

AI on a Pi

Julien Simon <@julsimon>

AI Evangelist, EMEA









All along,
what we really dreamed
about was...



Natural language
Conversational
Smart
Connected

AWS customers
are building this,
right now

A black, cylindrical Amazon Echo smart speaker with a blue light ring at the top, sitting on a white surface. The background is a blurred indoor setting with a person and a plant.

amazon
echo

25,000 skills



Liberty Mutual.

INSURANCE

“Alexa...

...tell Insurance Advisor I'd like
to find an agent in my area.”

...ask Insurance Advisor what total
vehicle loss is.”

...ask Insurance Advisor what types
of insurance Safeco offers.”

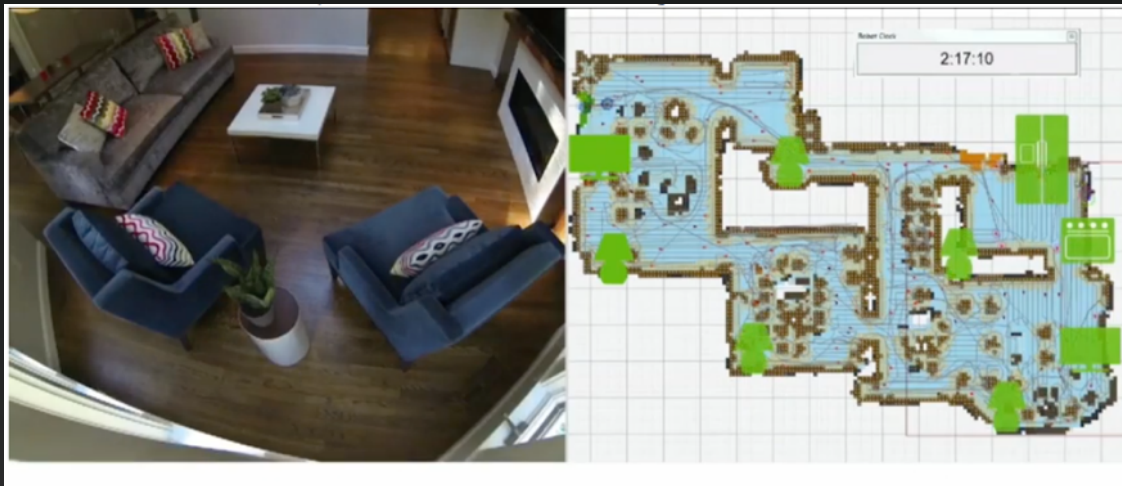




Connected Roomba
launched in 2015

More than 45 million
square meters mapped

Discovery and interaction
with Smart Home
devices





“Car as a Sensor”

Collect sensor data
from BMW 7 Series cars
to give drivers **dynamically
updated map information**

100,000 vehicles by 2018

Service launched
in **6 months**





As soon as 2018, Alexa will be your companion in BMWs

« Alexa, drive me home » can't be far away

图森 **tu** Simple



Last June, tuSimple drove an autonomous truck for 200 miles from Yuma, AZ to San Diego, CA



Amazon AI

Intelligent Services Powered By Deep Learning

Amazon AI for every developer

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

Polly: Life-like Speech Service



Converts text
to life-like speech



Fully managed



48 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
functionality



Easy to use
& thoughtfully integrated



Built for
production



Low
cost

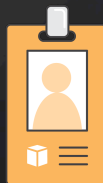
Rekognition: Search & Understand Visual Content



Real-time &
batch image
analysis



Object & Scene
Detection



Facial Detection



Facial Analysis



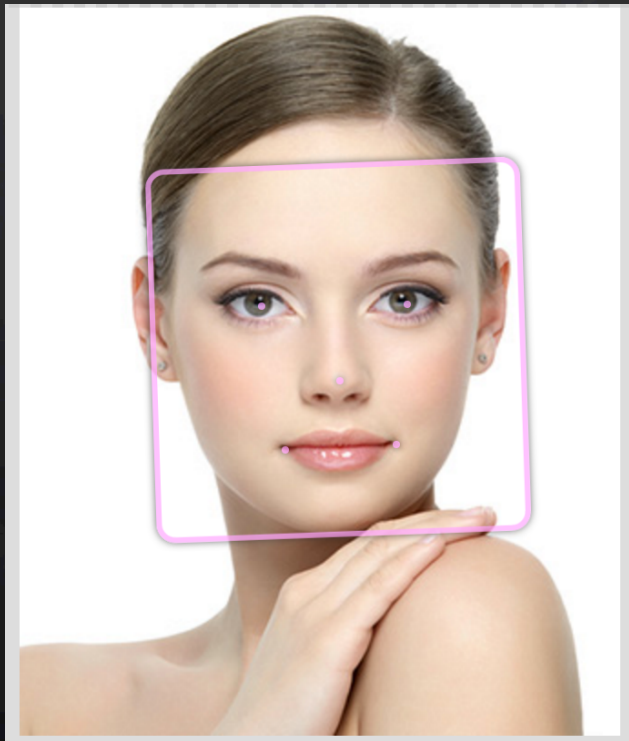
Face Search

Rekognition: Object & Scene Detection

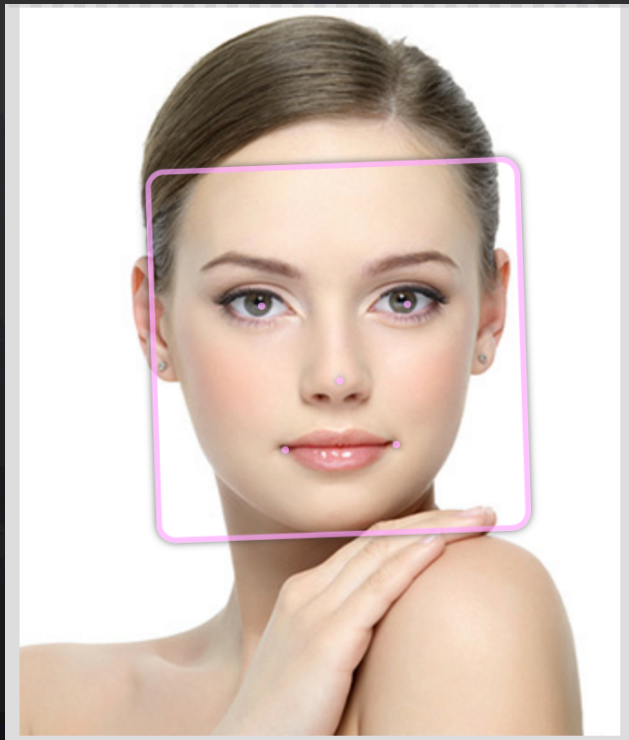


Beach	99%
Coast	99%
Outdoors	99%
Sea	99%
Water	99%
Palm Tree	98.5%
Plant	98.5%
Tree	98.5%
Landscape	51.5%
Nature	51.5%

Rekognition: Facial Detection



Rekognition: Facial Analysis



looks like a face	99.8%
appears to be female	100%
age range	26 - 43 years old
smiling	84.4%
appears to be happy	55%
appears to be calm	51.6%
not wearing glasses	99.9%
not wearing sunglasses	99.9%
eyes are open	99.9%
mouth is closed	99.7%
does not have a mustache	99.9%
does not have a beard	99.9%

Rekognition: Compare Faces



Similarity: 97.0%

Rekognition: Facial Search



Facial
verification

Compare two faces



Face
Search

Compare many faces



Visual Similarity
Search

Find similar faces



Celebrity
Detection

Sports, music, movies, etc.



Content
Moderation

Explicit, suggestive, etc.

Rekognition: Search & Understand Visual Content



High quality,
through
best-in-class
deep learning



Deep
functionality



Easy to use
& thoughtfully integrated



Built for
production



Low
cost

Amazon AI for every developer

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

Apache MXNet: Open Source library for Deep Learning



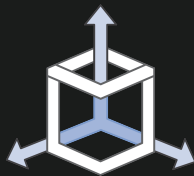
Programmable

Simple syntax,
multiple languages



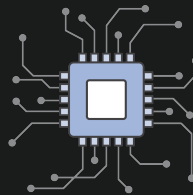
Most Open

Accepted into the
Apache Incubator



Portable

Highly efficient
models for mobile
and IoT



High Performance

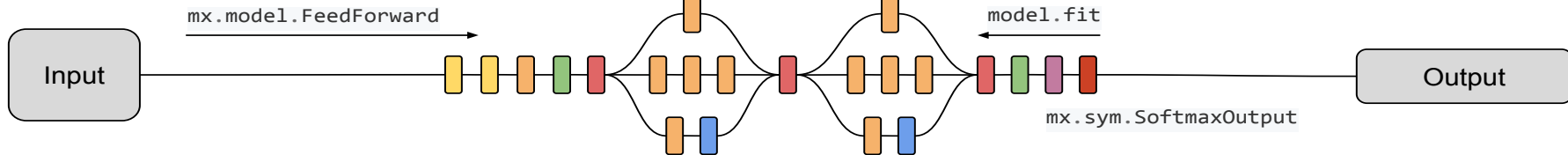
Near linear scaling
across hundreds of GPUs



Best On AWS

Optimized for
Deep Learning on AWS

More information at mxnet.io



Image



Video



Speech



Events

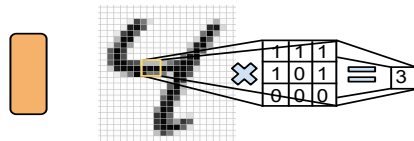
"People Riding Bikes"

Text

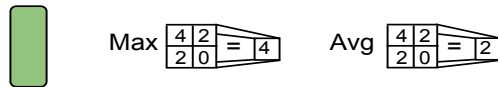
Input Weights

$$\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \otimes \begin{bmatrix} 0.2 \\ -0.1 \\ 0.7 \end{bmatrix} = \begin{bmatrix} 2 \end{bmatrix}$$

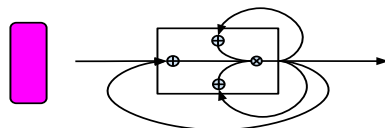
`mx.sym.FullyConnected(data, num_hidden=128)`



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

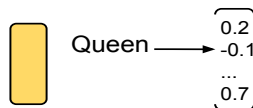


`mx.sym.Pooling(data, pool_type="max", kernel=(2,2),`



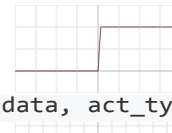
`stride=(2,2)`

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

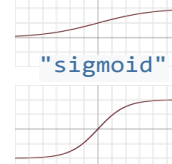


`mx.symbol.Embedding(data, input_dim, output_dim = k)`

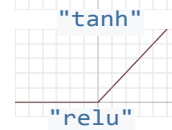
$$\cos(w, \text{queen}) = \cos(w, \text{king}) - \cos(w, \text{man}) + \cos(w, \text{woman})$$



"sigmoid"



"tanh"



"relu"



"softrelu"



Image Segmentation



Face Search



Neural Art

"People Riding Bikes"

Image Caption

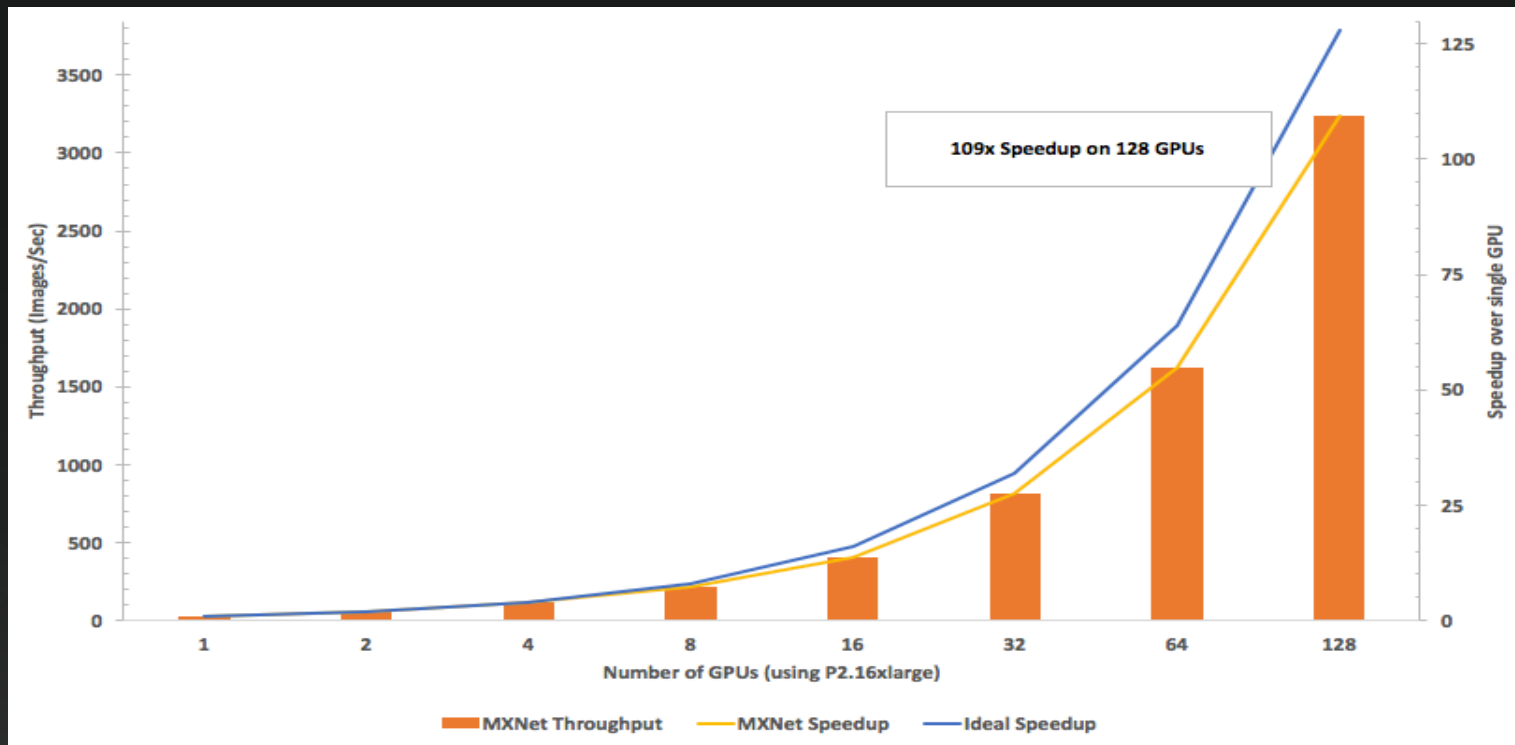
Bicycle, People, Road, Sport

Image Labels

"Οι άνθρωποι ιππασίας ποδήλατα"

Machine Translation

MXNet: near-linear training scalability



AWS Deep Learning AMI

Up to ~40k CUDA cores

Apache MXNet

TensorFlow

Theano

Caffe

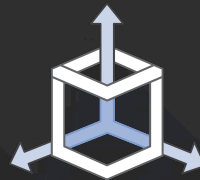
Torch

Pre-configured CUDA drivers

Anaconda, Python3

+ CloudFormation template

+ Container Image



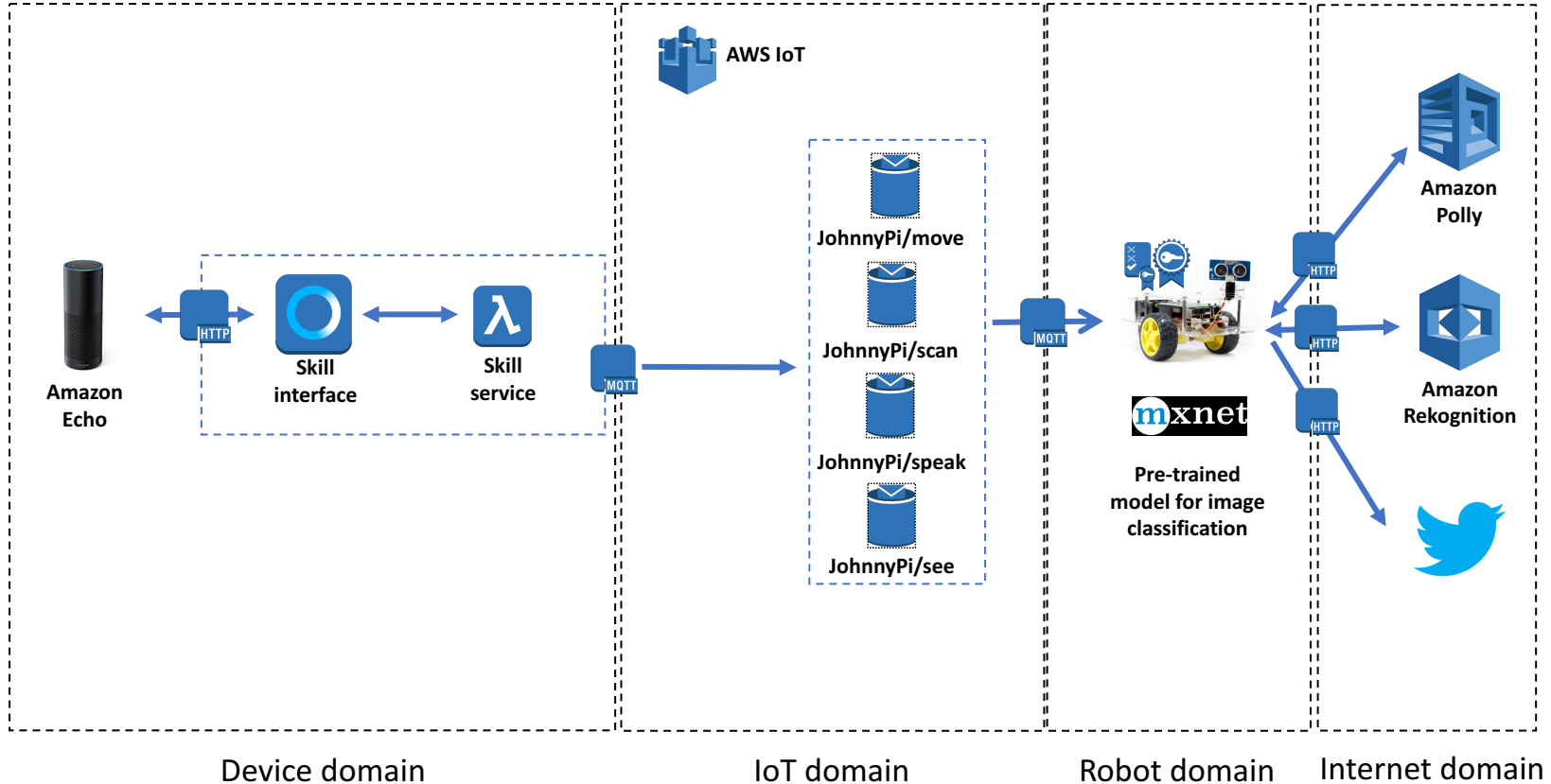
One-Click GPU Deep Learning

So, how about we build a
voice-powered robot that can
see and speak?

Voice-driven AI robot: 250€, 700 lines of code

<https://medium.com/@julsimon/johnny-pi-i-am-your-father-part-0-1eb537e5a36>

<https://github.com/juliensimon/johnnypi>



Who does what?

- Skill interaction model
 - Utterances (32)
 - Custom slots
 - Movement (7)
 - Target (2)
- Skill Lambda function
 - Python (225 LOC)
 - Handles session
 - Sends IoT messages with robot commands
- AWS IoT gateway
 - Zero code, configuration only
- Raspberry Pi app
 - Python server (400 LOC)
 - Receives IoT messages
 - Calls robot API for movement
 - Calls Polly API for speech
 - Calls Reko API for faces
 - Calls Twitter API
 - Uses local MXNet model for objects
 - Complex CNN: Inception v3
 - Pre-trained on Imagenet
 - Animals and objects, no humans
 - 1000 categories

Sample utterances

`DirectionIntent` move {Direction}
`DirectionIntent` go {Direction}
`DirectionIntent` turn {Direction}
`DirectionIntent` now move {Direction}
`DirectionIntent` now go {Direction}
`DirectionIntent` now turn {Direction}
`DirectionIntent` go {Direction} please
`DirectionIntent` move {Direction} please
`DirectionIntent` turn {Direction} please
`DirectionIntent` I want you to go {Direction}
`DirectionIntent` I want you to move {Direction}
`DirectionIntent` I want you to turn {Direction}
`DirectionIntent` {Direction}

`SeeIntent` Look at the {Target}
`SeeIntent` Take a look at the {Target}
`SeeIntent` Just look at the {Target}
`SeeIntent` Tell me about the {Target} you see
`SeeIntent` Tell me about the {Target} in front of you
`SeeIntent` What is the {Target} in front of you
`SeeIntent` Do you see the {Target} in front of you
`SeeIntent` Do you see an {Target}
`SeeIntent` Do you see {Target}
`SeeIntent` Describe the {Target} you see

Please suggest additional ones 😊

Anything you dream is fiction, and anything you accomplish is science, the whole history of mankind is nothing but science fiction.

Ray Bradbury



Get started at

<http://aws.amazon.com/ai>

<http://aws.amazon.com/iot>

<http://aws.amazon.com/evangelists/julien-simon>

@julsimon