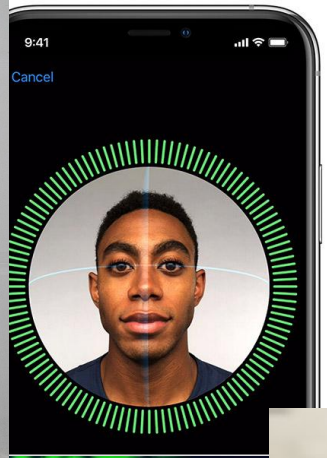
An Introduction to Machine Learning

—

Shrey Gupta, Duke University ‘20

Summer 2020


CL56282


9:41
Cancel


9:41
Monday, July 22


9:41
mastercard
Done


24,400
550,500
667,300
42,000
109,000
5.00
2.00
0.68
8.25

What is machine learning?

What is machine learning?

"Give computers the ability to learn without being explicitly programmed." -Arthur Samuel

What is (not) machine learning?

```
zip_code = input('what is your zip code?')

if zip_code in LIST_OF_NC_ZIPCODES:
    print 'user resides in North Carolina!'

if zip_code in LIST_OF_FL_ZIPCODES:
    print 'user resides in Florida!'

...
```

What is machine learning?

input (data)

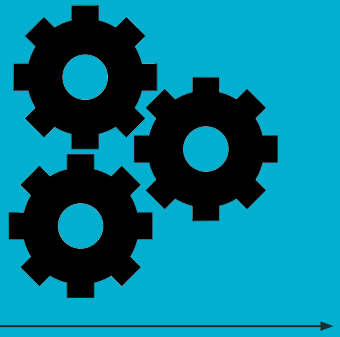
income

race

political affiliation

favorite grocery chain

...



output

state of residence

What is machine learning?

Training data: data used to train algorithm (i.e. create model).

<u>example data point</u>
income
race
political affiliation
favorite grocery chain
...

x 1,000

analyze examples for patterns →

model

What types of algorithms are there?

Grouped into two categories: **supervised** and **unsupervised** learning.

Supervised learning: classification

Data is labeled, and we want to predict a “class” or “category” as the output.

input (data)
feature #1
feature #2
...

output
category #1
OR category #2
OR ...

Example: classification

Given data about temperature, humidity, and wind speed, predict whether it will be sunny, cloudy, or raining.

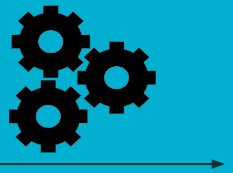
input (data)
temperature
humidity
wind speed

output
sunny
OR cloudy
OR raining

Example: classification

Predict whether the price of an equity will increase or decrease.

input (data)
P/E ratio
volatility
analyst sentiment
current price



output
increase
OR decrease
OR stay the same

Supervised learning: regression

Data is labeled, and we want to predict a continuous output.

input (data)
feature #1
feature #2
...

output
value

Example: regression

Predict the percentage increase or decrease in the price of an equity.

input (data)
P/E ratio
volatility
analyst sentiment
current price

output
returns
(percentage)

Example: regression

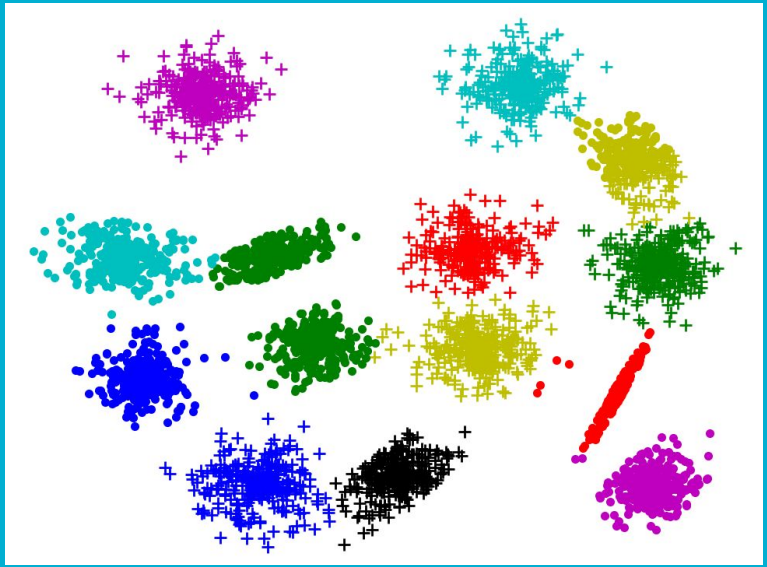
Given data about square footage, age, zip code, and housing demand, predict the selling price of a house.

input (data)
age
zip code
square footage
housing demand

output
selling price
(dollars)

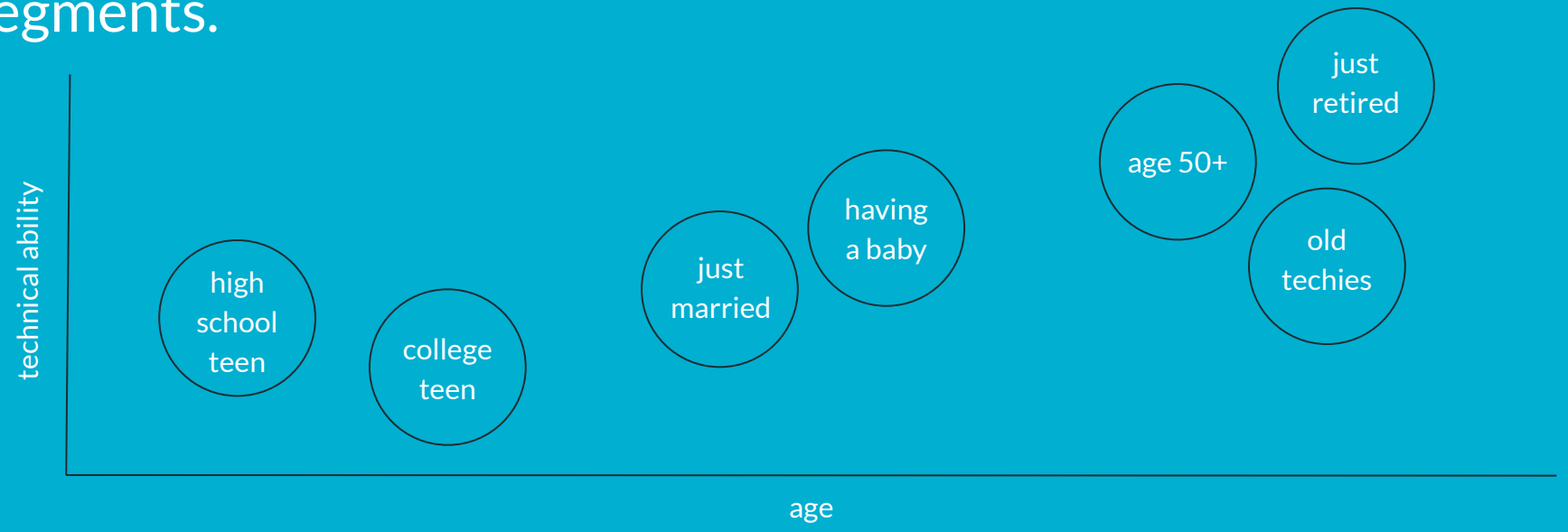
Unsupervised learning: clustering

Data is unlabeled, and we want to cluster the data points into groups.



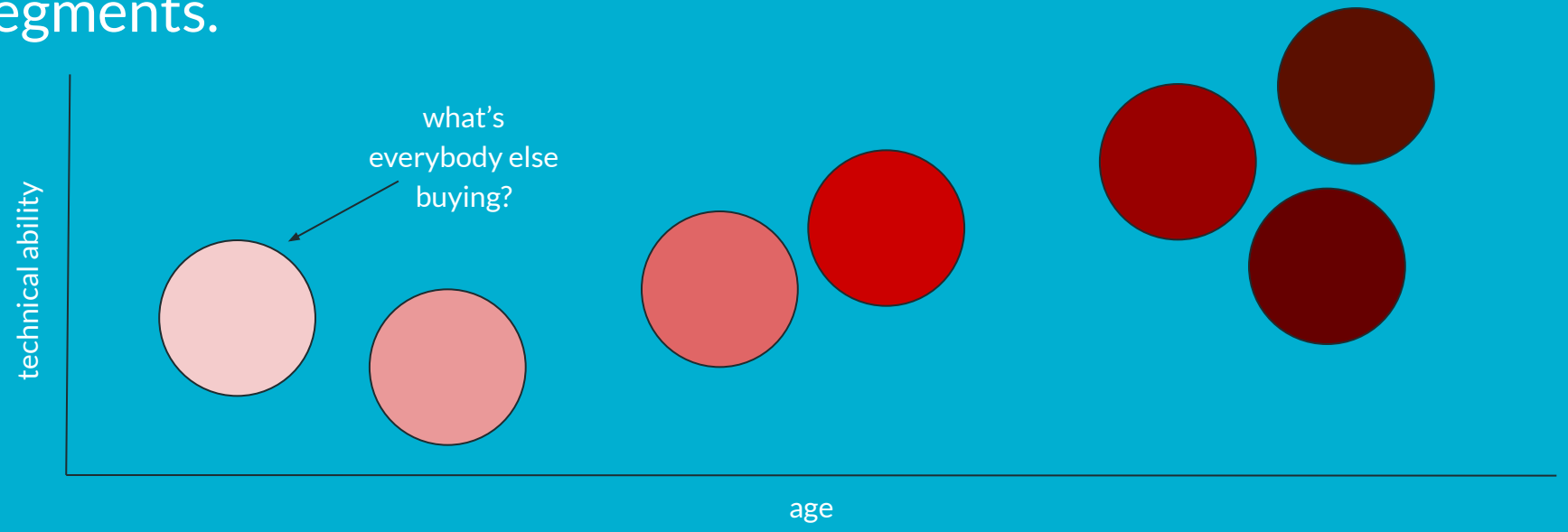
Example: clustering

Given consumption data, partition the consumers into market segments.



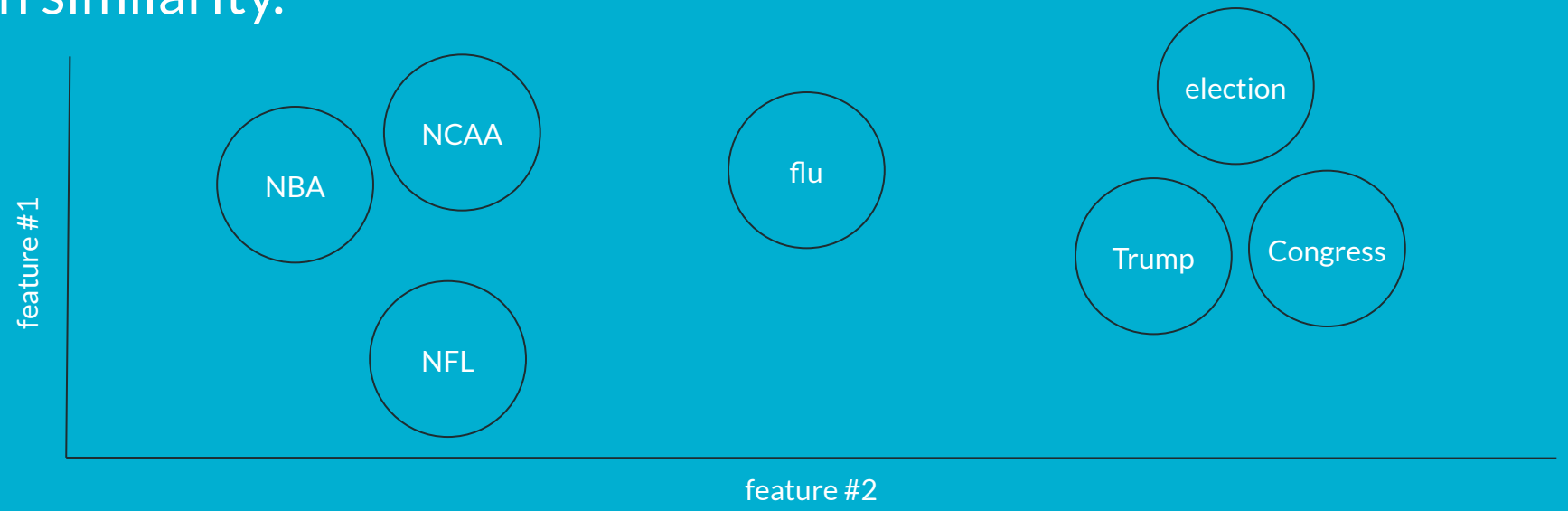
Example: clustering

Given consumption data, partition the consumers into market segments.



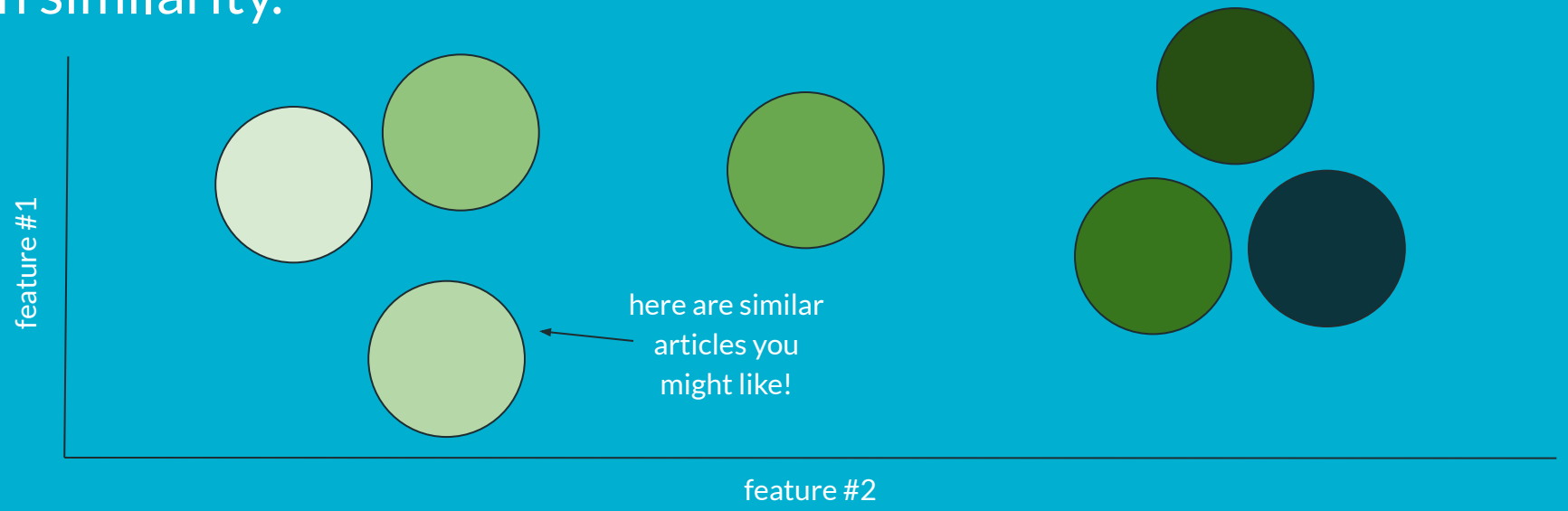
Example: clustering

Given several news articles (and their text), group them based on similarity.



Example: clustering

Given several news articles (and their text), group them based on similarity.

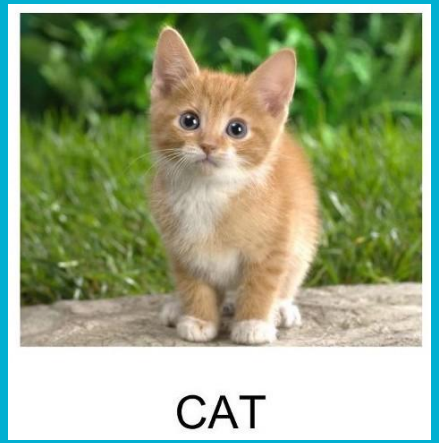


What is happening today in machine learning?

Computer vision

Computer vision is a related field that involves the understanding, processing, and reconstruction of 2- and 3-dimensional images.

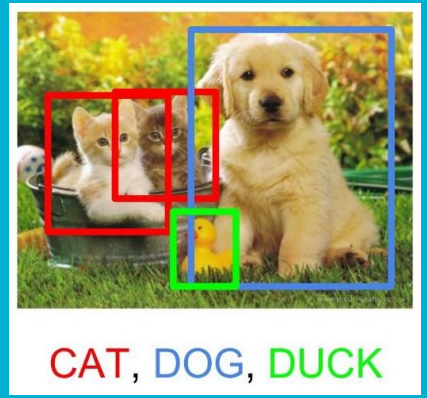
Common computer vision tasks in machine learning include **classification**, **localization**, **object detection**, and **landmark detection**.



classification

localization

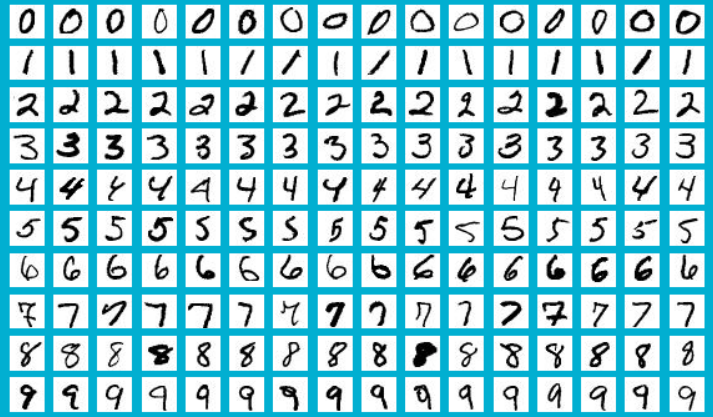
landmark detection



object detection

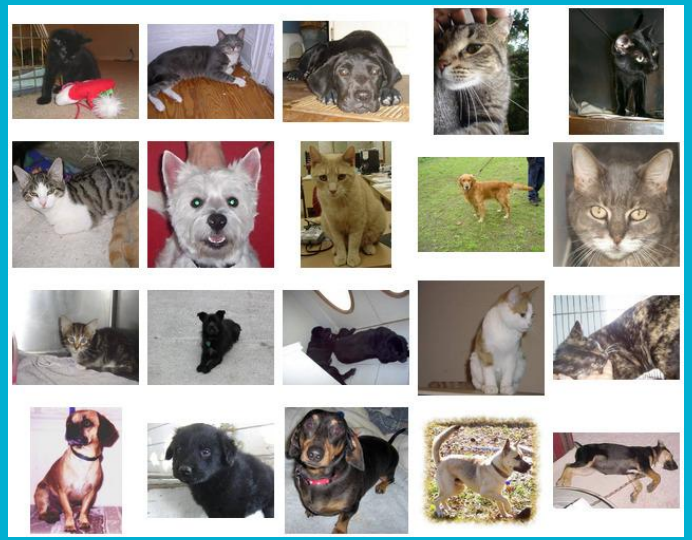
Computer vision

1998: Yann LeCun organizes the MNIST database of handwritten digits, and develops a model that can classify handwritten digits.



Computer vision

2012: Google Brain successfully trains a neural network to differentiate images of cats from dogs.

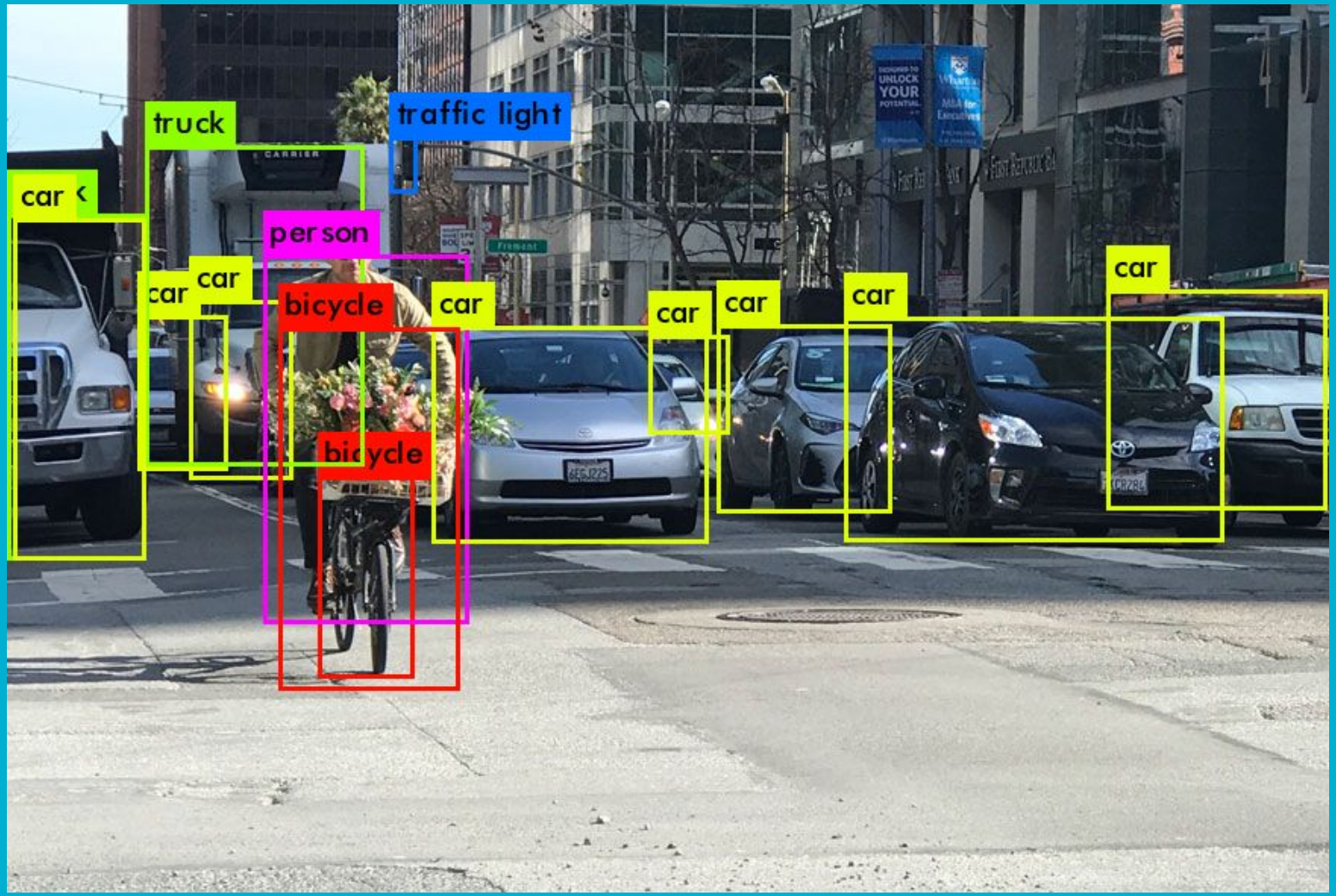


Computer vision

2014: Facebook's DeepFace successfully uses neural networks to perform facial recognition with over 97% accuracy.

Computer vision

2015: Joseph Redmon invents “You Only Look Once” (YOLO), performing real-time object detection with performance higher than ever before.


truck
traffic light
car
person
car
car
bicycle
car
car
car
car
car
bicycle

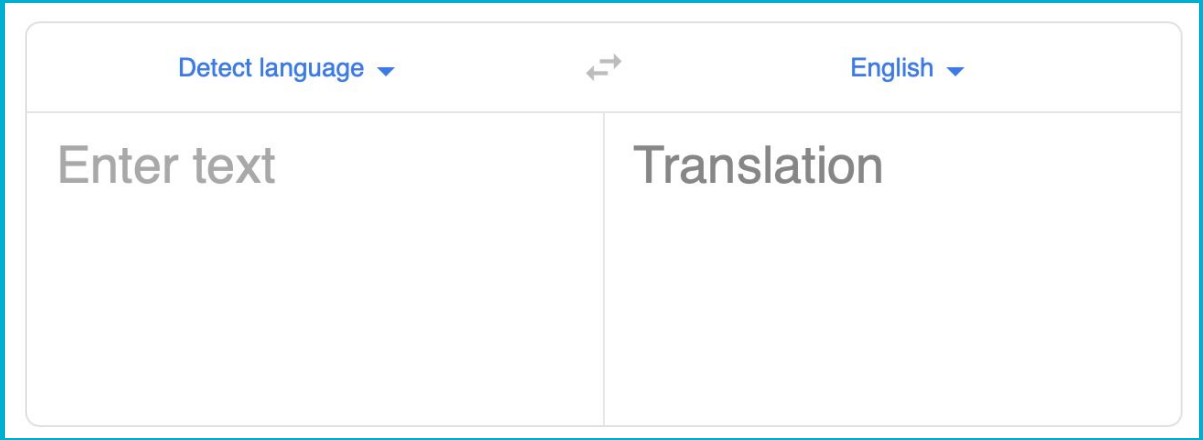
Natural language processing

Natural language processing is a subset of artificial intelligence concerned with understanding natural language, including text and speech.

Examples include **sentiment analysis**, **language translation**, **reading comprehension**, and **textual question-answering**.

Natural language processing

2006: Google Translate launches, allowing translation between multiple languages for free.



Natural language processing

2011: Siri, a natural language intelligent assistant, launches.

Other impressive achievements

1997: IBM's Deep Blue beats chess world champion Gary Kaspaov.

2009: The Netflix Prize is won for the best recommender system in predicting user film ratings.

2011: IBM's Watson is able to defeat human champions in Jeopardy!

Other impressive achievements

2014: The “Eugene Goostman” chatbot fools a third of judges in the Turing test.

2016: DeepMind develops AlphaGo and beats the top-ranked Go player. AlphaGo Zero, which is generalized to chess and other games, is developed the following year.

When is machine learning useful?

Power, complexity, and data

We have tons and tons of data, and huge amounts of compute power today.

More complex models need lots of data. Otherwise, the model might find patterns that don’t really exist.

Evaluation

Need to evaluate your model carefully.

Several metrics, such as **mean absolute error** for regression and **accuracy** and **precision** for classification, and methods, such as **cross-validation**.

Prediction and interpretability

Machine learning models are good for prediction, but don’t give underlying causation.

Complex models can be difficult to interpret.

Algorithmic bias

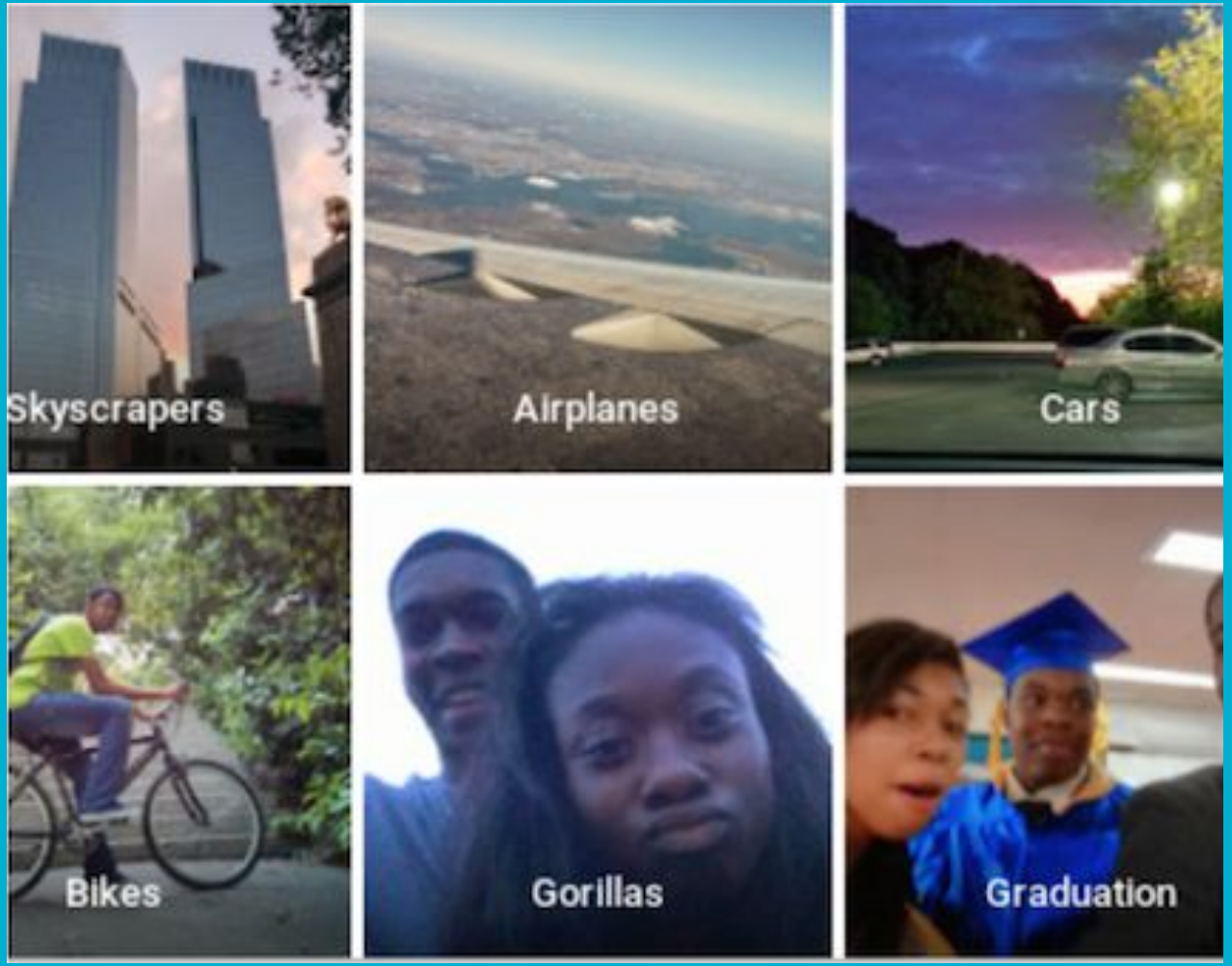
Machine learning is often used for high stakes decisions, such as determining whether to lend credit, facial recognition for criminals and terrorists, and recidivism.

Training data needs to be **representative** and **unbiased**.

Amazon Rekognition

FALSE MATCHES

28 current members of Congress


Skyscrapers
Airplanes
Cars
Bikes
Gorillas
Graduation

shreygupta.me/phoenix-ml