

# Model Asset Exchange (MAX)

Deep learning for everyone

Patrick Titzler

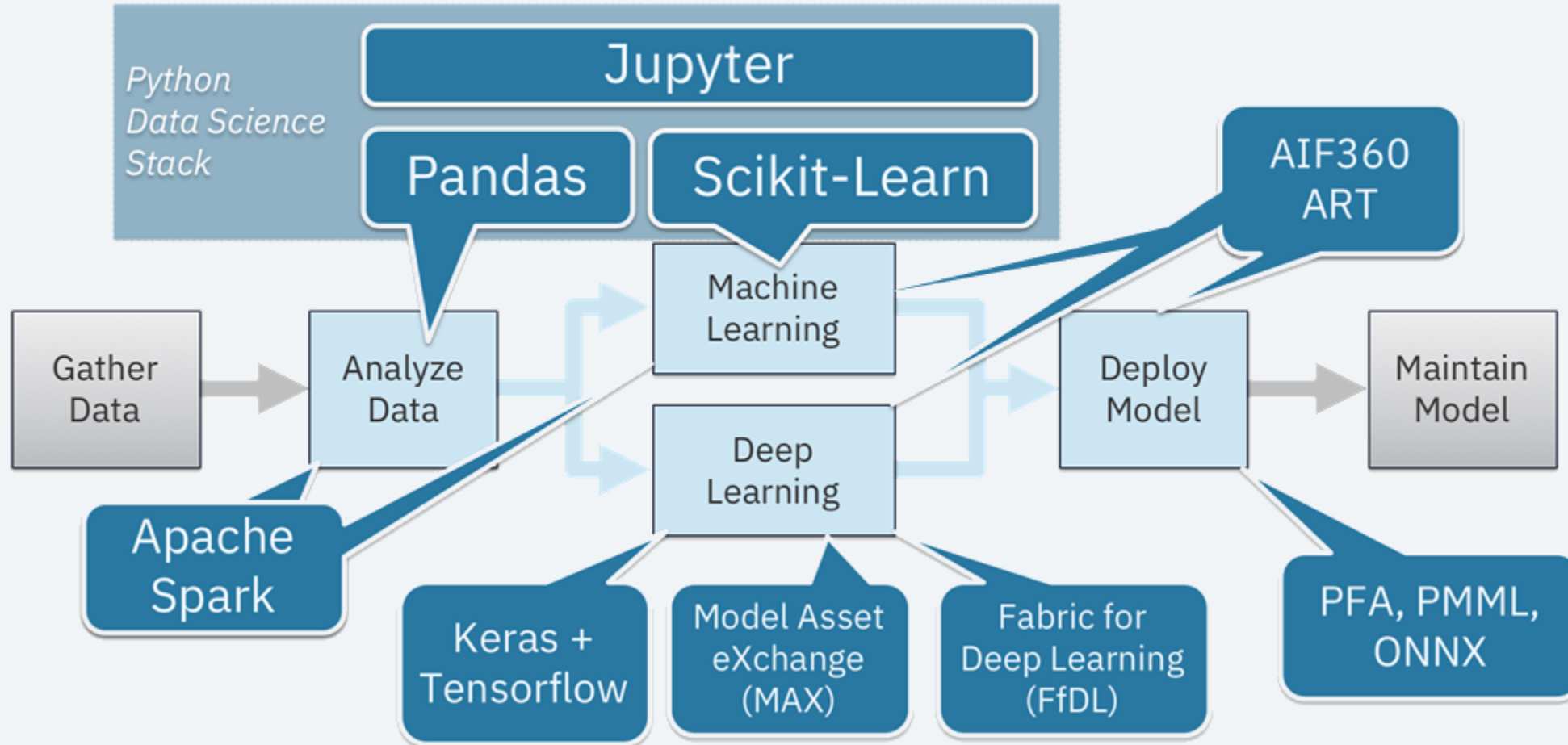
Developer Advocate

Center for Open-Source Data & AI Technologies

## About CODAIT.org

- ~40 data scientists and developers working on Open Source software
- 1,000+ IBMers contributing to Open Source software
- 62,000+ IBMers using Open Source software

# CODAIT: Improving Enterprise AI lifecycle in Open Source



<https://developer.ibm.com/code/open/centers/codait/projects/>

# Our MAX goal: broader adoption of AI

Think about your daily life

- ... know how to use a mobile phone
- ... know how to drive a car (take ride-share, ...)

You can do many things with a “minimum” amount of knowledge

**Make AI accessible to every developer:**

- Reduce need for expert skills
- Improve “Time to Value”

# Programming vs learning (1)

Programming (Developer):

- Implement *source code that produces desired outcome*
- Examples: web application, microservice, visualization, ...
- Not well suited to solve certain kinds of problems

Is there a fruit in the image?

What is the fruit?

Is it ripe?



# Programming vs learning (2)

Machine/deep learning (Data scientist, ML engineer):

- Build model [code]
- Train model using lots of *prepared* data: produces weights




- *Model code + weights yield (hopefully) desired outcome*

Trained model

- Model training and execution requires framework (TensorFlow,...)
- Examples: audio classification, object detection in images

# Example: suggest image tags

**Manage photos**



Exif **Tags** Description Location Visibility Settings

add a tag

Update

✓ Approved

Find model

Get code

Test, verify, fix

Train model

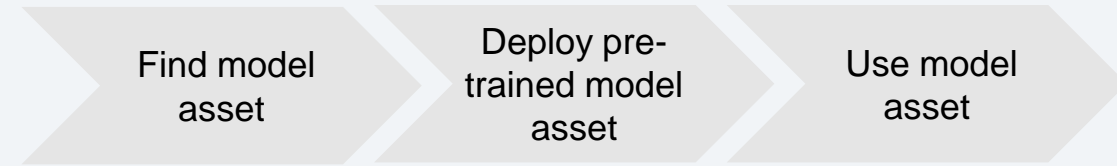
Deploy model

Use model

Requires time, expertise, and resources

Time to Value (days, weeks, months, ...)

# MAX: Reduces "Time to Value" for developers



Free, deployable, and trainable code.

A place for developers to find and use free and open source deep learning models.

[View all models >](#) [Try the tutorial >](#) [Join the community >](#)

- Audio classification
- Image classification
- Text classification
- Object detection
- Facial recognition
- Image-to-image translation
- Image-to-text translation
- Named entity recognition
- Text feature extraction
- ...

Deployable | Facial Recognition

### Facial Emotion Classifier

Detect faces in an image and predict the emotional state of each person

[View model >>](#)

Artificial intelligence Deep learning +

Deployable | Object Detection In Images

### Image Segmenter

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

[View model >>](#)

Artificial intelligence Deep learning +

Deployable | Object Detection In Images

### Object Detector

Localize and identify multiple objects in a single image.

[View model >>](#)

Artificial intelligence Deep learning +



# Pre-trained vs custom-trained model asset

- Pre-trained model might not be specific enough



- Custom training required to improve the results



# MAX model assets

- Cover 15+ application domains
  - Utilize state-of-the-art deep learning models
  - Validated and pre-trained
  - Free and open source
  - Fully documented
  - Cleared for personal and commercial usage
- 
- Are not just trained models.

## Model Metadata

Domain	Application	Industry	Framework	Training Data	Input Data Format
Vision	Object Detection	General	TensorFlow	<a href="#">COCO Dataset</a>	Image (RGB/HWC)

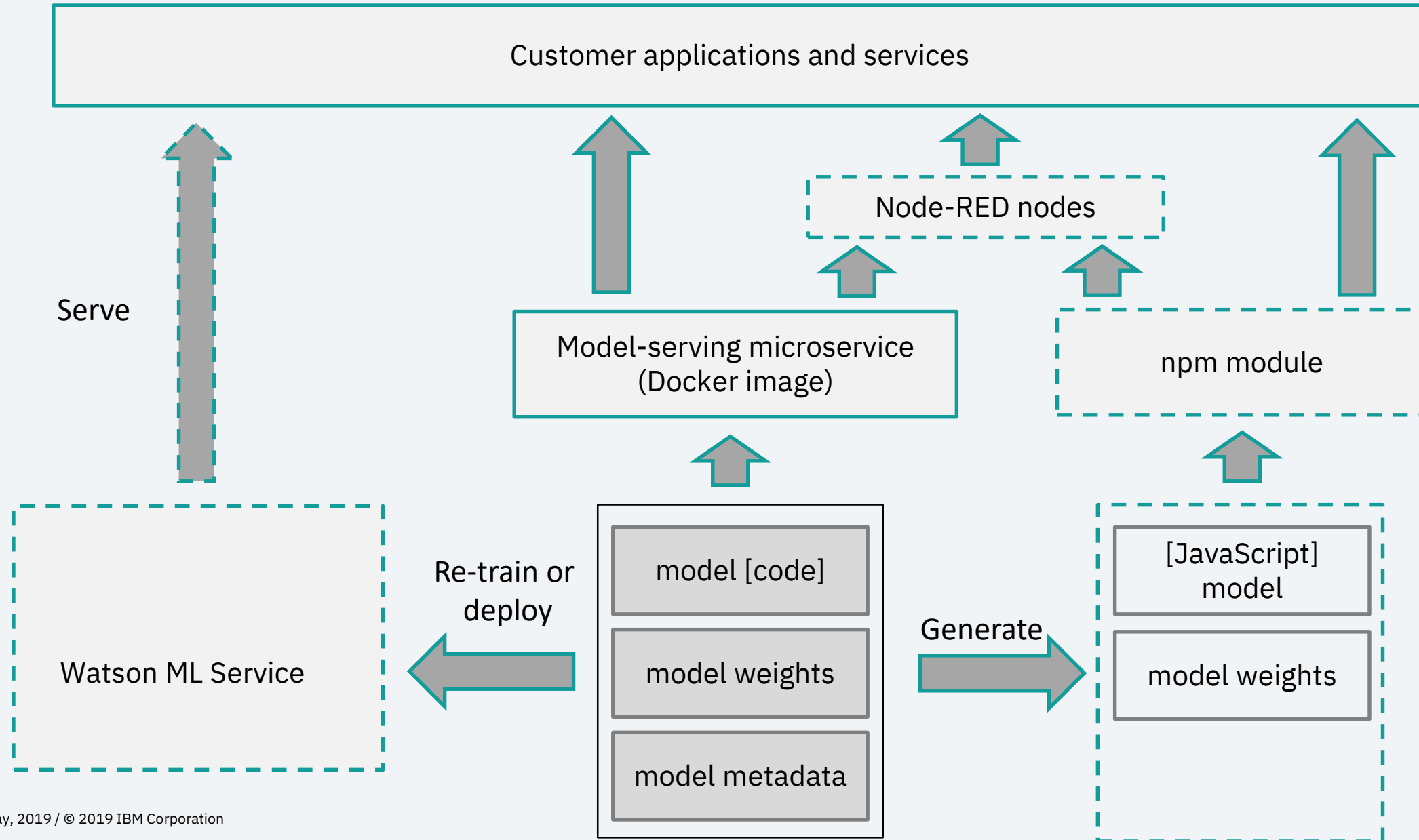
## References

- *J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors", CVPR 2017*
- *Tsung-Yi Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Lawrence Zitnick, P. Dollár, "Microsoft COCO: Common Objects in Context", arXiv 2015*
- *W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A. C. Berg, "SSD: Single Shot MultiBox Detector", CoRR (abs/1512.02325), 2016*
- *A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv 2017*
- [TensorFlow Object Detection GitHub Repo](#)

## Licenses

Component	License	Link
Model GitHub Repository	<a href="#">Apache 2.0</a>	<a href="#">LICENSE</a>
Model Weights	<a href="#">Apache 2.0</a>	<a href="#">TensorFlow Models Repo</a>

# What are MAX model assets?



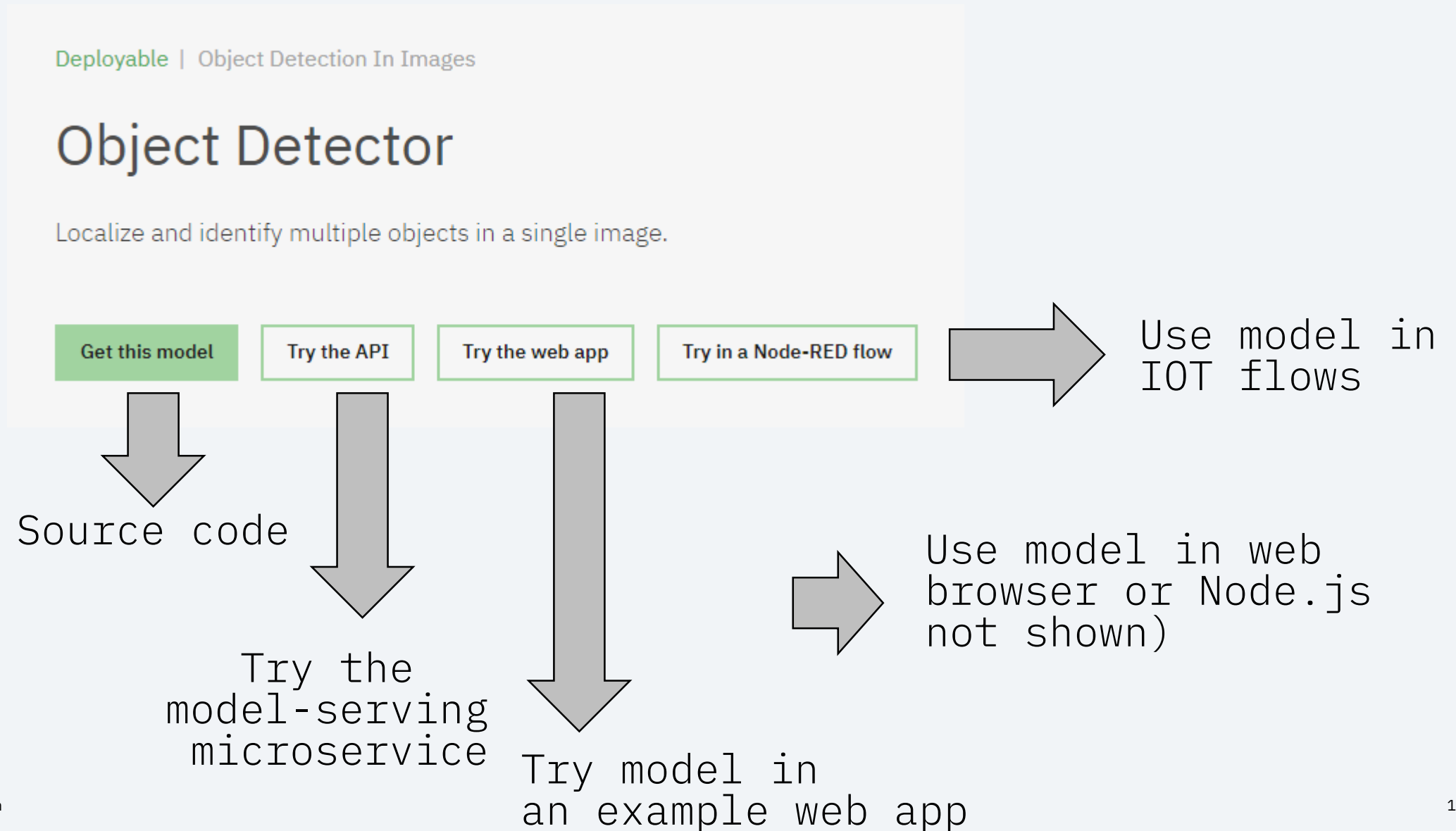
# MAX assets are “standardized”

- Application-friendly inputs and outputs
  - Examples:
    - Input is an image (png/jpg/...) not a multi-dimensional array of numbers
    - Output is JSON
- Consistent interfaces
- Consistent output formats across related models
  - Examples:
    - Bounding box coordinates ((x1,y1,h,w) vs (x1,y1,x2,y2) vs ...) for detection models

How can you “use” MAX?

[ibm.biz/model-exchange](https://ibm.biz/model-exchange)

# Exploring/Using model assets



# Model exploration/evaluation: OpenAPI spec

- Developers can try model-serving microservices (without any install)

**model** Model information and inference operations ▼

**GET** **/model/labels** Return the list of labels that can be predicted by the model

**GET** **/model/metadata** Return the metadata associated with the model

**POST** **/model/predict** Make a prediction given input data

**Parameters** Cancel

Name	Description
<b>image</b> * required file (formData)	An image file (encoded as JPEG, PNG or TIFF) <div><span>Choose File</span> No file chosen</div>

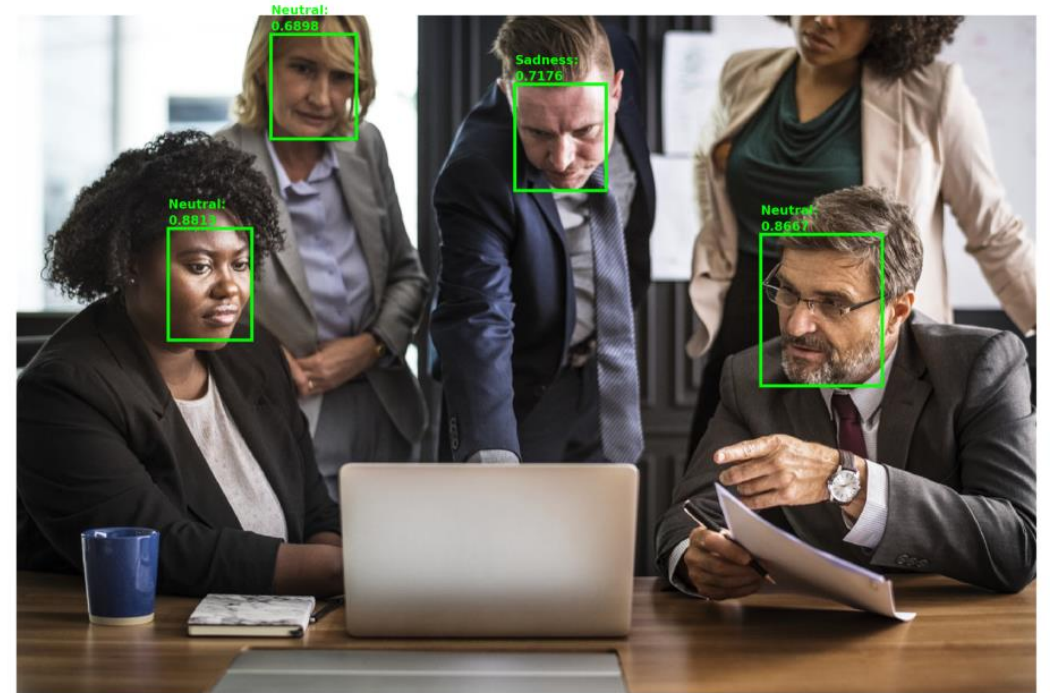
**Execute**

**Responses** Response content type application/json ▼

# Model exploration/evaluation: Jupyter notebooks

- Some image-processing models include a ready-to-run Python notebook
- Demonstrates:
  - How to invoke model service
  - How to visualize results
  - Sample use-cases (e.g. GDPR)
- Run locally or cloud

```
ax.add_patch(rectangle)  
# Plot the emotion class and probability text  
plt.text(x1, y1 - 25, emotion_label, fontsize=4, color=color, fontweight='bold')  
plt.text(x1, y1 - 5, emotion_probability, fontsize=4, color=color, fontweight='bold')  
  
plt.axis('off')  
plt.show()
```



In [ ]:



# Model exploration/evaluation: Sample apps

- Illustrate how to consume assets
- Published as code patterns

```
Nick's Terminal

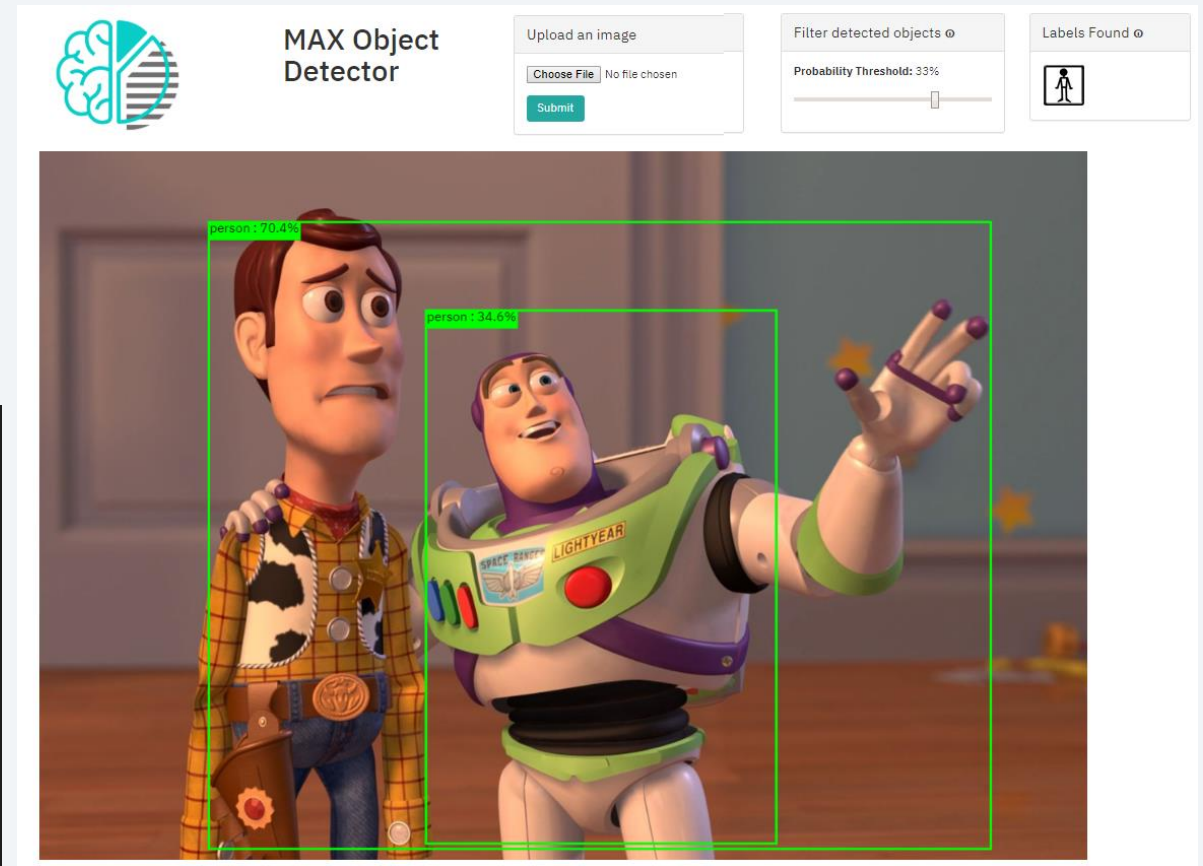
@kastentx~/Pictures> magicat my-images/ --contains dog --verbose
Scanning directory '/Users/nick/Pictures/my-images' for dog...

Dog found in:

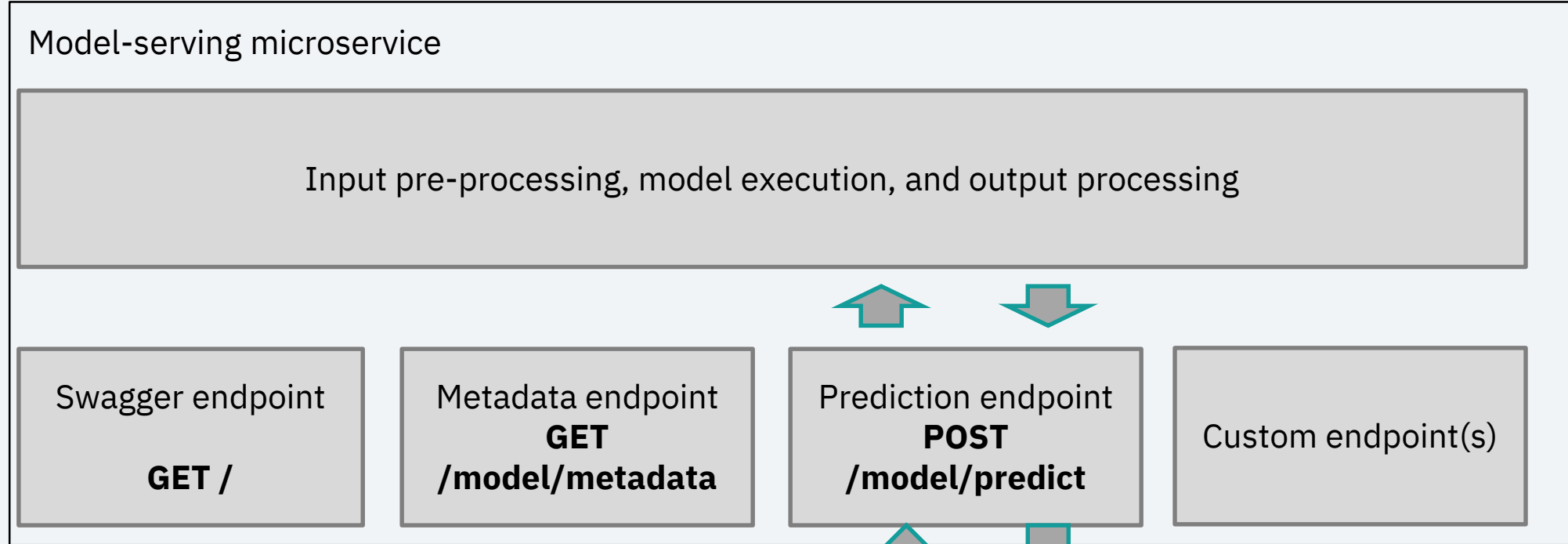
/Users/nick/Pictures/my-images/puppers.jpg
/Users/nick/Pictures/my-images/trails.jpg

No Dogs found in:

/Users/nick/Pictures/my-images/castle.jpg
/Users/nick/Pictures/my-images/cow.jpg
@kastentx~/Pictures> █
```



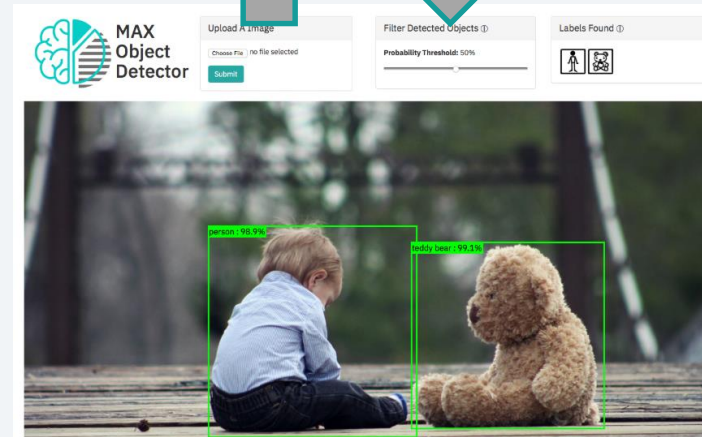
# Consumption: model-serving microservice



Docker, Kubernetes, cloud

Image

JSON



# Consumption: In JavaScript/Node.js

- Run prediction directly in the browser (or Node.js app)
- TensorFlow.js versions of a TensorFlow-based MAX models
- <https://github.com/CODAIT/max-tfjs-models>

TFJS Age Estimator Model



<https://ibm.github.io/fae-tfjs>

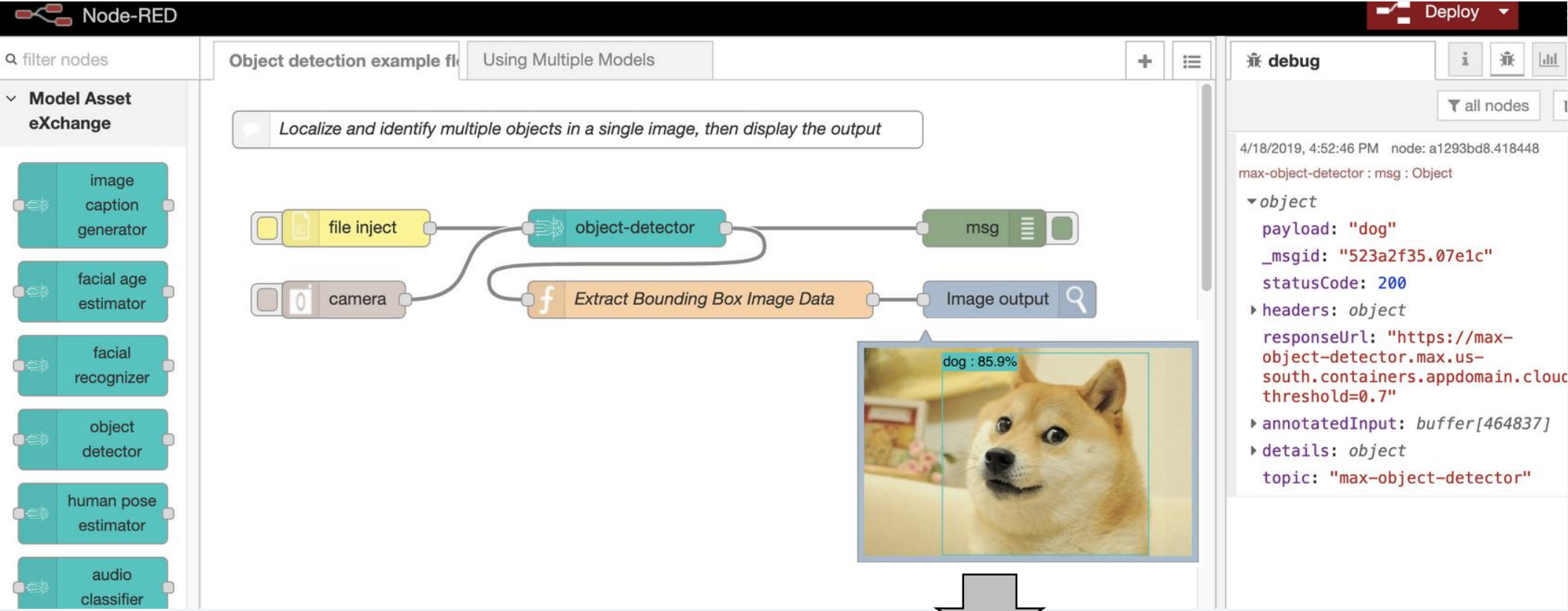


<https://ibm.biz/veremax>

# Consumption: In Node-RED (IOT) flows

- Nodes run on-device, PC, or the cloud
- <https://github.com/CODAIT/node-red-contrib-model-asset-exchange>

Model nodes



Object detection example flow: Using Multiple Models

Localize and identify multiple objects in a single image, then display the output

dog : 85.9%

Result preview

```
4/18/2019, 4:52:46 PM node: a1293bd8.418448
max-object-detector : msg : Object
▼ object
  payload: "dog"
  _msgid: "523a2f35.07e1c"
  statusCode: 200
  headers: object
    responseUrl: "https://max-object-detector.max.us-south.containers.appdomain.cloud/threshold=0.7"
  annotatedInput: buffer[464837]
  details: object
    topic: "max-object-detector"
```

# MAX summary

- Little/no AI expertise required to get started
- Open-source, fully vetted, ready to consume assets (short Time to Value)
  - Container-based microservice (programming language agnostic)
  - IOT support through Node-RED module
  - In-browser support through JavaScript/Node.js packages
- Low barrier to entry (no registration, no paid services required)
- No vendor lock-in (supports major ML frameworks)

# Resources

- MAX on IBM Developer: <https://ibm.biz/model-exchange>
- Learning path: <https://developer.ibm.com/series/create-model-asset-exchange/>
- Ecosystem status: <https://ibm.biz/max-status>
- Public Slack: <https://ibm.biz/max-slack>
- Twitter: <https://twitter.com/ibmcodait>    Medium: <https://medium.com/codait>