

Liberdade e Inovação: como ferramentas livres vão guiar a sua jornada como Desenvolvedor e Cientista de Dados

Gabriela de Queiroz

Sr. Machine Learning Manager, IBM California
Founder, R-Ladies & AI Inclusive

 [@gdequeiroz](https://twitter.com/gdequeiroz) | k-roz.com

slides: bit.ly/tdc-gabriela



Gabriela de Queiroz

Sr. Engineering and Data Science Manager, IBM

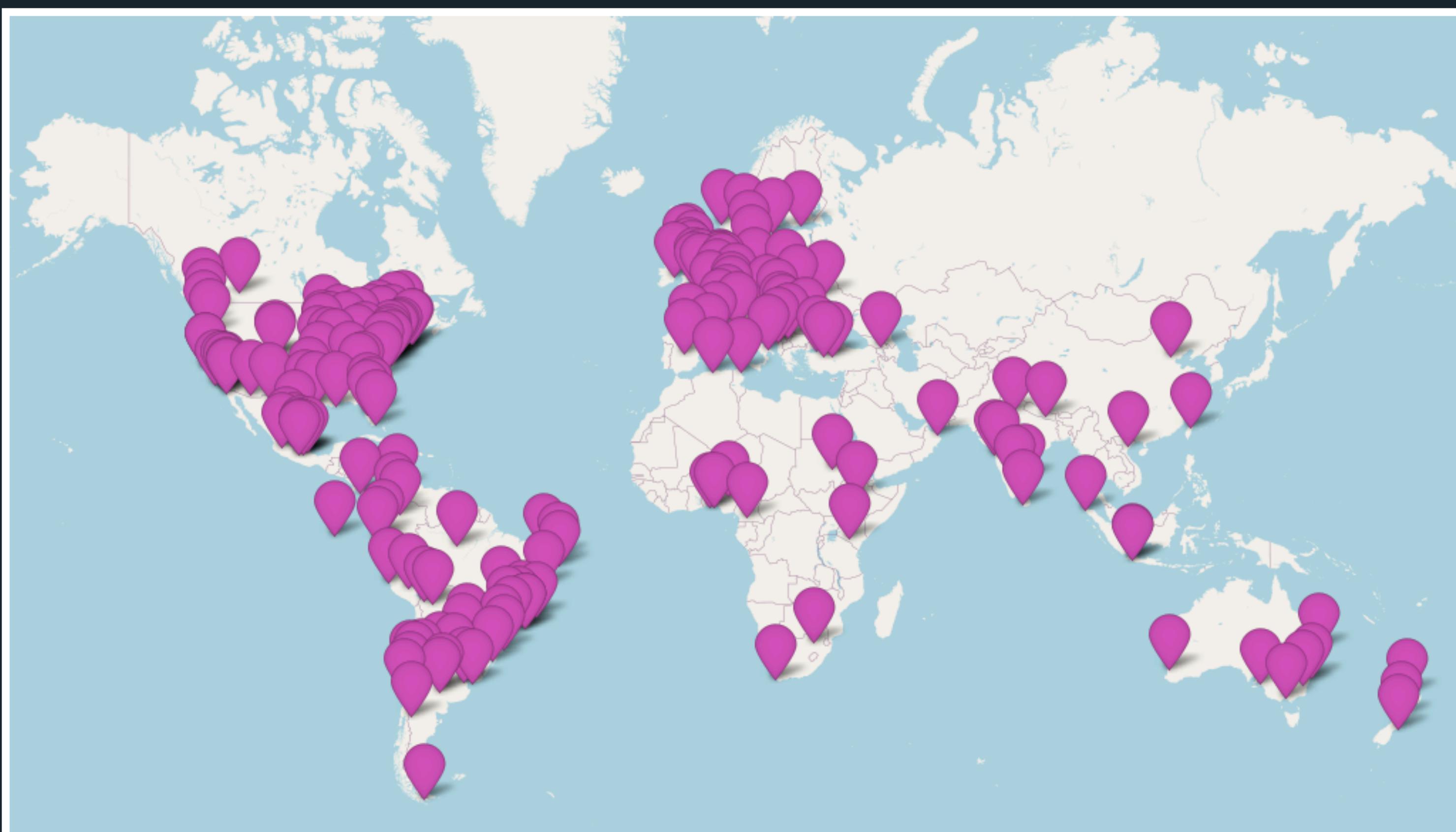
- Fundadora do R-Ladies (rladies.org)
- Fundadora do **AI Inclusive** (ai-inclusive.org)
- Graduação em Estatística (UERJ)
- Mestrado em Epidemiologia (ENSP/Fiocruz)
- Mestrado em Estatística (CSUEB)



**Data Scientist + Developer Advocate + Open Source Developer + Manager +
Statistician + Epidemiologist + Community Builder + Mentor + Speaker**



rladies.org



193

R-Ladies Cities



54

R-Ladies Countries



71618

R-Ladies members on meetup.com





Missão

Aumentar a **representatividade** e
participação de minorias de gênero
em Inteligência Artificial

- **Website:** ai-inclusive.org
- **Twitter:** bit.ly/ai-inclusive-twitter
- **Facebook:** bit.ly/ai-inclusive-facebook
- **Instagram:** bit.ly/ai-inclusive-instagram
- **Youtube:** bit.ly/ai-inclusive-youtube

Se tiver interesse em criar um capítulo, mande-nos um email:
info@ai-inclusive.org



Center for Open Source Data and AI Technologies (CODAIT)

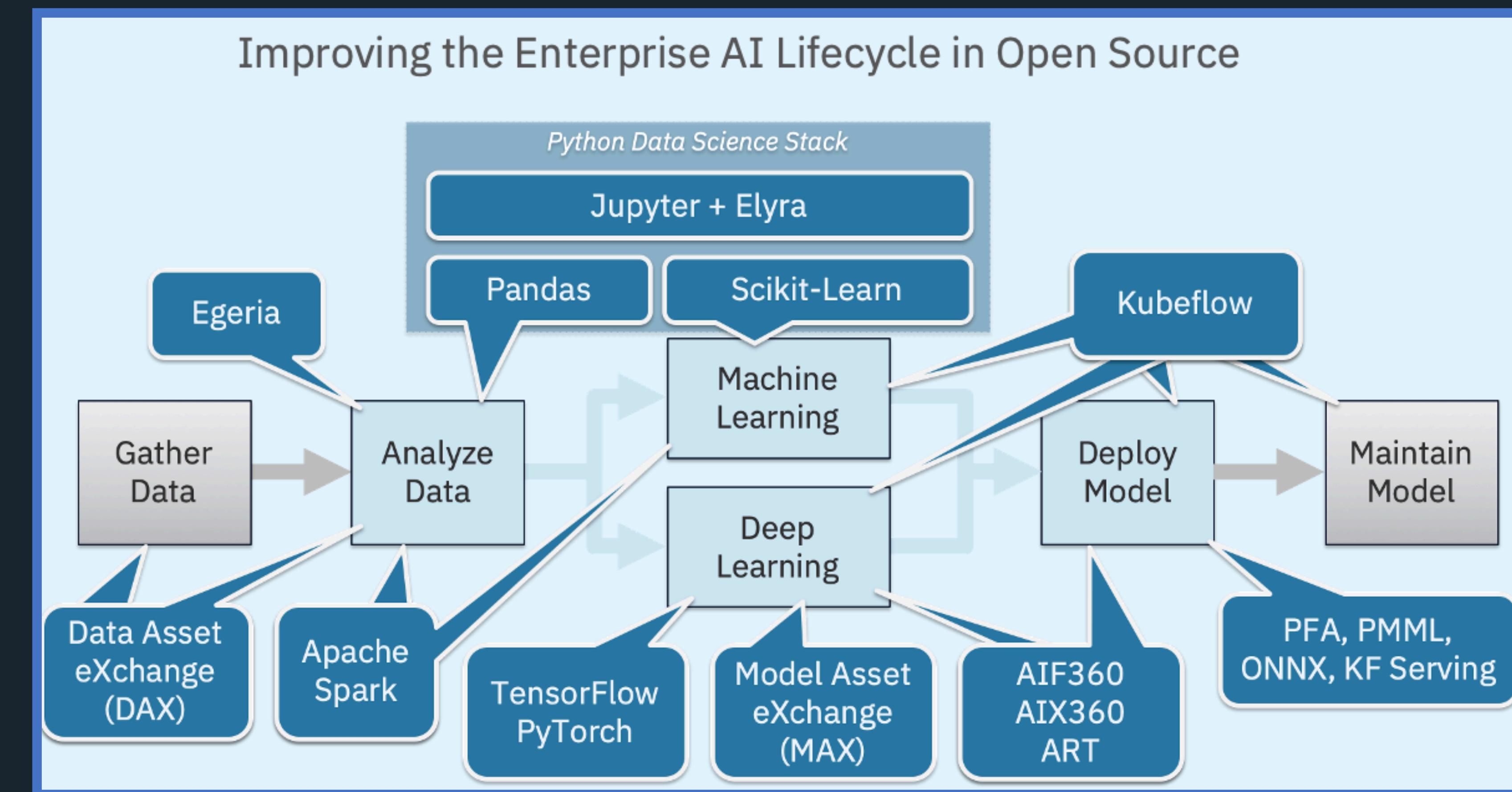
Open Source @ IBM



**Build tools to make
AI accessible to all**

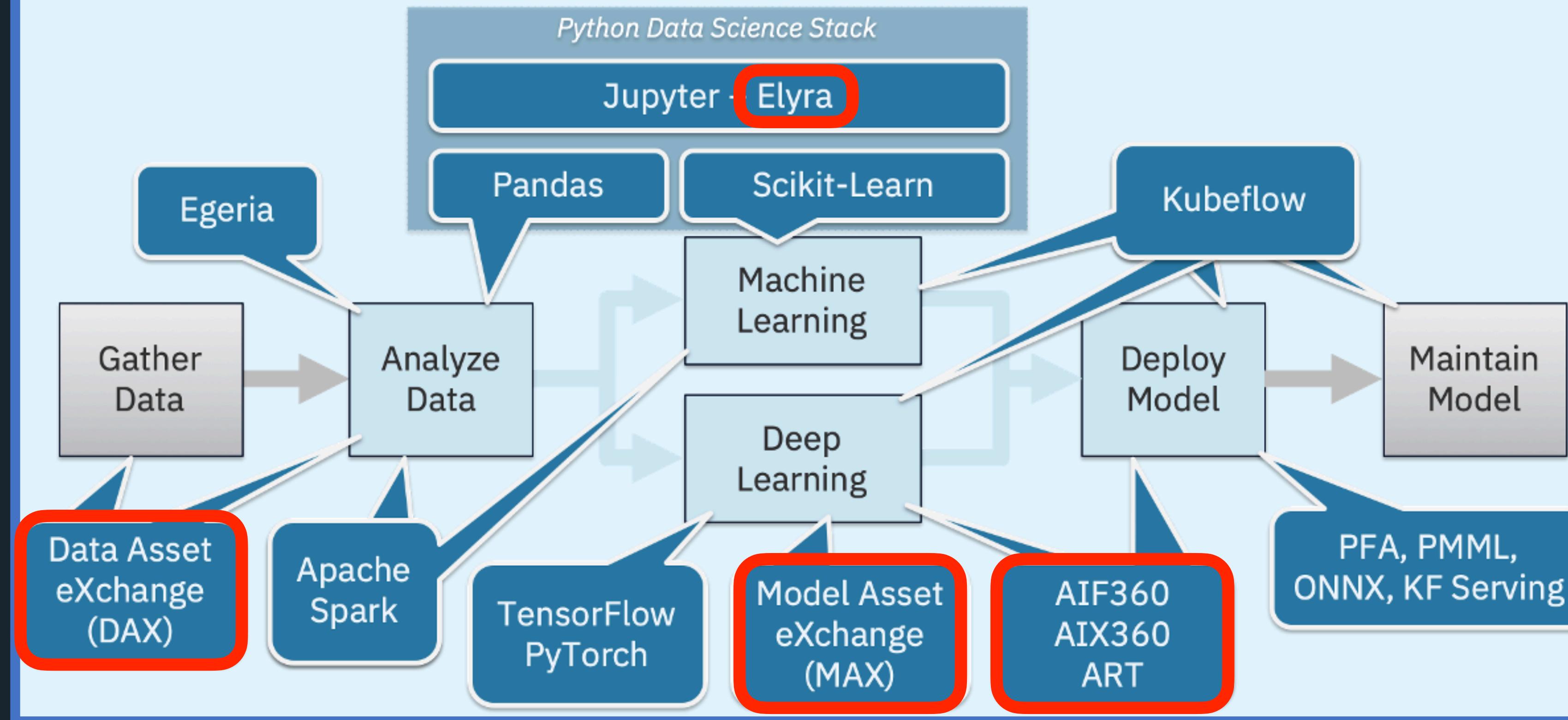
We contribute to and advocate for
the open-source technologies

30+ open source developers!



Center for Open Source Data and AI Technologies (CODAIT)

Improving the Enterprise AI Lifecycle in Open Source



Open Source @ IBM

Alguns Projetos:

- Model Asset eXchange (ibm.biz/model-exchange)
- Data Asset eXchange (ibm.biz/model-exchange)
- Elyra (ibm.biz/elyra-ai)

- AI Fairness 360 (aif360.mybluemix.net/)
- AI Explainability 360 (aix360.mybluemix.net/)
- Adversarial Robustness Toolbox (art-demo.mybluemix.net/)
- FactSheets (aif360.mybluemix.net/)

Você já é um **cientista de dados** ou você quer se tornar um?



The 8 Data Science Skills That Will Get You Hired

- Programming Skills. ...
- Statistics. ...
- Machine Learning. ...
- Multivariable Calculus & Linear Algebra. ...
- Data Wrangling. ...
- Data Visualization & Communication. ...
- Software Engineering. ...
- Data Intuition.

What data scientist should know?

9 Must-have skills you need to become a Data Scientist, updated

- By Simplilearn.
- Education.
- R Programming.
- Python Coding.
- Hadoop Platform.
- SQL Database/Coding.
- Apache Spark.
- Machine Learning and AI.

A data scientist needs strong math skills, particularly in multivariable calculus and linear algebra.

- Identifying Algorithms.
- Creating and Maintaining Algorithms.
- Information Retrieval Data Sets.
- Linear Algebra.
- Machine Learning Models.
- Machine Learning Techniques.
- Multivariable Calculus.
- Statistics.





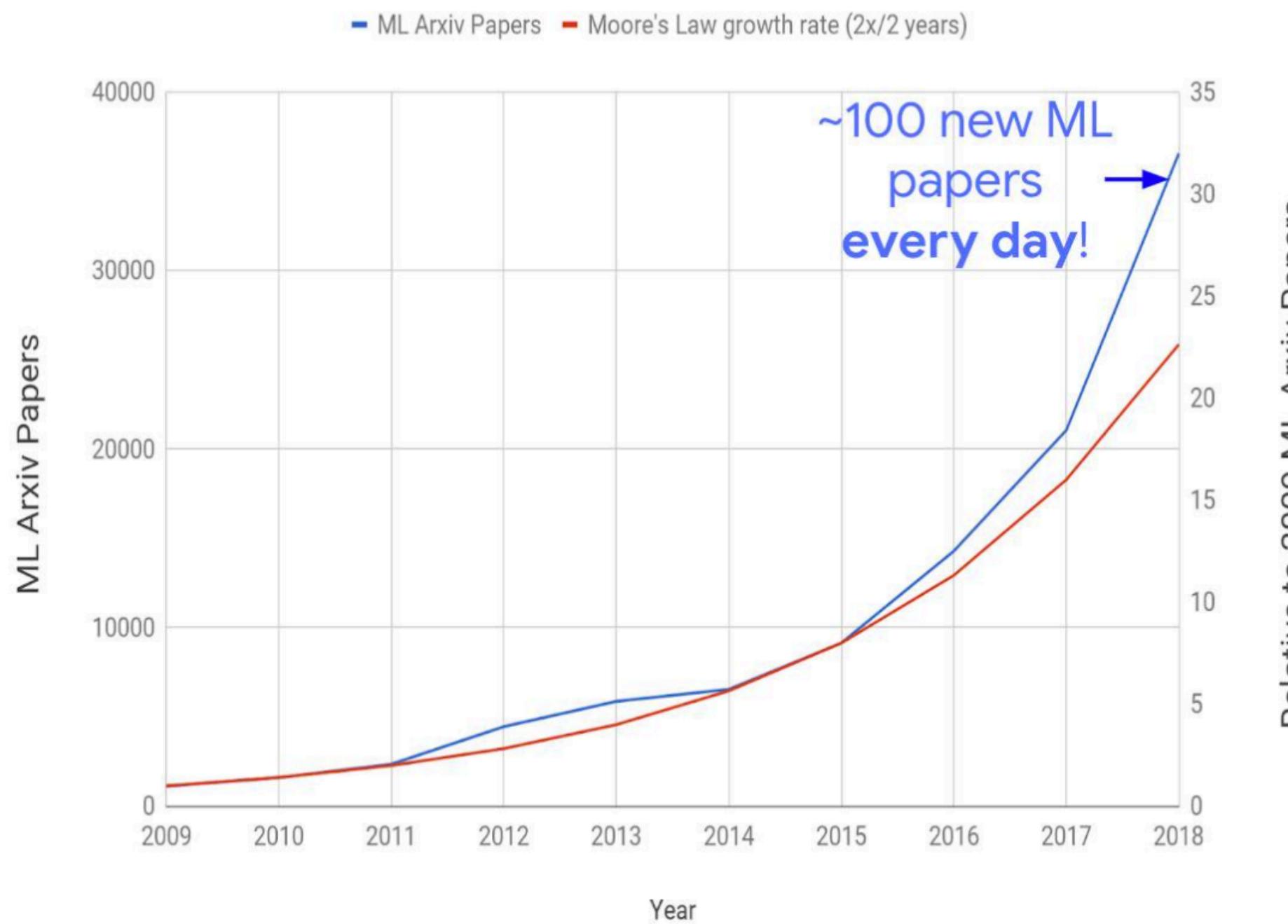
Jeff Dean ✅
@JeffDean

Arxiv ML papers/day

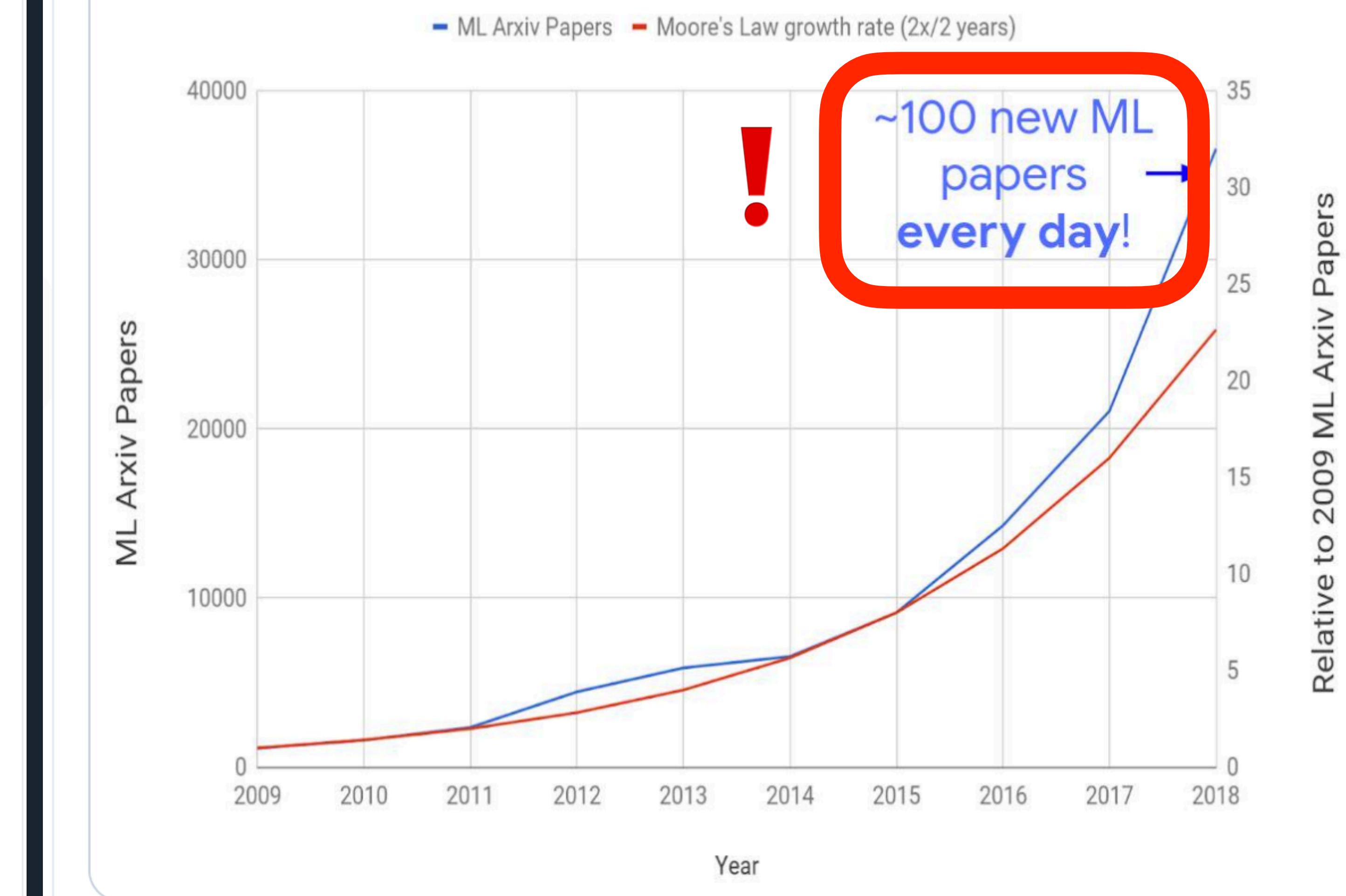
I have some bad news for you, [@deliprao](#). I collect data on the "ML subset" of [@arxiv_org](#) and its growth.

Absolute numbers are approx ~3000 new papers/month at end of 2018 (100/day!) & likely higher now, well past the bioRxiv 2500/month that you feared!

Machine Learning Arxiv Papers per Year



Machine Learning Arxiv Papers per Year



<https://twitter.com/JeffDean/status/1135114657344237568?s=20>



Articles

About 4,500,000 results (0.05 sec)

> 4 million results!

deep learning courses



All

News

Videos

Images

Shopping

More

Settings

Tools

About 183,000,000 results (0.73 seconds)

> 183 million results!



HELP!

Model Asset eXchange

Place for developers/data scientists to find and use
free and **open source** deep learning models

ibm.biz/model-exchange

Model Asset eXchange

[Try the tutorial](#)



[Join the community](#)



Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019



Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019



Model | Deployable, Trainable

Image Segmenter

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

Sep 21, 2018



Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018



Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018



Model | Deployable

Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018



[View all models](#)

Model Deployable, Trainable Question Answering Answer questions on a given corpus of text Sep 17, 2019 → Jul 09, 2019	Model Deployable Text Summarizer Generate a summarized description of a body of text Jun 04, 2019 → Jul 09, 2019	Model Deployable Toxic Comment Classifier Detect 6 types of toxicity in user comments Jun 04, 2019 →	Model Deployable Image Colorizer Adds color to black and white images. Sep 21, 2018 →	Model Deployable, Trainable Image Segmenter Identify objects in an image, additionally assigning each pixel of the image to a particular object. Sep 21, 2018 →	Model Deployable Image Classifier - Inception ResNet v2 Identify objects in images using a third-generation deep residual network. Sep 21, 2018 →
Model Deployable Chinese Phonetic Similarity Estimator Estimate the phonetic distance between Chinese words and get similar sounding candidate words. May 28, 2019 →	Model Deployable Image Resolution Enhancer Upscale an image by a factor of 4, while generating photo-realistic details. Mar 29, 2019 →	Model Deployable, Trainable Text Sentiment Classifier Detect the sentiment captured in short pieces of text Mar 29, 2019 →	Model Deployable News Text Generator Generate English-language text similar to the news articles in the One Billion Words data set. Sep 21, 2018 →	Model Deployable Object Detector Localize and identify multiple objects in a single image. Sep 21, 2018 →	Model Deployable, Trainable Image Classifier - ResNet50 Identify objects in images using a first-generation deep residual network. Sep 21, 2018 →
Model Deployable Nucleus Segmenter Identify nuclei in a microscopy image and assign each pixel of the image to a particular nucleus Mar 28, 2019 →	Model Deployable Speech to Text Converter Converts spoken words into text form. Mar 27, 2019 →	Model Deployable Human Pose Estimator Detect humans in an image and estimate the pose for each person. Dec 12, 2018 →	Model Deployable Review Text Generator Generate English-language text similar to the text in the Yelp® review data set. Sep 21, 2018 →	Model Deployable Scene Classifier Classify images according to the place/location labels in the Places365 data set. Sep 21, 2018 →	Model Deployable Sports Video Classifier Classify sporting activities in videos. Sep 21, 2018 →
Model Deployable, Trainable Named Entity Tagger Locate and tag named entities in text. Oct 31, 2018 →	Model Deployable Fast Neural Style Transfer Generate a new image that mixes the content of a source image with the style of another image. Sep 21, 2018 →	Model Deployable Image Caption Generator Generate captions that describe the contents of images. Sep 21, 2018 →	Model Deployable Weather Forecaster Predict hourly weather features given historical data for a specific location Sep 21, 2018 →	Model Trainable Word Embedding Generator Generates word embedding vectors from text files. Sep 21, 2018 →	Model Deployable Audio Classifier Identify sounds in short audio clips. Sep 21, 2018 →

30 ready to use deep learning models

Model Asset eXchange (MAX)

- Wide variety of domains (text, audio, image, etc)
- Multiple deep learning frameworks (TensorFlow, PyTorch, Keras)
- **Trainable** and **Deployable** versions

ibm.biz/model-exchange

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019

Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019

Model | Deployable, Trainable

Image Segmente

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

Sep 21, 2018

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018

Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018

Model | Deployable

Image Caption Generator

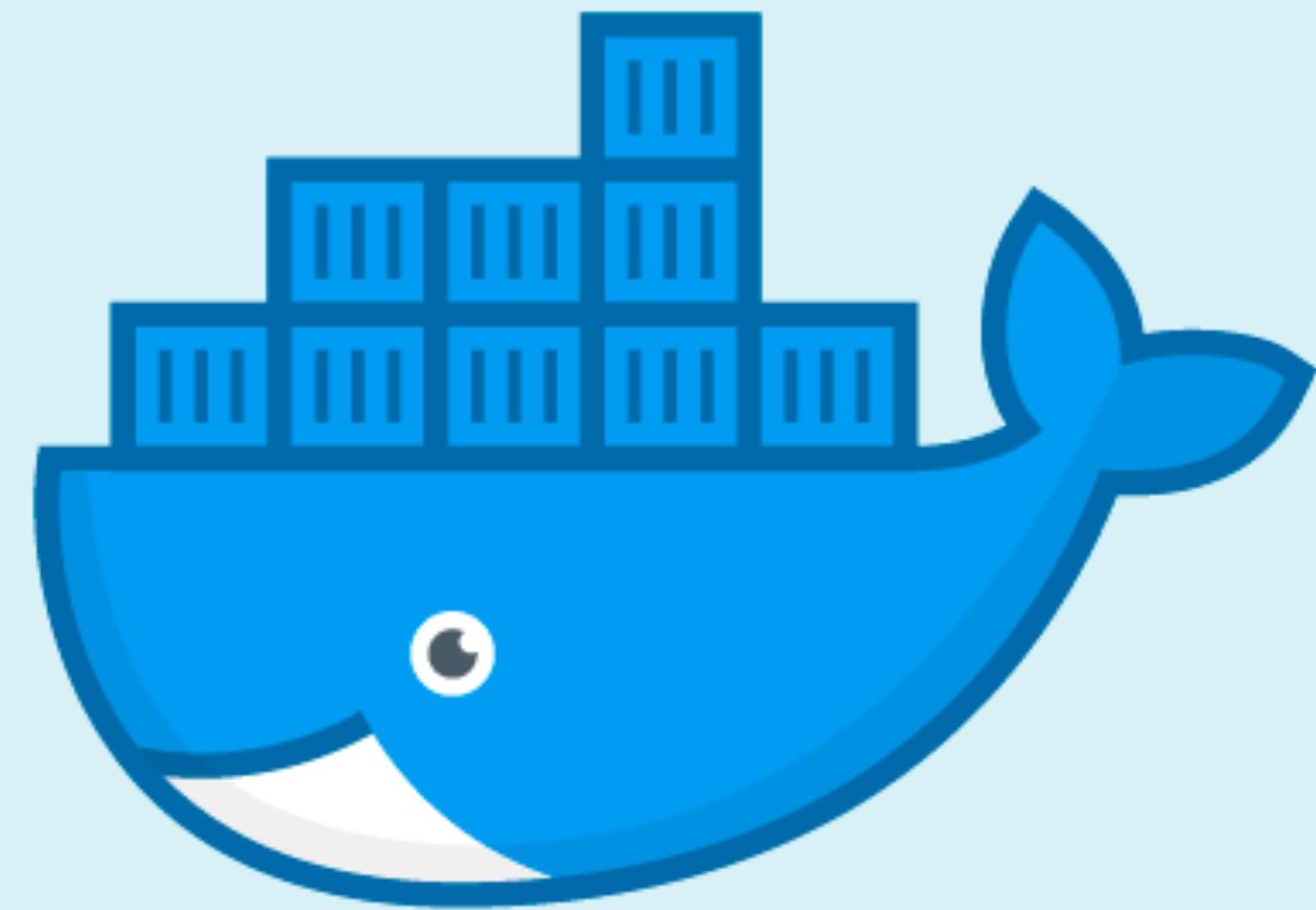
Generate captions that describe the contents of images.

Sep 21, 2018

[View all models](#)

What do I need to get started?

START
HERE.

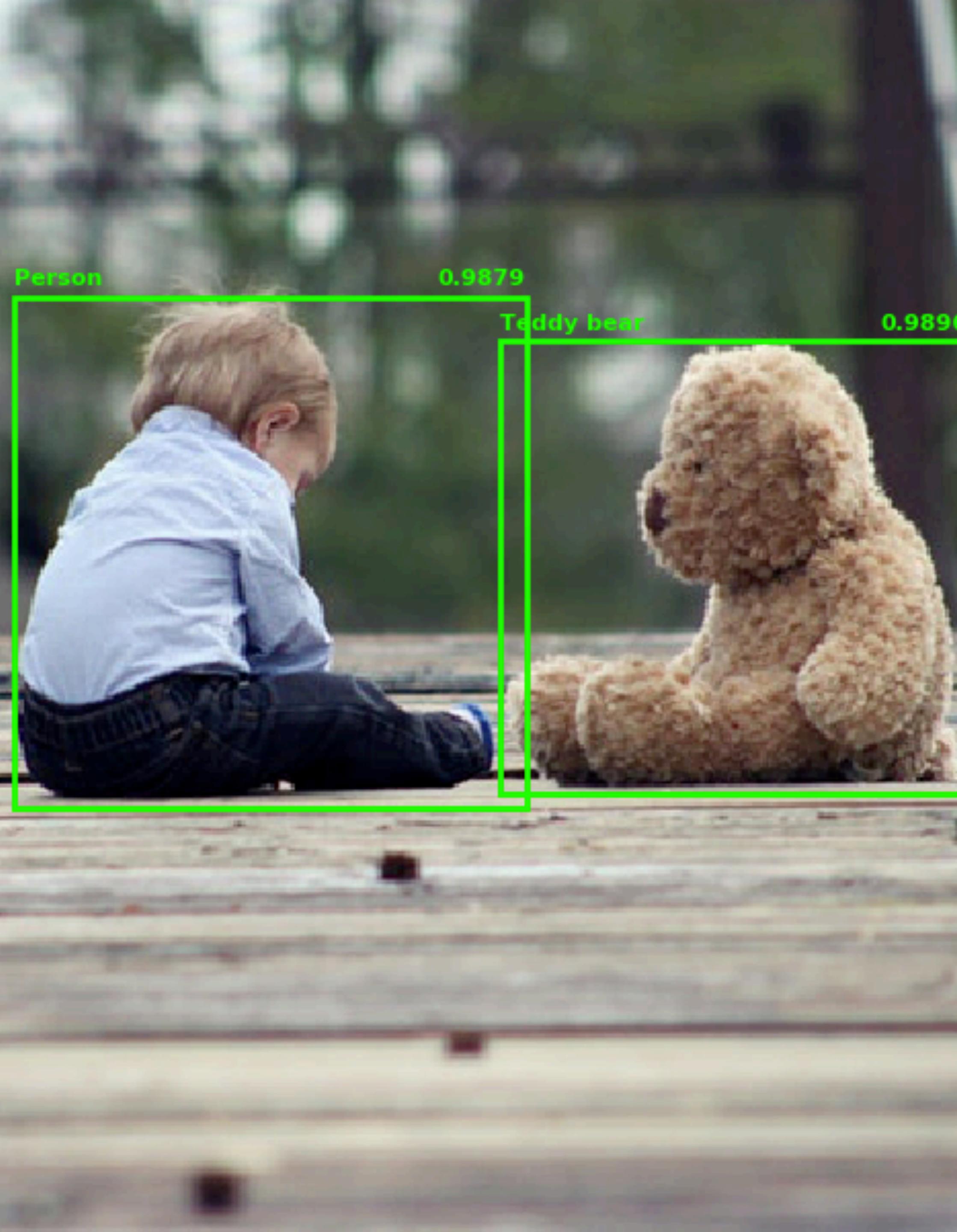


docker

<https://www.docker.com>

Ways of accessing the models





OBJECT DETECTOR

Localize and identify multiple objects in a single image

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →
Join the community →

Featured Deployable Trainable

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019 →

Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019 →

Model | Deployable, Trainable

Image Segmente

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

Mar 29, 2019 →

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018 →

Model | Deployable, Trainable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018 →

Model | Deployable

Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018 →

ibm.biz/model-exchange

Model Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Get this model



→

Try the API

→

Try the web app

→

Try in a Node-RED flow

→

Object Detector

Localize and identify multiple objects in a single image.

Get this model



Try the API



Try the web app



Try in a Node-RED flow



[IBM / MAX-Object-Detector](#)

Watch 37 Star 120 Fork 98

Code Issues 6 Pull requests Actions Projects Security Insights

master ▾ Go to file Code ▾

SSaishruthi Training Update (#145) ... ✓ 22 days ago 106

Folder	Commit Message	Date
api	Add license headers (#55)	13 months ago
core	code cleanup (#64)	13 months ago
docs	[ImgBot] Optimize images (#59)	13 months ago
protos	code cleanup (#64)	13 months ago
samples	Add training test sample image (#119)	5 months ago
tests	Training Update (#145)	22 days ago
training	Training Update (#145)	22 days ago
utils	code cleanup (#64)	13 months ago

About
Localize and identify multiple objects in a single image.

[🔗 developer.ibm.com/exchan...](#)

docker-image
machine-learning
machine-learning-models
coco-dataset
tensorflow-model

Readme
Apache-2.0 License

Deployment options

- [Deploy from Docker Hub](#)
- [Deploy on Red Hat OpenShift](#)
- [Deploy on Kubernetes](#)
- [Run Locally](#)

Access the API via Swagger

Model | Deployable, Trainable

Object Detector

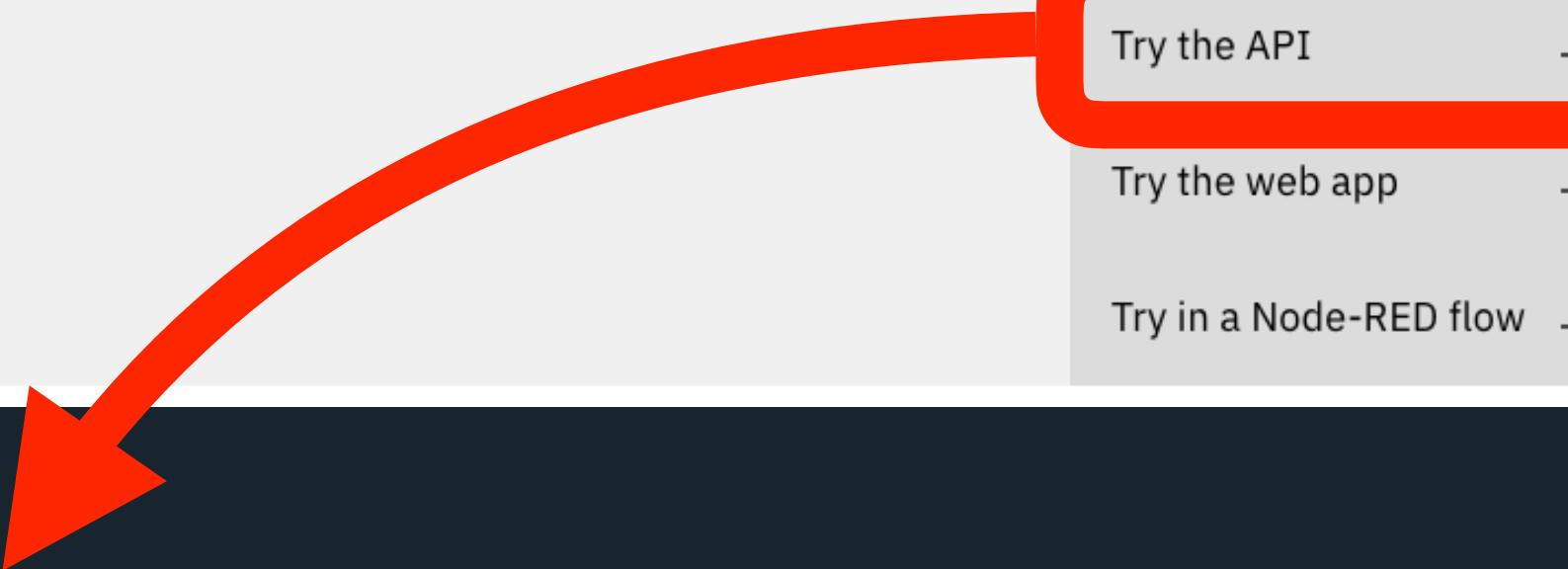
Localize and identify multiple objects in a single image.

Get this model →

Try the API → **Try the API** →

Try the web app →

Try in a Node-RED flow →



POST /model/predict Make a prediction given input data

Parameters

Name	Description
image * required	An image file (encoded as PNG or JPG/JPEG)
file (formData)	<input type="button" value="Choose File"/> traffic.jpeg
threshold number (query)	Probability threshold for including a detected object in the response in the range [0, 1] (default: 0.7). Lowering the threshold includes objects the model is less certain about. 0.7

Execute **Clear**

Responses

Response content type application/json

Curl

```
curl -X POST "http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/model/predict?threshold=0.7" -H "accept: application/json" -H "Content-Type: multipart/form-data" -F "image=@traffic.jpeg;type=image/jpeg"
```

Request URL

```
http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/model/predict?threshold=0.7
```

Server response

Code Details

200 Response body

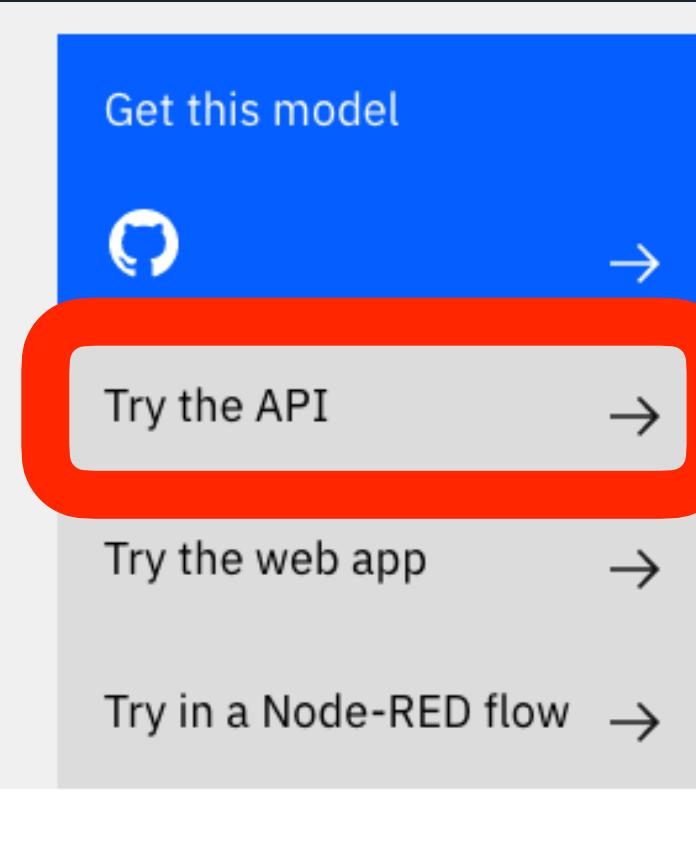
```
{
  "status": "ok",
  "predictions": [
    {
      "label_id": "3",
      "label": "car",
      "probability": 0.9741649627685547,
      "detection_box": [
        0.4575640559196472,
        0.4168500304222107,
        0.6725007891654968,
        0.9029390215873718
      ]
    },
    {
      "label_id": "1",
      "label": "person",
      "probability": 0.8824045658111572,
      "detection_box": [
        0.6231690645217896,
        0.30522748827934265,
        0.8462619781494141,
        0.4034259617328644
      ]
    },
    {
      "label_id": "4",
      "label": "motorcycle",
      "probability": 0.8141902685165405,
      "detection_box": [
        0.3583410978317261,
        0.11843161284923553,
        0.5586568117141724,
        0.37292349338531494
      ]
    }
  ]
}
```

Access the API via Python

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.



```
# Model
url = 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09
db52-0000.us-east.containers.appdomain.cloud/'
model_endpoint = 'model/predict'
complete_url = url + model_endpoint

# Upload an image to the MAX model's rest API
path_to_input_image = 'baby-bear.jpg'

with open(path_to_input_image, 'rb') as file:
    file_form = {'image': (path_to_input_image, file, 'image/jpeg')}
    # Post the image to the rest API using the requests library
    r = requests.post(url=complete_url, files=file_form)
    # Return the JSON
    response = r.json()

IPython.display.Image(path_to_input_image, width = 450)
```

response

```
{'status': 'ok',
'predictions': [{label_id': '88',
'label': 'teddy bear',
'probability': 0.9896332025527954,
'detection_box': [0.27832502126693726,
0.5611844062805176,
0.643224835395813,
0.8432191610336304]},
{'label_id': '1',
'label': 'person',
'probability': 0.9879012107849121,
'detection_box': [0.24251867830753326,
0.26926860213279724,
0.655893087387085,
0.5768759250640869]}]}
```

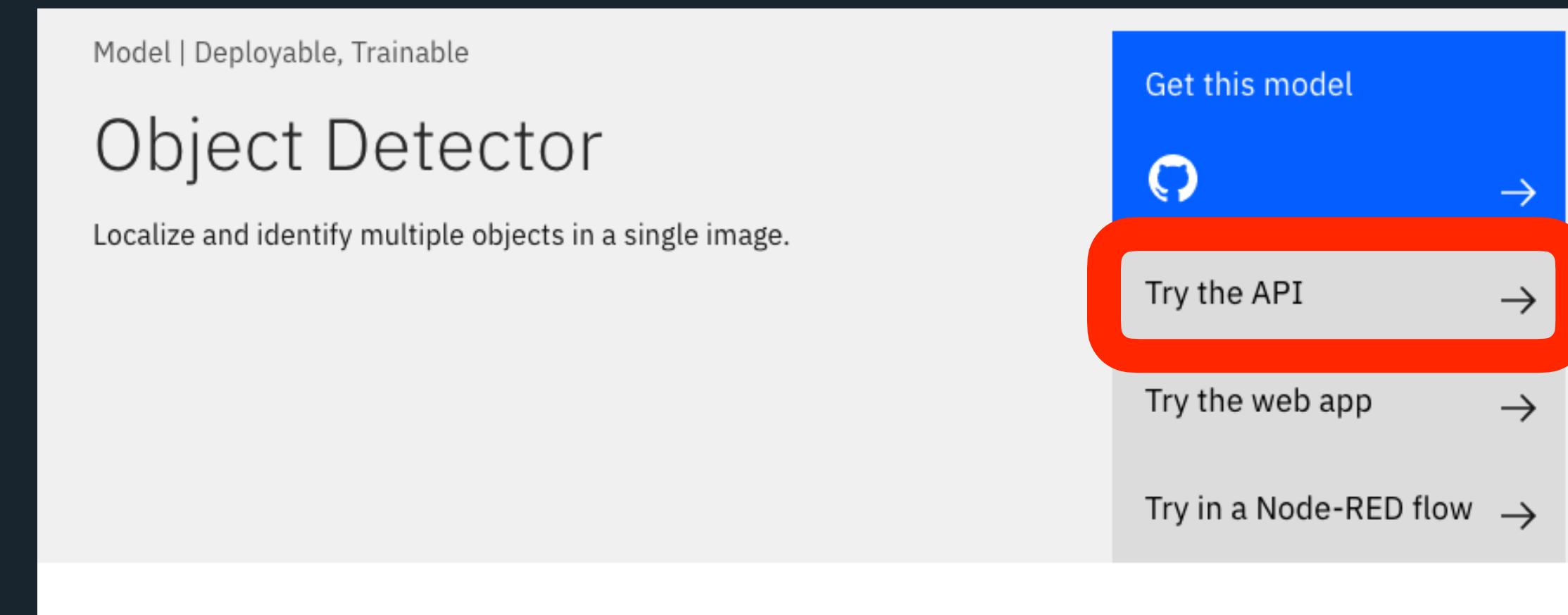
Try yourself here:
ibm.biz/max-notebook



IBM Cloud

Se quiser rodar esse notebook na cloud,
registre-se aqui: ibm.biz/tdc-sao-paulo

Access the API via R



```
library(dplyr)
library(httr)

# Endpoint
endpoint <- 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/'
# endpoint <- 'http://localhost:5000' # if running docker locally or docker hub

object_detector <- function(path_to_img, endpoint) {
  model_endpoint <- paste0(endpoint, 'model/predict') # Model endpoint
  # POST
  response <- httr::POST(url = model_endpoint,
                          body = list(image = upload_file(path_to_img,
                                                          type = "image/jpeg")),
                          encode = c("multipart"))
  ) %>% content()
  response$predictions
}

# Get the image file from GH
download.file(url = "http://github.com/IBM/MAX-Object-Detector/blob/master/samples/baby-bear.jpg?raw=true",
              'baby-bear.jpg', mode = 'wb')

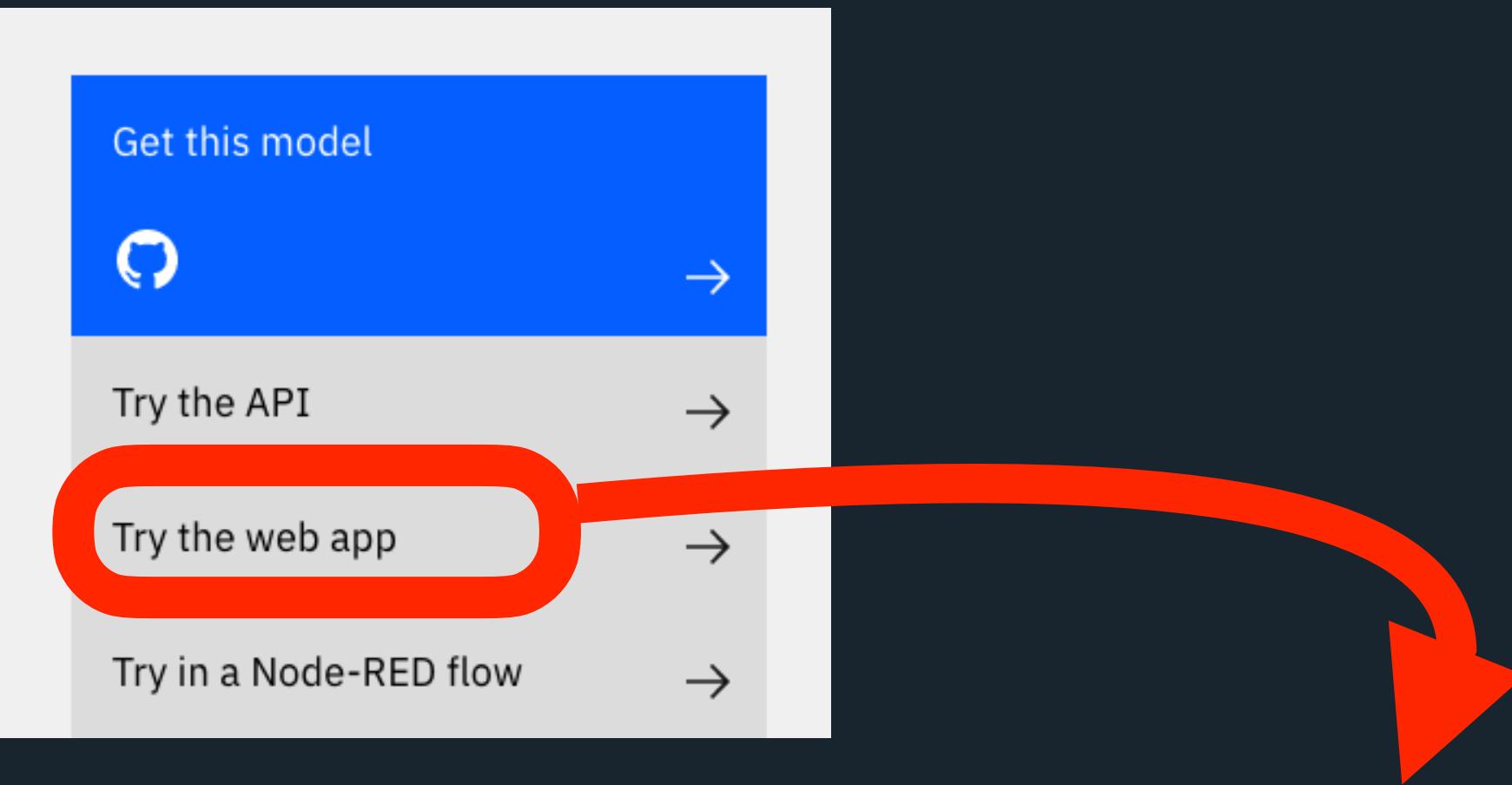
object_detector("baby-bear.jpg", endpoint)
```

Access the API via Web App

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

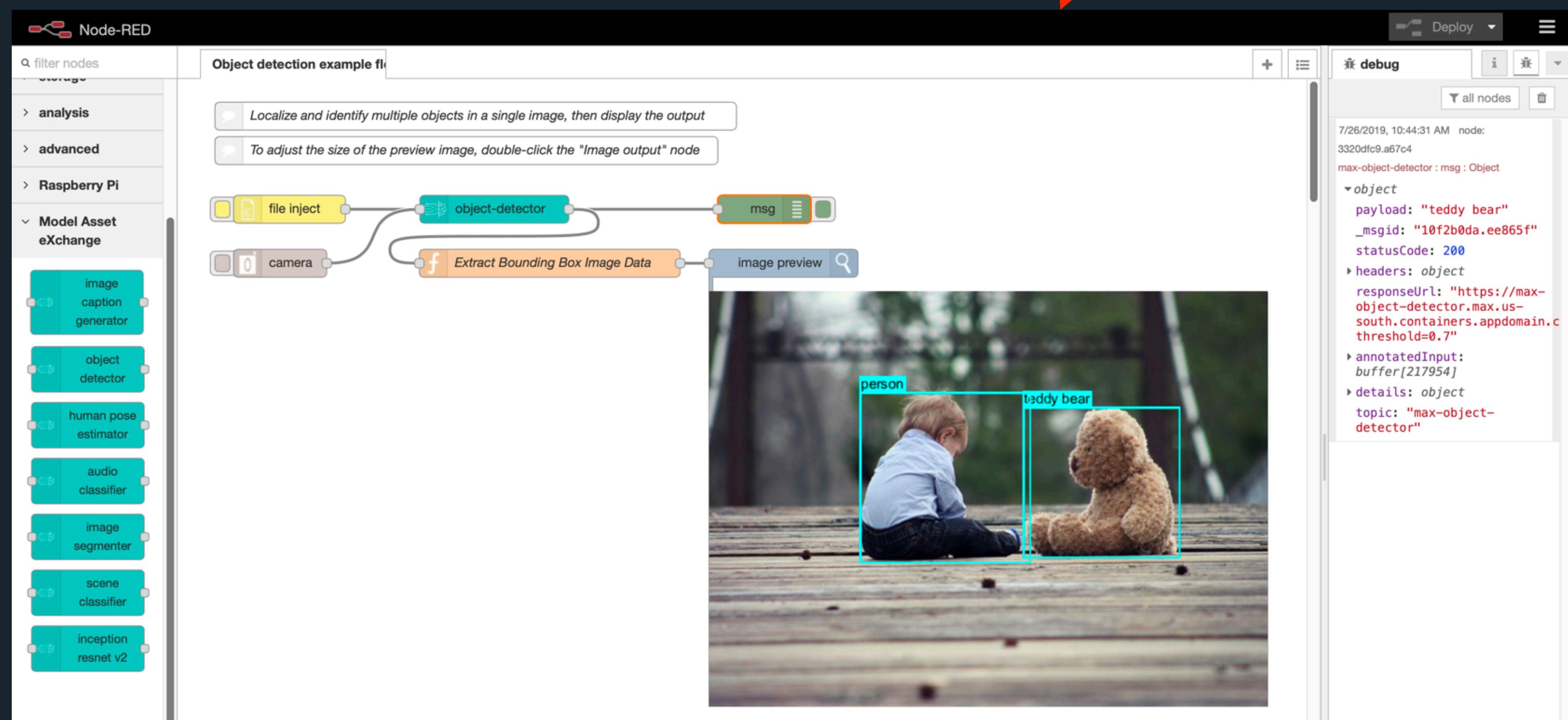
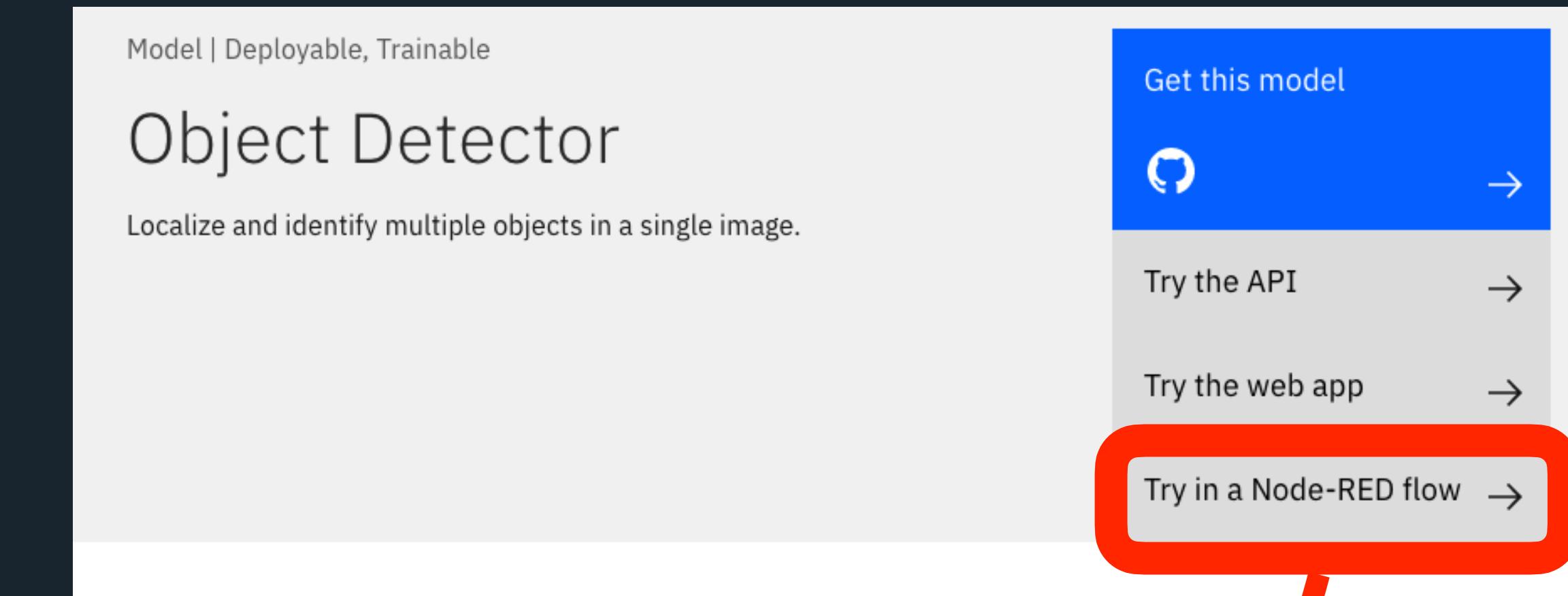


The screenshot shows the MAX Object Detector web application. It features a logo with a brain and a bar chart, followed by the text "MAX Object Detector". On the left, there's an "Upload an image" section with a "Choose File" button (No file chosen) and a "Submit" button. Next to it is a "Filter detected objects" section with a "Probability Threshold" slider set at 70%. On the right, there's a "Labels Found" section showing icons for a person and a dog. Below these controls is a preview image of a woman sitting on the grass with a dog. Two green bounding boxes are drawn around them, with text labels indicating the detected objects and their confidence scores: "person : 88.9%" and "dog : 81.2%".

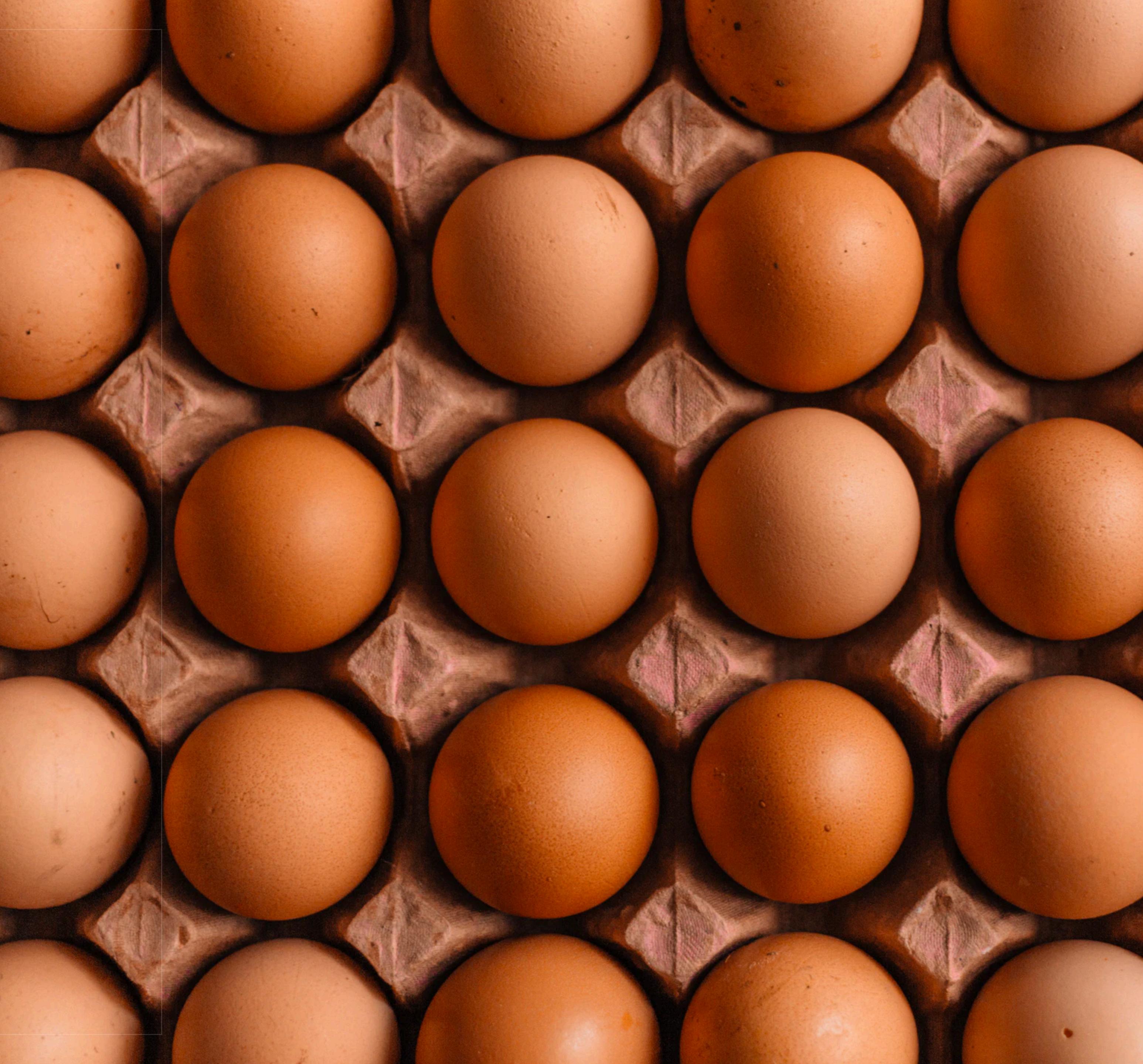
Try yourself here:
ibm.biz/object-detector-webapp

Other ways...

- using cURL
- Node-RED flow
- CodePen
- Serverless app



All this in a
standardized way



Overview

This model recognizes the objects present in an image from the 80 different high-level classes of objects in the [COCO Dataset](#). The model consists of a deep convolutional net base model for image feature extraction, together with additional convolutional layers specialized for the task of object detection, that was trained on the COCO data set. The input to the model is an image, and the output is a list of estimated class probabilities for the objects detected in the image.

The model is based on the [SSD Mobilenet V1 object detection model for TensorFlow](#).

Model Metadata

Domain	Application	Industry	Framework	Training Data	Input Data Format
Vision	Object Detection	General	TensorFlow	COCO Dataset	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	Apache 2.0	LICENSE
Model Weights	Apache 2.0	TensorFlow Models Repo
Model Code (3rd party)	Apache 2.0	TensorFlow Models Repo
Test Assets	CC0	Samples README

BEHIND THE SCENES

Find* a state-of-art open source deep learning model specific to domain

Validate license terms

Perform model health check & code clean up

Wrap models in MAX framework and provide REST API

Publish the deployable model as Docker images on Docker Hub

Use the MAX training framework to create an image for custom model training

Review and Continuous Integration

* or build from scratch

And if you are
feeling
adventurous...



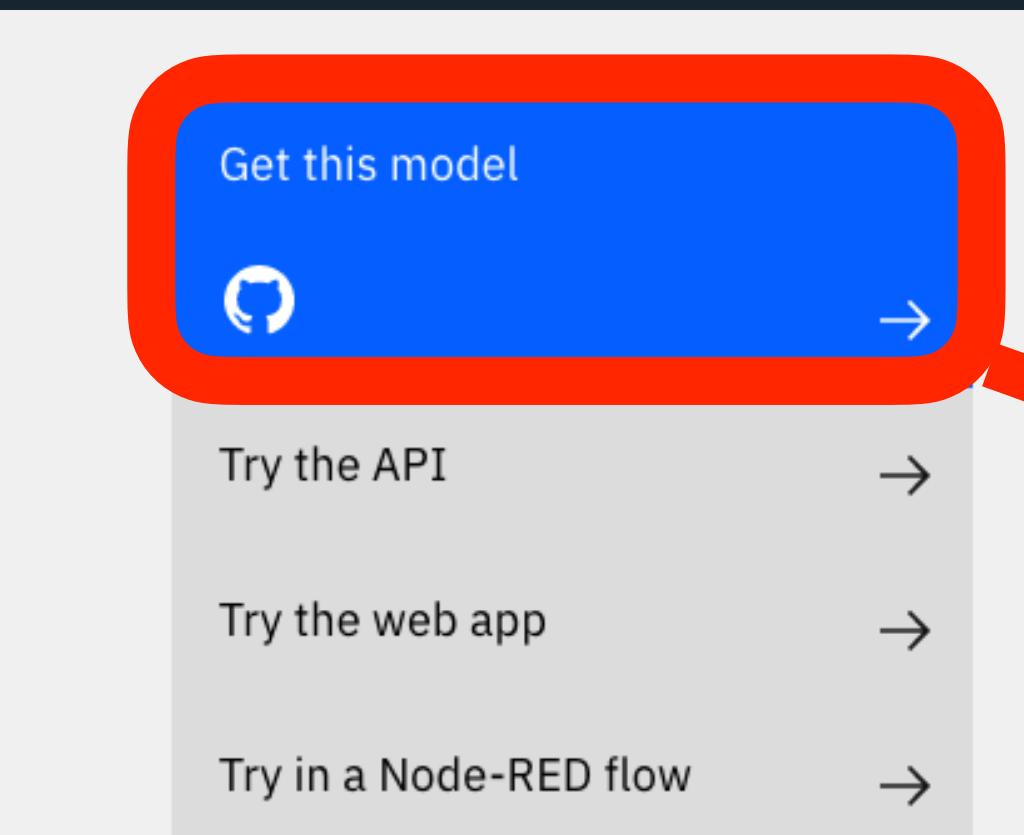
You can train your model using your own data



Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.



IBM / MAX-Object-Detector

Unwatch 33 Unstar 53 Fork 33

Code Issues 3 Pull requests 1 Actions Projects 0 Wiki Security Insights Settings

Localize and identify multiple objects in a single image. <https://github.com/IBM/MAX-Object-Detector> Edit

docker-image machine-learning machine-learning-models coco-dataset tensorflow-model Manage topics

62 commits 3 branches 0 packages 4 releases 17 contributors Apache-2.0

Branch: master New pull request Create new file Upload files Find file Clone or download

ptitzler	use training framework package (#78)	Latest commit 0bb2727 5 hours ago
api	Add license headers (#55)	4 months ago
core	code cleanup (#64)	3 months ago
docs	[ImgBot] Optimize images (#59)	4 months ago
protos	code cleanup (#64)	3 months ago
samples	WML Training (#61)	2 months ago
tests	WML Training (#61)	2 months ago
training	use training framework package (#78)	5 hours ago
utils	code cleanup (#64)	3 months ago



How to Train Object Detector Model Using Your Own Data

- [Collect Data for Training](#)
- [Train the Model](#)
- [Rebuild the Model-Serving Microservice](#)

Collect Data for Training

Collect RGB images encoded as jpeg or png containing objects that need to be detected. Make sure the training images have large variations in angle, resolution, lighting and background so that they generalize well with the test data. Use a reasonably large number of images per class to provide better results.

Train the Model

- [Install Local Prerequisites](#)
- [Run the Setup Script](#)
- [Prepare Data for Training](#)
- [Customize Training](#)
- [Train the Model Using Watson Machine Learning](#)



How do I get started?

ibm.biz/max-tutorial

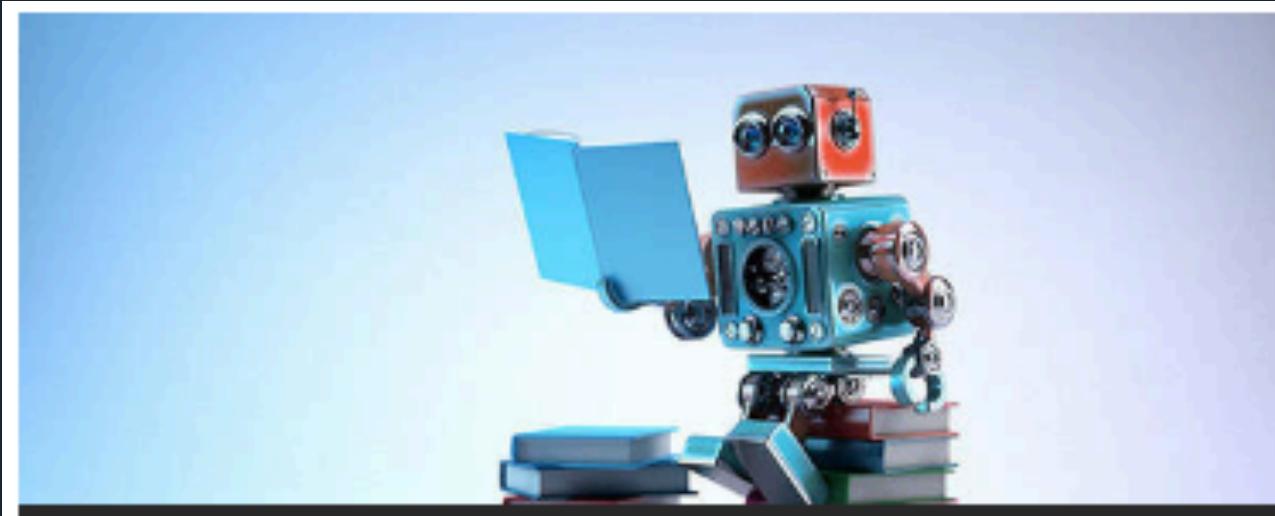
Series

Learning Path: An introduction to the Model Asset Exchange

Learn how to use state-of-the-art deep learning models in your applications or services

Examples on how to easily consume MAX models

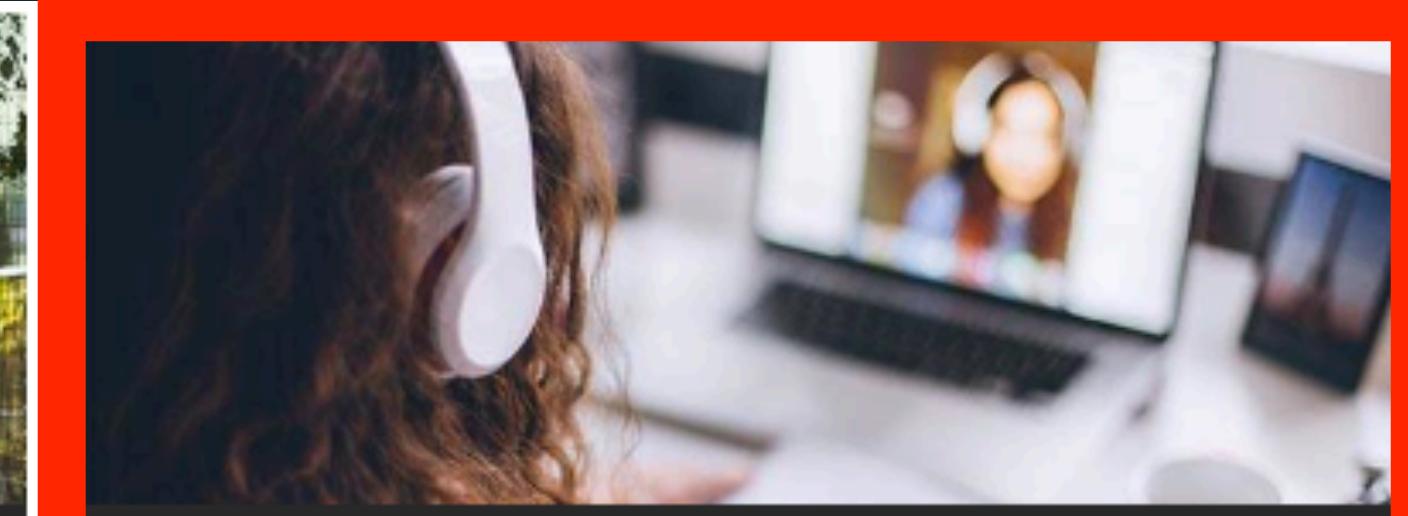
ibm.biz/max-code-patterns



Code Pattern
Create a machine learning powered web app to answer questions
Nov 05, 2019 →



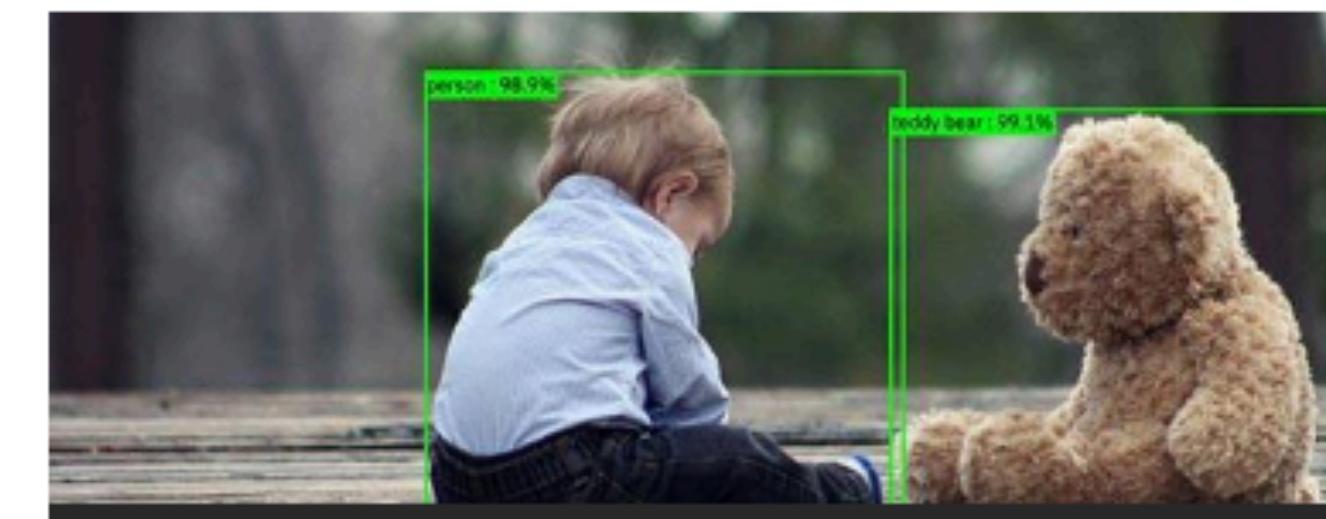
Code Pattern
Build a web app that recognizes yoga poses using a model from the Model Asset Exchange
Oct 03, 2019 →



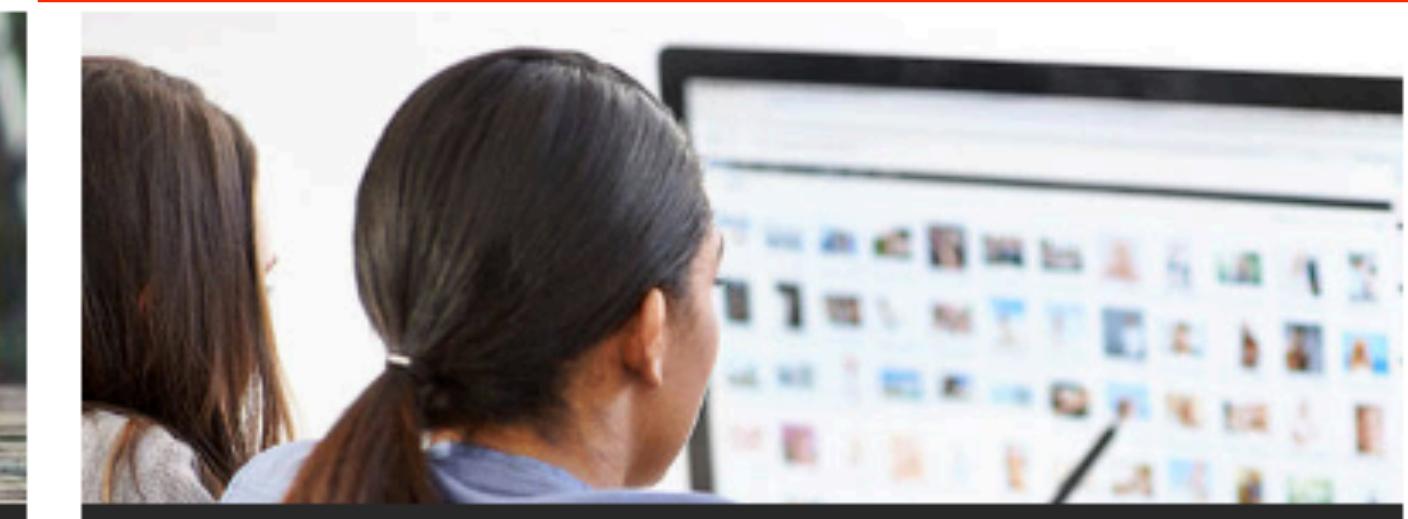
Code Pattern
Use your arms to make music
Apr 22, 2019 →



Code Pattern
Create a web app to interact with machine learning generated image captions
Mar 28, 2019 →



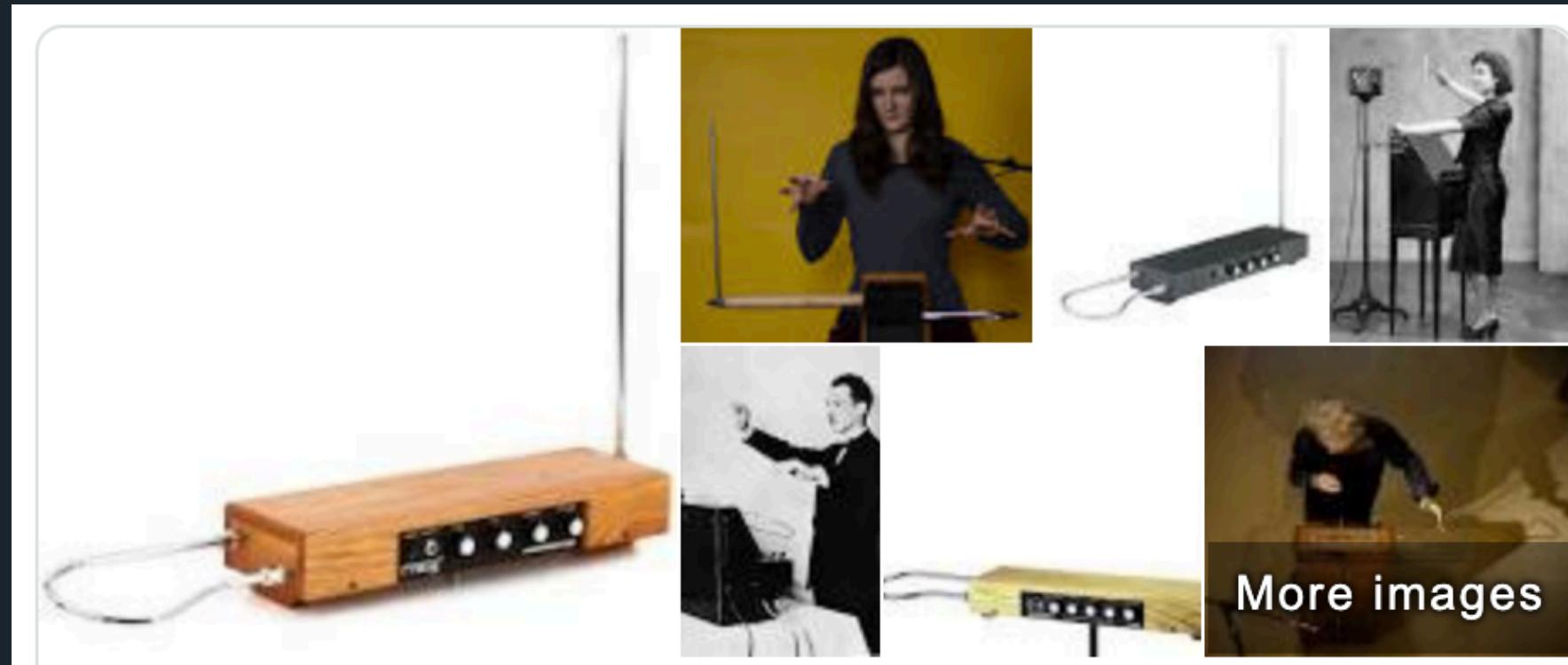
Code Pattern
Create a web app to visually interact with objects detected using machine learning
Mar 28, 2019 →



Code Pattern
Deploy a deep learning-powered 'Magic cropping tool'
Mar 28, 2019 →

Use your arms to make music

Create music with your arms using the Model Asset eXchange (MAX) **human pose estimator** model and **TensorFlow**



Theremin
Musical instrument

The theremin is an electronic musical instrument controlled without physical contact by the thereminist. It is named after its inventor, Léon Theremin, who patented the device in 1928. [Wikipedia](#)

Instrument family: Electronic Musical Instruments, Musical Keyboards

Invented: 1920

Related instrument: Ondes Martenot, Electro-Theremin

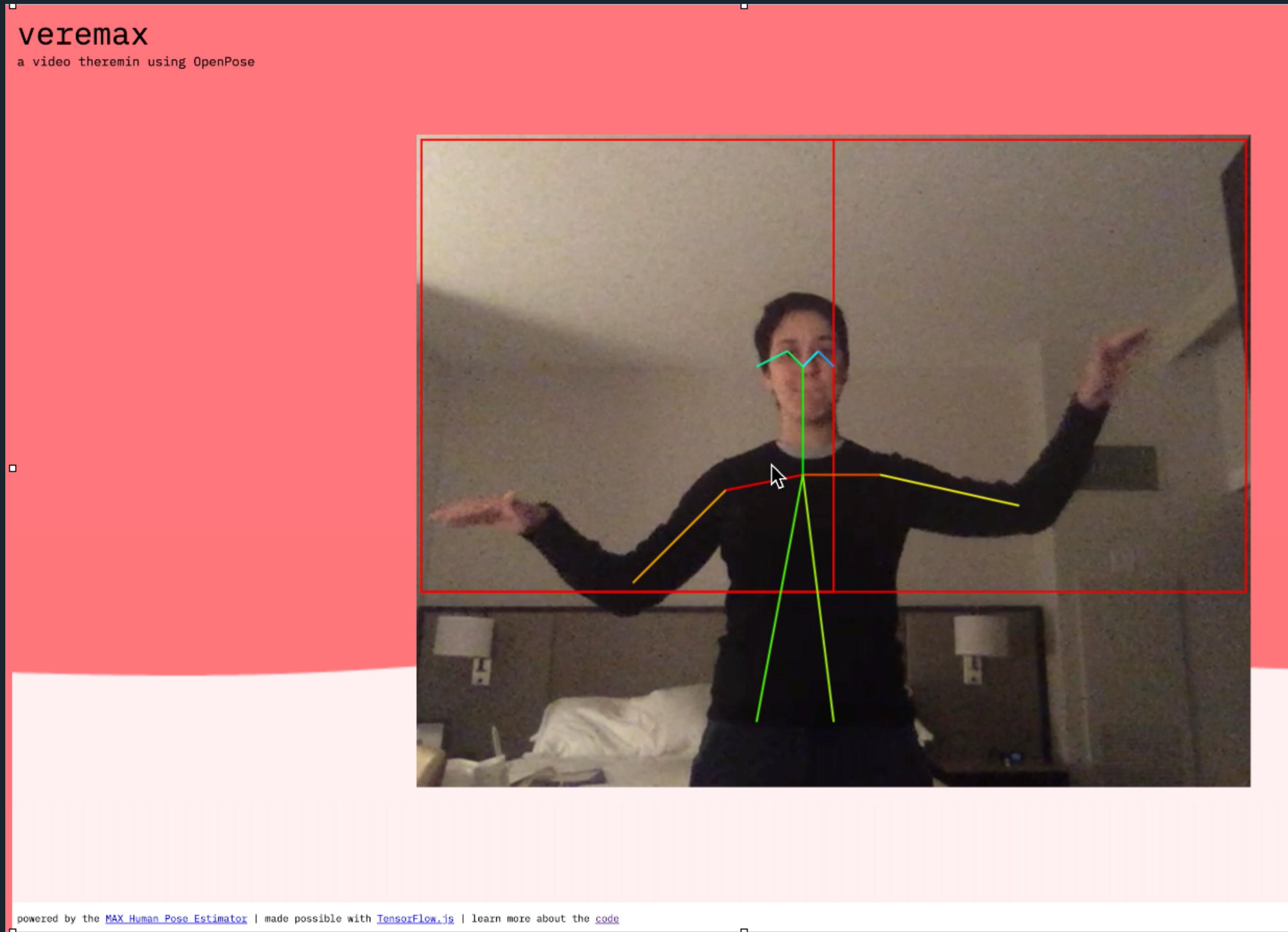
Inventor: Léon Theremin

Hornbostel–Sachs classification: 531.1; (Electrophone)



Use your arms to make music

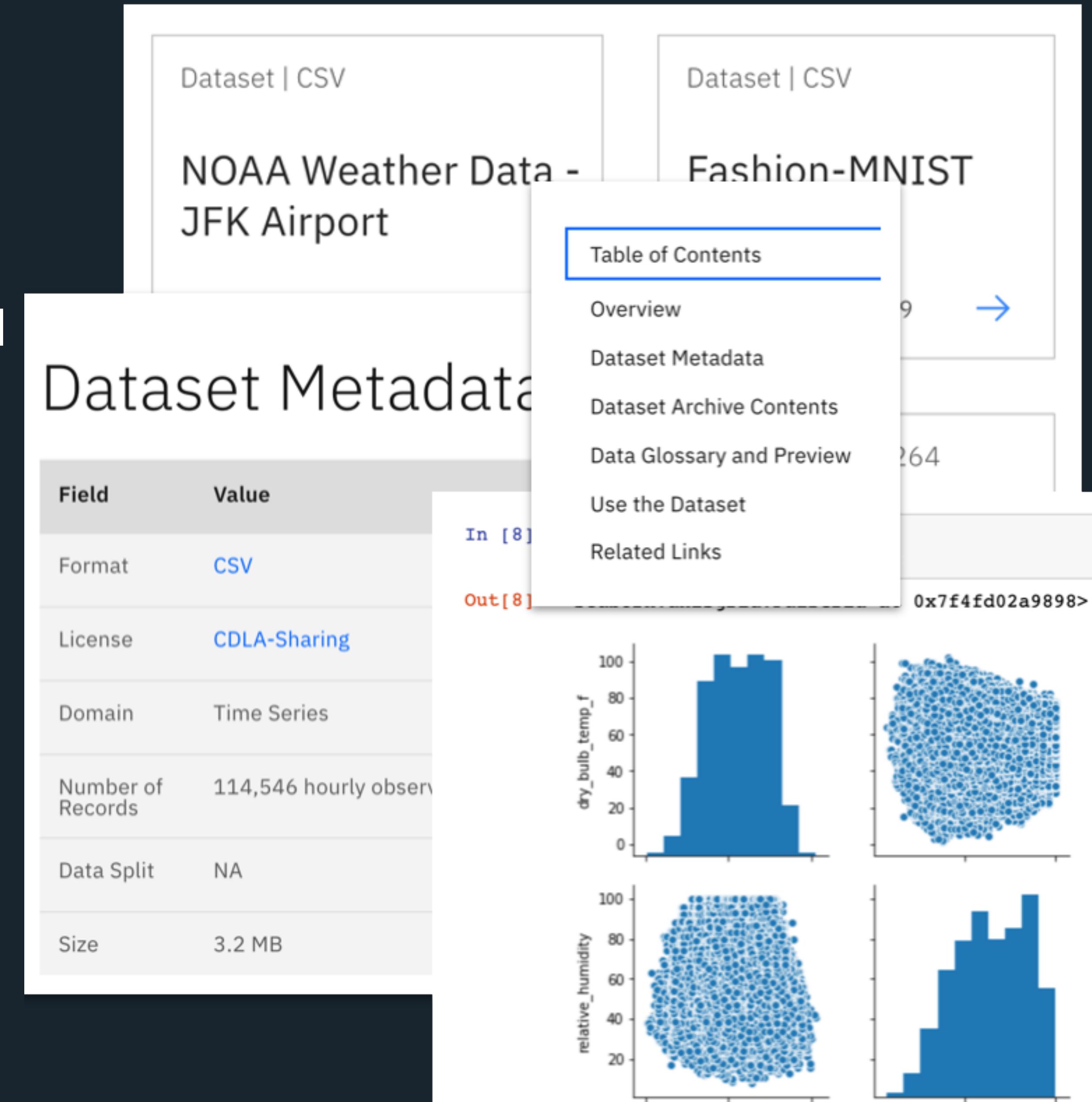
Create music with your arms using the Model Asset eXchange (MAX) **human pose estimator** model and **TensorFlow**



Data Asset eXchange (DAX)

- Curated repository for **open** datasets from IBM Research and third-parties
- Published under data friendly licenses
- Standardized dataset formats and metadata
- Many data sets include starter notebooks (cleansing, data exploration, analysis)

ibm.biz/data-exchange



Data Asset eXchange

Explore useful and relevant data sets for enterprise data science

ibm.biz/data-exchange

Dataset CSV NOAA Weather Data - JFK Airport August 11, 2020 →	Dataset IOB format Groningen Meaning Bank - Modified May 14, 2020 →	Dataset CSV Fashion-MNIST August 17, 2020 →	Dataset CSV, H.264 Double Pendulum Chaotic September 12, 2019 →	Dataset CSV IBM Debater® Concept Abstractness July 29, 2019 →	Dataset XLSX, TXT IBM Debater® Sentiment Composition Lexicons August 1, 2019 →
Dataset JPG, JSON PubLayNet August 15, 2020 →	Dataset WAV TensorFlow Speech Commands August 17, 2020 →	Dataset PNG, JSON PubTabNet August 11, 2020 →	Dataset CSV IBM Debater® Claim Sentences Search August 1, 2019 →	Dataset ANN IBM Debater® Mention Detection Benchmark July 29, 2019 →	Dataset CSV IBM Debater® Wikipedia Category Stance August 1, 2019 →
Dataset JSON, HDF5 Oil Reservoir Simulations August 11, 2020 →	Dataset CoNLL-U Finance Proposition Bank August 11, 2020 →	Dataset CoNLL-U Contracts Proposition Bank August 11, 2020 →	Dataset CSV IBM Debater® Wikipedia Oriented Relatedness August 1, 2019 →	Dataset CSV IBM Debater® Thematic Clustering of Sentences August 3, 2019 →	Dataset XML Forum Classify September 12, 2019 →
Dataset XML Forum Summarization September 12, 2019 →	Dataset CSV Expert in the Loop AI - Polymer Discovery June 15, 2020 →	Dataset Parquet Mono Lake Surface Water Extent Landsat8 Data May 6, 2020 →	Dataset TSV, TXT WebQSP Relation Detection May 7, 2020 →	Dataset CSV IBM Debater® Multi Word Term Relatedness Benchmark July 29, 2019 →	Dataset WAV, TXT IBM Debater® Recorded Debating #1 June 29, 2020 →
Dataset Text WikiText-103 June 23, 2020 →	Dataset TSV, TXT SimpleQuestions Relation Detection May 7, 2020 →	Dataset CSV Taranaki Basin Curated Well Logs May 10, 2020 →	Dataset WAV, CSV, TXT IBM Debater® Recorded Debating #2 August 3, 2019 →	Dataset WAV, CSV, TXT IBM Debater® Recorded Debating #3 August 12, 2019 →	Dataset TSV IBM Debater® Sentiment Lexicon of IDiomatic Expressions (SLIDE) September 5, 2019 →
Dataset MP4, CSV Video-Text Compliance June 29, 2020 →	Dataset CSV, JSON Nutch July 16, 2019 →	Dataset TSV Wikipedia Entity Graph April 21, 2020 →	Dataset TXT IBM Debater® Labeled Emphasized Words in Speech July 29, 2019 →	Dataset JSON Lines MedNLI September 17, 2019 →	Dataset JSON VizWiz - Visual Question Answering March 26, 2020 →

NOAA Weather Data – JFK Airport

Local climatological data originally collected at JFK airport.

Save Like

- Get this dataset →
- Run dataset notebooks →
- Preview the data & notebooks →

NOAA Weather Data – JFK Airport

Dataset Metadata

Dataset Preview

Dataset Glossary

Format	CSV
License	CDLA-Sharing
Domain	Time Series
Number of Records	114,546 hourly observations
Data Split	NA
Size	3.2 MB
Data Origin	National Oceanic and Atmospheric Administration (NOAA)
Dataset Version	Version 2 – September 12, 2019 Version 1 – July 16, 2019
Dataset Coverage	Location: New York City Dates: 2010-01-01 through 2018-07-27 Note: To download raw data from NOAA for a different region or date span, follow the steps outlined in the data archive's README.txt.
Agriculture	Detect unseasonal temperature change and alert farmers about potential damage to plants. Energy Regulate solar cell charging hours based on weather type condition and temperature. Regulate wind turbine operation based on wind speed and wind direction. Generate energy demand alerts based on temperature. Remotely adjust air conditioning configs to boost energy efficiency based on temperature shifts.

Business Use Case	Retail Estimate outdoor retail foot traffic based on weather condition and temperature predictions.
-------------------	--

NOAA Weather Data – JFK Airport

Part 1 - Data Cleaning

Part 2 - Data Analysis

Part 3 - Time Series Forecasting

```
In [1]: # @hidden_cell
# The project token is an authorization token that is used to access project resources like data sources, connections, and used by platform APIs.
from project_lib import Project
project = Project(project_id='...', project_access_token='...')
```

Cleaning NOAA Weather Data of JFK Airport (New York)

This notebook relates to the NOAA Weather Dataset - JFK Airport (New York). The dataset contains 114,546 hourly observations of 12 local climatological variables (such as temperature and wind speed) collected at JFK airport. This dataset can be obtained for free from the IBM Developer [Data Asset Exchange](#).

In this notebook, we clean the raw dataset by:

- removing redundant columns and preserving only key numeric columns
- converting and cleaning data where required
- creating a fixed time interval between observations (this aids with later time-series analysis)
- filling missing values
- encoding certain weather features

Table of Contents:

- [0. Prerequisites](#)
- [1. Read the Raw Data](#)
- [2. Clean the Data](#)
 - [2.1 Select data columns](#)
 - [2.2 Clean up precipitation column](#)
 - [2.3 Convert columns to numerical types](#)
 - [2.4 Reformat and process data](#)
 - [2.5 Create a fixed interval dataset](#)
 - [2.6 Feature encoding](#)
 - [2.7 Rename columns](#)



Se quiser rodar esse notebook na cloud, registre-se aqui:
ibm.biz/tdc-sao-paulo

Show me more!



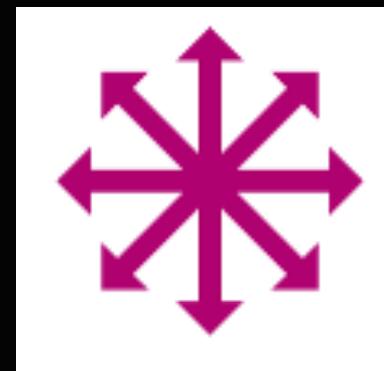
Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application



IBM and LFAI move forward on
trustworthy and responsible AI
IBM donates Trusted AI toolkits to the Linux Foundation AI

Did anyone tamper
with it?



ROBUSTNESS

Adversarial Robustness 360
↳ (ART)

DEMO: art-demo.mybluemix.net

Is it fair?



FAIRNESS

AI Fairness 360
↳ (AIF360)

DEMO: aif360.mybluemix.net

Is it easy to understand?



EXPLAINABILITY

AI Explainability 360
↳ (AIX360)

DEMO: aix360.mybluemix.net

Is it accountable?



LINEAGE

AI FactSheets 360

DEMO: aifs360.mybluemix.net

Trusted-AI

This GitHub org hosts LF AI Foundation projects in the category of Trusted and Responsible AI.

IBM
@LFAI_Foundation
info@lfaifoundation.org

Repositories 4 Packages People Projects

Pinned repositories

adversarial-robustness-toolbox
Adversarial Robustness Toolbox (ART) - Python Library for Machine Learning Security - Evasion, Poisoning, Extraction, Inference
Python ⭐ 1.7k 480

AIF360
A comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets and models.
Python ⭐ 1k 340

AIX360
Interpretability and explainability of data and machine learning models
Python ⭐ 621 139

AI Fairness 360 (AIF360) R Package
CRAN 0.1.0 CRAN 0.1.0

Available in R too!



THE DEVELOPER'S CONFERENCE

Obrigada! ❤



LAURA DAMACENO



AHIRTON LOPES



YARA SENGER



FACES
OF
OPEN
SOURCE

Photo by Peter Adams -
<http://www.facesofopensource.com/>

Thank you!

→ K-ROZ.COM

🐦 @GDEQUEIROZ

CONTACTKROZ@GMAIL.COM



ai-inclusive.org



INFO@AI-INCLUSIVE.ORG

Bonus Slides

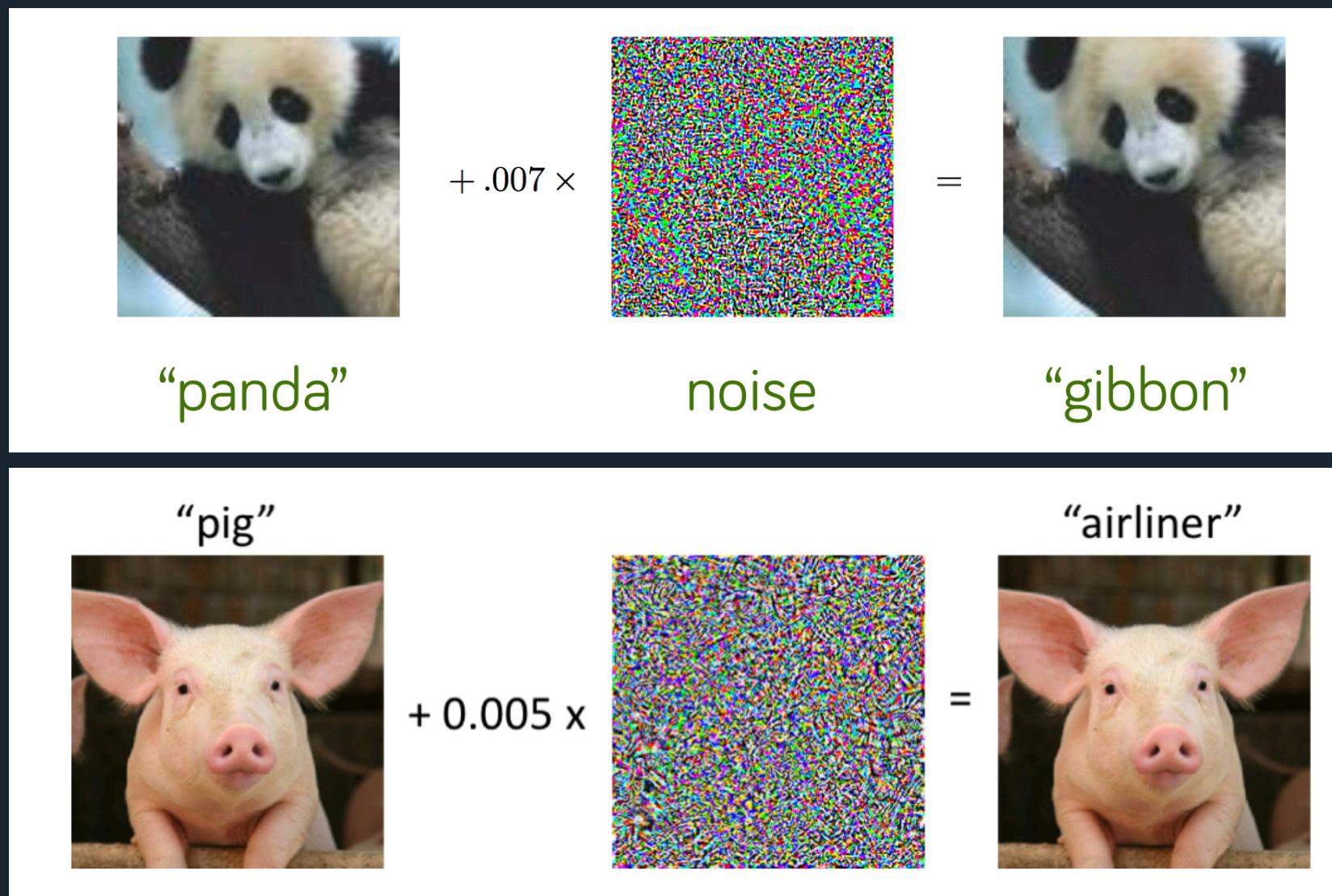
Adversarial Machine Learning

Adversarial machine learning can be used to “trick” machine learning models into providing incorrect predictions, often with devastating consequences
e.g. self driving vehicles



Adding small black and white stickers to stop signs, they could make them invisible to computer vision algorithms

Adversarial Attacks

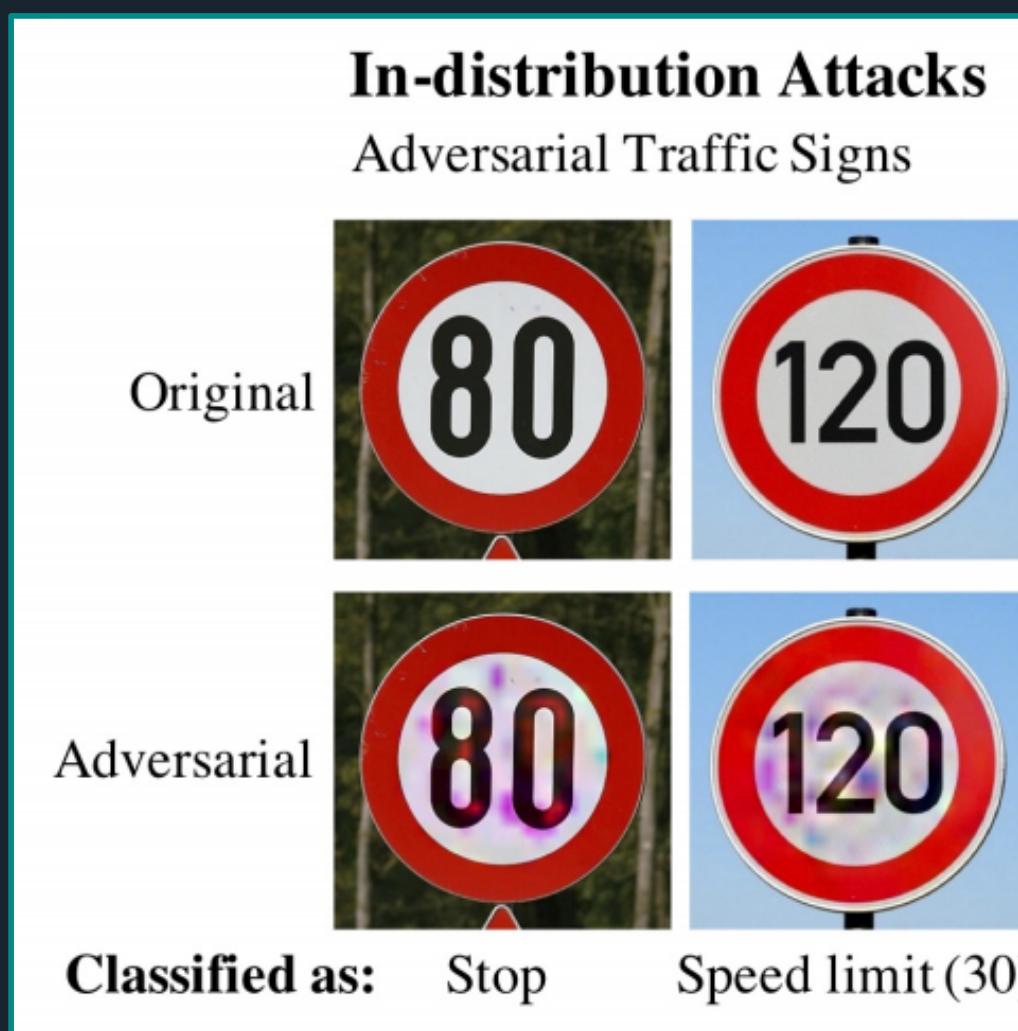


<https://arxiv.org/abs/1412.6572>



<https://arxiv.org/pdf/1712.09665.pdf>

Video: <https://www.youtube.com/watch?v=i1sp4X57TL4&feature=youtu.be>



<https://arxiv.org/pdf/1802.06430.pdf>



Adversarial Robustness Toolbox 360

↳ (ART)

ART is a Python library for machine learning security.

Its purpose is to allow rapid crafting and analysis of **attack and defense methods** for ML models.

Applicable domains include finance, self driving vehicles etc.

ART provides an implementation for many state-of-the-art methods for attacking and defending classifiers.

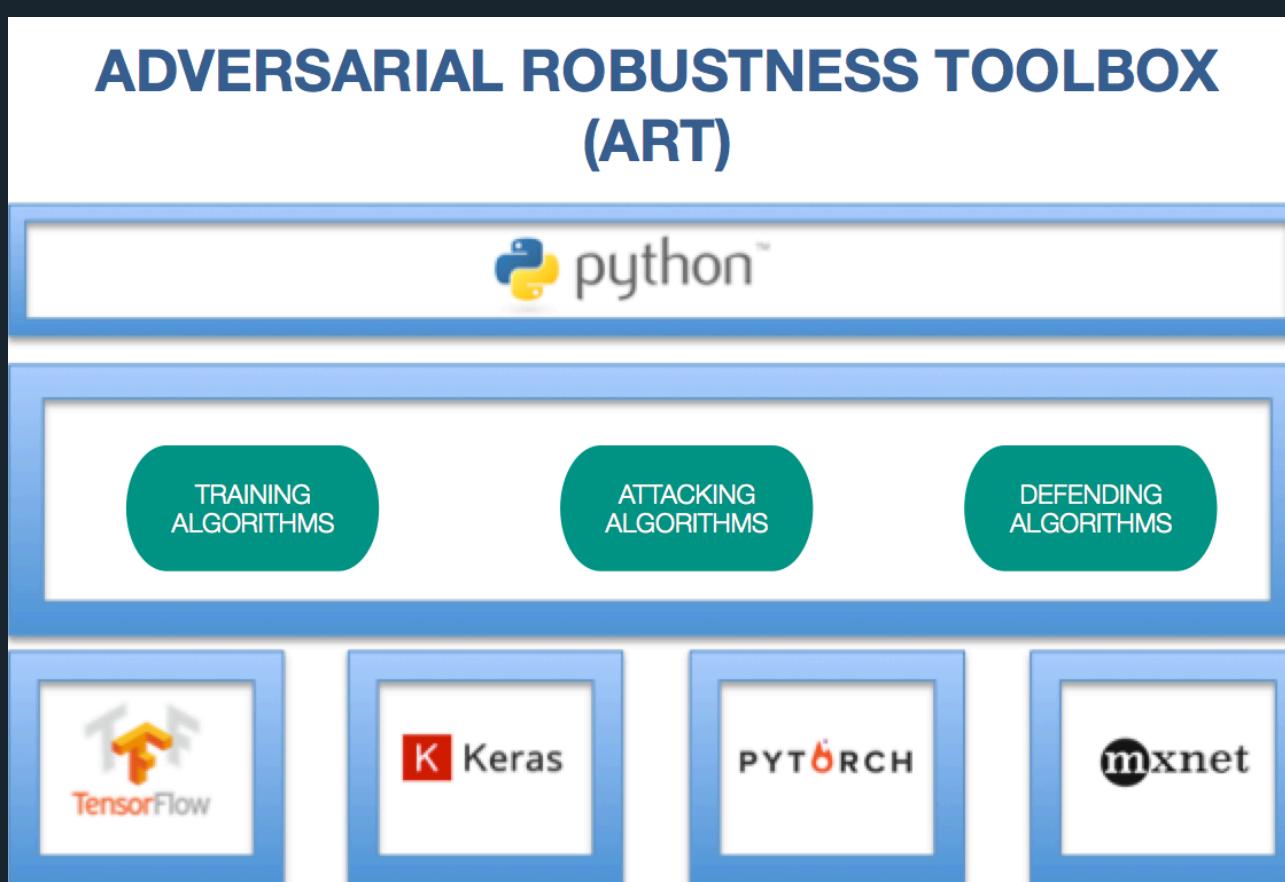
Toolbox

Evasion attacks

Defense methods

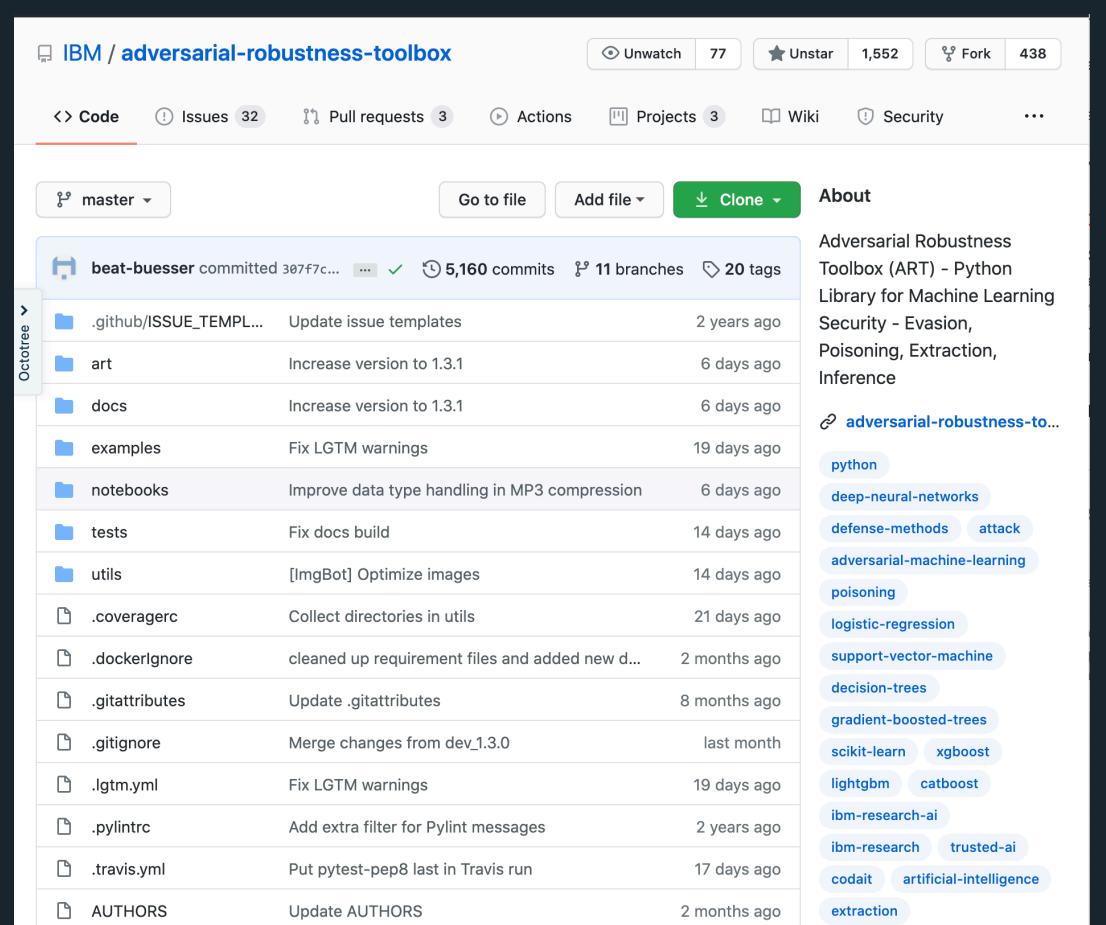
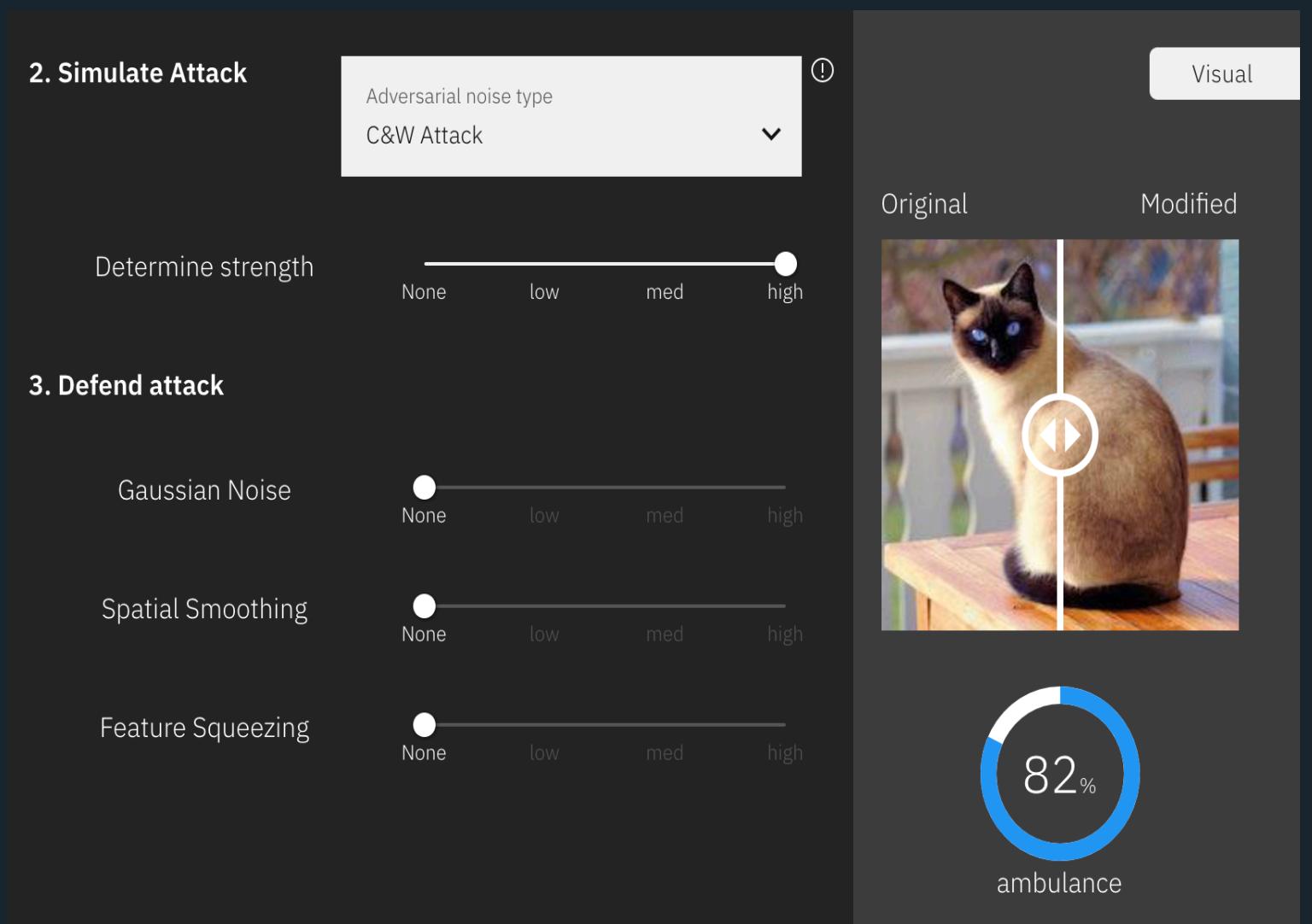
Detection methods

Robustness metrics



This screenshot shows the "Try it out" section of the ART web interface. It displays four target images: a mousetrap, a car wheel, a lightbulb, and a Siamese cat. Below these are sections for "2. Simulate Attack" and "3. Defend attack", each with sliders for "Determine strength" and dropdown menus for "Adversarial noise type". A circular progress bar indicates a success rate of 92% for the Siamese cat image.

<https://art-demo.mybluemix.net/>



AI Fairness 360

↳ **(AIF360)**

AIF360 toolkit is an open-source library to help detect and mitigate bias in machine learning models.

Applicable domains include finance, human capital management, healthcare, and education.

The AI Fairness 360 Python & R package includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.

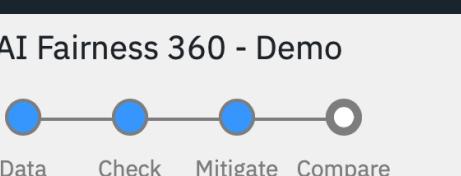
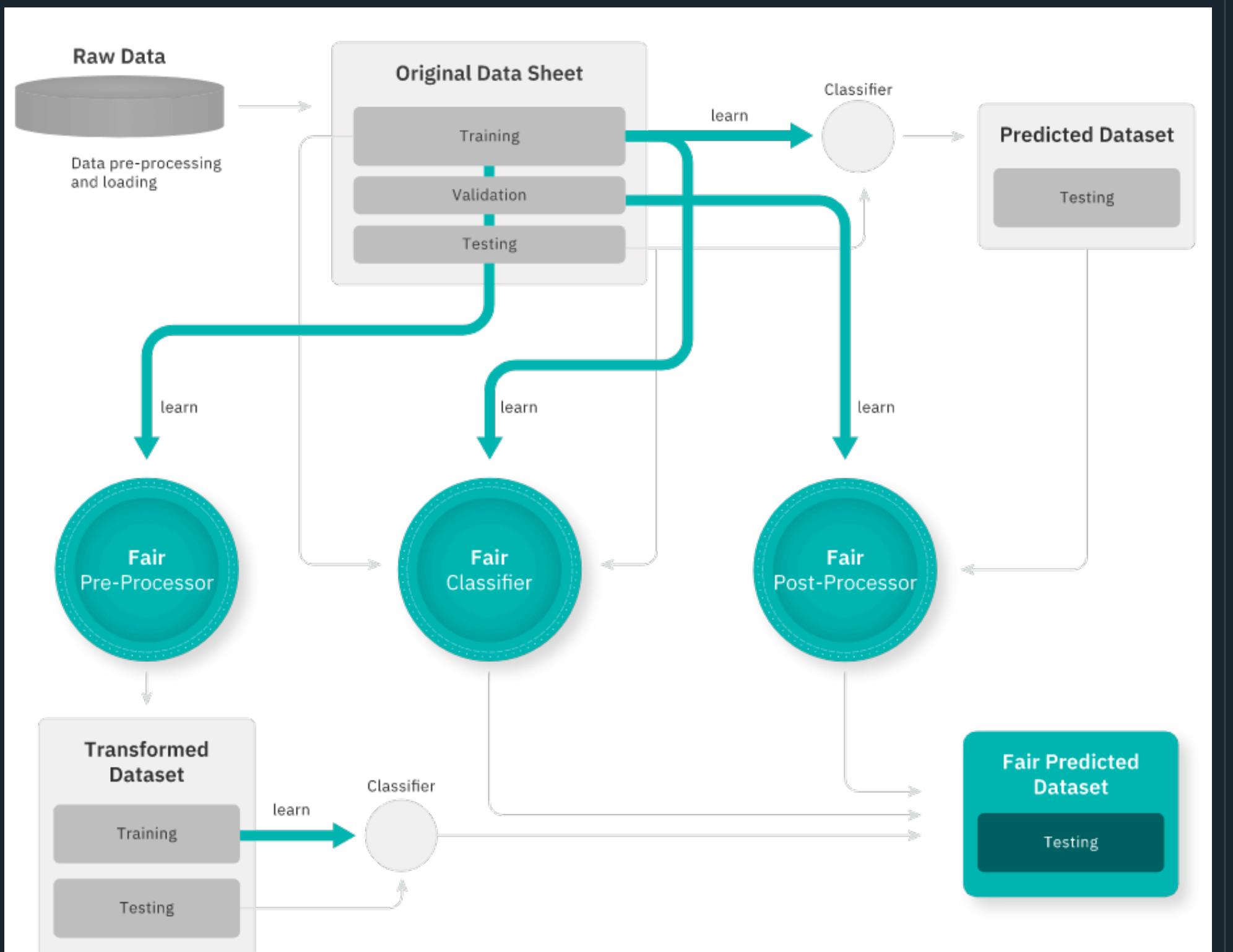
Toolbox

Fairness metrics (70+)

Fairness metric explanations

Bias mitigation algorithms (10+)

AIF360 Demo: <http://aif360.mybluemix.net>

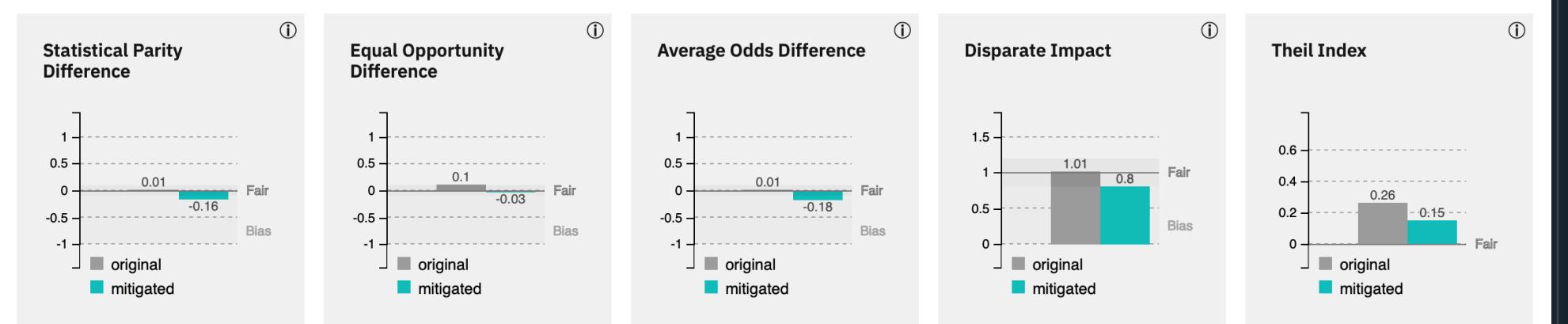


4. Compare original vs. mitigated results

Dataset: German credit scoring
Mitigation: Adversarial Debiasing algorithm applied

Protected Attribute: Sex

Privileged Group: **Male**, Unprivileged Group: **Female**
Accuracy after mitigation changed from 76% to 70%
Bias against unprivileged group unchanged after mitigation (0 of 5 metrics indicate bias)



AI Explainability 360

↳ (AIX360)

AIX360 toolkit is an open-source library to help **explain** AI and ML models and their predictions.

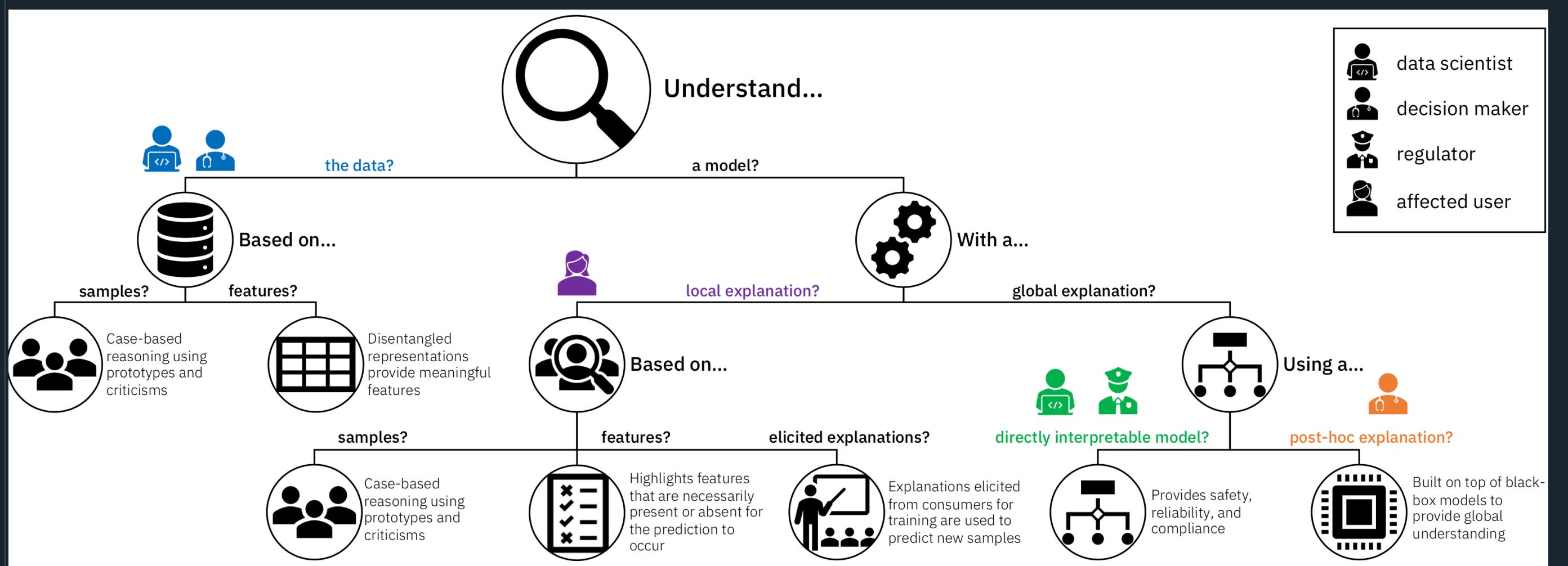
This includes 3 classes of algorithms: local post-hoc, global post-hoc, and directly interpretable explainers for models that use image, text, and structured/tabular data.

Toolbox

Local post-hoc

Global post-hoc

Directly interpretable



AI Explainability 360 - Demo

Data Consumer Explanation

A Bank Customer wants to understand:

Why was my application rejected?
What can I improve to increase the likelihood my application is accepted?

Providing Contrastive Explanations for Insight into Loan Application Outcomes

The Bank Customer wants to know how and why the decision was made to accept or reject their loan application. The explanation given will help them understand if they've been treated fairly, and also provide insight into what – if their application was rejected – they can improve in order to increase the likelihood it will be accepted in the future. To help provide that insight and suggest avenues for improvement, we will use the **Contrastive Explanations Method (CEM)** algorithm available in AI Explainability 360. This algorithm sits on top of an existing predictive model and helps detect both the features that a bank customer could improve (e.g., amount of time since last credit inquiry, average age of accounts), and also further detects the features that will increase the likelihood of approval and those that are within reach for the customer. See examples below.

Select a customer asking for explanations

Jason Denied

Ann Denied

Julia Denied

AIX360 Demo: <http://aix360.mybluemix.net>

FactSheets 360

↳ (AIFs360)

Transparency in AI Marketplace

Problems

Consumers of AI **services/models** have insufficient information about the service

Goal

Increase transparency (and trust) about the service/model by providing appropriate information

Challenge

How to increase transparency without mandating access to all of the “code”?

- Full “code” access is not always possible
- The code can be millions of lines
- Some “code” is really complex mathematical equations that are **not** understandable by even their developer

In 2018, IBM Research proposed the concept of AI FactSheets

FactSheets: Increasing Trust in AI Services through Supplier’s Declarations of Conformity

What is the **intended use** of the service output?
What **algorithms** or techniques does this service implement?

Which datasets was the service **tested** on?
Describe the **testing methodology** and **test results**.

Are you aware of possible examples of **bias**, **ethical issues**, or other **safety risks** as a result of using the service?

Are the service outputs **explainable** and/or interpretable?

For each dataset used by the service:

- Was the dataset checked for **bias**?
- What efforts were made to ensure that it is **fair** and **representative**?
- Does the service implement and perform any **bias detection** and **remediation**?

What is the **expected performance** on unseen data or data with different distributions?

Was the service checked for **robustness against adversarial attacks**?

When were the models last updated?



IBM researchers propose ‘factsheets’ for AI transparency

The screenshot shows the IBM Research AI FactSheets 360 website. The top navigation bar includes links for Home, Introduction, Methodology, Governance, Examples, Overview, Audio Classifier, Object Detector, Image Caption Generator, Resources, Our Papers, Related Work, Events, Videos, Slack Community, Glossary, and FAQs. The main content area features a section titled "AI FactSheets 360" with a sub-section "Learn More" containing links to "Introduction to FactSheets", "A Methodology for Creating AI FactSheets", and "AI Lifecycle Governance". Below this is a "Examples" section with cards for "Audio Classifier", "Object Detector", and "Image Caption Generator", each with sub-links like "Author Notes", "Ensemble", and "Robustness". A sidebar on the right contains a graphic of three overlapping documents.

<https://aifs360.mybluemix.net/>