

Harvard

Yá'át'ééh 🖐️

# CRESTLEX 3.0

**CR**eating **E**ffective **ST**em  
Learning **EX**periences

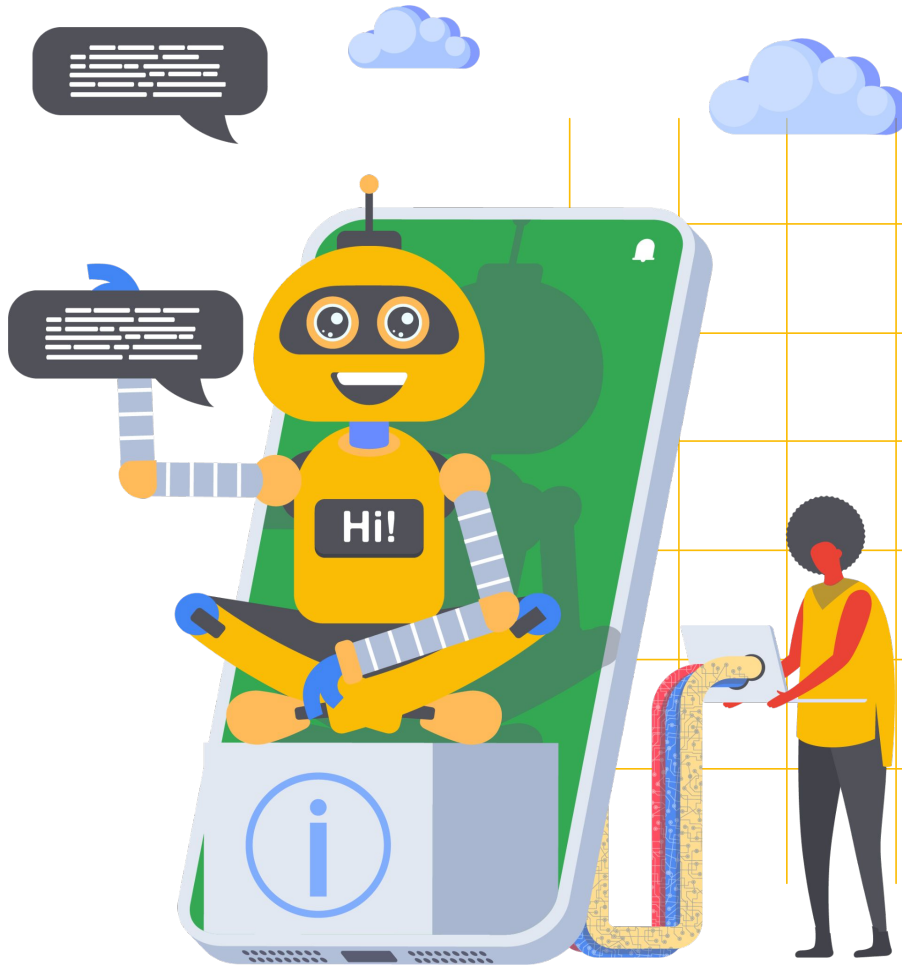
with Navajo Tech



Harvard

# How ML works?

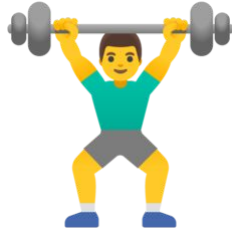
with Professor VJ



# Review what we've learned

Machine learning provides a computer with data, **rather than explicit instructions**. Using these data, the computer learns to **recognize patterns** and becomes able to execute tasks on its own.

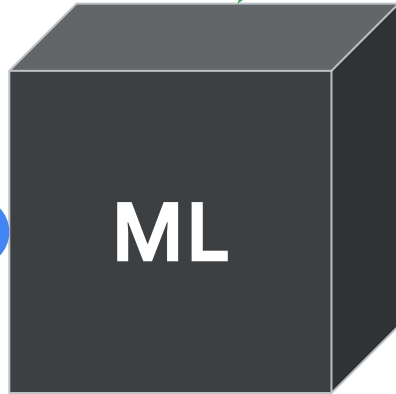
# Training the machine



WE PROVIDE

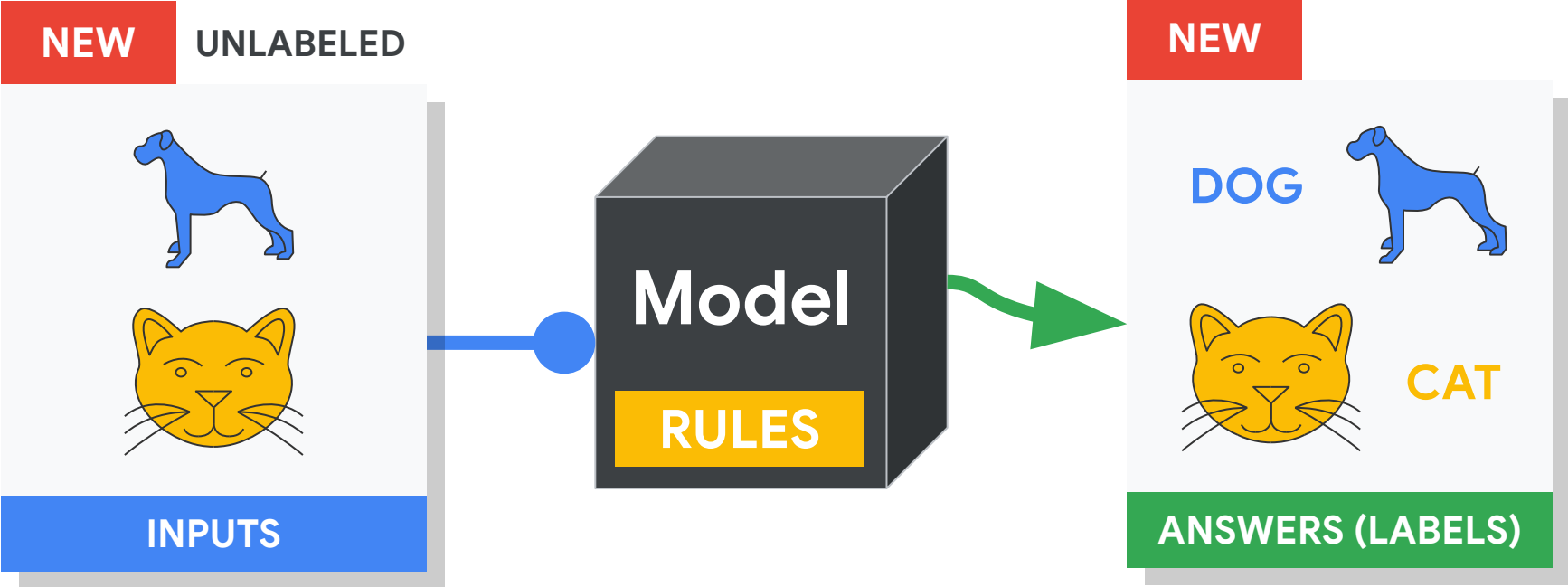
**LABELS**

**INPUTS**

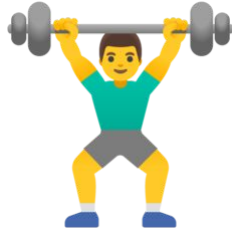


**RULES**

# Making predictions:



# Training the machine



WE PROVIDE

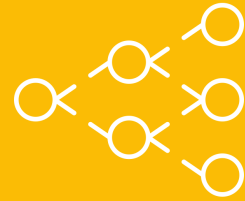
**ANSWERS**



Artificial Intelligence

Machine Learning

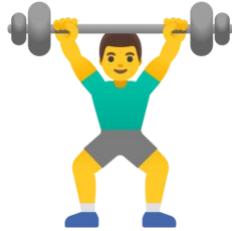
Deep Learning



a lot of  
**DATA**



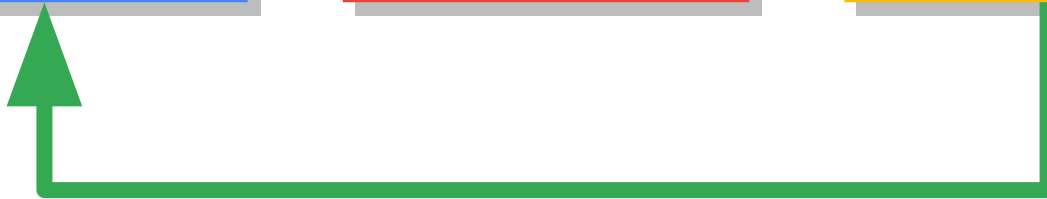
# Training the machine



**Make a  
Guess!**

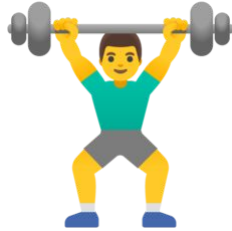
**Check  
and count  
mistakes**

**Improve  
your guess**

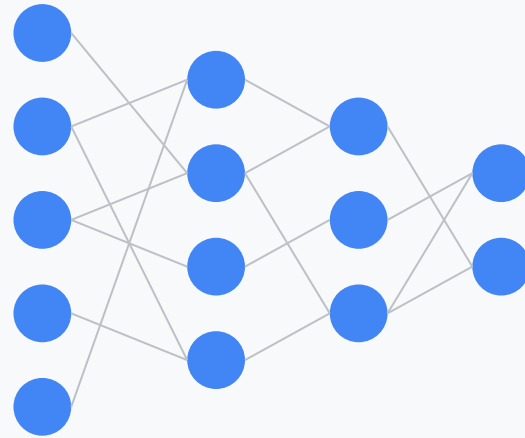




# Training the machine



**Make a  
Guess!**

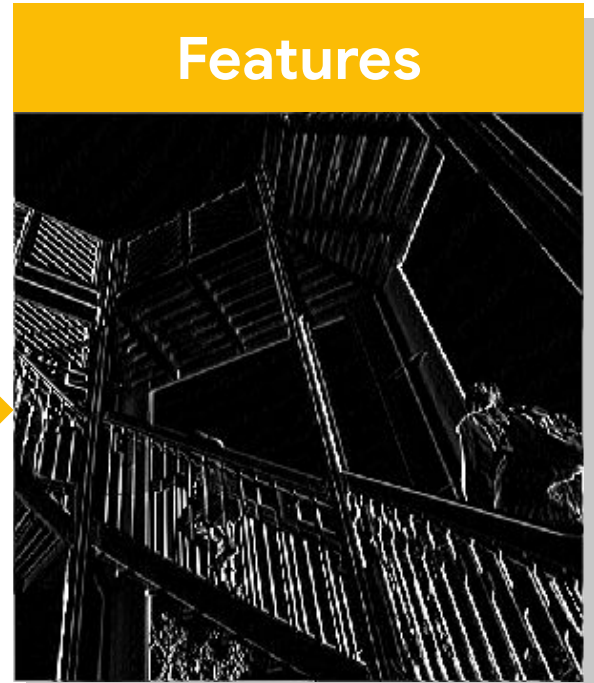


*Neural Network*

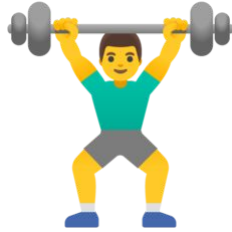
# Transformation: Convolution



-1	0	1
-2	0	2
-1	0	1



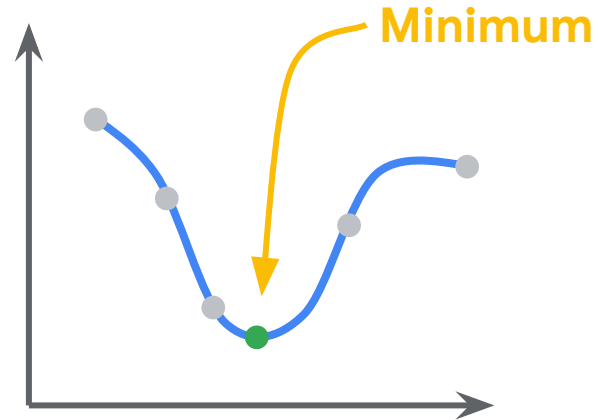
# Training the machine



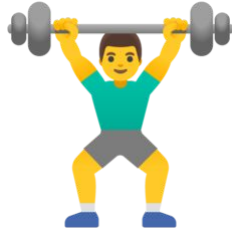
**Loss**  
number of mistakes

**Check**  
and count  
mistakes

**Improve**  
your guess



# Training the machine



to make our machine **smarter** and stronger

a lot of **practice (training) data**

when we  
come back!

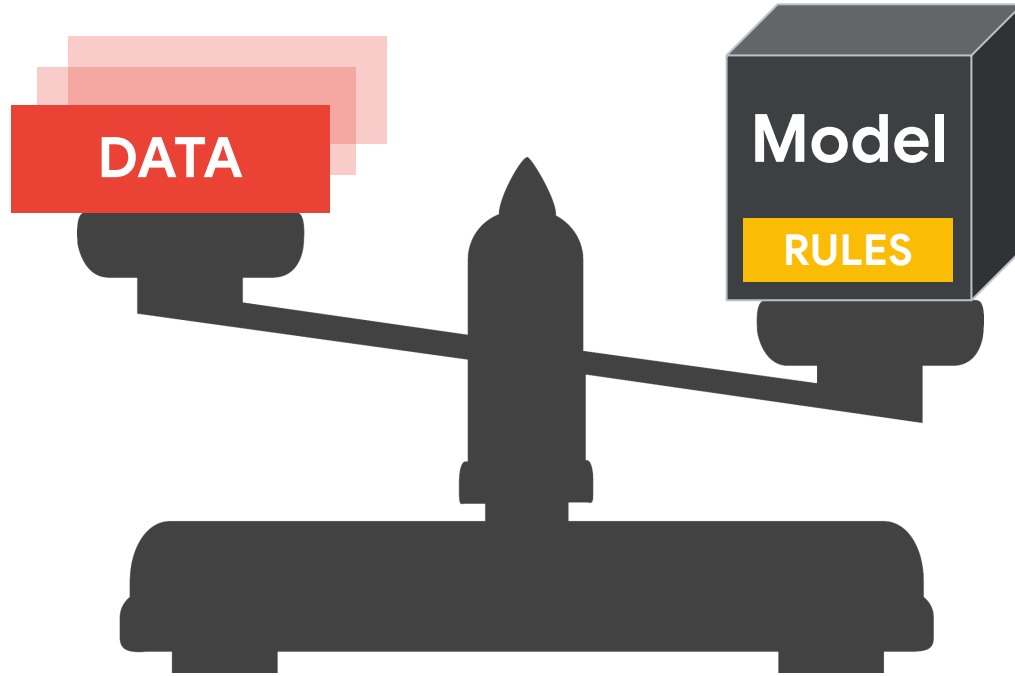
# Tips from **Laurence**

- Learn how to code in **Python**
- Build strong **data skills**
  - collecting data (gather)
  - cleaning data (format)
  - managing data
- Develop well-rounded **testing skills**
  - think about products as a **whole**
  - consider the **diversity** of your users

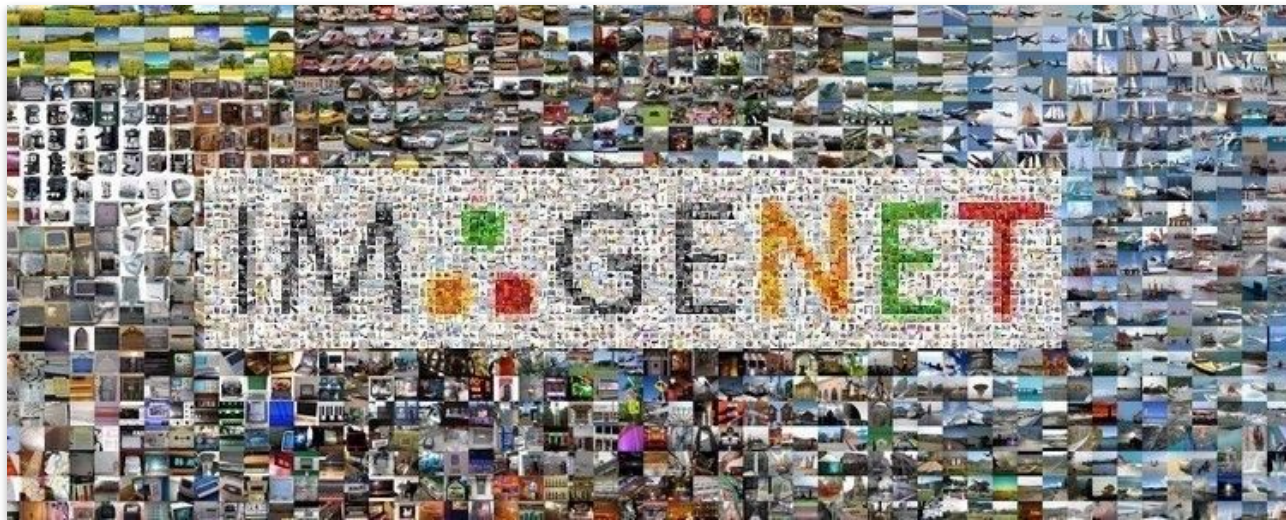


Laurence Moroney  
@lmoroney

# Better model or better data?



# Training Pipeline: **Need Lots of Data**



1000 Classes

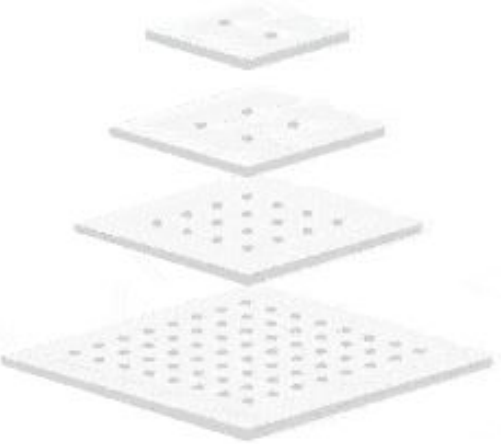
1000 Images / Class

# Image Classification



↓

CAT DOG





# Sources of data

We are starting to connect more and more **smart** devices

**sensors**

# Sensors

## Acoustic Sensors

Ultrasonic, Microphones,  
Geophones, Vibrometers

## Image Sensors

Thermal, Image

## Motion Sensors

Gyroscope, Radar,  
Accelerometer, IMU

# Endpoint Devices



Google Assistant



# Good Data is Necessary for Accuracy

What problem are you trying to *solve*?

- Your data must contain useful features
- Can a human (expert) distinguish between examples of each class?
- How will you measure performance?

# Good Data is Necessary for Accuracy

What problem are you trying to *solve*?

- Your data must contain useful features
- Can a human (expert) distinguish between examples of each class?
- How will you measure performance?

Both *quantity* and *quality* will influence your model's performance

- Wide variety of training examples
- Correct labels (answers)
- Good Balance (e.g., dog, cat, random)

QUICK, DRAW!



[quickdraw.withgoogle.com](https://quickdraw.withgoogle.com)

Our neural net figured out 6 of your doodles.  
Select one to see how it figured it out.



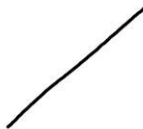
✓ birthday cake



✓ computer



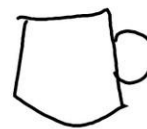
✓ remote control



✓ line



✓ moustache



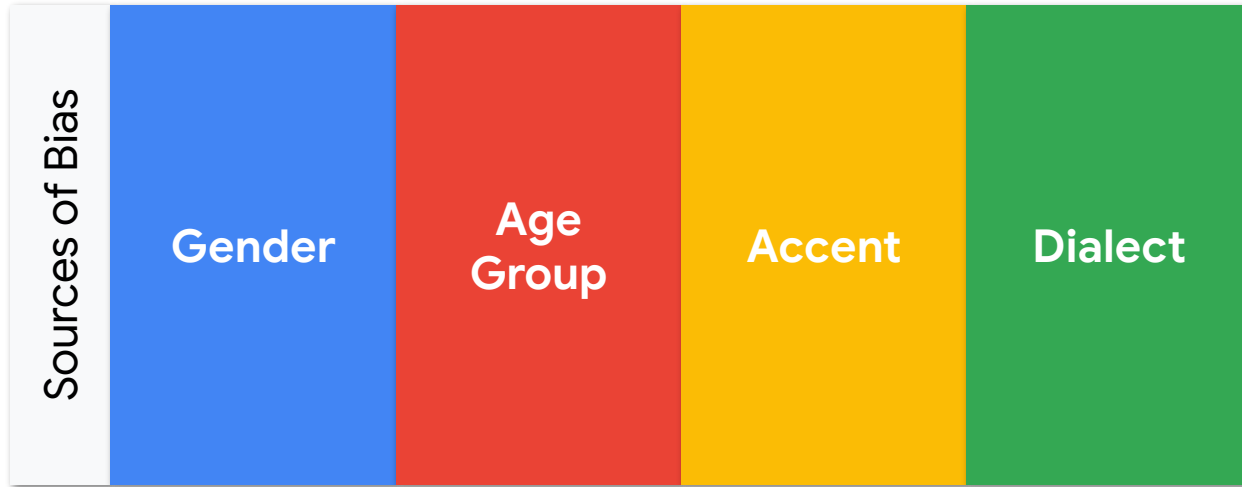
✓ mug

# What do 50 million drawings look like?

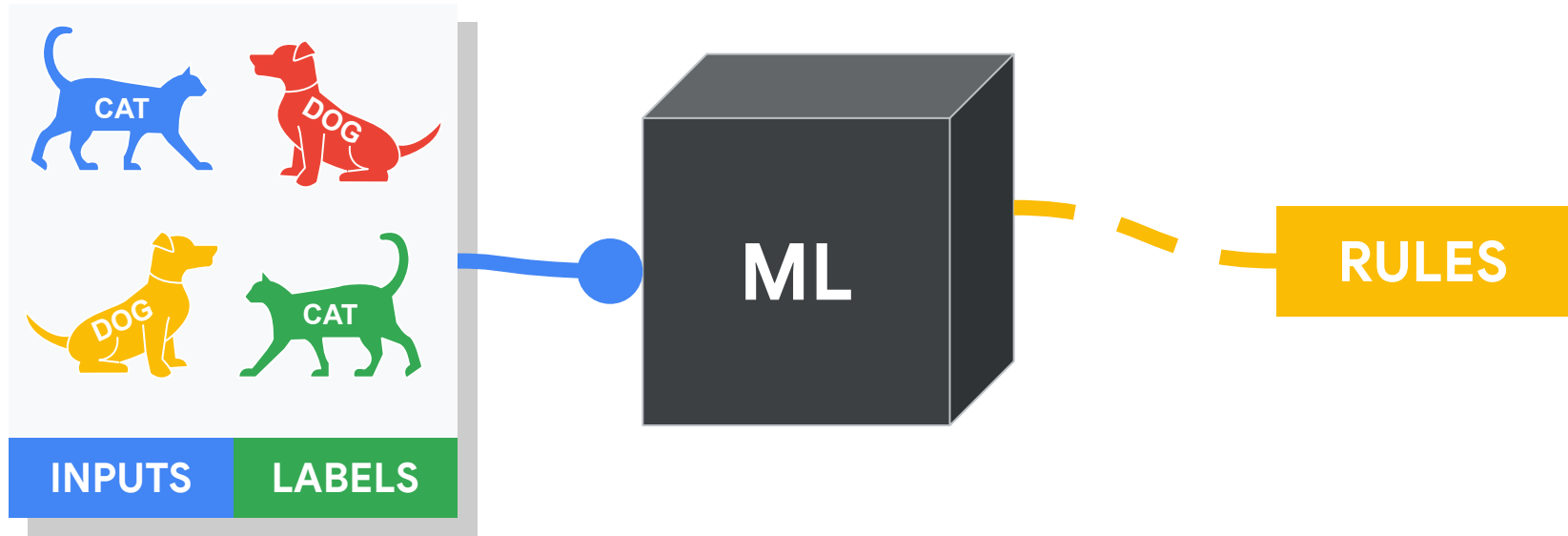
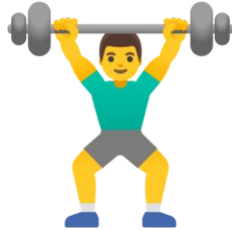




# Potential **Bias** in Speech Recognition



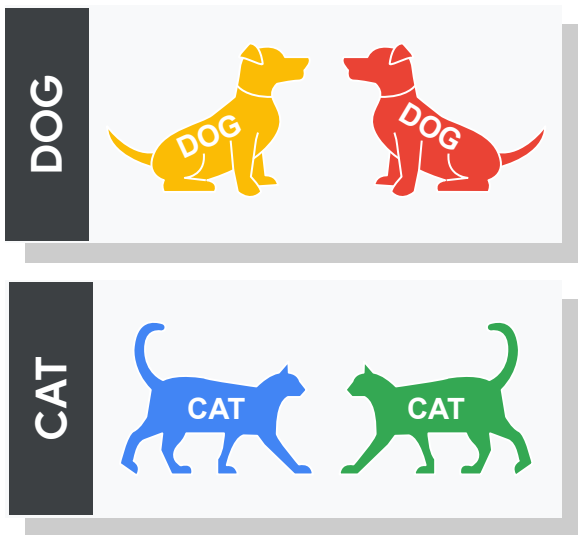
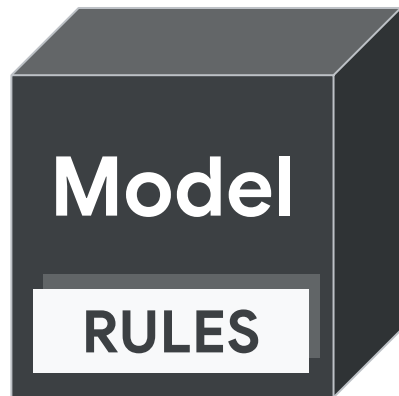
# Training the machine



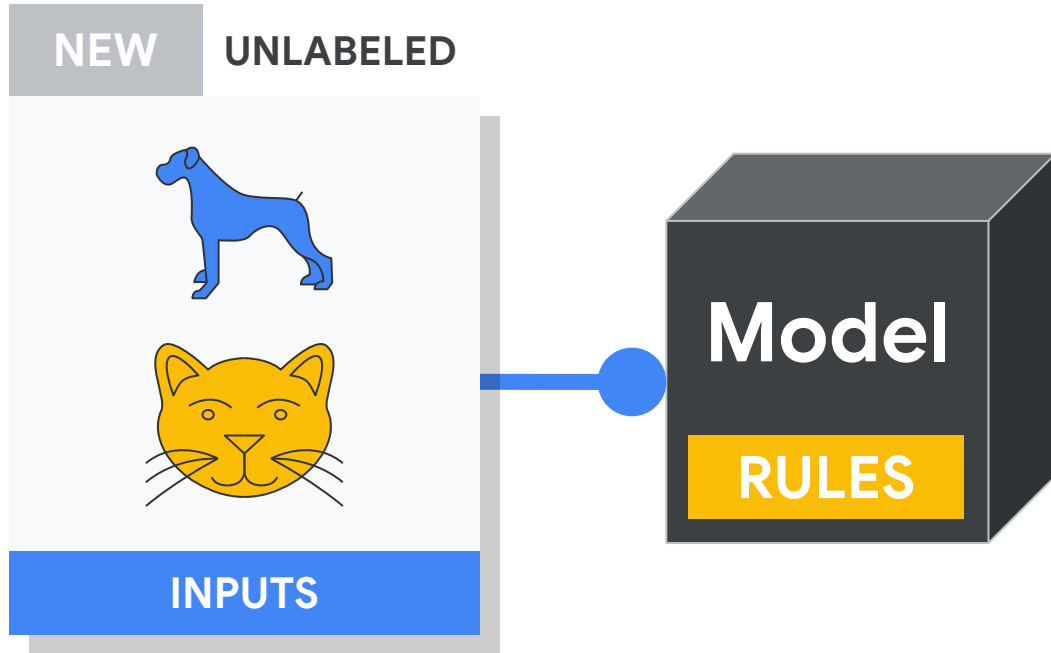
Uh, oh...  
**OVERFITTING**



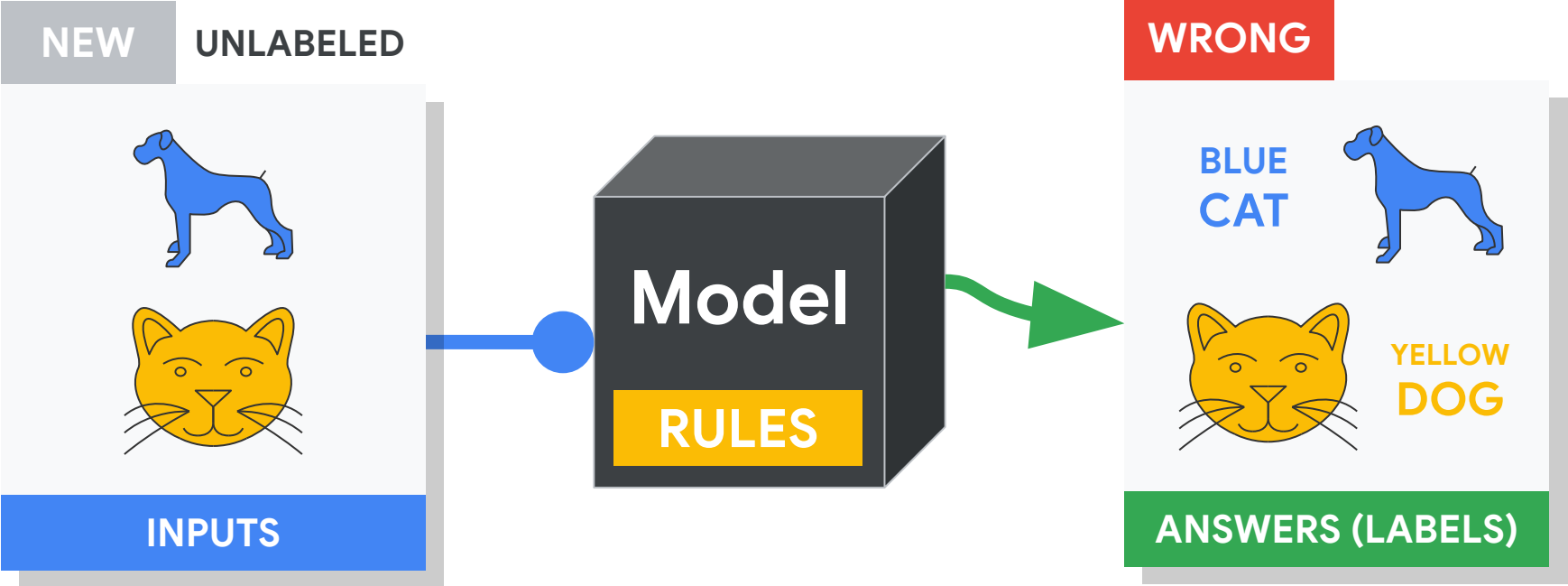
**we trained  
too much**



# Prediction after **overfitting**:



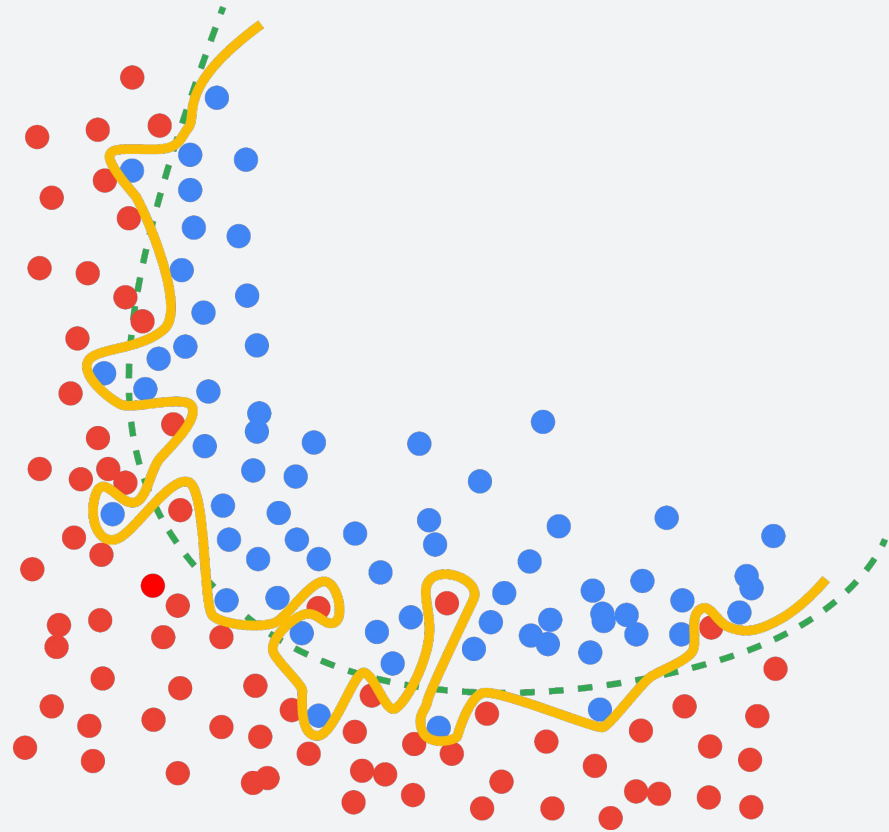
# Prediction after **overfitting**:



# Overfitting

**Yellow line** gives us **no errors** on training data, but will it **generally** do well on new data?

**Green line** has a **few errors** on training data. Is that **okay**?



Don't memorize. **Generalize.**