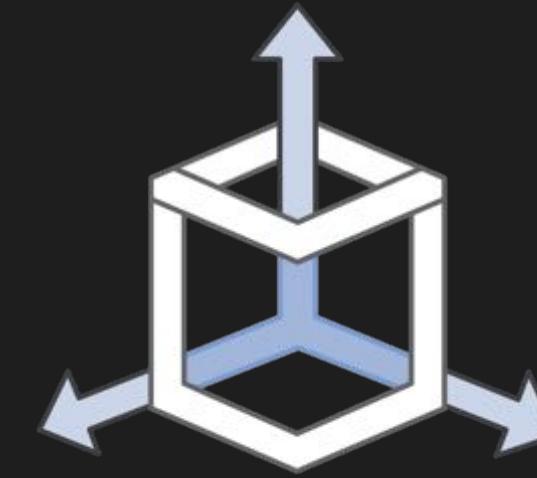


Expedia®



Amazon AI Ecosystem





One-Click Deep Learning



Available in the AWS
Marketplace

AWS Deep Learning AMIs

Amazon Linux & Ubuntu

Up to ~40k CUDA cores

Apache MXNet

TensorFlow

Theano

Keras

Caffe

CNTK

Torch

Pre-configured CUDA drivers

Anaconda, Python3

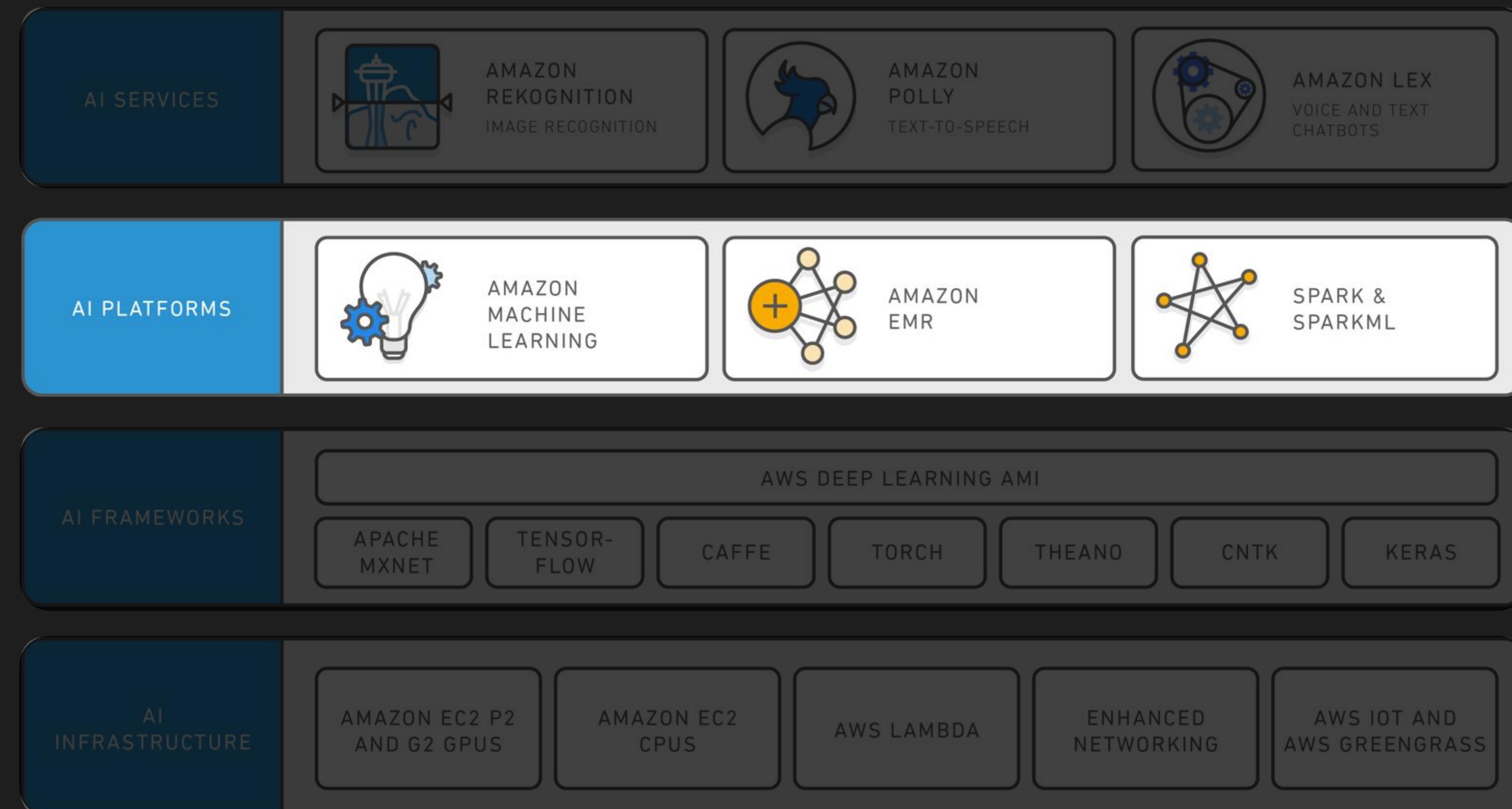
Out-of-the-box Tutorials

+ CloudFormation template

+ Container Image



Amazon AI Ecosystem

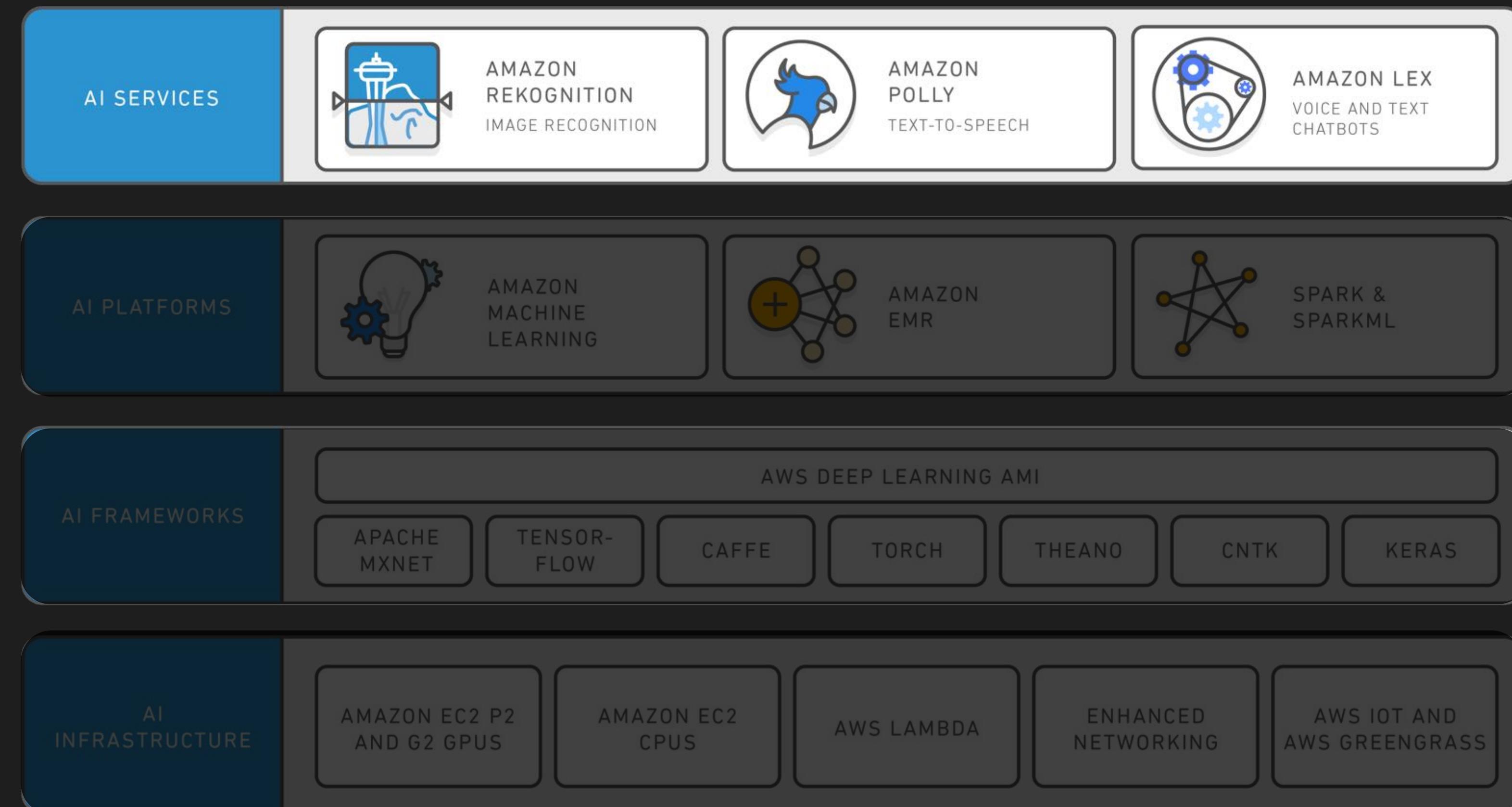




Can We Help Customers
Put Intelligence At The Heart Of
Every Application & Business?



Amazon AI Ecosystem





Artificial Intelligence Services on AWS



Polly

Text-to-Speech



Rekognition

Image Analysis



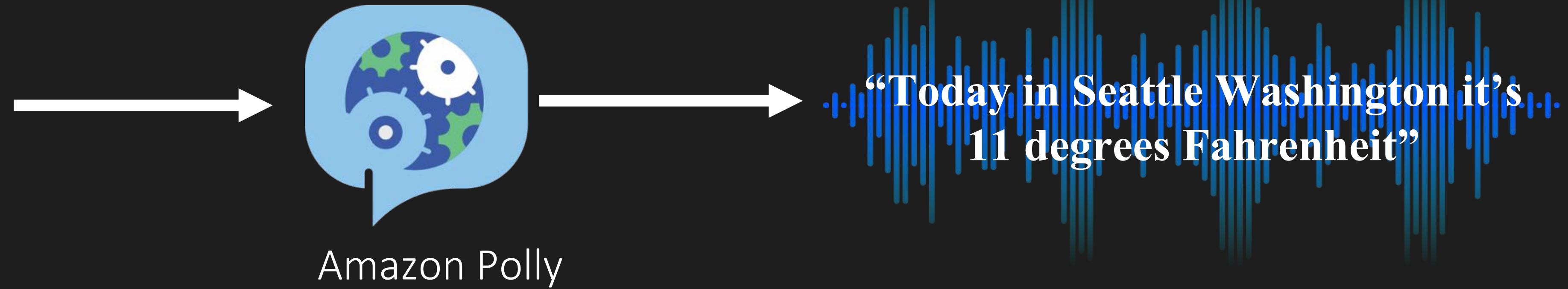
Lex

Conversation Bots



Text In, Life-like Speech Out

“Today in Seattle, WA it’s
11°F”





A Focus On Voice Quality & Pronunciation

1. Automatic, Accurate Text Processing



"Today in Seattle, WA, it's 11°F"



"We live for the music" live from the Madison Square Garden.'



A Focus On Voice Quality & Pronunciation

1. Automatic, Accurate Text Processing



2. Intelligible and Easy to Understand



A Focus On Voice Quality & Pronunciation

<https://www.w3.org/TR/speech-synthesis/>

```
<speak>
  The spelling of my name is
    <prosody rate='x-slow'>
      <say-as interpret-as="characters">Boaz</say-as>
    </prosody>
</speak>
```





A Focus On Voice Quality & Pronunciation

1. Automatic, Accurate Text Processing

2. Intelligible and Easy to Understand

3. Add Semantic Meaning to Text



“Richard’s number is 2122341237”



“Richard’s number is 2122341237”

Telephone Number



A Focus On Voice Quality & Pronunciation

1. Automatic, Accurate Text Processing
2. Intelligible and Easy to Understand
3. Add Semantic Meaning to Text
4. Customized Pronunciation



“My daughter’s name is Kaja.”



“My daughter’s name is Kaja.”



Duolingo voices its language learning service Using Polly

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

Severin Hacker
CTO, Duolingo

- Spoken language crucial for language learning
 - Accurate pronunciation matters
 - Faster iteration thanks to TTS
 - As good as natural human speech

duolingo

Duolingo is a free language learning service where users help translate the web and rate translations.

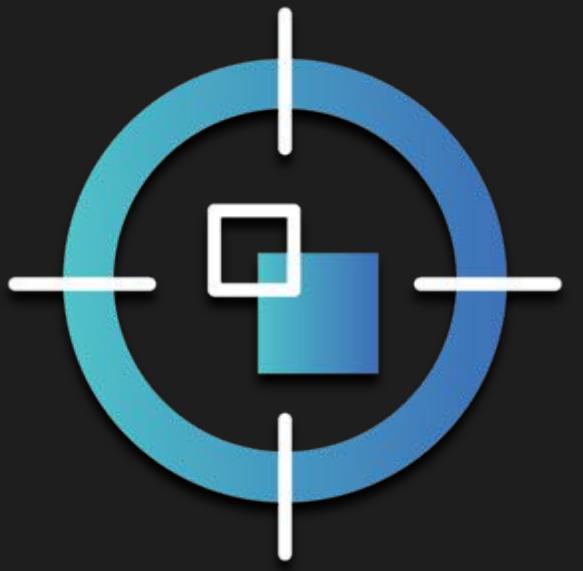


Artificial Intelligence Services on AWS



Polly

Text-to-Speech



Rekognition

Image Analysis

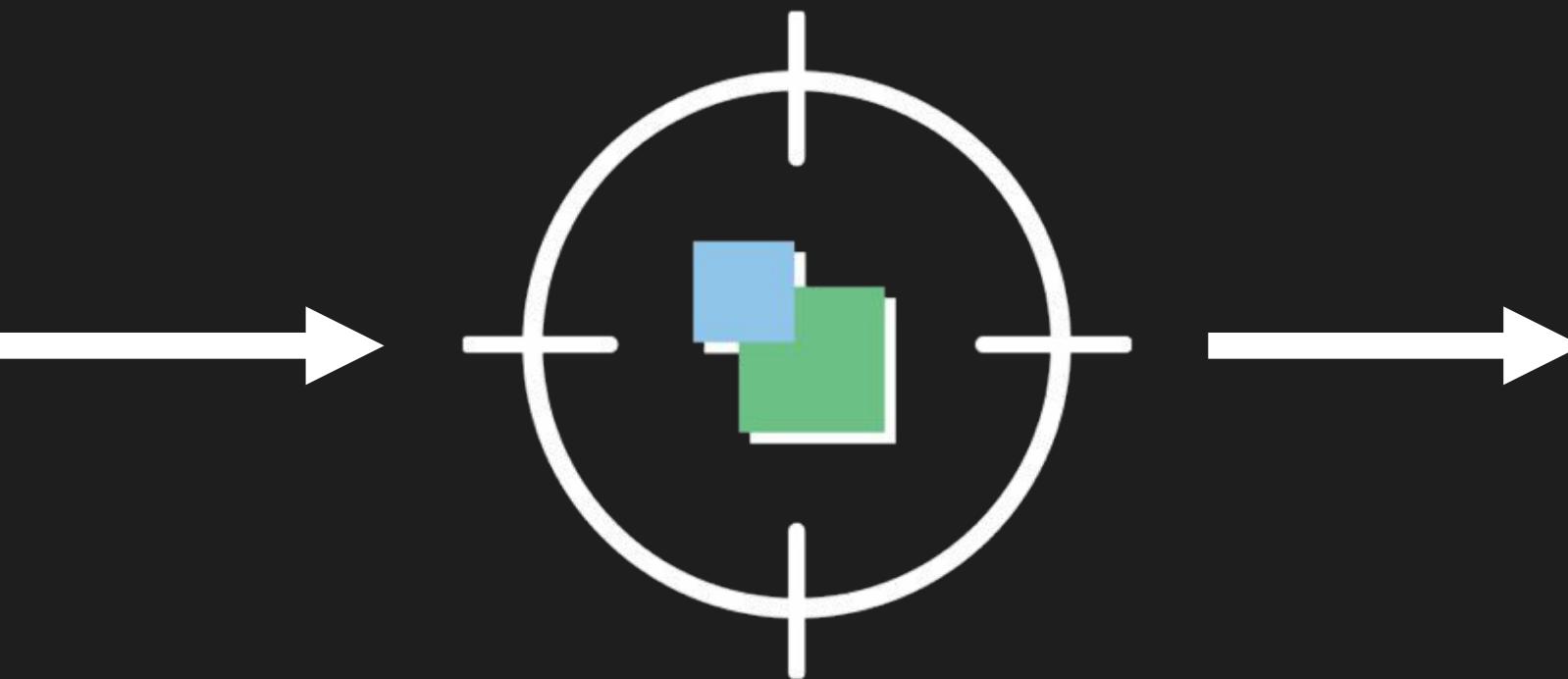


Lex

Conversation Bots



Amazon Rekognition: Images In, Rich Metadata Out



Amazon
Rekognition

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

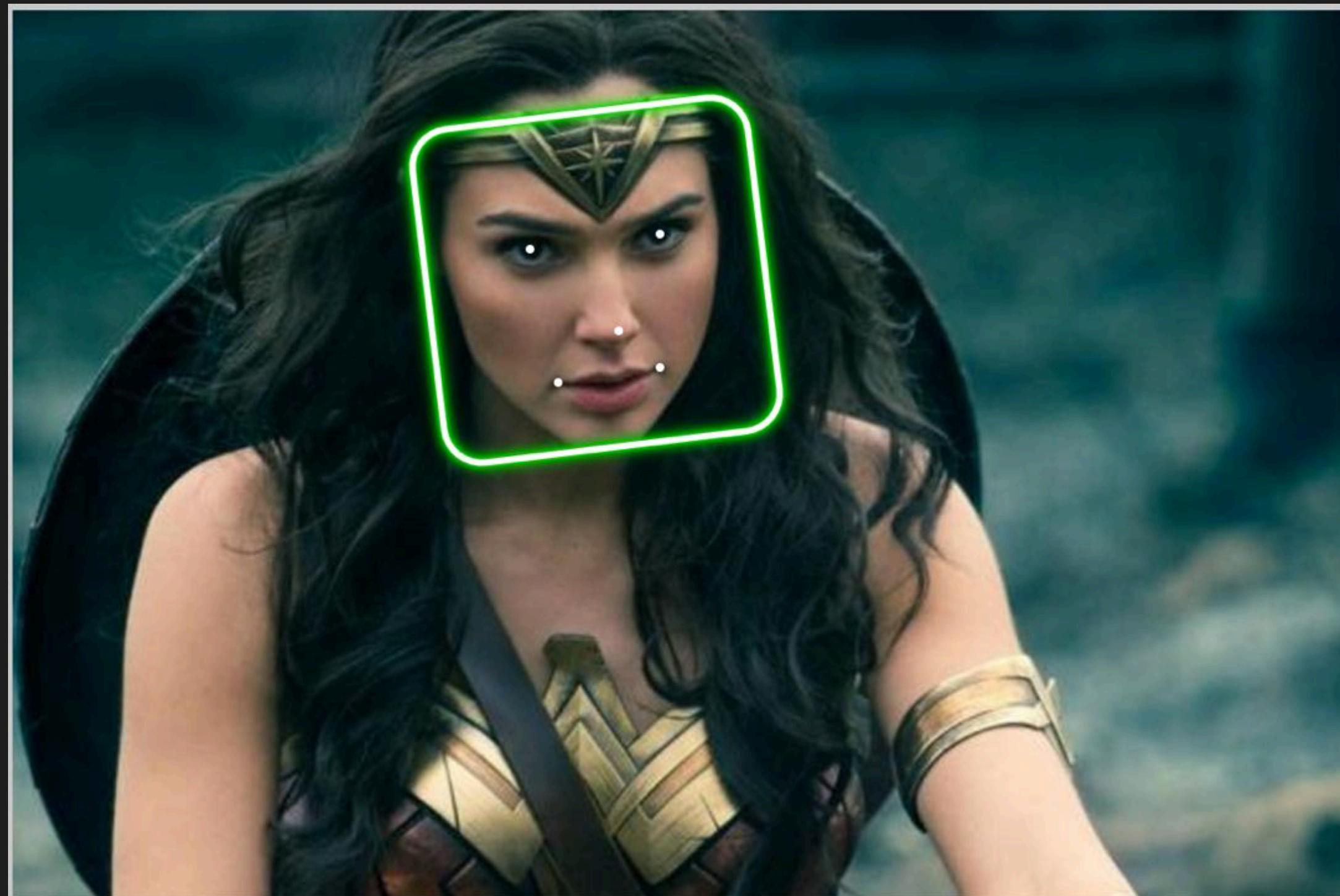


Object & Scene Detection



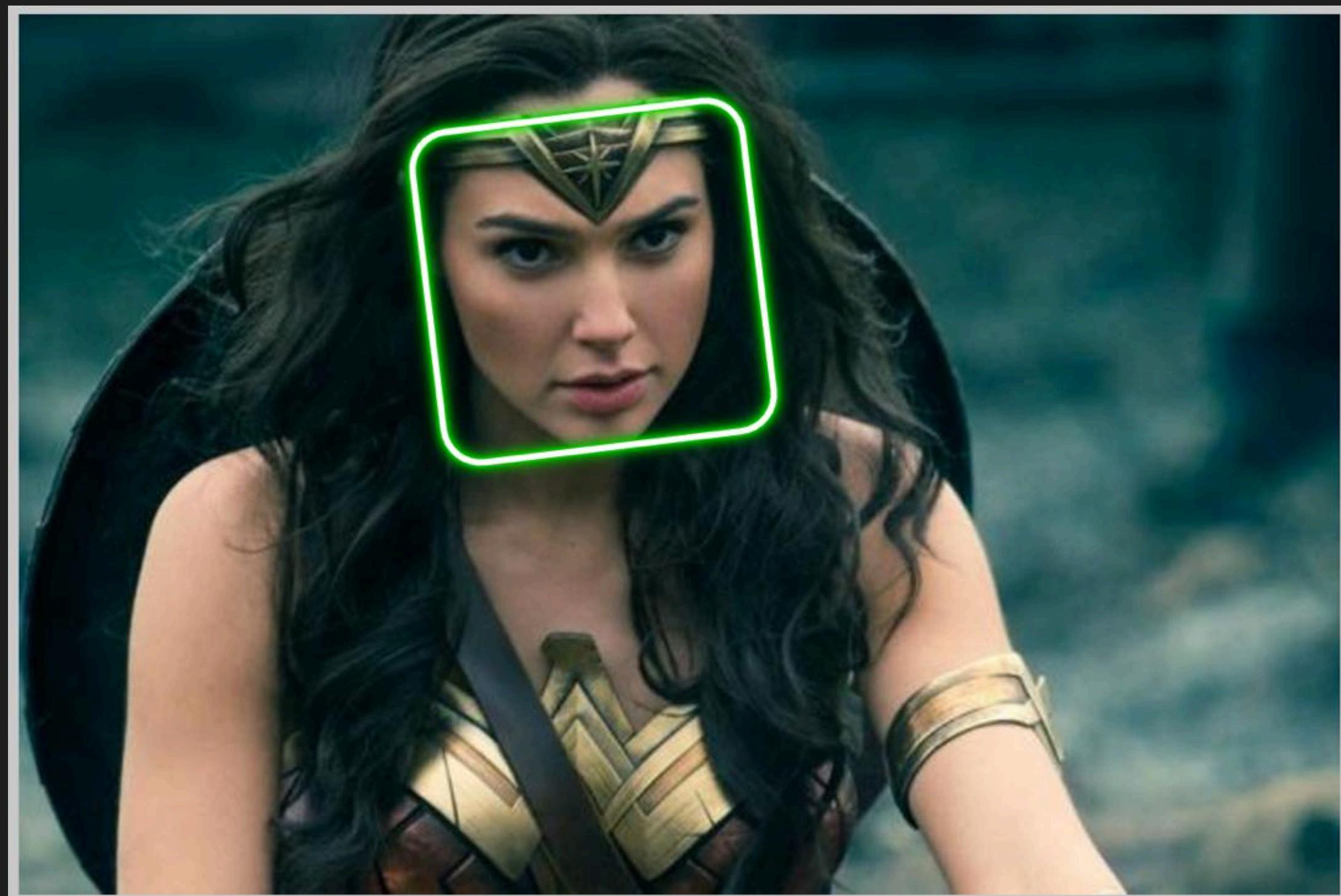
City	91.4%
Downtown	91.4%
Metropolis	91.4%
Urban	91.4%
Building	56.2%
High Rise	56.2%
Dock	55.6%
Pier	55.6%
Dawn	52.4%
Dusk	52.4%
Sky	52.4%
Sunrise	52.4%
Sunset	52.4%
Housing	51.4%
Coast	51.1%
Outdoors	51.1%

Facial Analysis



▼ Results	
	
looks like a face	99.9%
appears to be female	100%
age range	27 - 44 years old
not smiling	86.4%
not wearing glasses	78.8%
not wearing sunglasses	98.3%
eyes are open	99.8%
mouth is closed	99.7%
does not have a mustache	99.8%
does not have a beard	99.9%

Celebrity Recognition



▼ Results

 Gal Gadot [Learn More](#)

Match confidence 100%

A screenshot of a celebrity recognition results page. It shows a small portrait of Gal Gadot and her name, "Gal Gadot". Below the name is a link "Learn More". At the bottom of the card, it says "Match confidence" followed by "100%".



Facial Search

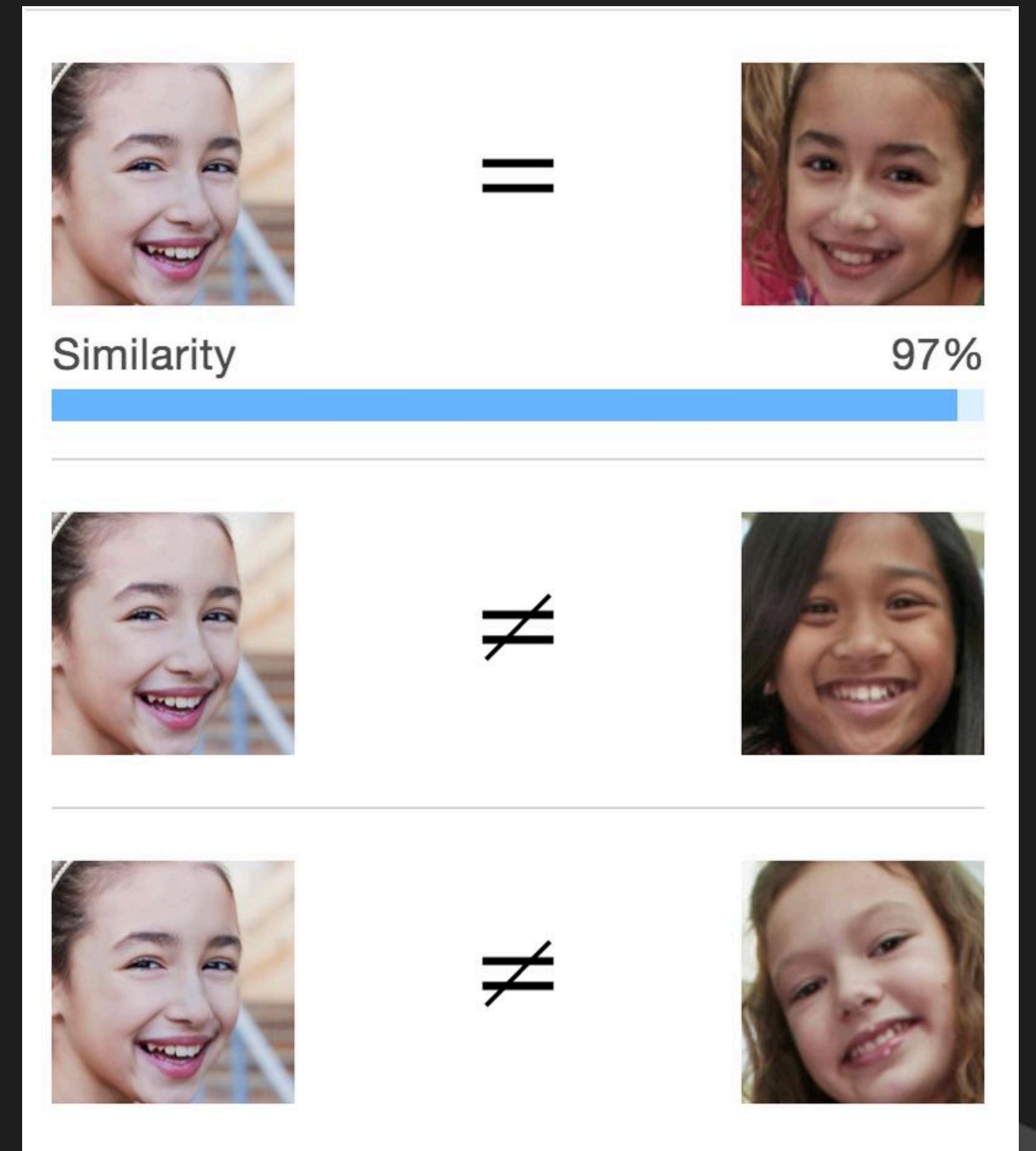
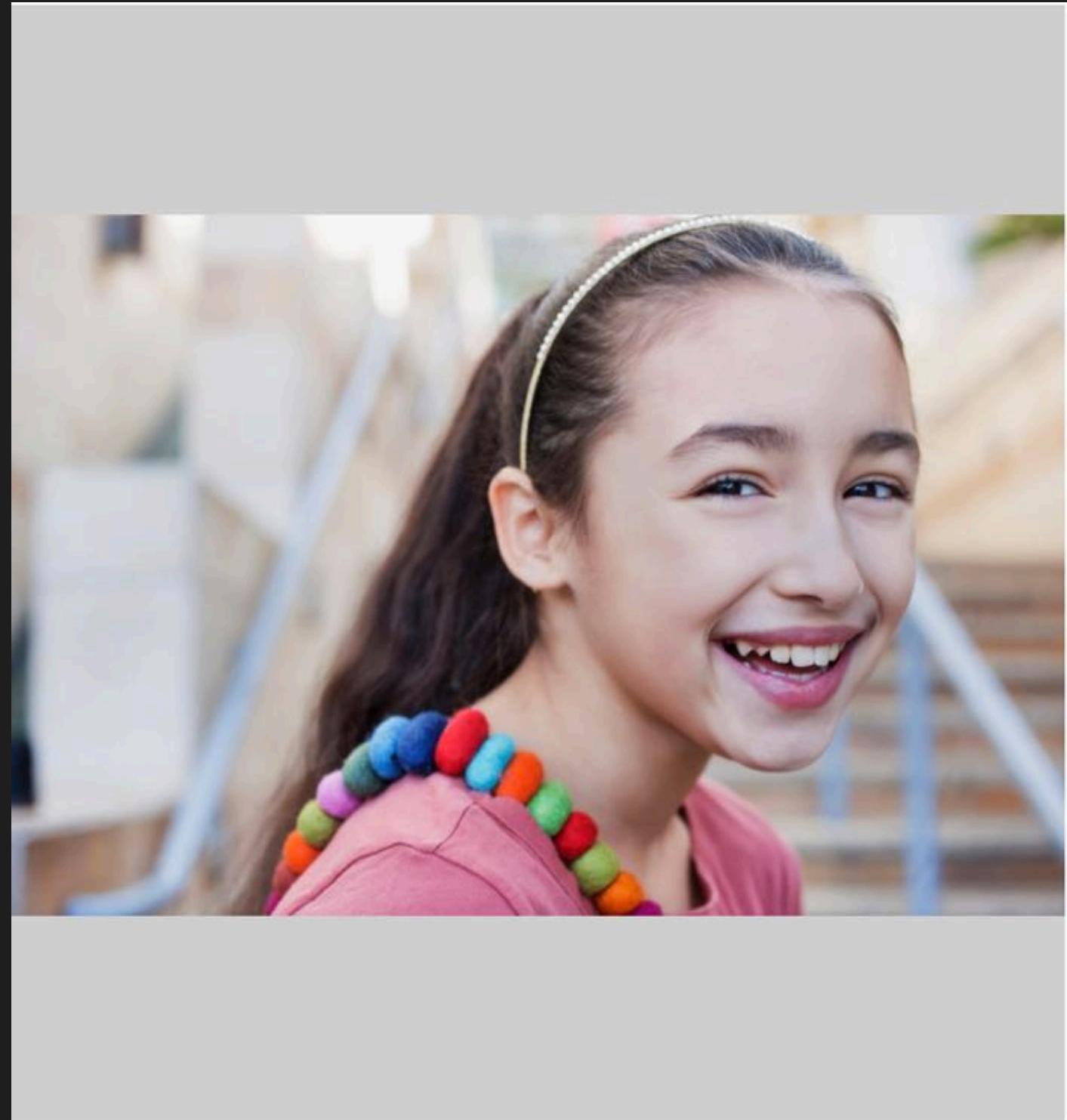


Image moderation



▼ 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

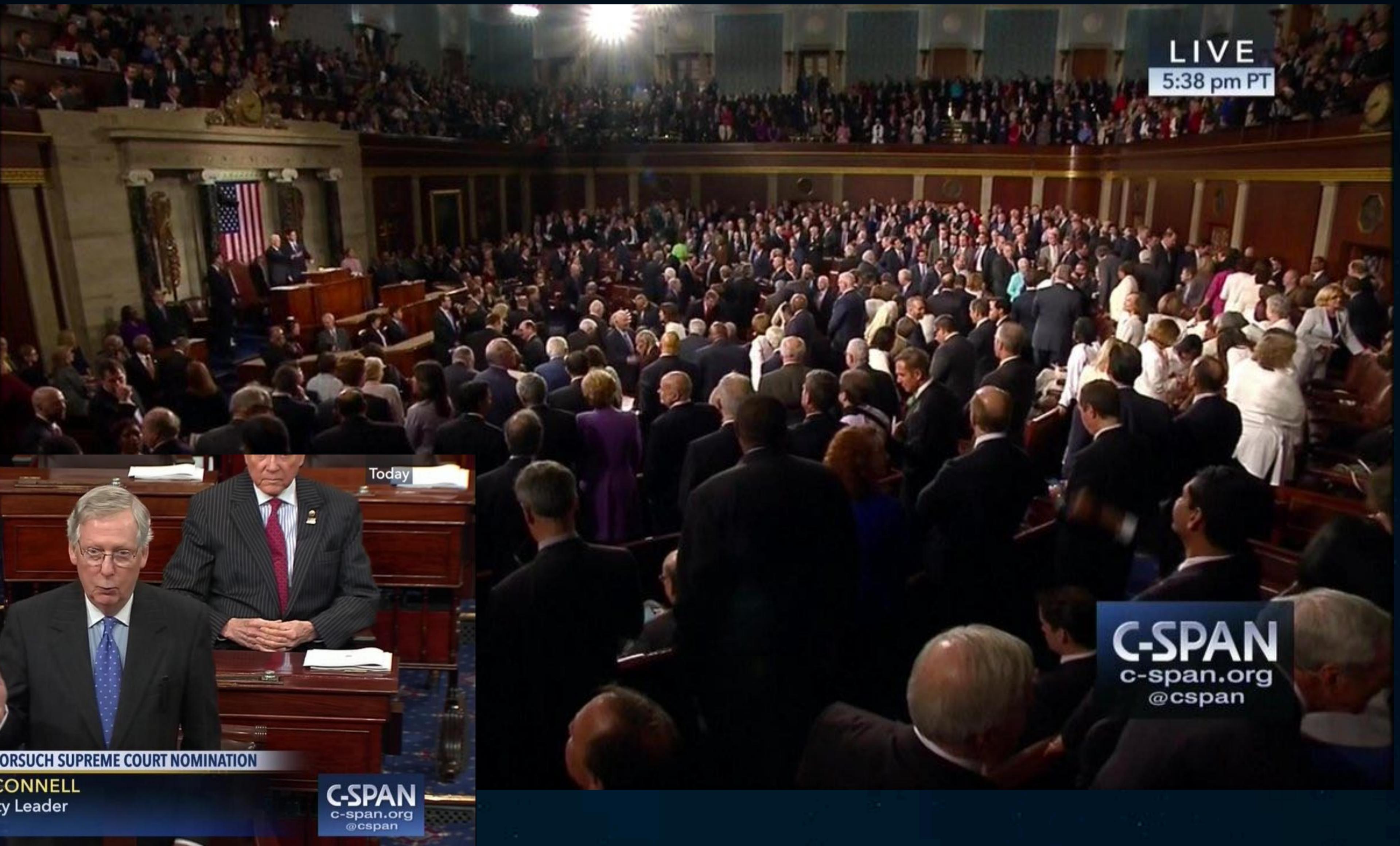
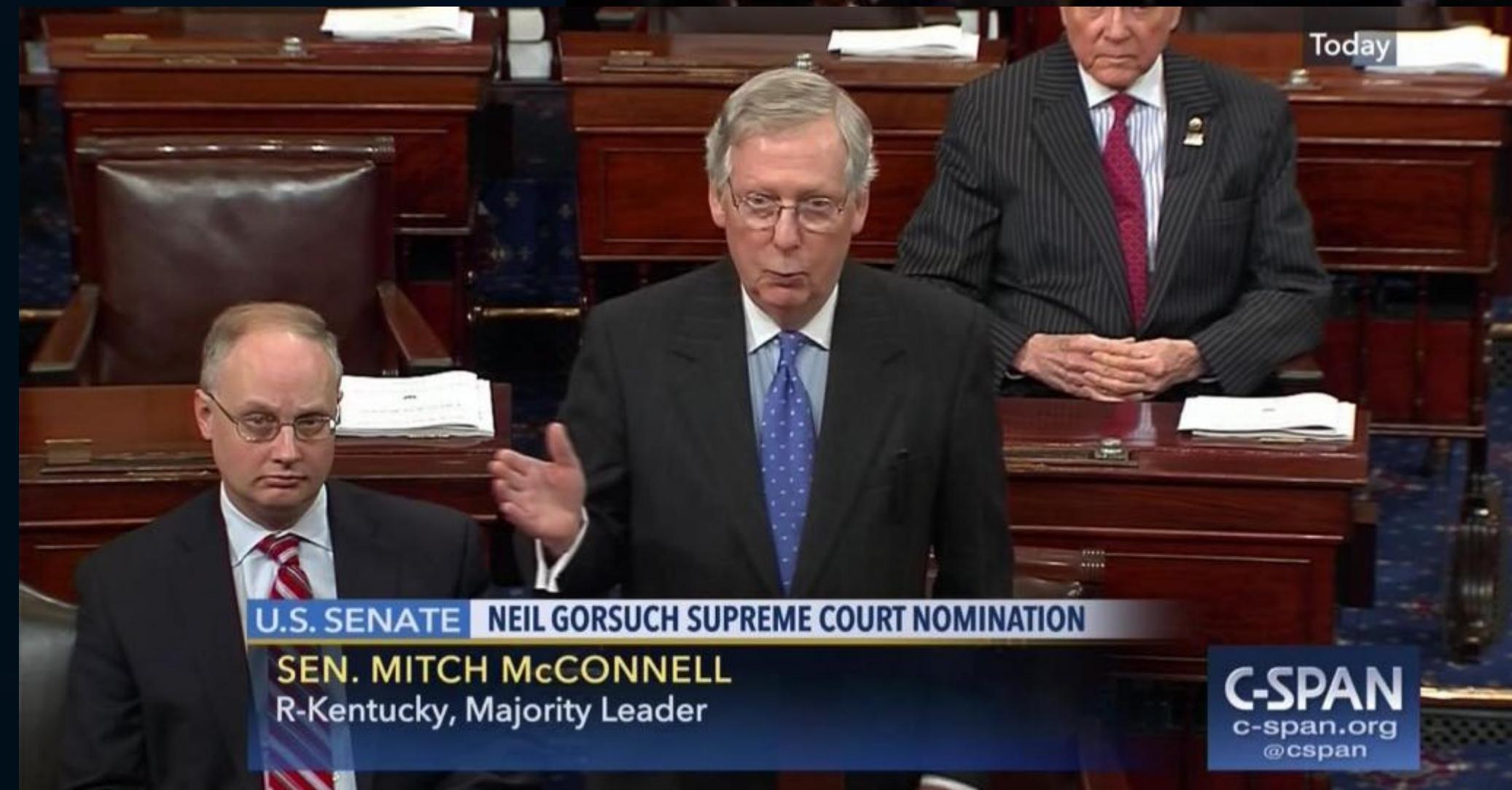


Amazon Rekognition Customers

- Digital Asset Management
- Media and Entertainment
- Travel and Hospitality
- Influencer Marketing
- Systems Integration
- Digital Advertising
- Consumer Storage
- Law Enforcement
- Public Safety
- eCommerce
- Education



C-SPAN





Artificial Intelligence Services on AWS



Polly

Text-to-Speech



Rekognition

Image Analysis

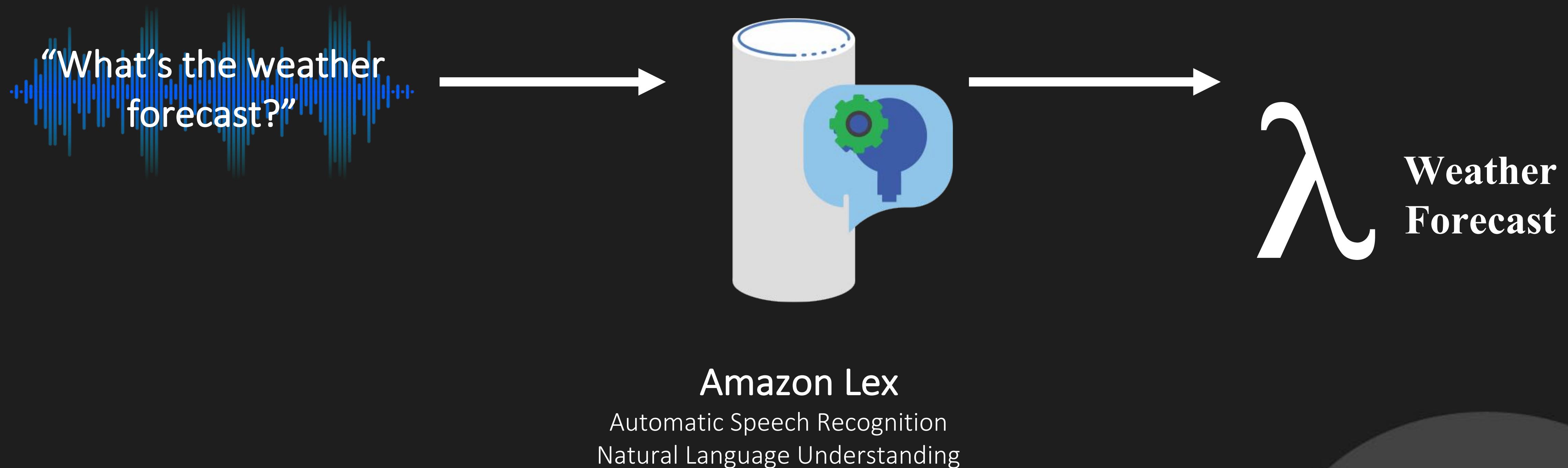


Lex

Conversation Bots

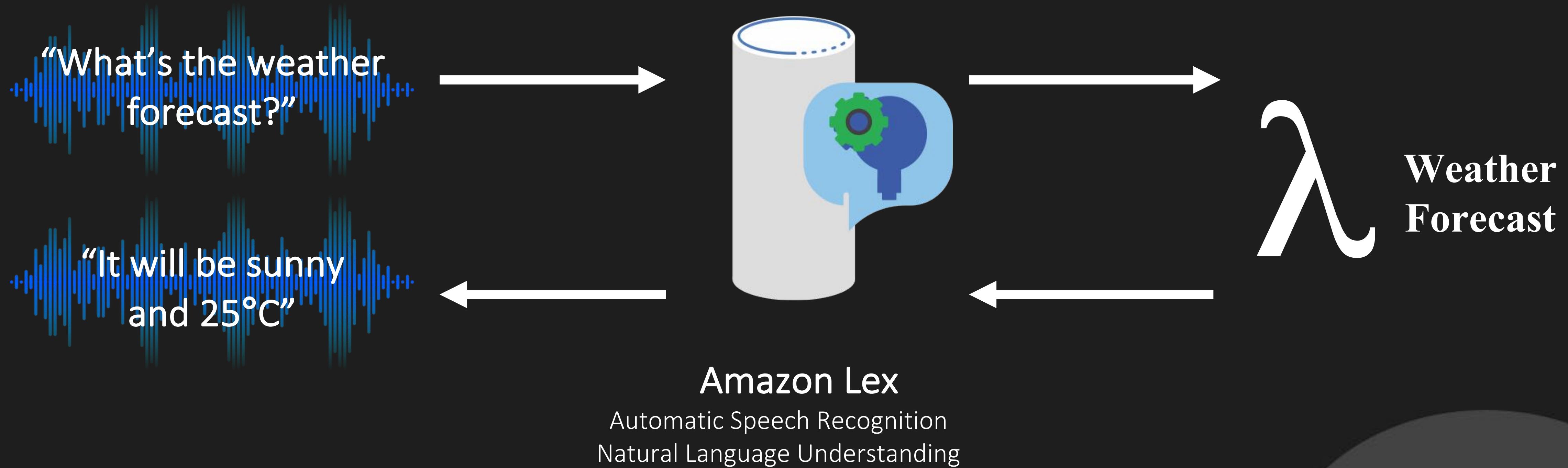


Speech Recognition & Natural Language Understanding



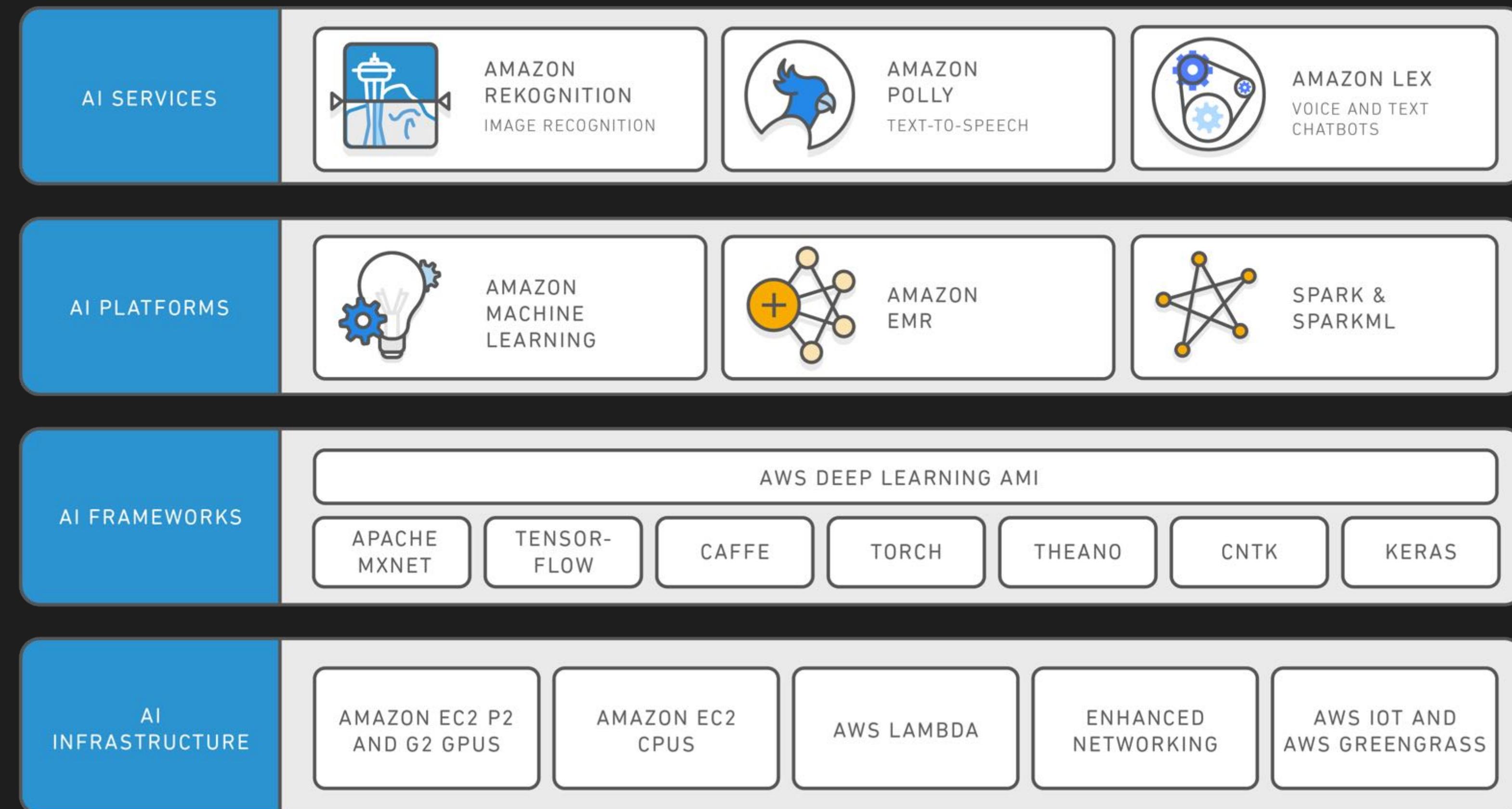


Speech Recognition & Natural Language Understanding





Wrap up



Follow us for more
events



Thank You!

Boaz Ziniman - Technical Evangelist, AWS

 @ziniman

 @AWScloud for Global AWS News & Announcements

Local Events: <https://aws.amazon.com/events/aws-israel/>