



# Amazon AI for every developer

Julien Simon

AI Evangelist, EMEA

@julsimon

# Artificial Intelligence At Amazon

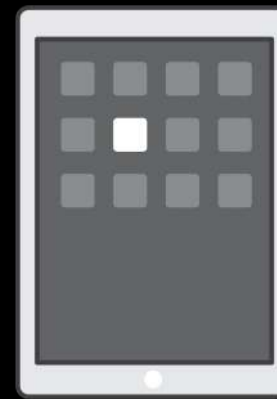
Thousands Of Employees Across The Company Focused  
on AI



Discovery  
&  
Search



Fulfilment  
&  
Logistics



Enhance  
Existing  
Products



Define New  
Categories  
Of  
Products



Bring  
Machine  
Learning To  
All



# Amazon AI for every developer

Services	<div><div>Chat</div><div>Amazon Lex</div></div> <div><div>Speech</div><div>Amazon Polly</div></div> <div><div>Vision</div><div>Amazon Rekognition</div></div>					
Platforms	Amazon ML	Spark & EMR	Kinesis	Batch	ECS	
Engines	MXNet	TensorFlow	Caffe	Theano	Pytorch	CNTK
Infrastructure	GPU	CPU	IoT	Mobile		

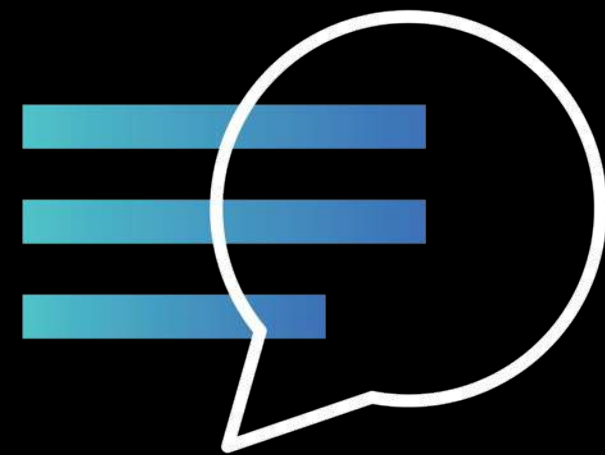
# Amazon AI: Three New Deep Learning Services



**Polly**

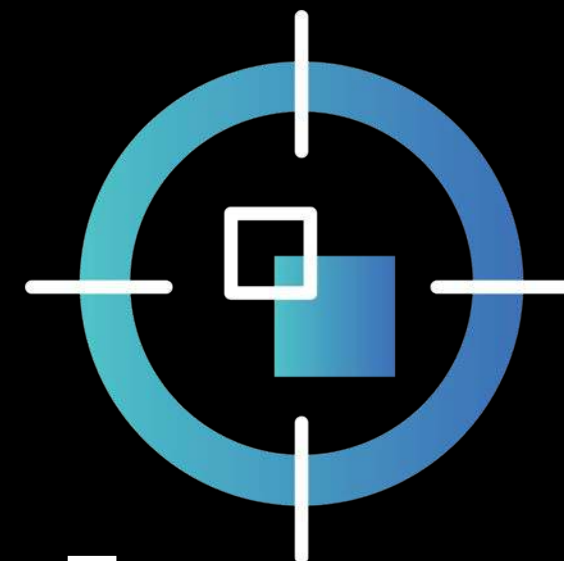
Life-Like  
Speech

# Amazon AI: Three New Deep Learning Services



**Polly**

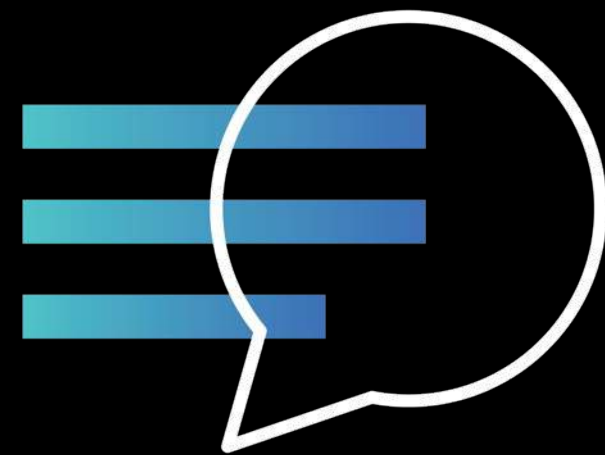
Life-Like  
Speech



**Rekognition**

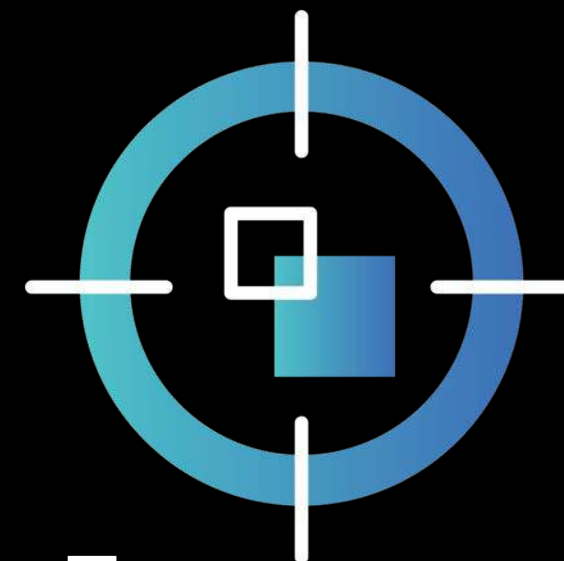
Image  
Analysis

# Amazon AI: Three New Deep Learning Services



**Polly**

Life-Like  
Speech



**Rekognition**

Image  
Analysis



**Lex**

Conversational  
AI  
Engine

# Amazon AI: Three New Deep Learning Services



**Polly**

Life-Like  
Speech



**Rekognition**

Image  
Analysis



**Lex**

Conversational  
AI  
Engine

# Polly: Life-like Speech Service



Converts text  
to life-like speech



Fully  
managed



50  
voices



24  
languages



Low  
latency,  
real time



# **Polly:** 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.’

# **Polly:** A Focus On Voice Quality & Pronunciation

1. Automatic, Accurate Text  
Processing



**2. Intelligible and Easy to  
Understand**

# **Polly:** 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**

# **Polly:** 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.”

# Polly: Life-like Speech Service



High quality,  
through  
best-in-class  
deep  
learning



Deep  
functionalit  
y



Easy to use  
& thoughtfully  
integrated



Built for  
productio  
n



Lo  
w  
cos  
t

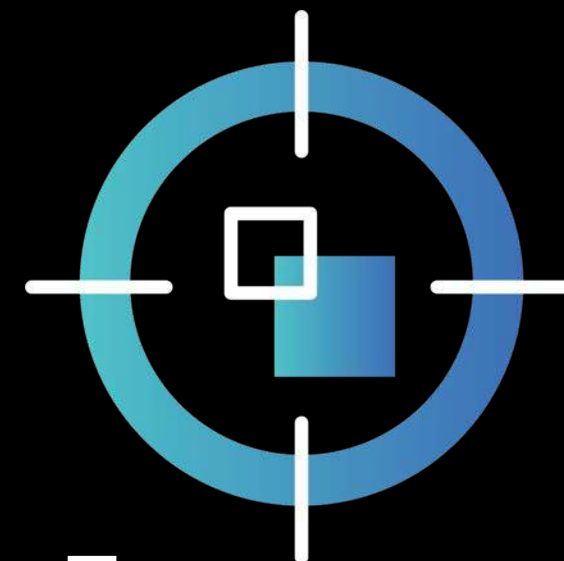


# Amazon AI: Three New Deep Learning Services



**Polly**

Life-Like  
Speech



**Rekognition**

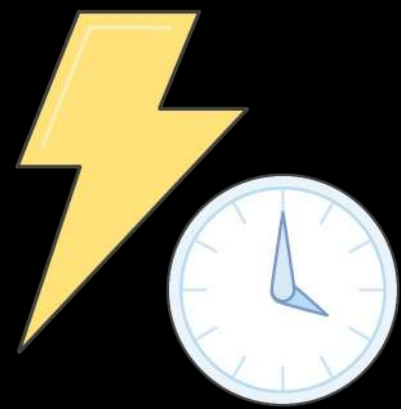
Image  
Analysis



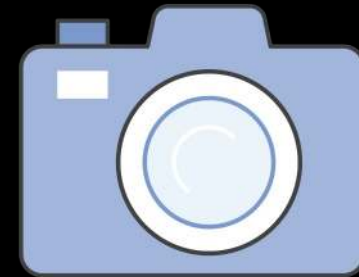
**Lex**

Conversational  
AI  
Engine

# Rekognition: Search & Understand Visual Content



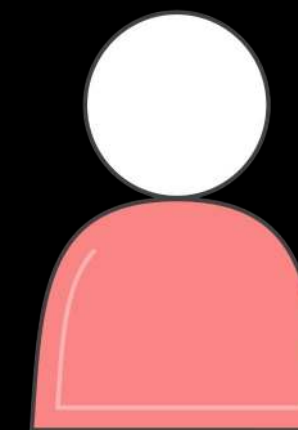
Real-time &  
batch image  
analysis



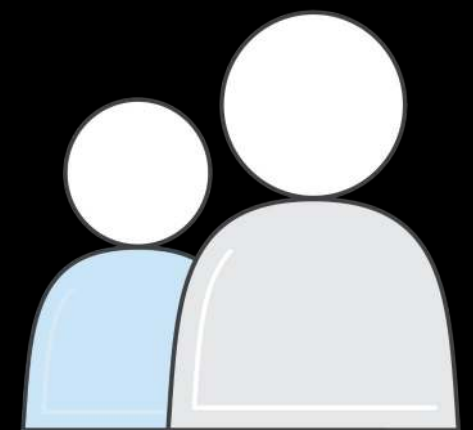
Object &  
Scene  
Detection



Facial  
Detection



Facial  
Analysis



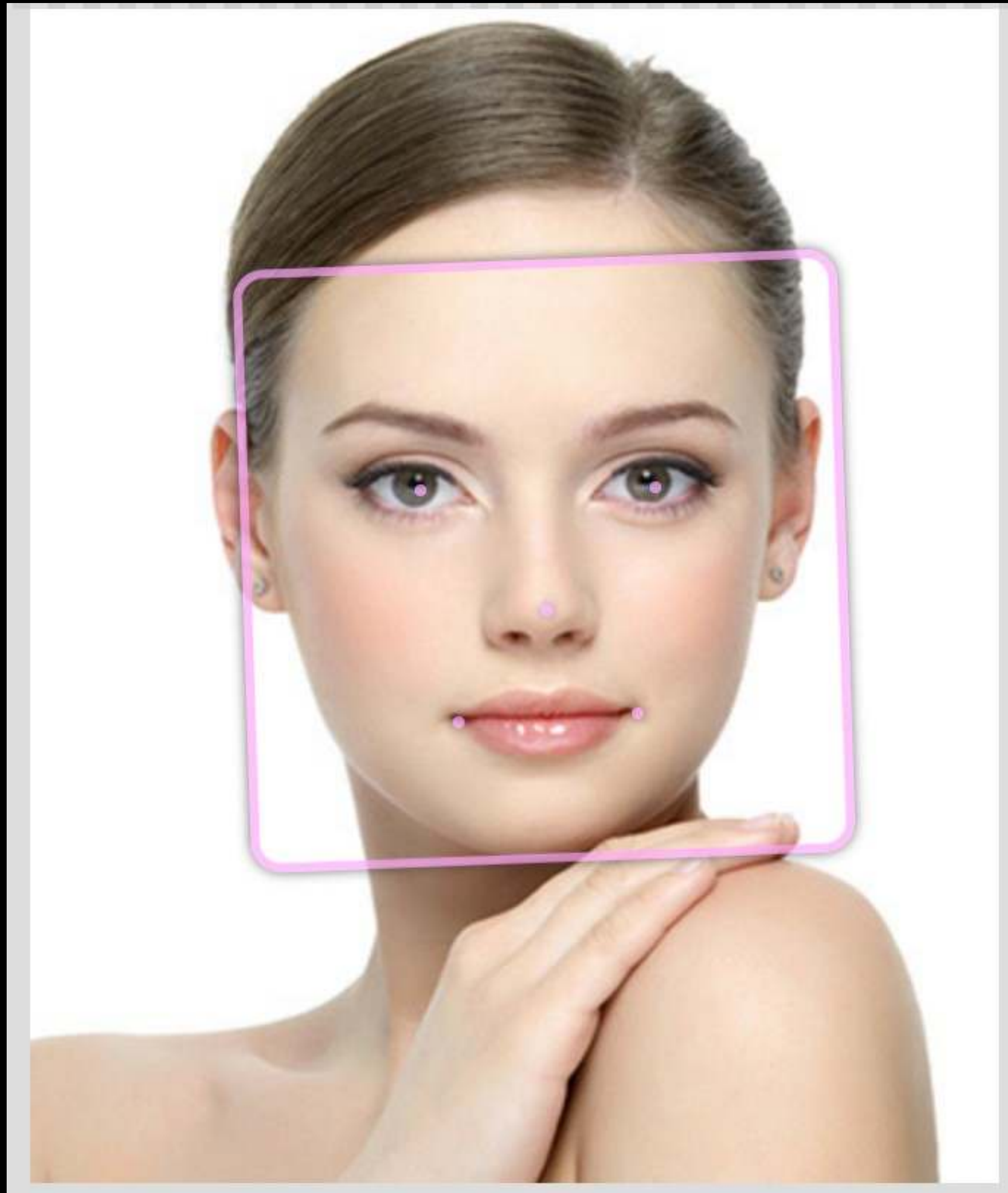
Face  
Search

# Rekognition: Object & Scene Detection

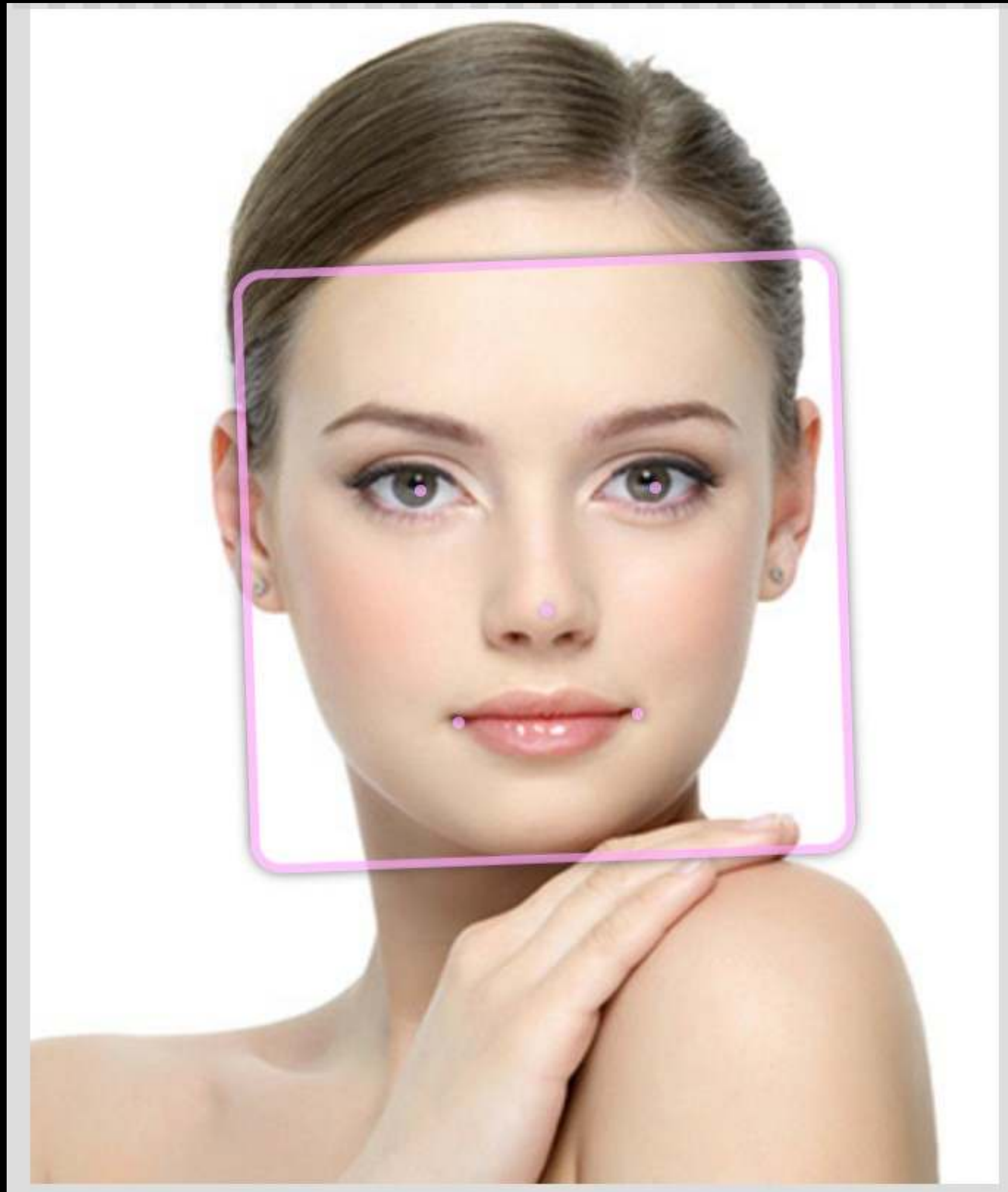


Category	Confidence
Bay	99.18%
Beach	99.18%
Coast	99.18%
Outdoors	99.18%
Sea	99.18%
Water	99.18%
Palm_tree	99.21%
Plant	99.21%
Tree	99.21%
Summer	58.3%
Landscape	51.84%
Nature	51.84%
Hotel	51.24%

# Rekognition: Facial Detection



# Rekognition: Facial Analysis



**Emotion:** calm: 73%

**Sunglasses:** false (value: 0)

**Gender:** female (value: 0)

**Mouth open wide:** 0% (value: 0)

**Eye closed:** open (value: 0)

**Glasses:** no glass (value: 0)

**Mustache:** false (value: 0)

**Beard:** no (value: 0)



# Rekognition: Compare Faces

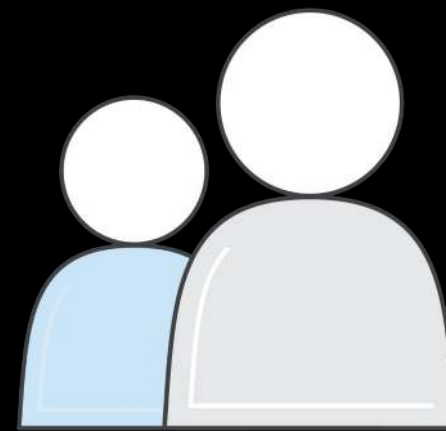


Similarity: 97.0%

# Rekognition: Facial Search



Facial  
verification  
n  
(compare two faces)



Face  
Search  
h  
(compare many faces)



Visual  
Similarity  
Search  
(find similar faces)

# Rekognition: Search & Understand Visual Content



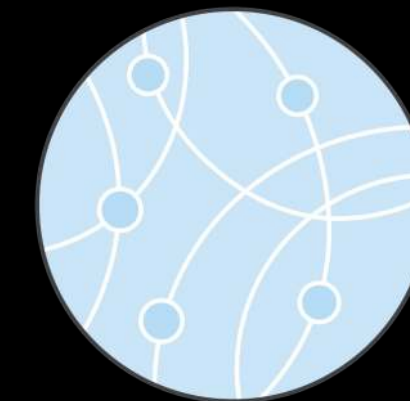
High quality,  
through  
best-in-class  
deep  
learning



Deep  
functionalit  
y



Easy to use  
& thoughtfully  
integrated



Built for  
productio  
n



Lo  
w  
cos  
t

# Amazon AI: Three New Deep Learning Services



**Polly**

Life-Like  
Speech



**Rekognition**

Image  
Analysis



**Lex**

Conversational  
Engine

# Lex: Build Natural, Conversational Interactions In Voice & Text



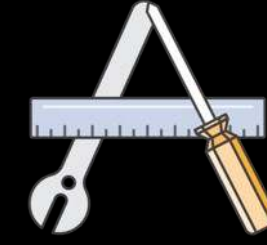
Voice & Text  
“Chatbots”



Power  
s  
Alexa



Voice  
interactions  
on mobile, web  
& devices



Text interaction  
with Slack &  
Messenger  
(with more  
coming)



Enterprise  
Connector

S  
Salesforce  
Microsoft  
Dynamics  
Marketo  
Zendesk  
Quickbooks  
Hubspot



✈️ Flight Booking

Origin

Destination

Departure Date



“Book a flight to  
London”

✈ Flight Booking	
Origin	
Destination	
Departure Date	



“Book a flight to  
London”

Automatic  
Speech Recognition

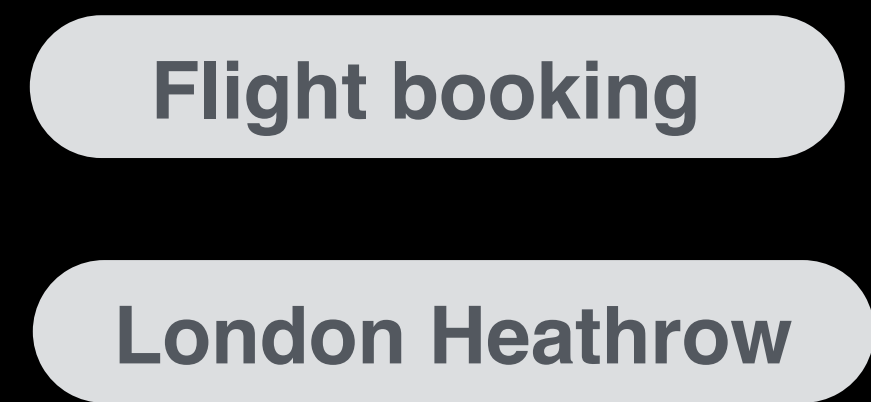
✈️ Flight Booking	
Origin	
Destination	
Departure Date	



“Book a flight to  
London”



Automatic  
Speech Recognition



Natural Language  
Understanding



Knowledge  
Graph



Grammar  
Graph



Utterances



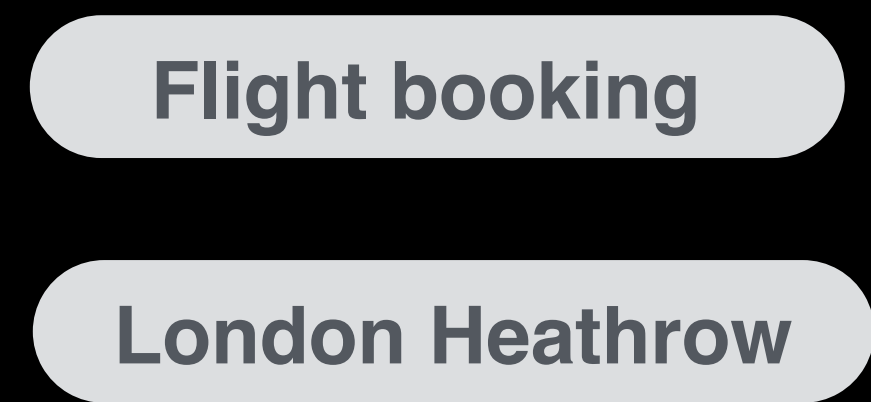
✈️ Flight Booking	
Origin	
Destination	
Departure Date	



“Book a flight to  
London”



Automatic  
Speech Recognition



Natural Language  
Understanding



Knowledge  
Graph



Grammar  
Graph

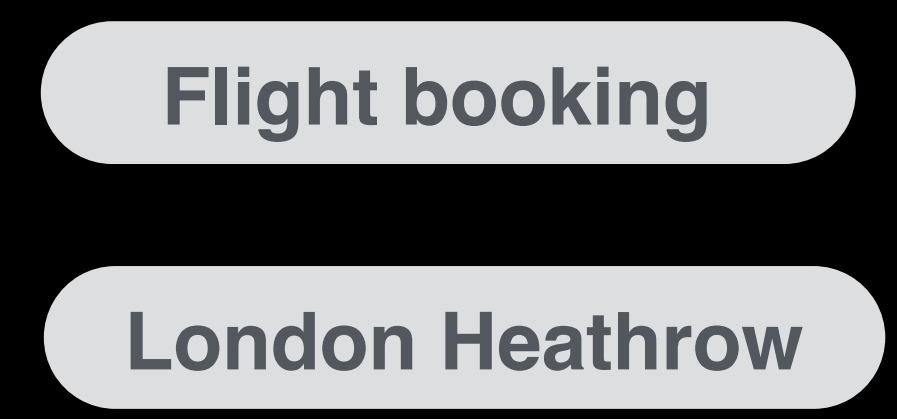
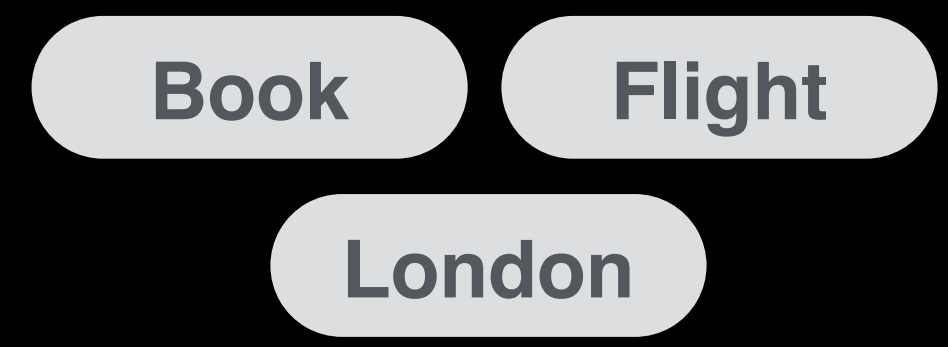


Utterances



✈️ Flight Booking	
Origin	
Destination	London Heathrow
Departure Date	





Knowledge  
Graph

“Book a flight to  
London”

Automatic  
Speech Recognition

Natural Language  
Understanding



Grammar  
Graph

Location

Location



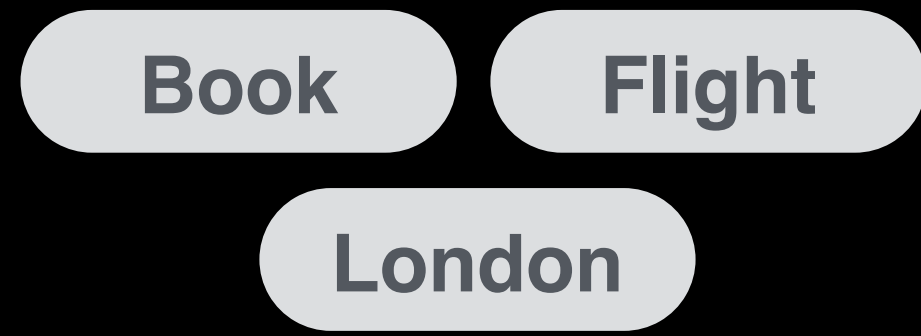
Utterances  
↑

✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	



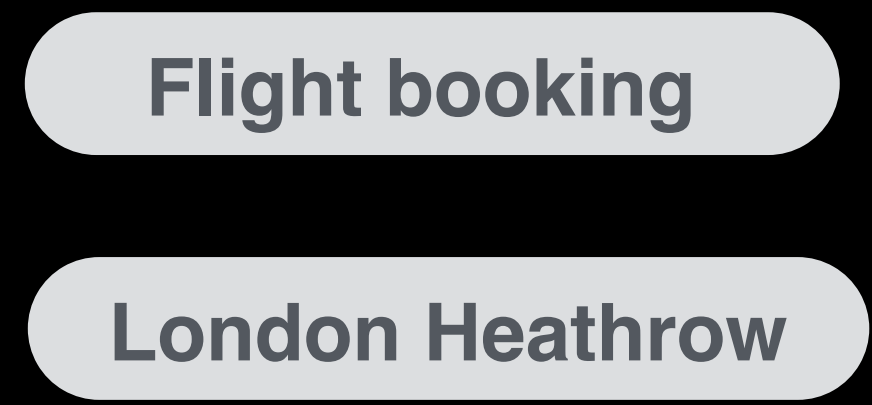
“Book a flight to  
London”

Location



Automatic  
Speech Recognition

Location



Natural Language  
Understanding



Knowledge  
Graph



Grammar  
Graph



Utterances

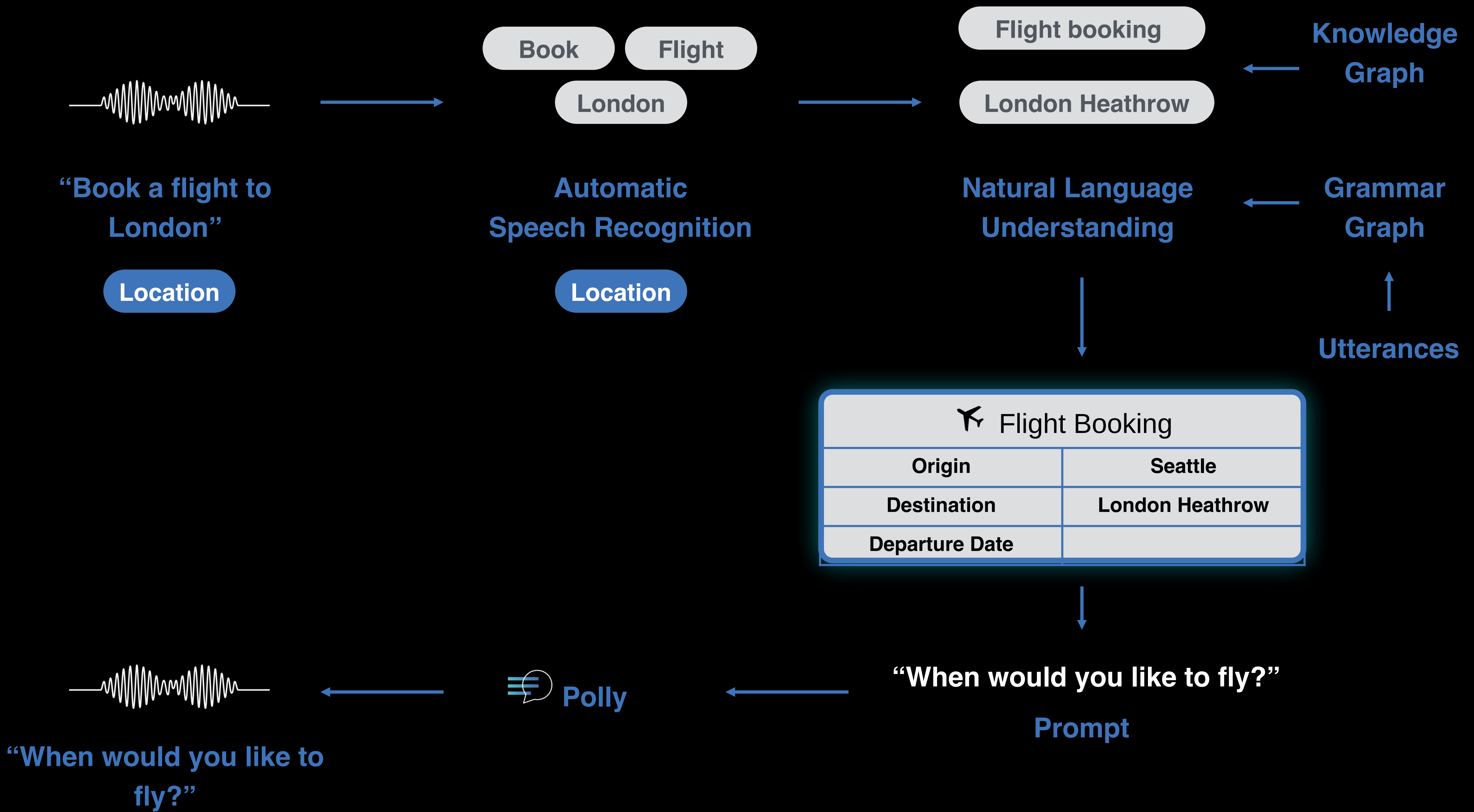


✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	



“When would you like to fly?”

Prompt





“Next Friday”



“When would you like to  
fly?”

✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	



Next Friday

“Next Friday”

Automatic  
Speech Recognition

✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	





“Next Friday”



Next

Friday

Automatic  
Speech Recognition



Flight booking

11/18/2016

Natural Language  
Understanding



✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	11/18/2016



Knowledge  
Graph



Grammar  
Graph



Utterances



“Next Friday”



Next

Friday

Automatic  
Speech Recognition



Flight booking

11/18/2016

Natural Language  
Understanding



✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	11/18/2016



$\lambda$



Knowledge  
Graph



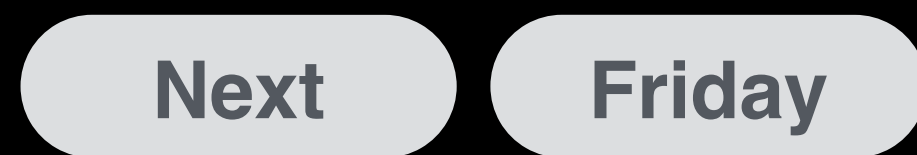
Grammar  
Graph



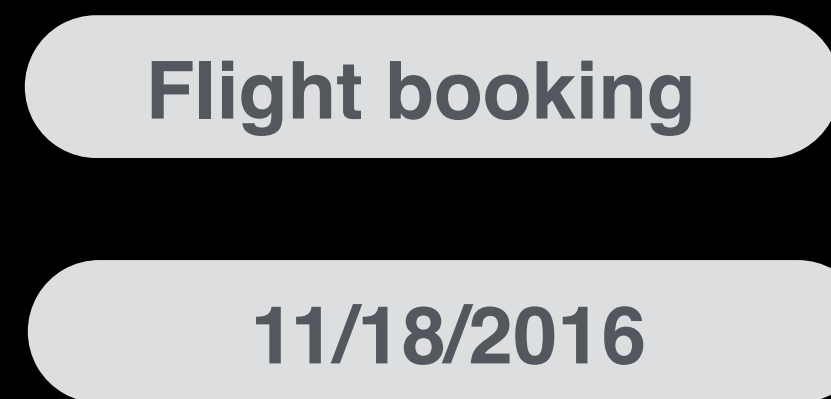
Utterances



“Next Friday”



Automatic  
Speech Recognition



Natural Language  
Understanding



Knowledge  
Graph



Grammar  
Graph



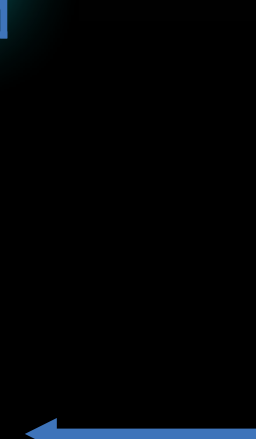
✈ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	11/18/2016

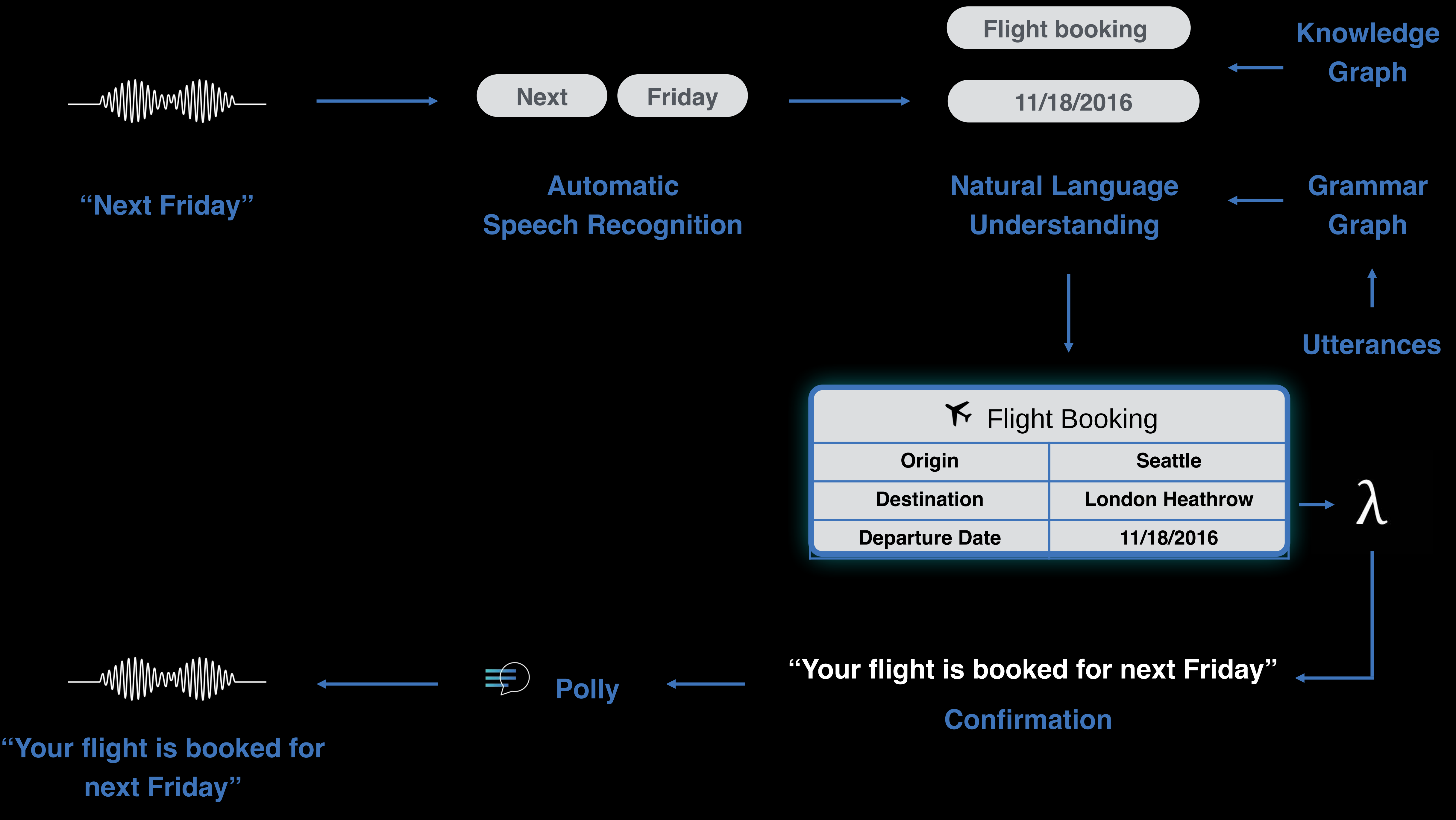
Utterances  
↑

→  $\lambda$

“Your flight is booked for next Friday”

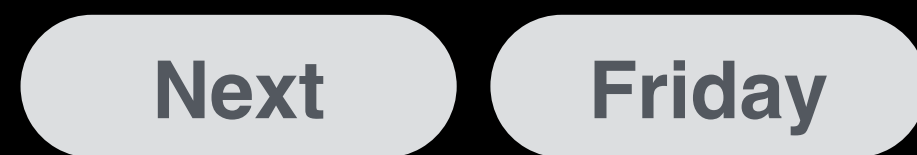
Confirmation



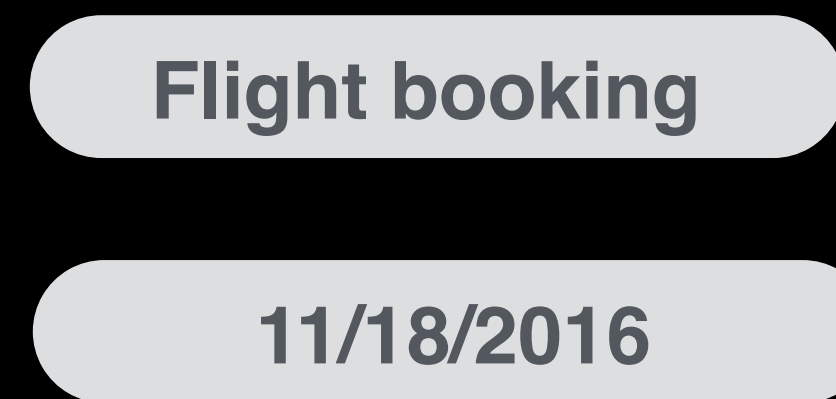




“Next Friday”



Automatic  
Speech Recognition



Natural Language  
Understanding



Knowledge  
Graph



Grammar  
Graph



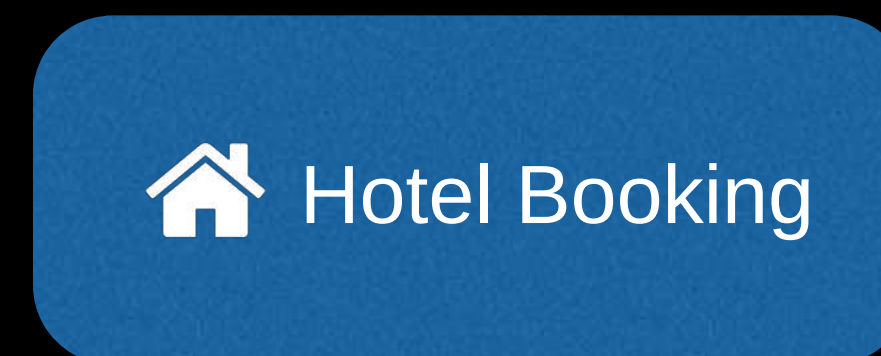
Utterances



✈️ Flight Booking	
Origin	Seattle
Destination	London Heathrow
Departure Date	11/18/2016



$\lambda$



# Lex: Build Natural, Conversational Interactions In Voice & Text



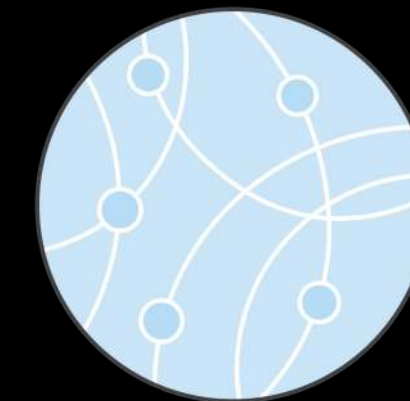
High quality,  
through  
best-in-class  
deep  
learning



Deep  
functionalit  
y



Easy to use  
& thoughtfully  
integrated

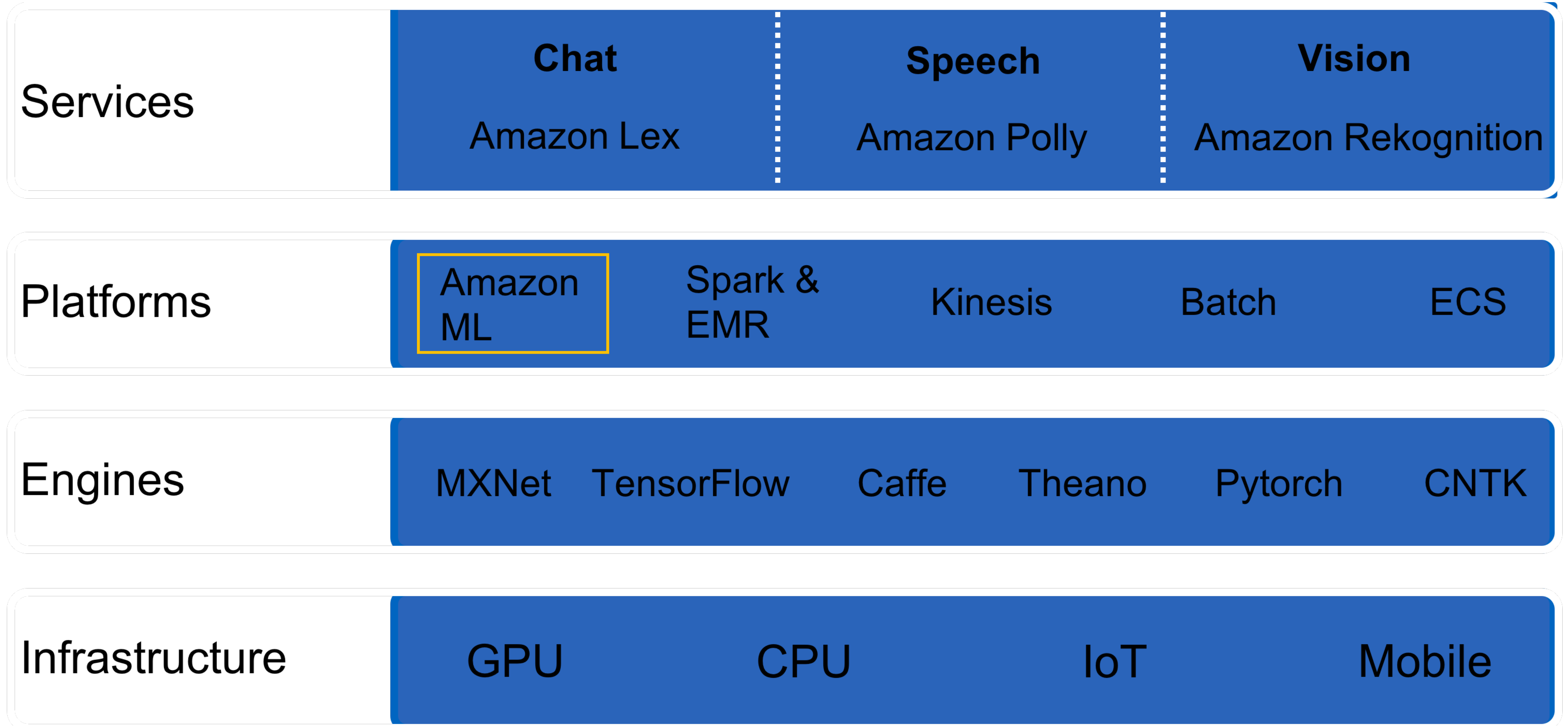


Built for  
productio  
n



Lo  
w  
cos  
t

# Amazon AI for every developer





# Amazon Machine Learning

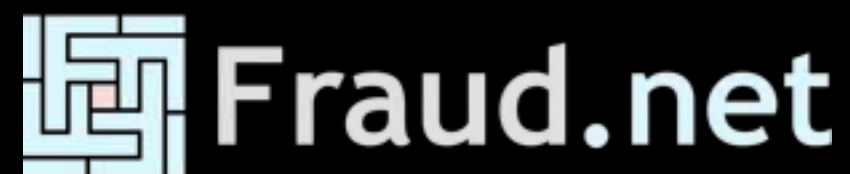
- **Easy-to-use, managed** machine learning service built for developers
- Robust, powerful machine learning technology based on **Amazon's** internal systems
- Create **prediction** and **classification** models using your data already stored in the AWS Cloud
- Deploy models to production in **seconds**

# Fraud.net Uses AWS to Quickly, Easily Detect Online Fraud



Amazon Machine Learning helps us **reduce complexity** and make sense of emerging fraud patterns.

**Oliver  
Clark**  
CTO,  
Fraud.net



Fraud.net is the world's leading crowdsourced fraud prevention platform.



- Needed to build and train a **larger number** of **more targeted** machine-learning models
- Uses Amazon Machine Learning to provide more than **20** models
- Easily builds and trains models to effectively detect online payment fraud
- Reduces complexity and makes sense of **emerging** fraud patterns
- Saves clients **\$1 million** weekly by helping them detect and prevent fraud

# Upserve Uses AWS to Help Restaurants Predict Business

“

Using Amazon Machine Learning, we can **predict** the total number of **customers** who will walk through a restaurant's doors in a night.

**Bright Fulton**

Director of Infrastructure  
Engineering,

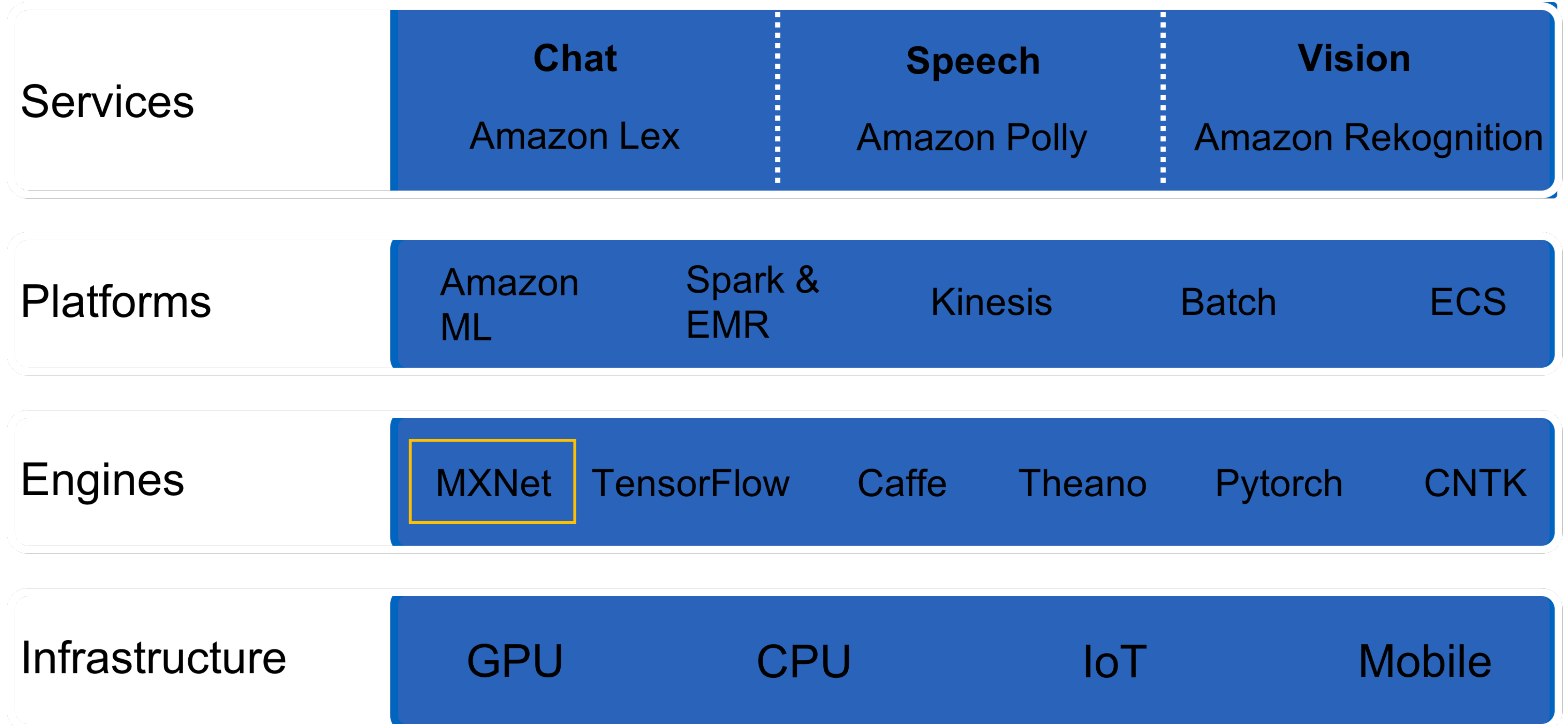


Upserve provides online payment and analytical software to thousands of restaurant owners throughout the U.S.

”

- Needed its restaurant management platform to provide more **predictive analytics**
- Builds and trains more than **100 machine learning models weekly**
- Streams restaurant sales and menu item data in **real time**
- Helps restaurateurs **predict** nightly business

# Amazon AI for every developer



# Apache MXNet: Open Source library for Deep Learning



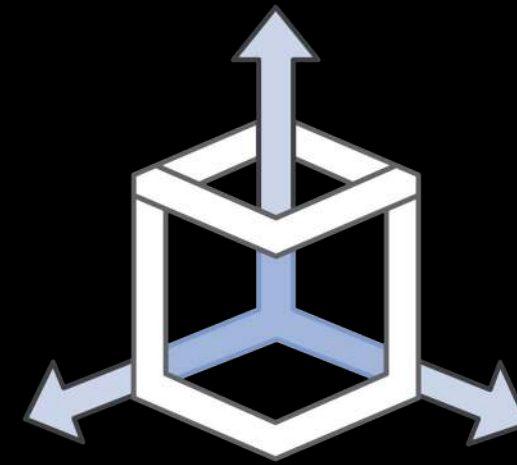
## Programmable

Simple syntax,  
multiple languages



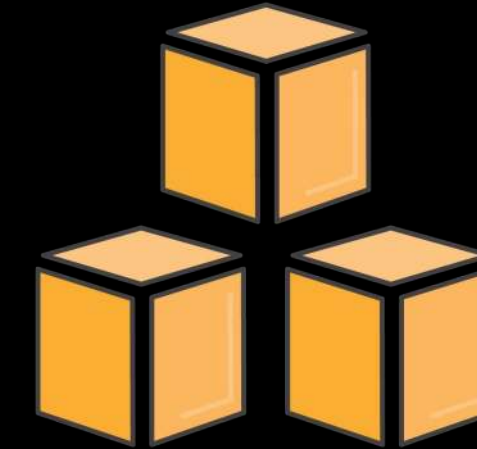
## Most Open

Accepted into the  
Apache Incubator



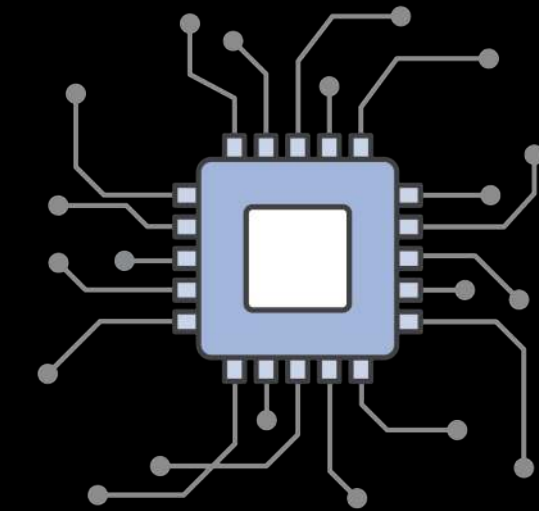
## Portable

Highly efficient  
models for mobile  
and IoT



## Best On AWS

Optimized for  
Deep Learning on AWS



## High Performance

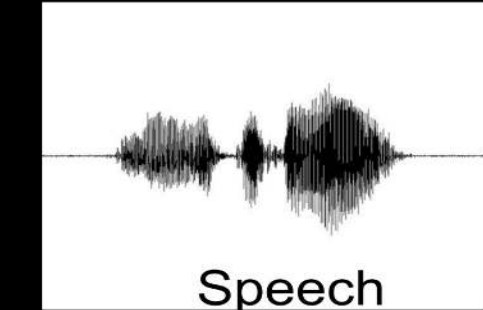
Near linear scaling  
across hundreds of GPUs

More information at  
[mxnet.io](https://mxnet.io)



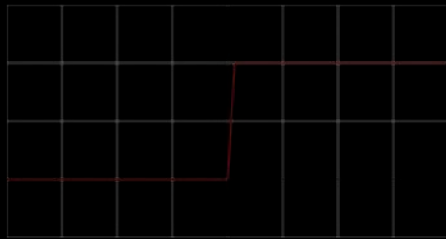
Input

Output

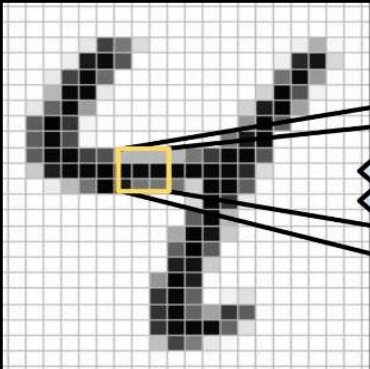


$\times =$

`mx.sym.Activation(data, act_type="xxxx")`



`mx.sym.FullyConnected(data, num_hidden=128)`

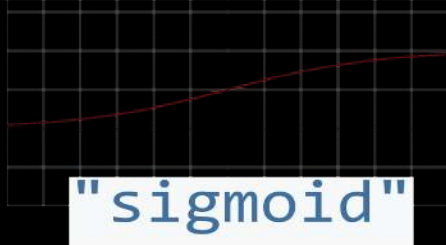


$\times =$

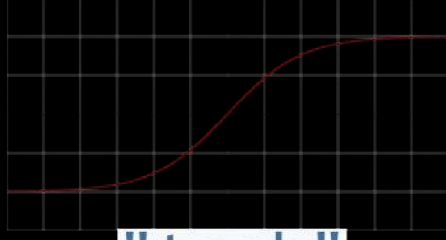
`mx.sym.Convolution(data, kernel=(5,5), num_filter=20)`



"sigmoid"



"tanh"



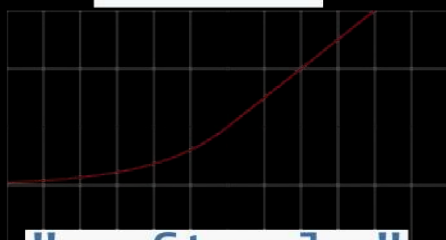
"relu"



`mx.sym.Pooling(data, pool_type="max", kernel=(2,2),`

`stride=(2,2)`

"softrelu"



$\oplus$   
 $\oplus$   
 $\otimes$

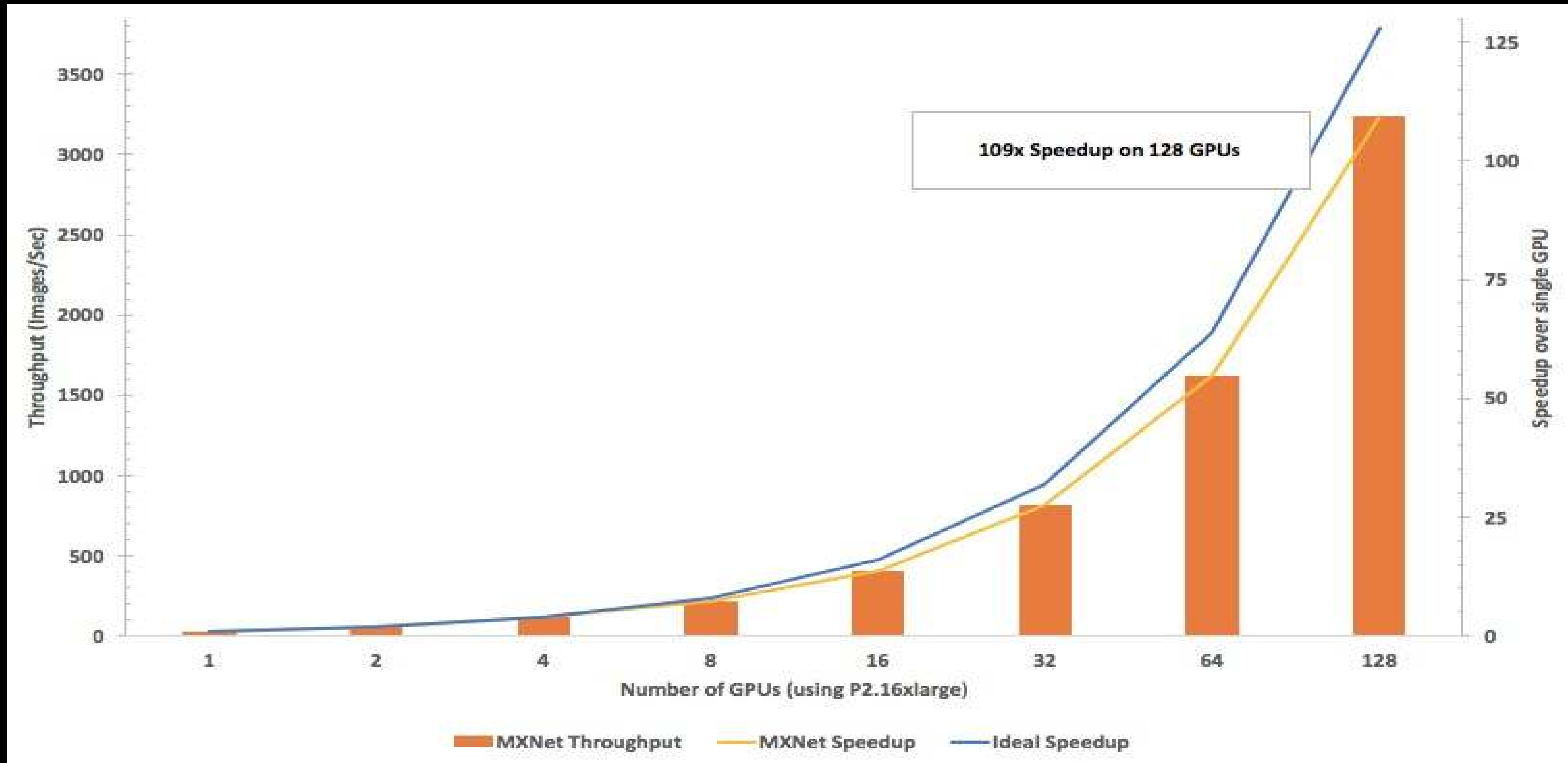
`lstm.lstm_unroll(num_lstm_layer, seq_len, len, num_hidden, num_embed)`



$$\cos(w, queen) = \cos(w, king) - \cos(w, man) + \cos(w, woman)$$

`mx.symbol.Embedding(data, input_dim, output_dim = k)`

# MXNet: Scalable Deep Learning Framework





## **AWS Deep Learning AMI**

Up to~40k CUDA cores

Apache MXNet

TensorFlow

Theano

Caffe

Torch

Pre-configured CUDA drivers

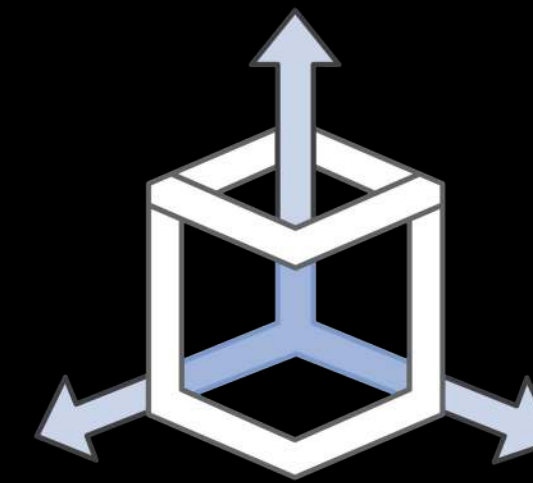
Anaconda, Python3

**+ CloudFormation**

**template**

**+ Container**

**Image**



# One-Click GPU Deep Learning

# Gluon: Deep Learning gets even easier

- Announced October 11<sup>th</sup>
- Available now in MXNet, soon in Microsoft Cognitive Toolkit
- Developer-friendly **high-level API**
- Dynamic networks can be **modified** during training
- No compromise on **performance**
- Extensive **model zoo**

<https://github.com/gluon-api/>

<https://aws.amazon.com/blogs/aws/introducing-gluon-a-new-library-for-machine-learning-from-aws-and-microsoft/>

# Gluon Model Zoo

vgg11	VGG-11 model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg13	VGG-13 model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg16	VGG-16 model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg19	VGG-19 model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg11_bn	VGG-11 model with batch normalization from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg13_bn	VGG-13 model with batch normalization from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg16_bn	VGG-16 model with batch normalization from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
vgg19_bn	VGG-19 model with batch normalization from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
VGG	VGG model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.
get_vgg	VGG model from the “ <a href="#">Very Deep Convolutional Networks for Large-Scale Image Recognition</a> ” paper.

resnet18_v1	ResNet-18 V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
resnet34_v1	ResNet-34 V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
resnet50_v1	ResNet-50 V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
resnet101_v1	ResNet-101 V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
resnet152_v1	ResNet-152 V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
resnet18_v2	ResNet-18 V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
resnet34_v2	ResNet-34 V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
resnet50_v2	ResNet-50 V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
resnet101_v2	ResNet-101 V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
resnet152_v2	ResNet-152 V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
ResNetV1	ResNet V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
ResNetV2	ResNet V2 model from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
BasicBlockV1	BasicBlock V1 from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
BasicBlockV2	BasicBlock V2 from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
BottleneckV1	Bottleneck V1 from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.
BottleneckV2	Bottleneck V2 from “ <a href="#">Identity Mappings in Deep Residual Networks</a> ” paper.
get_resnet	ResNet V1 model from “ <a href="#">Deep Residual Learning for Image Recognition</a> ” paper.

mobilenet1_0	MobileNet model from the “ <a href="#">MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications</a> ” paper, with width multiplier 1.0.
mobilenet0_75	MobileNet model from the “ <a href="#">MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications</a> ” paper, with width multiplier 0.75.
mobilenet0_5	MobileNet model from the “ <a href="#">MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications</a> ” paper, with width multiplier 0.5.
mobilenet0_25	MobileNet model from the “ <a href="#">MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications</a> ” paper, with width multiplier 0.25.
MobileNet	MobileNet model from the “ <a href="#">MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications</a> ” paper.

densenet121	Densenet-BC 121-layer model from the “ <a href="#">Densely Connected Convolutional Networks</a> ” paper.
densenet161	Densenet-BC 161-layer model from the “ <a href="#">Densely Connected Convolutional Networks</a> ” paper.
densenet169	Densenet-BC 169-layer model from the “ <a href="#">Densely Connected Convolutional Networks</a> ” paper.
densenet201	Densenet-BC 201-layer model from the “ <a href="#">Densely Connected Convolutional Networks</a> ” paper.
DenseNet	Densenet-BC model from the “ <a href="#">Densely Connected Convolutional Networks</a> ” paper.

inception_v3	Inception v3 model from “ <a href="#">Rethinking the Inception Architecture for Computer Vision</a> ” paper.
Inception3	Inception v3 model from “ <a href="#">Rethinking the Inception Architecture for Computer Vision</a> ” paper.

alexnet	AlexNet model from the “ <a href="#">One weird trick...</a> ” paper.
AlexNet	AlexNet model from the “ <a href="#">One weird trick...</a> ” paper.

squeezenet1_0	SqueezeNet 1.0 model from the “ <a href="#">SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and &lt;0.5MB model size</a> ” paper.
squeezenet1_1	SqueezeNet 1.1 model from the <a href="#">official SqueezeNet repo</a> .
SqueezeNet	SqueezeNet model from the “ <a href="#">SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and &lt;0.5MB model size</a> ” paper.

VGG  
ResNet  
AlexNet  
DenseNet  
SqueezeNet  
Inception  
MobileNet

[https://mxnet.incubator.apache.org/versions/master/api/python/gluon/model\\_zoo.html](https://mxnet.incubator.apache.org/versions/master/api/python/gluon/model_zoo.html)

# Amazon AI for every developer

Services	<div>Chat</div> <div>Amazon Lex</div>	<div>Speech</div> <div>Amazon Polly</div>	<div>Vision</div> <div>Amazon Rekognition</div>			
Platforms	Amazon ML	Spark & EMR	Kinesis	Batch	ECS	
Engines	MXNet	TensorFlow	Caffe	Theano	Pytorch	CNTK
Infrastructure	GPU	CPU	IoT	Mobile		

# **AWS is the Center of Gravity for Artificial Intelligence**

<http://aws.amazon.com/ai>  
<https://reinvent.awsevents.com/>





# Thank you!

Julien Simon

AI Evangelist, EMEA

@julsimon

<http://medium.com/@julsimon>