# Machine Learning

Lecture 1
Intro

Anna Kuzina
https://github.com/AKuzina/ml_se

# Logistics: Main Links

All the materials: https://akuzina.github.io/ml_se
( including links below)

Assignments in AnyTask: https://anytask.org/course/769

Questions / Announcements / Discussion:
Telegram Chat: https://t.me/ml_se21

Feedback to the course team:
Google form: https://forms.gle/KeGbnntmsPcQXzhX6

# Logistics: Team

## Lecturers


Evgenii Egorov
egorov.evgenyy@ya.ru

Anna Kuzina
av.kuzina@ya.ru


## Teachers

Maria Tikhonova

Maksim Karpov


Polina Polunina

Vadim Kokhtev

# Logistics: Course Plan

- 15 Lectures
- 15 Practical Sessions
- 6 Assignments

# Logistics: Course Plan

## Supervised ML [7 lectures]

- Gradient Optimization
- Linear Regression
- Linear Classification
- Decision Trees
- Bagging, Random Forest, Gradient boosting

## Unsupervised ML [2 lectures]

- Clustering and Anomaly Detection
- EM and PCA

## Advanced Topics [4 lectures]

- Bayesian Linear Regression and Gaussian Processes
- MLP and DNN
- Deep Generative Models

\+ Introduction and Summary Lectures

# Logistics: Grades

- Assignments
  - 5 compulsory
  - 1 optional (to get bonus points)
- Exam
  - Oral exam
  - Covers all the topics from lectures and seminars

Final grade = 0.7*HW + 0.3*Exam

# Logistics: Grades

- Assignments
  - 5 compulsory
  - 1 optional (to get bonus points)
- Exam
  - Oral exam
  - Covers all the topics from lectures and seminars

Final grade = 0.7*HW + 0.3*Exam

You can skip the exam if your average grade for the assignments is not smaller than 6:

Final grade = HW >=6

You can choose this option only before the exam date.

# Logistics

QUESTIONS?

# Learning Outcomes

After this lecture you should know:

- Course goals

- Key objects of the course
  types of ML problems, features, target variables, etc.

- Connection of course contents with applications

# Course Goals

- distinguish major problems of data analysis
- recognise, apply and understand advantages and disadvantages of major algorithms
- be able to use data analysis libraries from python - numpy, scipy, pandas, matplotlib and scikit-learn
- know, how to transform data to make it more suitable for machine learning algorithms

# What is Machine Learning?

“Field of study that gives computers the ability to learn without being explicitly programmed”
- Arthur Samuel

https://en.wikipedia.org/wiki/Arthur_Samuel

# Task: Convert Hours into Minutes

# Task: Convert Hours into Minutes

x - hours

$f(x) = 60x$

# Task: Time to the Ground

- I throw an object from the height $x$
- How long will it be falling?

# Task: Time to the Ground

- I throw an object from the height $x$
- How long will it be falling?
- If I'm Galileo:
  I need to conduct an experiment



https://en.wikipedia.org/wiki/Galileo%27s_Leaning_Tower_of_Pisa_experiment

# Task: Time to the Ground

- I throw an object from the height $x$
- How long will it be falling?
- If I'm Galileo:
  I need to conduct an experiment
- Me in the 21st century:

$$f(x) = \sqrt{\frac{2x}{g}}$$

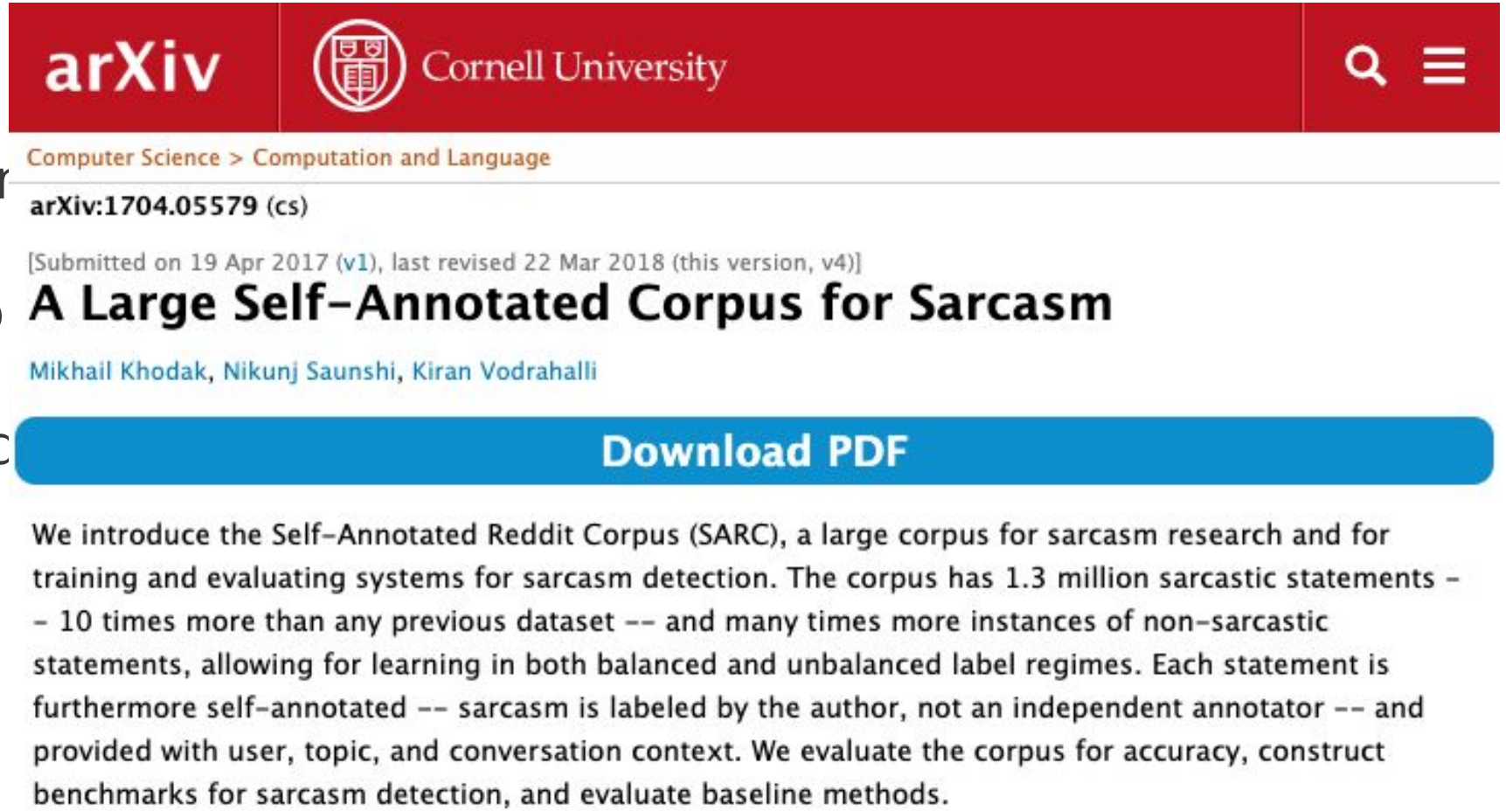
# Task: Who is on the Picture?

# Task: Text Sentiment

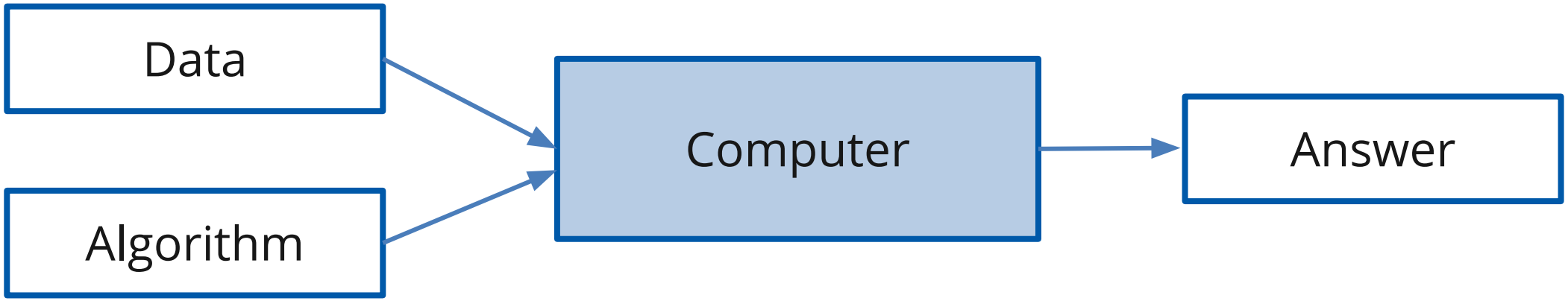
- “I can not say that I like this soup, but it is better than nothing”
- “I do not like this banana”
- “I did not dislike that steak”

# Task: Text Sentiment

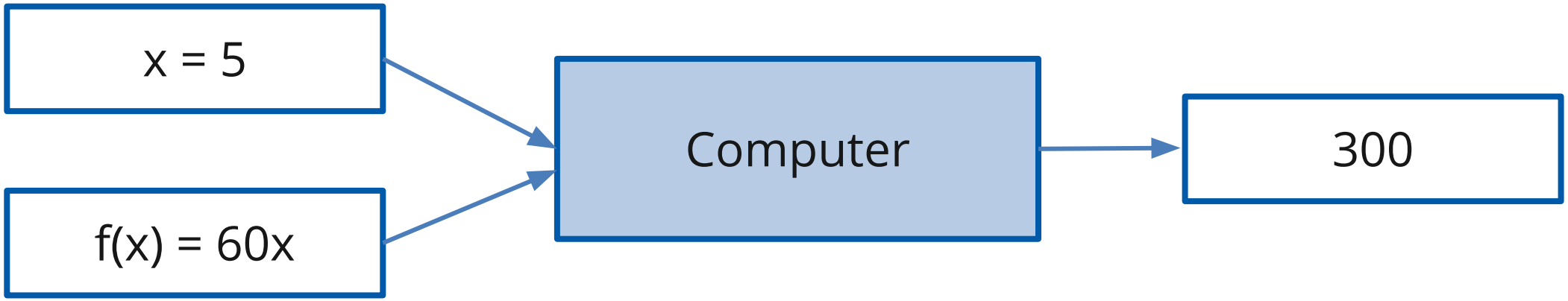
- “I car
- “I do
- “I dic



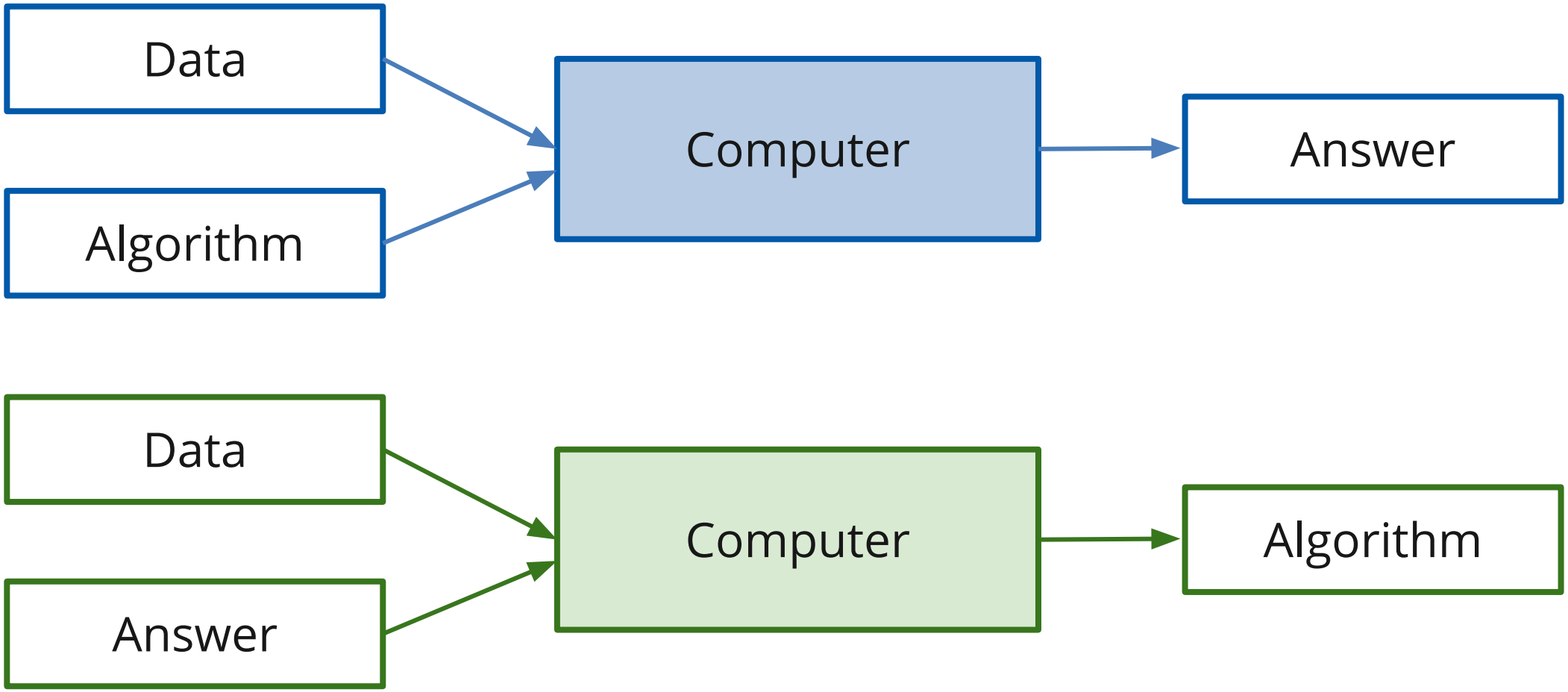
# Machine Learning vs Traditional Programming



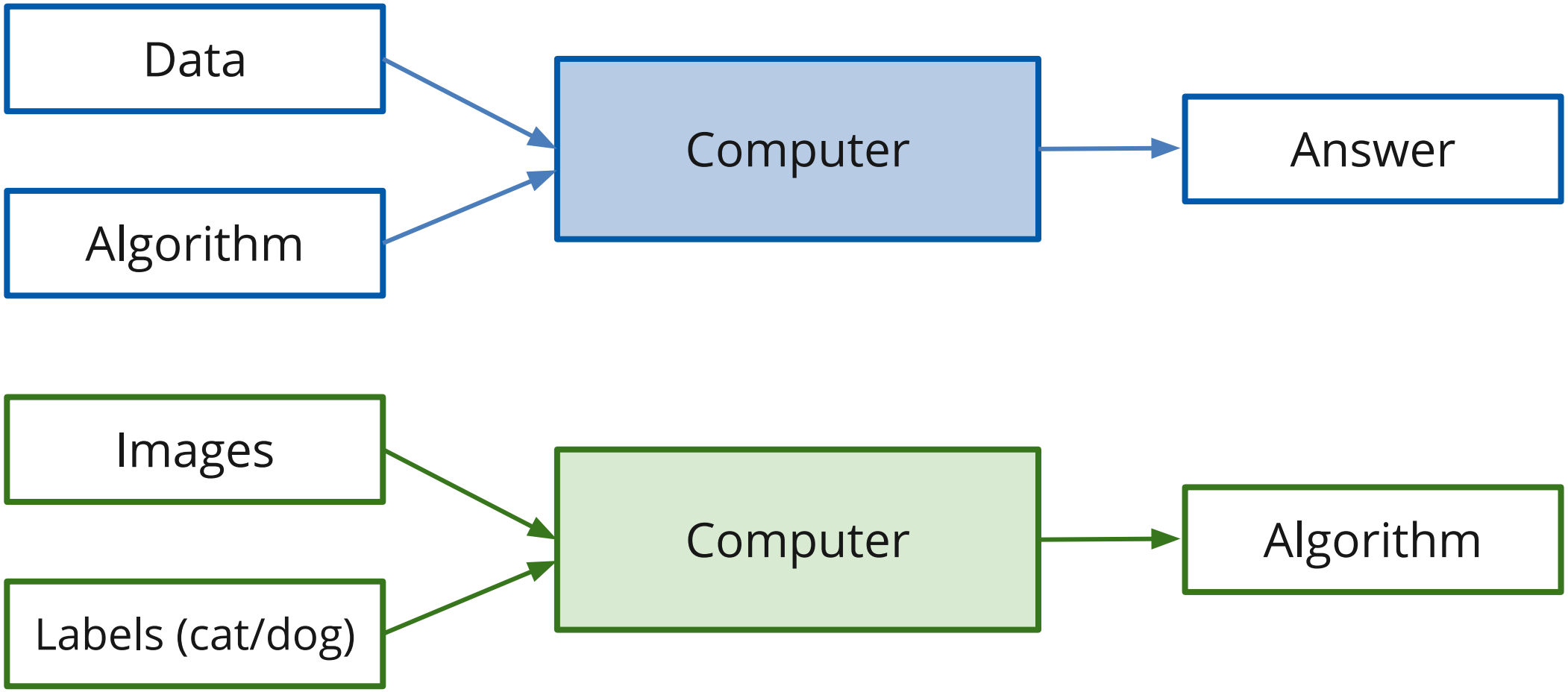
# Machine Learning vs Traditional Programming



# Machine Learning vs Traditional Programming



# Machine Learning vs Traditional Programming



# Chess vs Go

Deep Blue vs Kasparov (1985)

~ 10^123 possible moves


AlphaGo vs Lee Sedol (2016)

~ 10^360 possible moves


https://deepmind.com/research/case-studies/alphago-the-story-so-far

# Why Now?

Deep Blue vs Kasparov (1985)

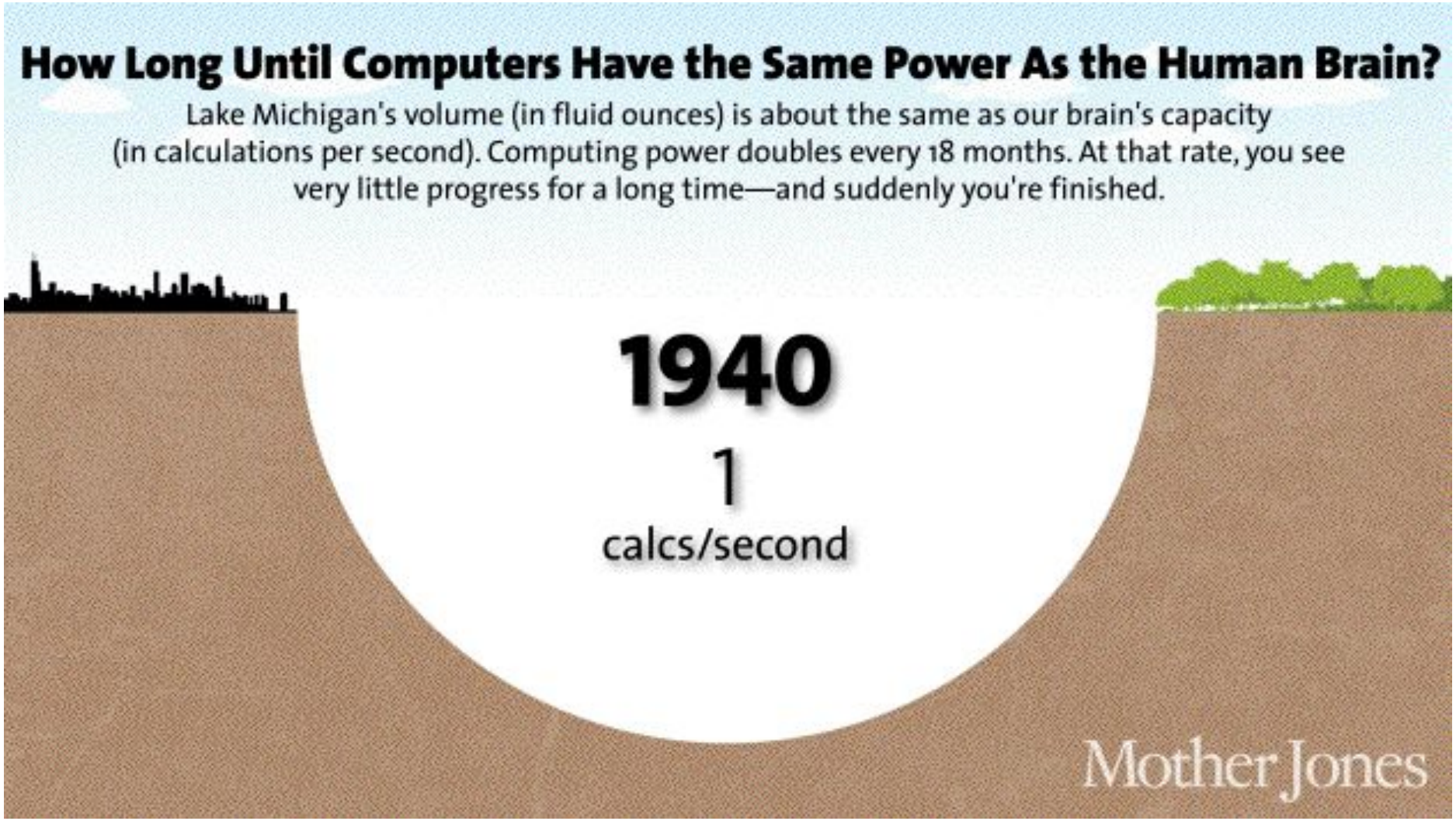
~ 10^123 possible moves


AlphaGo vs Lee Sedol (2016)

~ 10^360 possible moves

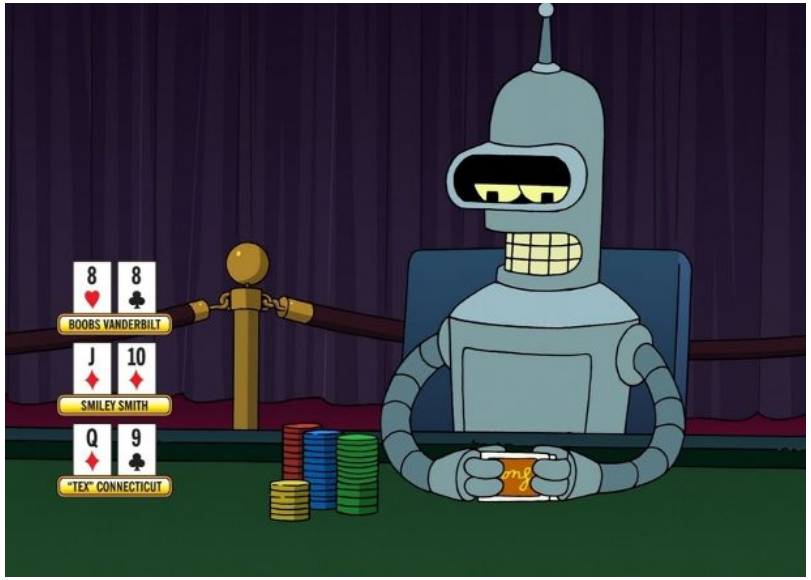

https://deepmind.com/research/case-studies/alphago-the-story-so-far

# Why Now?


How Long Until Computers Have the Same Power As the Human Brain?
Lake Michigan's volume (in fluid ounces) is about the same as our brain's capacity (in calculations per second). Computing power doubles every 18 months. At that rate, you see very little progress for a long time—and suddenly you're finished.
1940
1
calcs/second
Mother Jones

# More Examples

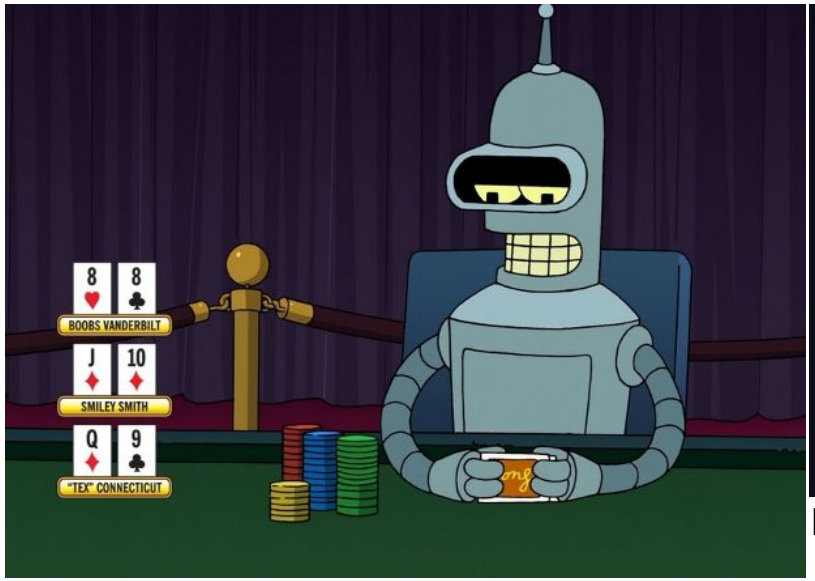
2017




https://www.wired.com/2017/01/mystery-ai-just-crushed-best-human-players-poker/

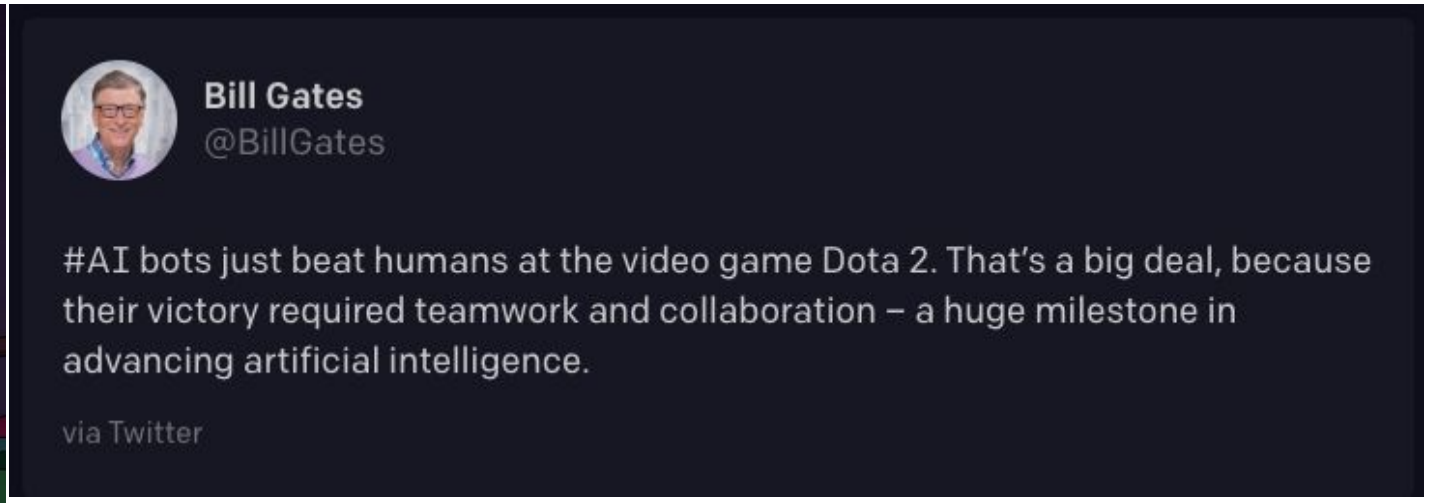
# More Examples

2017




https://www.wired.com/2017/01/mystery-ai-just-crushed-best-human-players-poker/

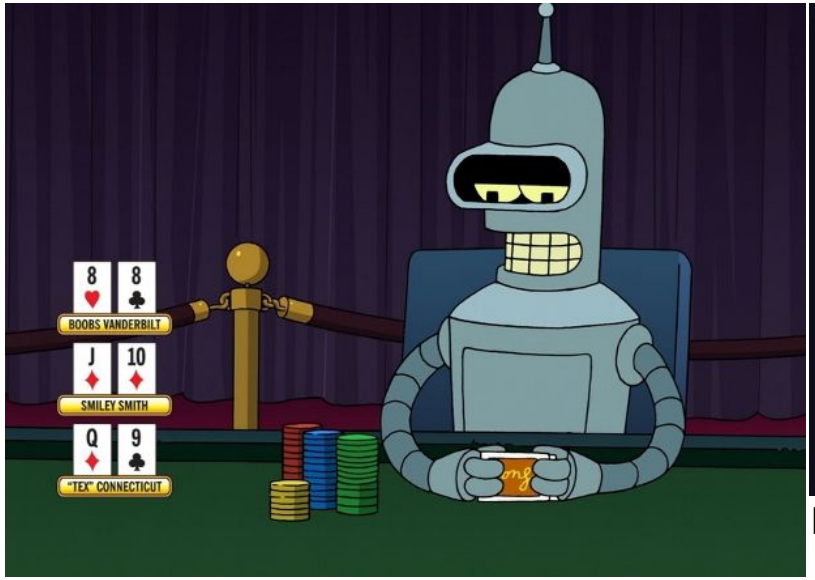
2018




https://openai.com/projects/five/

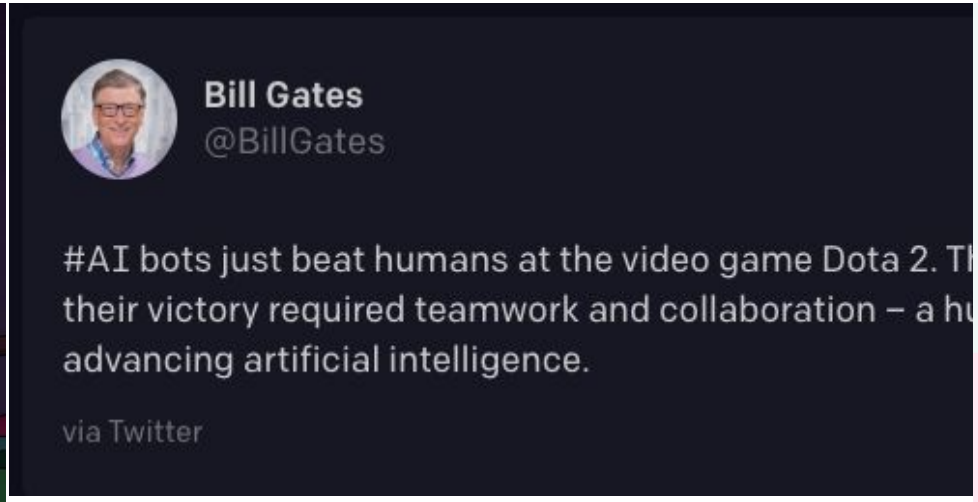
# More Examples

2017




https://www.wired.com/2017/01/mystery-ai-just-crushed-best-human-players-poker/

2018




https://openai.com/projects/five/

2019


https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii

# Not Only Games

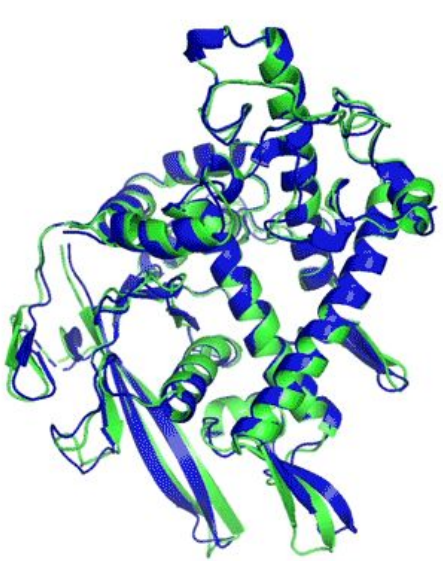
What does these people have in common?

# Not Only Games

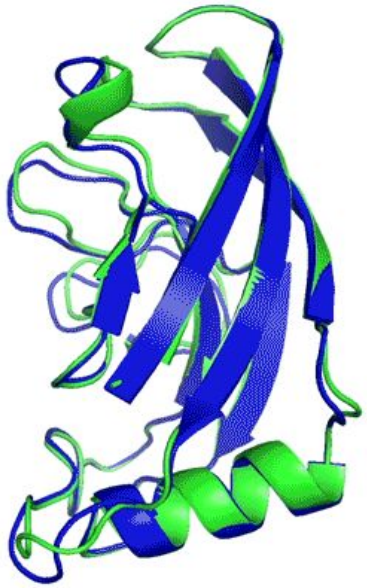
What does these people have in common?

**https://thispersondoesnotexist.com/**

# Not Only Games Protein Folding



**T1037 / 6vr4**
90.7 GDT
(RNA polymerase domain)



**T1049 / 6y4f**
93.3 GDT
(adhesin tip)

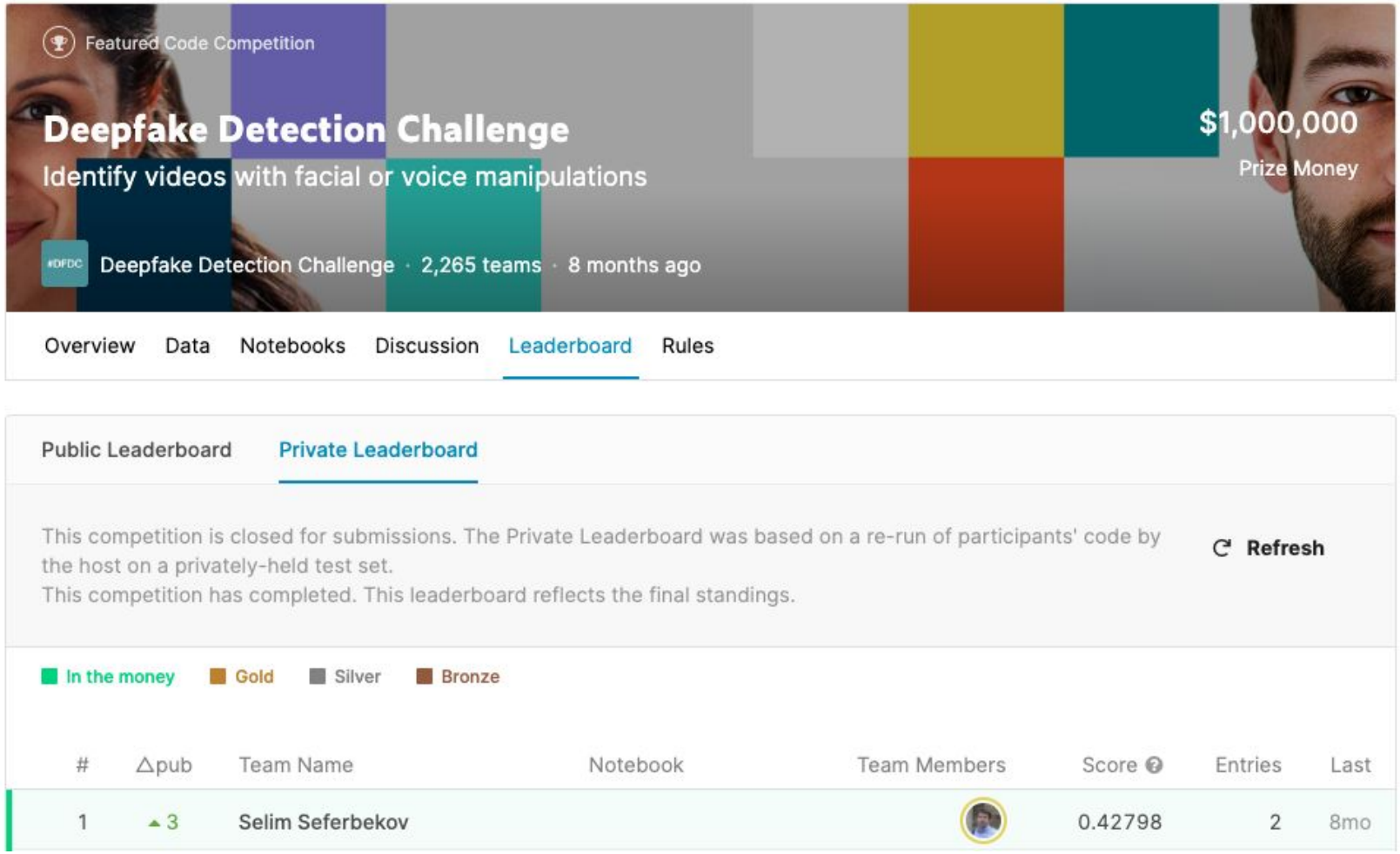
Experimental result
Computational prediction

https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

# Not Only Games Deep Fakes


https://www.youtube.com/watch?v=2jp4M1cIJ5A

# Not Only Games Deep Fakes Detection

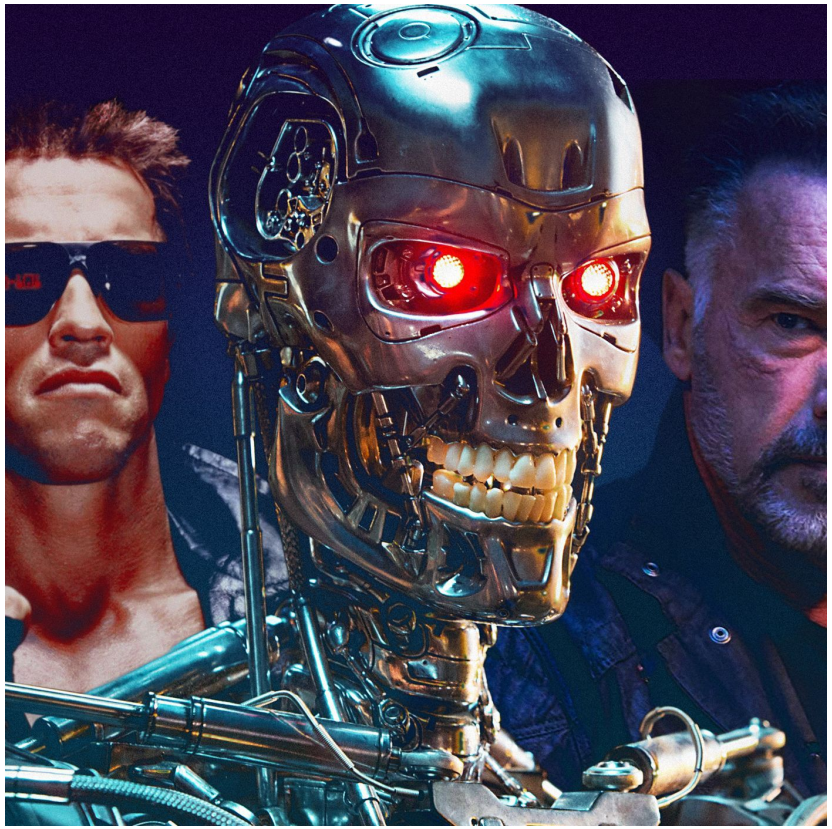



https://www.kaggle.com/c/deepfake-detection-challenge/leaderboard

# Artificial Intelligence?

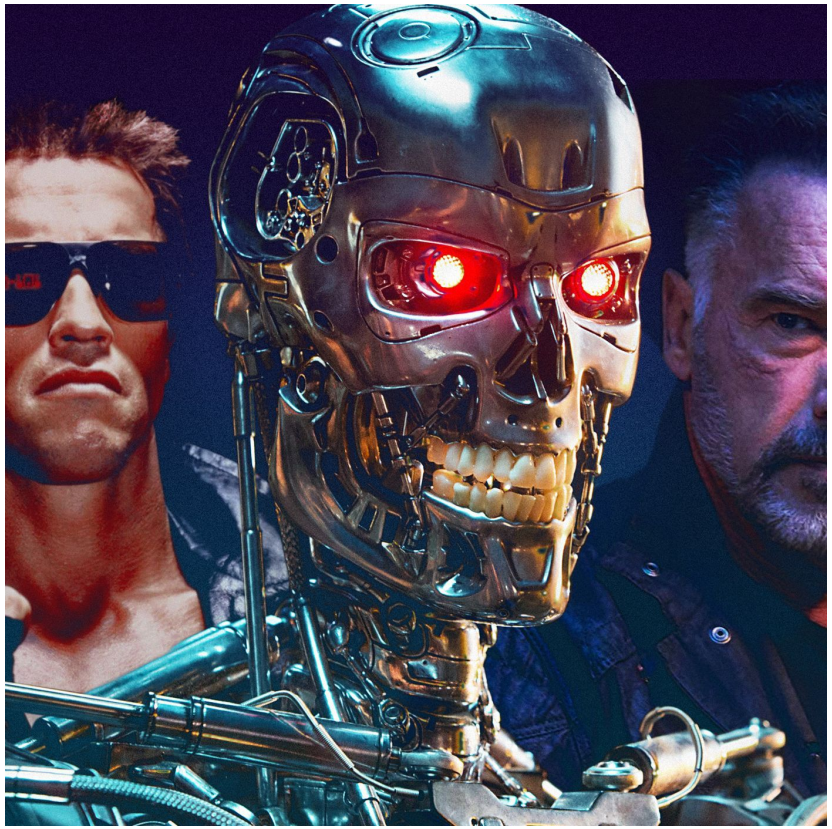
- Week


- General



# Artificial Intelligence?

- Week
- ~~General~~

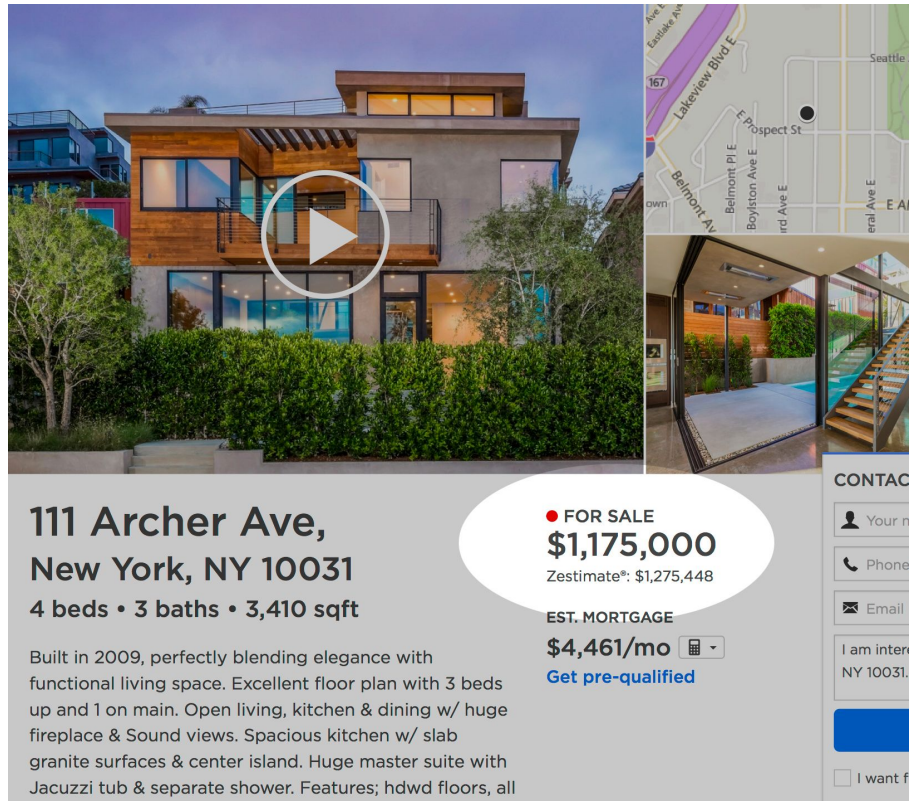



# Main Definitions

# Running Example Predicting House Prices

Kaggle.com competitions:

- Zillow's Home Value Prediction ($25'000)
- Sberbank Russian Housing Market ($12'000)



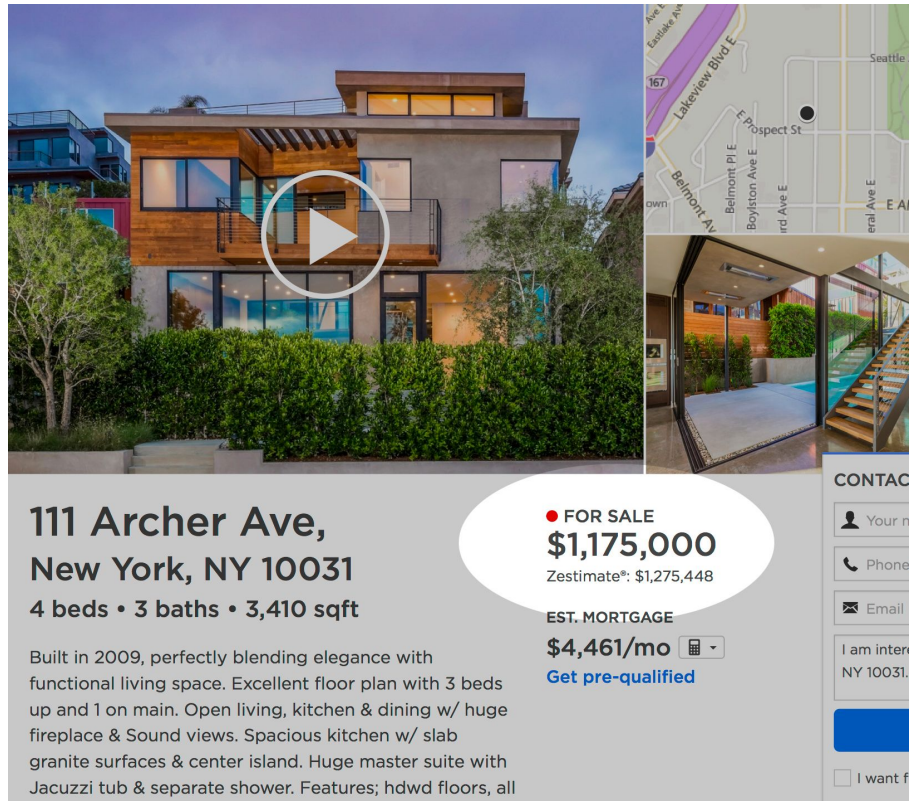

https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Running Example Predicting House Prices

Kaggle.com competitions:

- Zillow's Home Value Prediction ($25'000)
- Sberbank Russian Housing Market ($12'000)

We'd like to learn to predict price of the house…




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

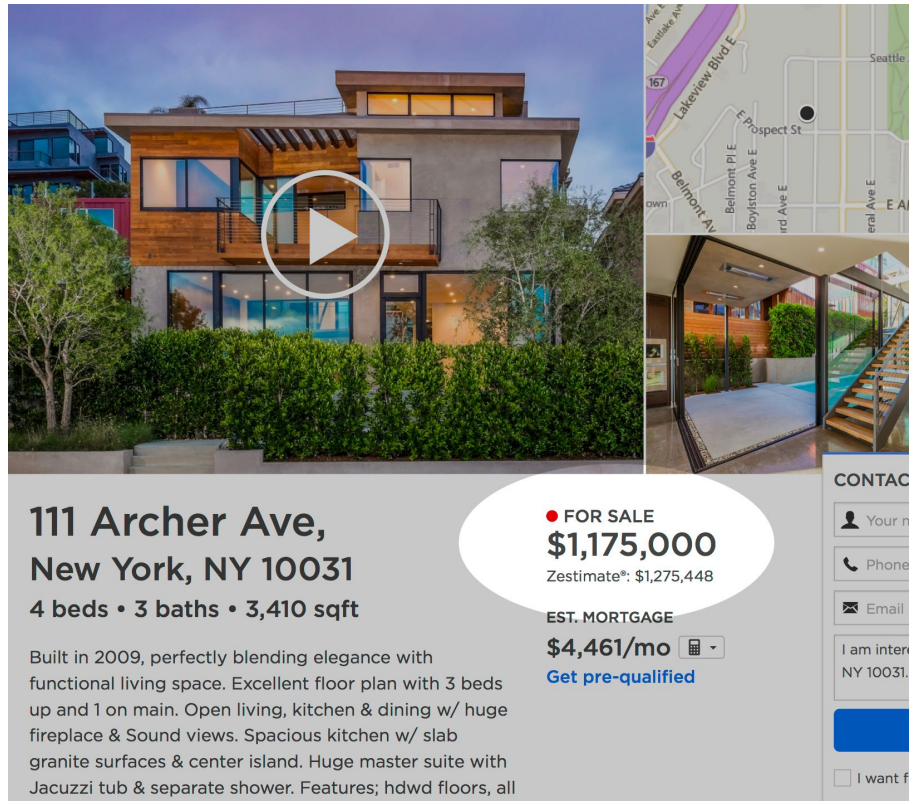
# Running Example Predicting House Prices

Kaggle.com competitions:

- Zillow's Home Value Prediction ($25'000)
- Sberbank Russian Housing Market ($12'000)

We'd like to learn to predict price of the house…

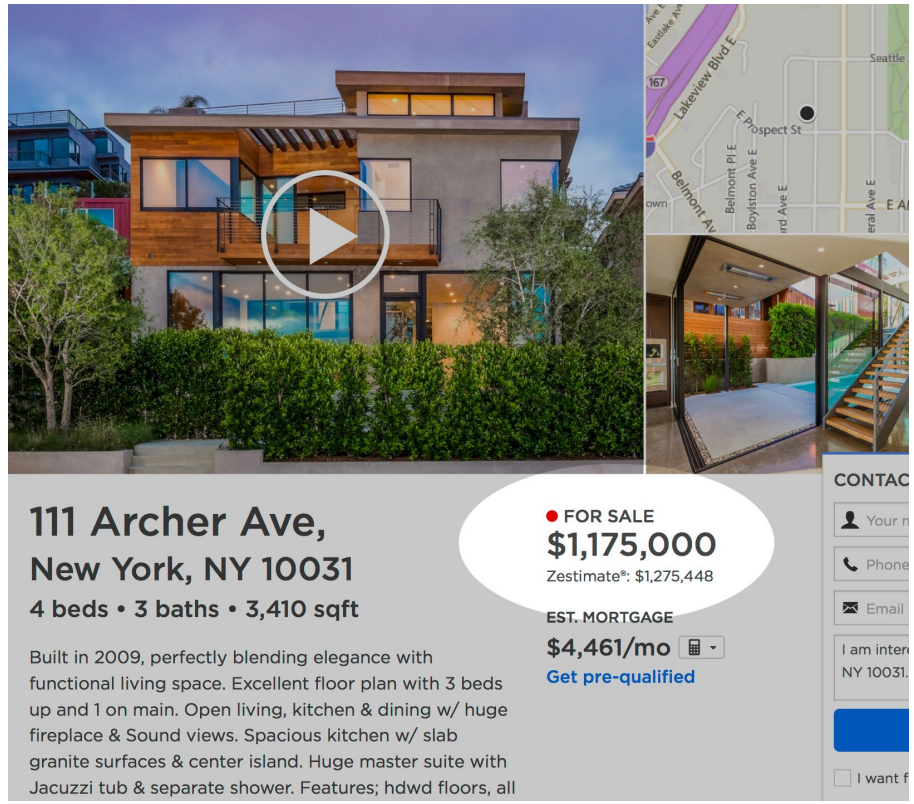
… by looking at 100500 houses and their prices




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions

- $\boldsymbol{x}$ — sample / object / observation
  the house, its description
- $\mathbb{X}$ — space of all objects
  all possible houses we are considering
- $\boldsymbol{y}$ — target
  price of the house
- $\mathbb{Y}$ — space of all possible target values
  positive, real valued numbers




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Feature

- $\boldsymbol{x}$ — sample

We will describe each sample using a set of numbers — features.

$$x = (x_1, x_2, \ldots, x_d)$$

# Definitions Feature

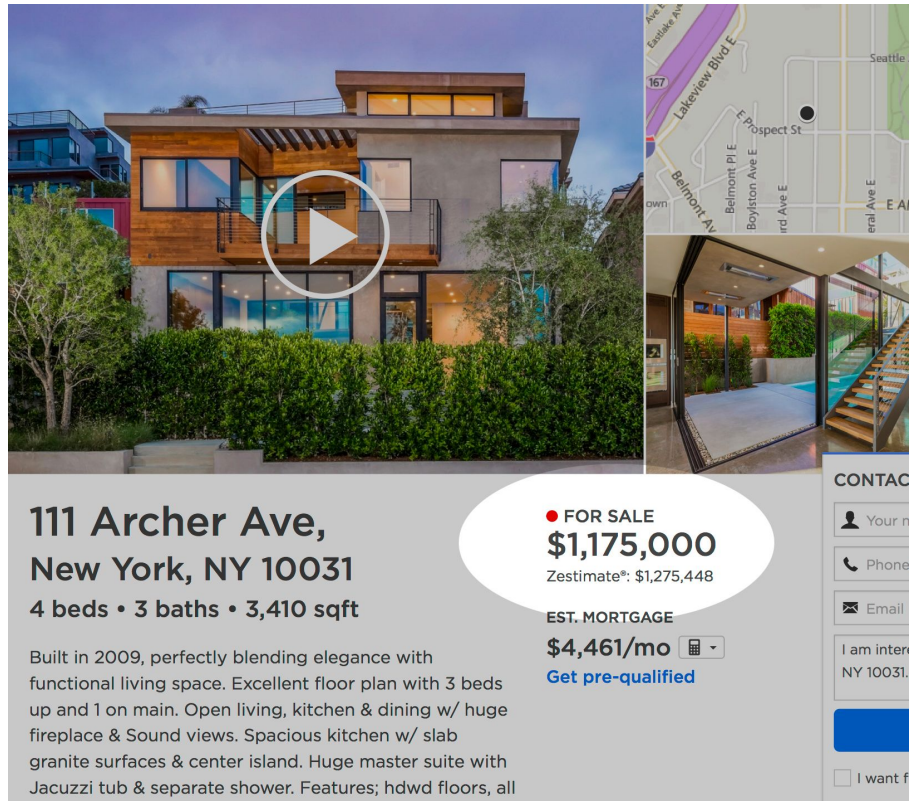
- $\boldsymbol{x}$ — sample

We will describe each sample using a set of numbers — features.

$$x = (x_1, x_2, \ldots, x_d)$$

Features of the house:

- Area
- Num. bedrooms
- District
- …



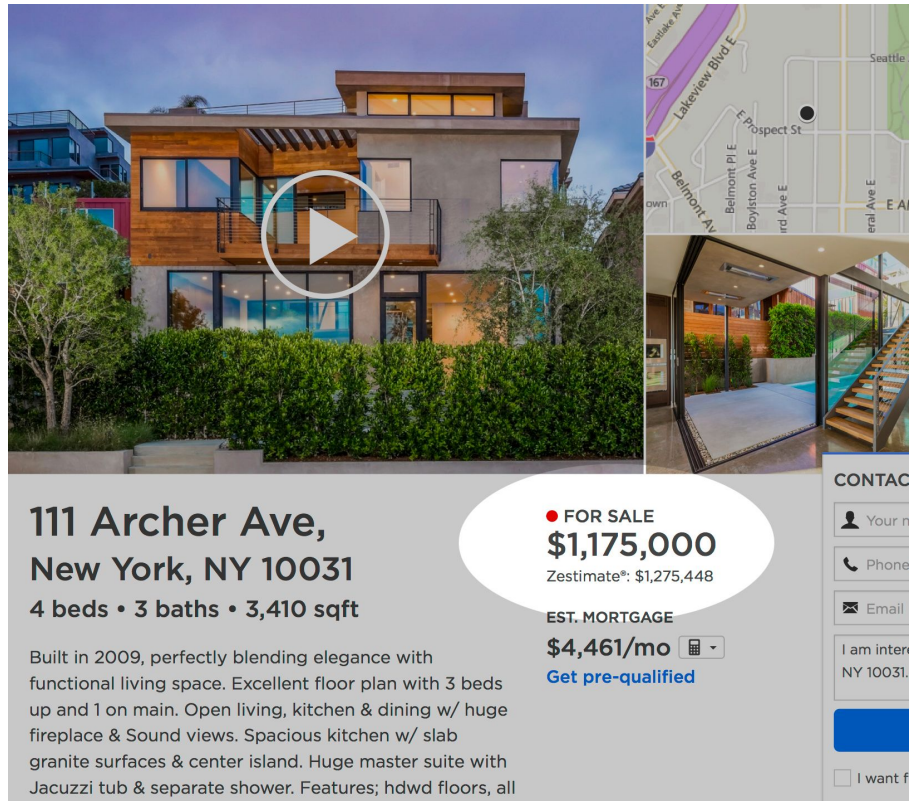

https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Dataset

A set of objects and target variables

$$\{(x_n, y_n)\}_{n=1}^{N}$$

N — sample size




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Algorithm

A function which predicts target variable for any object $\boldsymbol{x}$:

$$a(x) : \mathbb{X} \to \mathbb{Y}$$

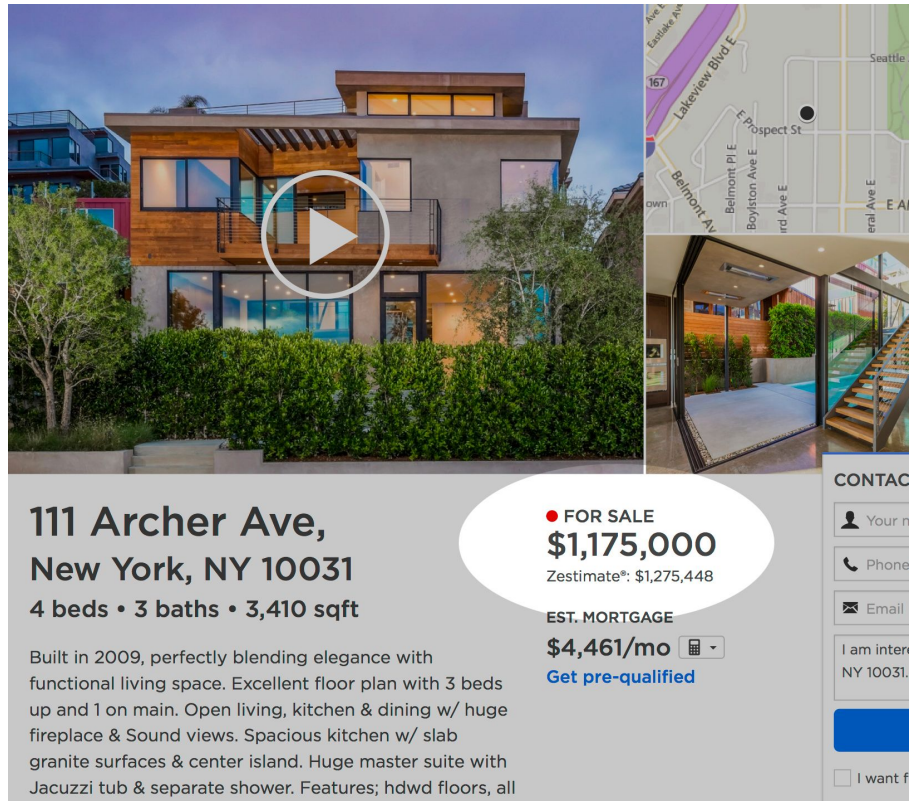
# Definitions Algorithm

A function which predicts target variable for any object $\boldsymbol{x}$:

$$a(x) : \mathbb{X} \to \mathbb{Y}$$

E.g. Linear Model:

$$a(x) = 5000 + 1000(\text{area}) - 500(\text{distance to center})$$




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Parameters and Hyperparameters

Model — function which predicts target variable for any object $x$:

$$a(x) : \mathbb{X} \to \mathbb{Y}$$

Parameters

What we learn about a(x) from the data

Hyperparameters

External configuration, determined before training

# Definitions Loss Function

A function which measures how far prediction is from the real value for a given object:

$$L(a, x)$$

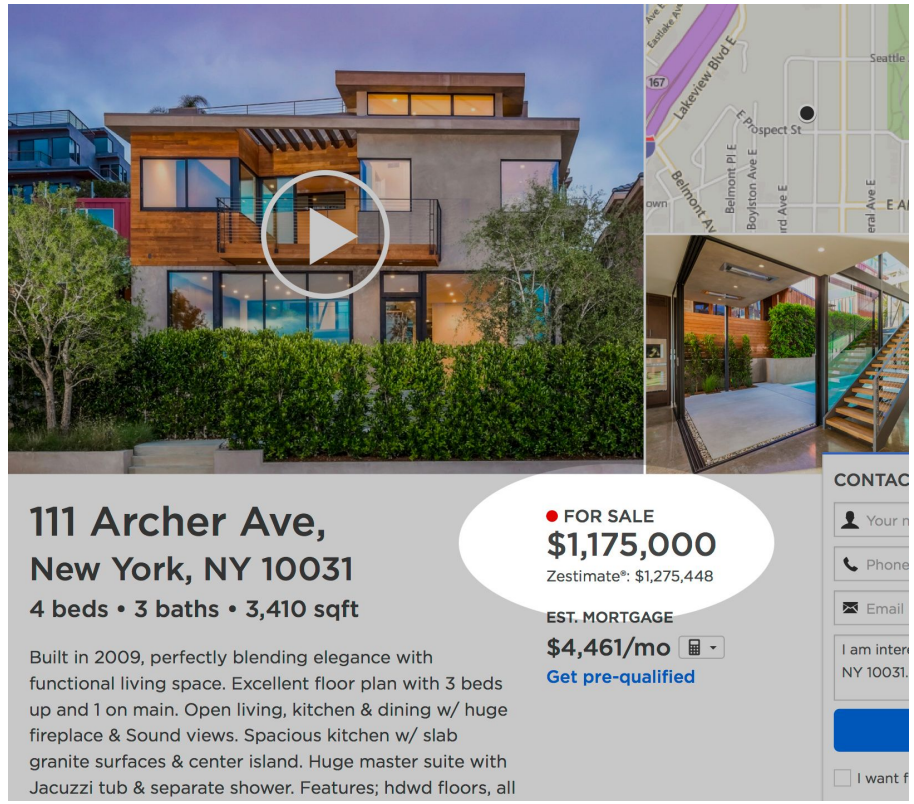
# Definitions Loss Function

A function which measures how far prediction is from the real value for a given object:

$$L(a, x)$$

E.g. when we predict real value, it is common to use squared difference

$$L(a, x) = (a(x) - y)^2$$




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Cost Function

A function which measures how far prediction is from the real value for a given dataset

$$L(a, x) \quad \Rightarrow \quad \mathcal{L}(a, X)$$

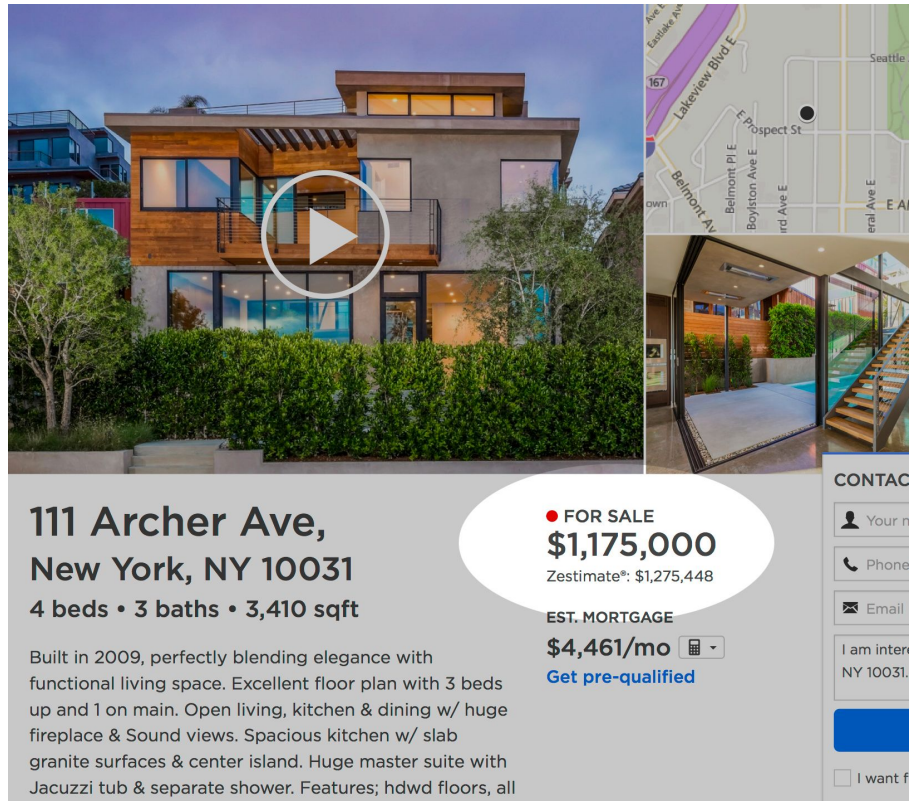
# Definitions Cost Function

A function which measures how far prediction is from the real value for a given dataset

$$L(a, x) \quad \Rightarrow \quad \mathcal{L}(a, X)$$

Squared difference -> Mean Squared Error (MSE)

$$\mathcal{L}(a, X) = \frac{1}{N} \sum_{n=1}^{N} (a(x_n) - y_n)^2$$




https://www.kaggle.com/c/zillow-prize-1/overview/competition-overview

# Definitions Training an Algorithm

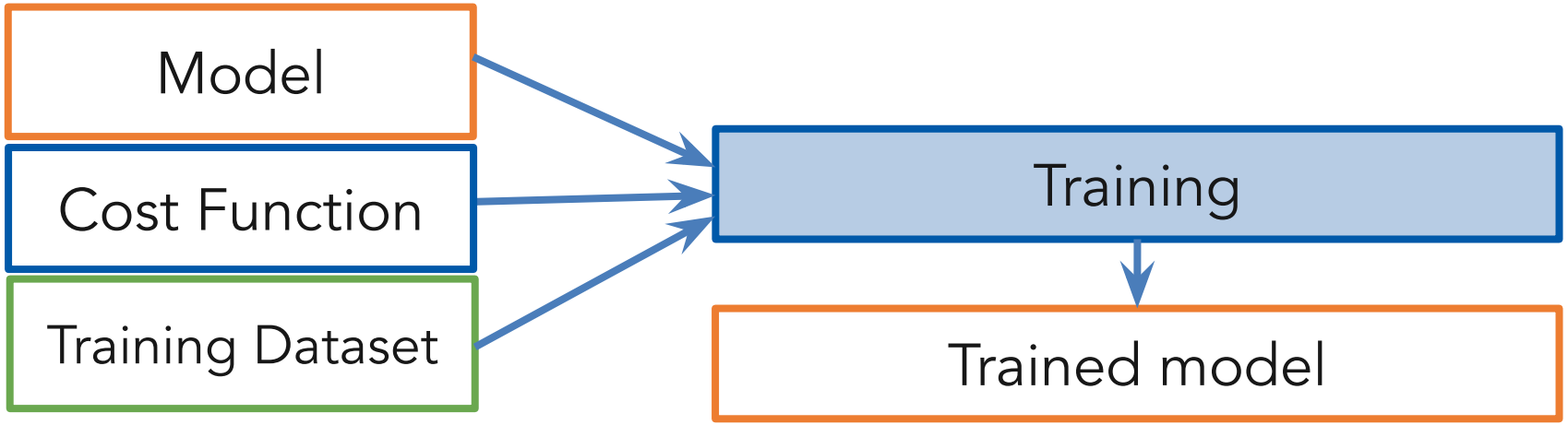
- We have a dataset and cost function

- Family of models $\mathscr{A}$
  - What we are choosing the model from
    - E.g. a linear model with d features
    - $\mathcal{A} = \{w_0 + w_1 x_1 + \cdots + w_d x_d \mid w_0, w_1, \ldots, w_d \in \mathbb{R}\}$

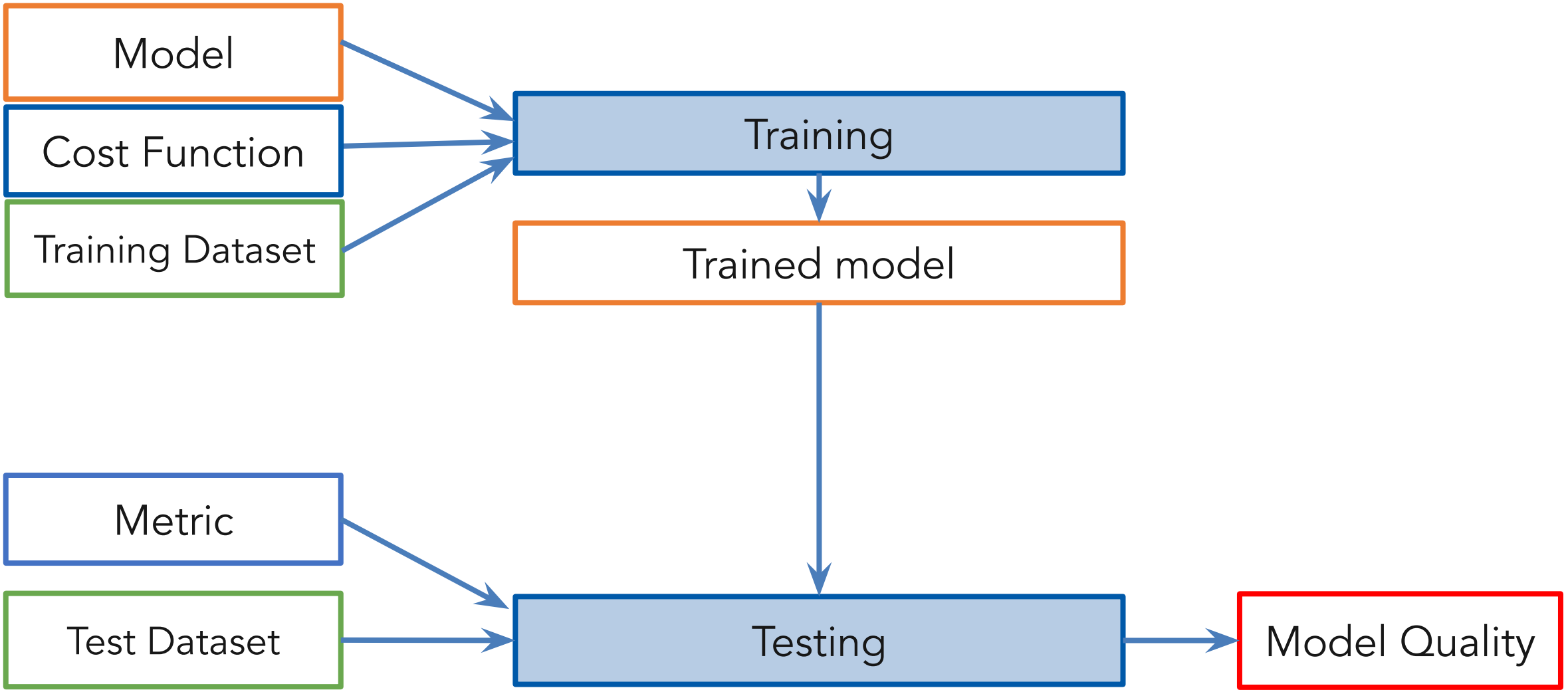
- Training: search of an optimal model in term of a chosen cost function

# Definitions

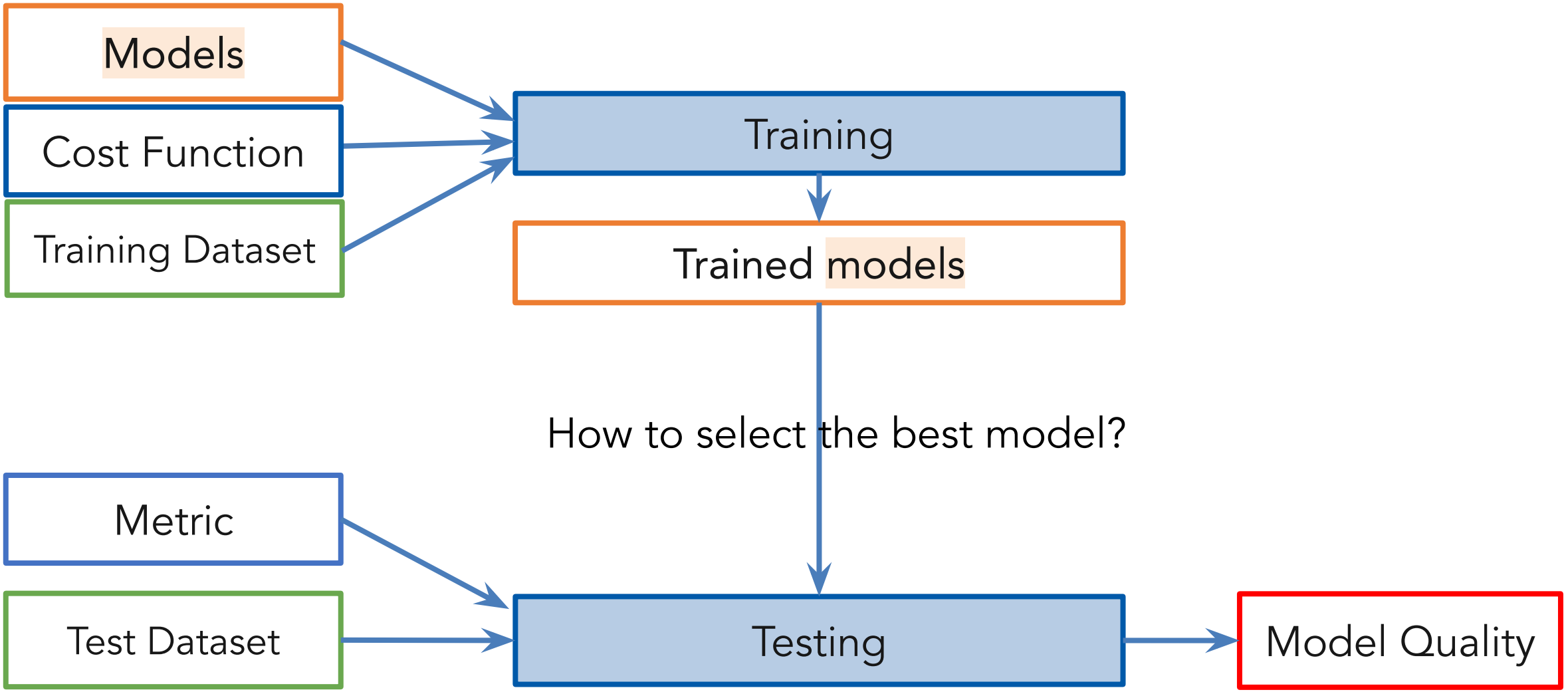
## Training an Algorithm



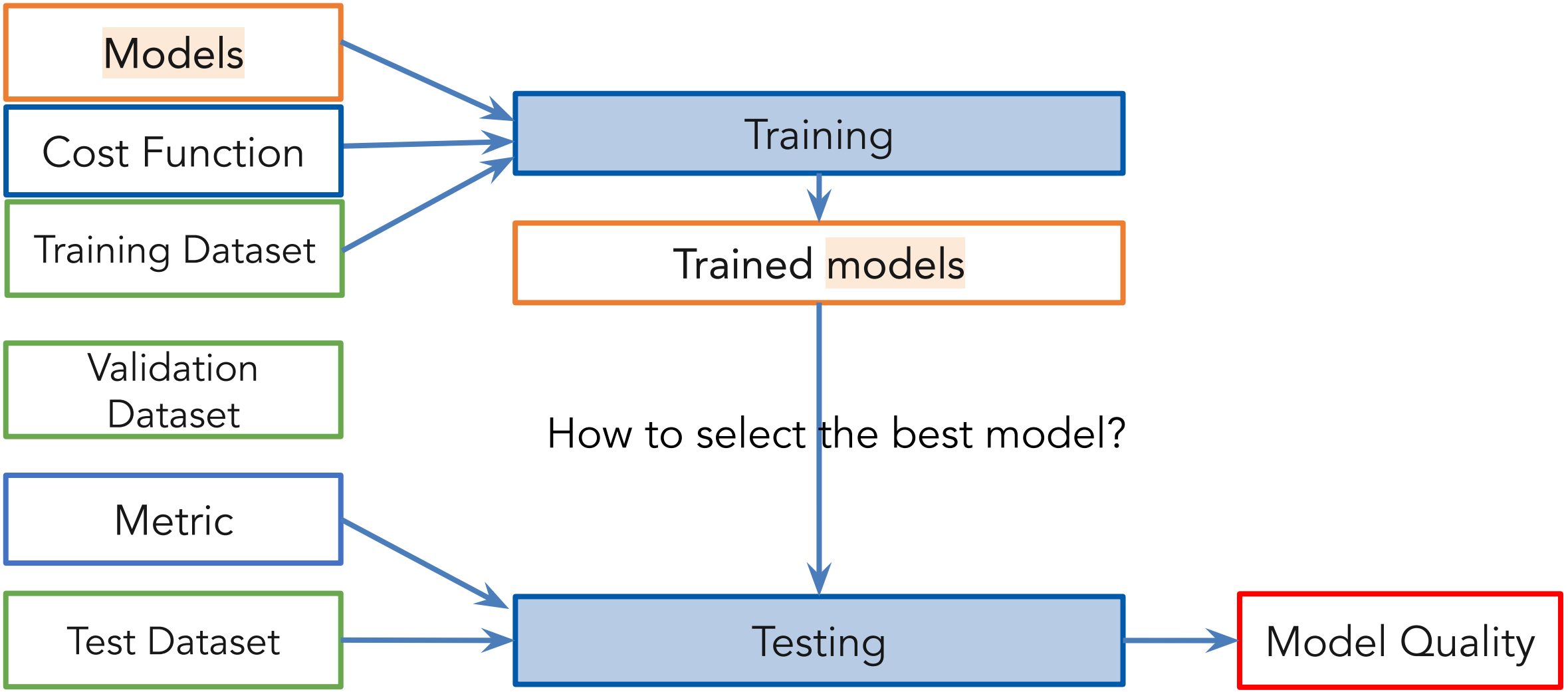
# ML Pipeline


Model
Cost Function
Training Dataset
Training
Trained model
Metric
Test Dataset
Testing
Model Quality

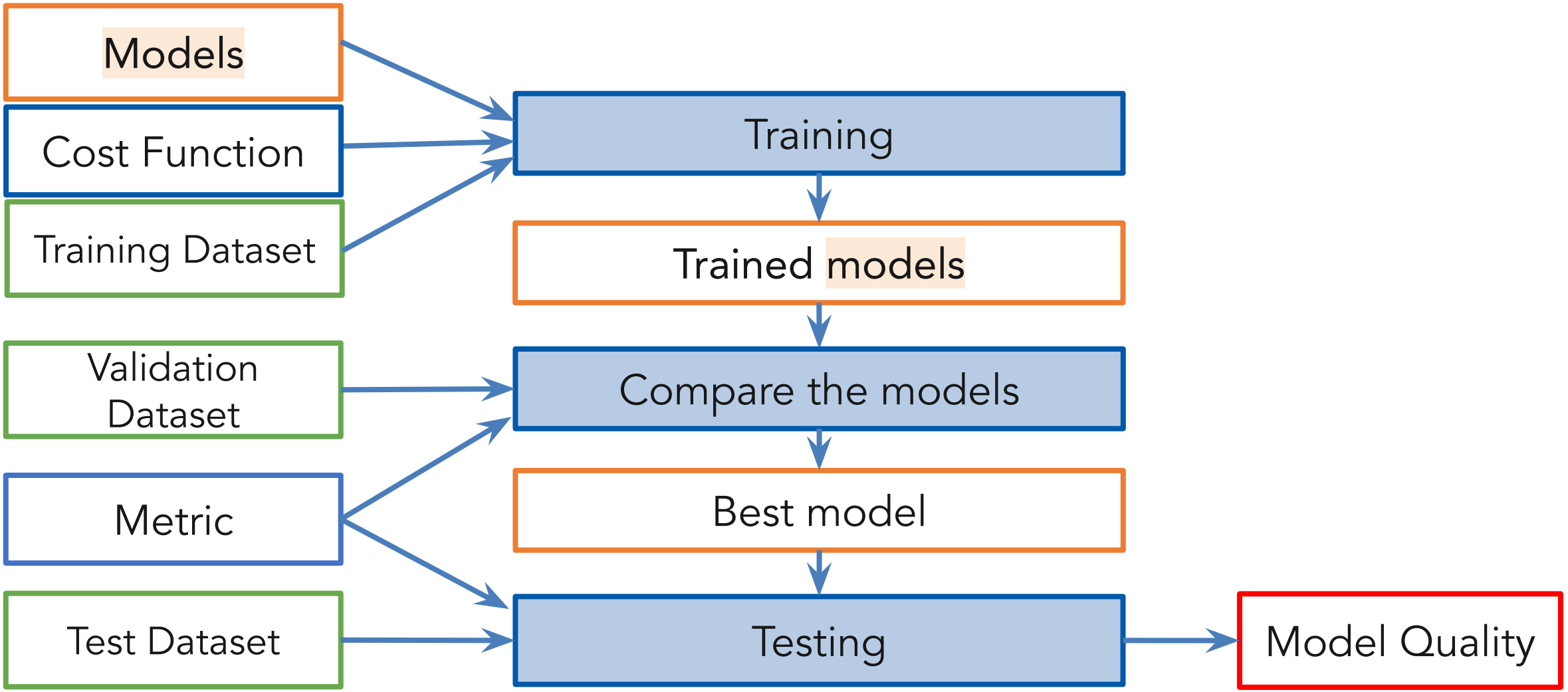
# ML Pipeline


Models
Cost Function
Training Dataset
Training
Trained models
How to select the best model?
Metric
Test Dataset
Testing
Model Quality

# ML Pipeline


Models
Cost Function
Training Dataset
Training
Trained models
Validation Dataset
How to select the best model?
Metric
Test Dataset
Testing
Model Quality

# ML Pipeline

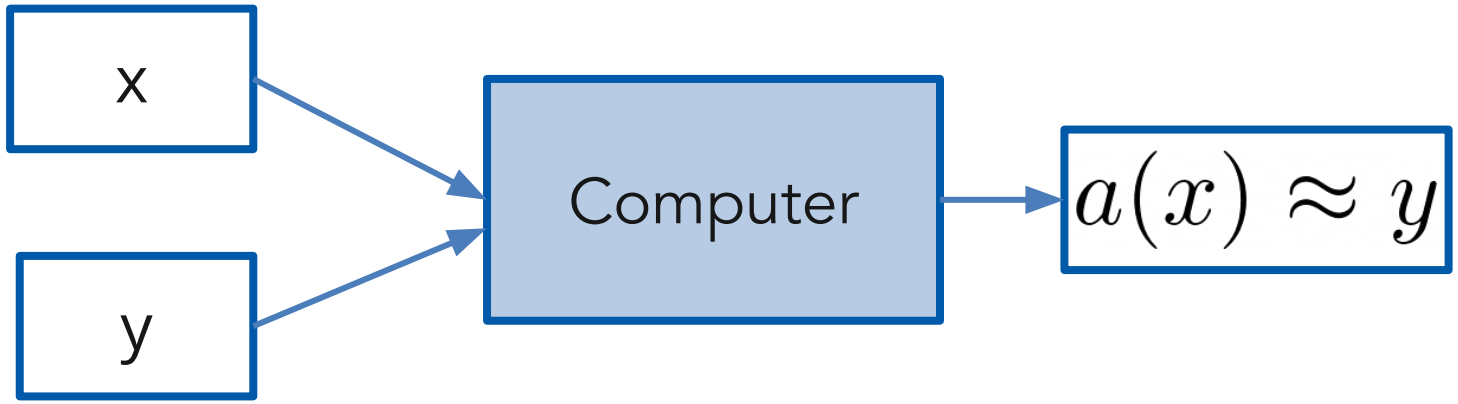

Models
Cost Function
Training Dataset
Training
Trained models
Validation Dataset
Compare the models
Metric
Best model
Test Dataset
Testing
Model Quality

# Types of Tasks

# Supervised and Unsupervised

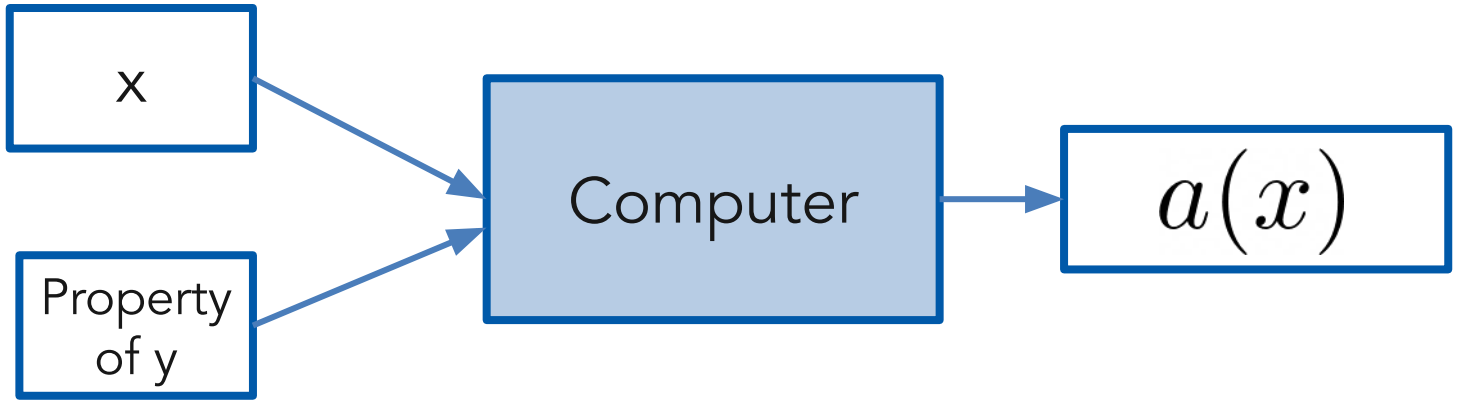
Supervised learning

- We have “correct answers” in training dataset




Unsupervised learning

- We have only objects



# Classification $\mathbb{Y} = \{1, \cdots, K\}$

# 2006


Asirra is a human interactive proof that asks users to identify photos of cats and dogs. It's powered by over three million photos from our unique partnership with Petfinder.com. Protect your web site with Asirra ◆ free!

Please select all the cat photos:

Score Test

You're a **bot**!

Computer accuracy = 60%  Probability to guess= $0.6^{12} = 0.00217$

# 2014

Completed • Swag • 215 teams

## Dogs vs. Cats

Wed 25 Sep 2013 – Sat 1 Feb 2014 (8 months ago)

Dashboard

Private Leaderboard - Dogs vs. Cats

This competition has completed. This leaderboard reflects the final standings.

See someone

| # | Δ1w | Team Name * in the money | Score | Entries | Last Submission UTC (Best – Last |
|---|---|---|---|---|---|
| 1 | — | Pierre Sermanet * | 0.98914 | 5 | Sat, 01 Feb 2014 21:43:19 (- |
| 2 | ↑26 | orchid * | 0.98309 | 17 | Sat, 01 Feb 2014 23:52:30 |

Computer accuracy = 98%  Probability to guess = $0.98^{12} = 0.875$

# Example 1: Credit scoring

https://www.kaggle.com/c/home-credit-default-risk

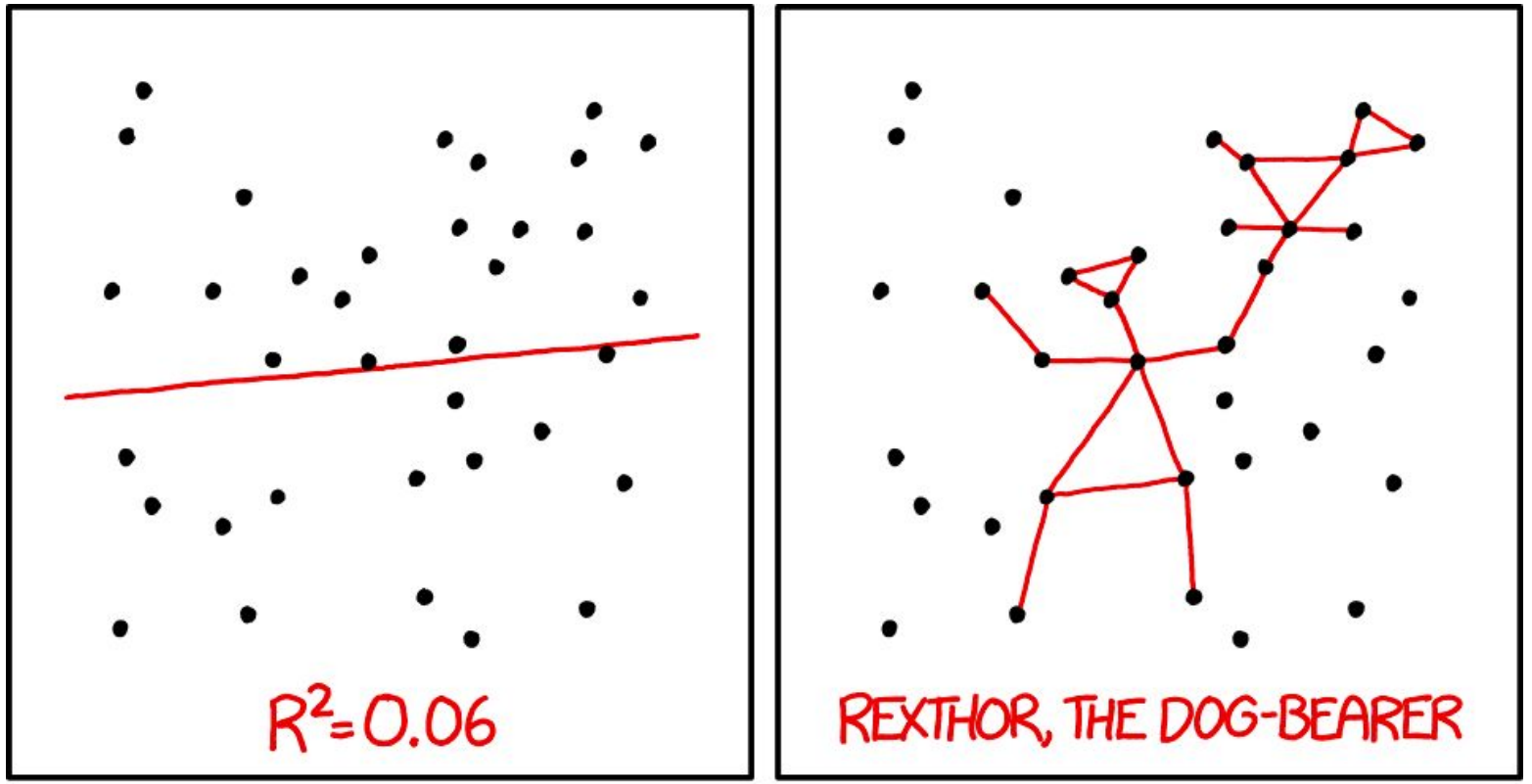
# Example 2: Sentiment analysis

https://pypi.org/project/dostoevsky/
https://www.kaggle.com/c/tweet-sentiment-extraction

# Example 3: Image Segmentation

https://ai.facebook.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/
https://www.med.upenn.edu/cbica/brats2020/tasks.html

# Regression $\mathbb{Y} = \mathbb{R}$




I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

# Example 1: Predicting age

https://www.kaggle.com/c/trends-assessment-prediction
https://www.how-old.net

# Example 2: Detection

https://ai.facebook.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/

# Clustering

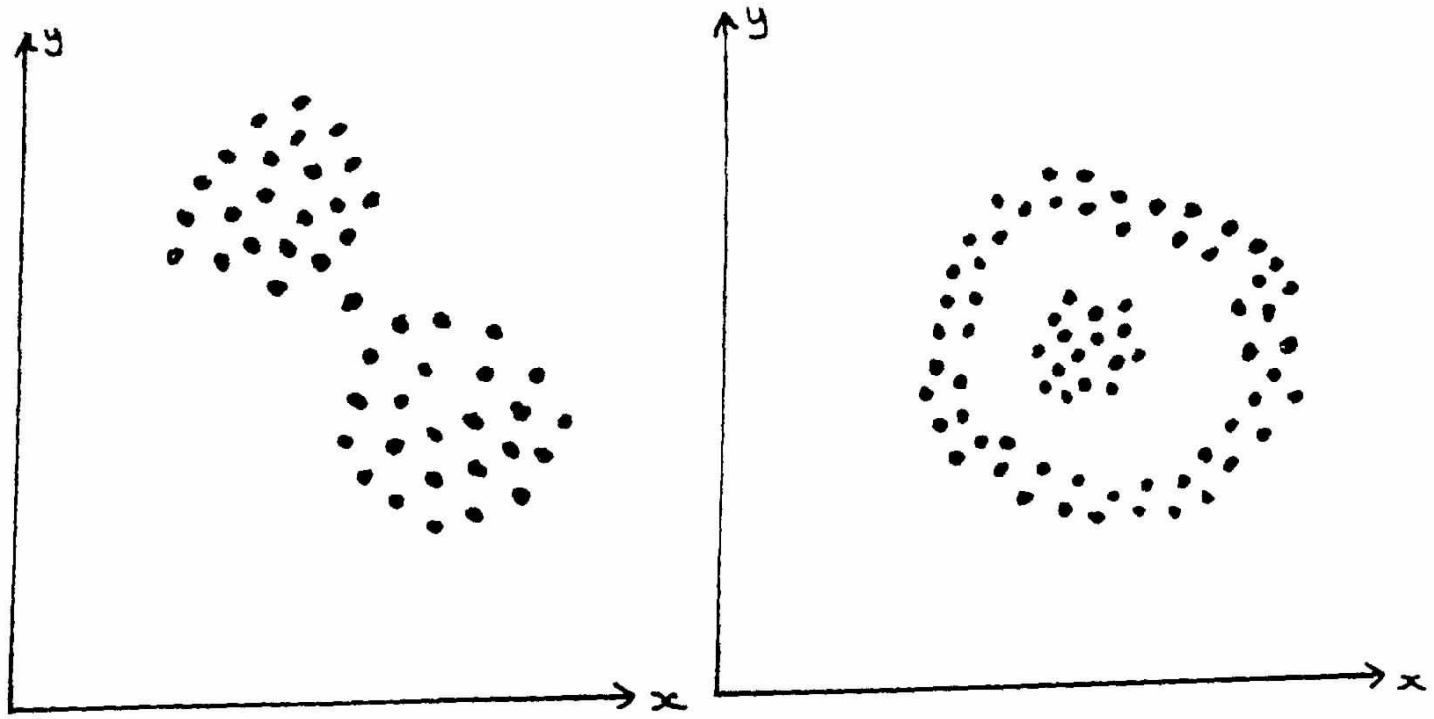
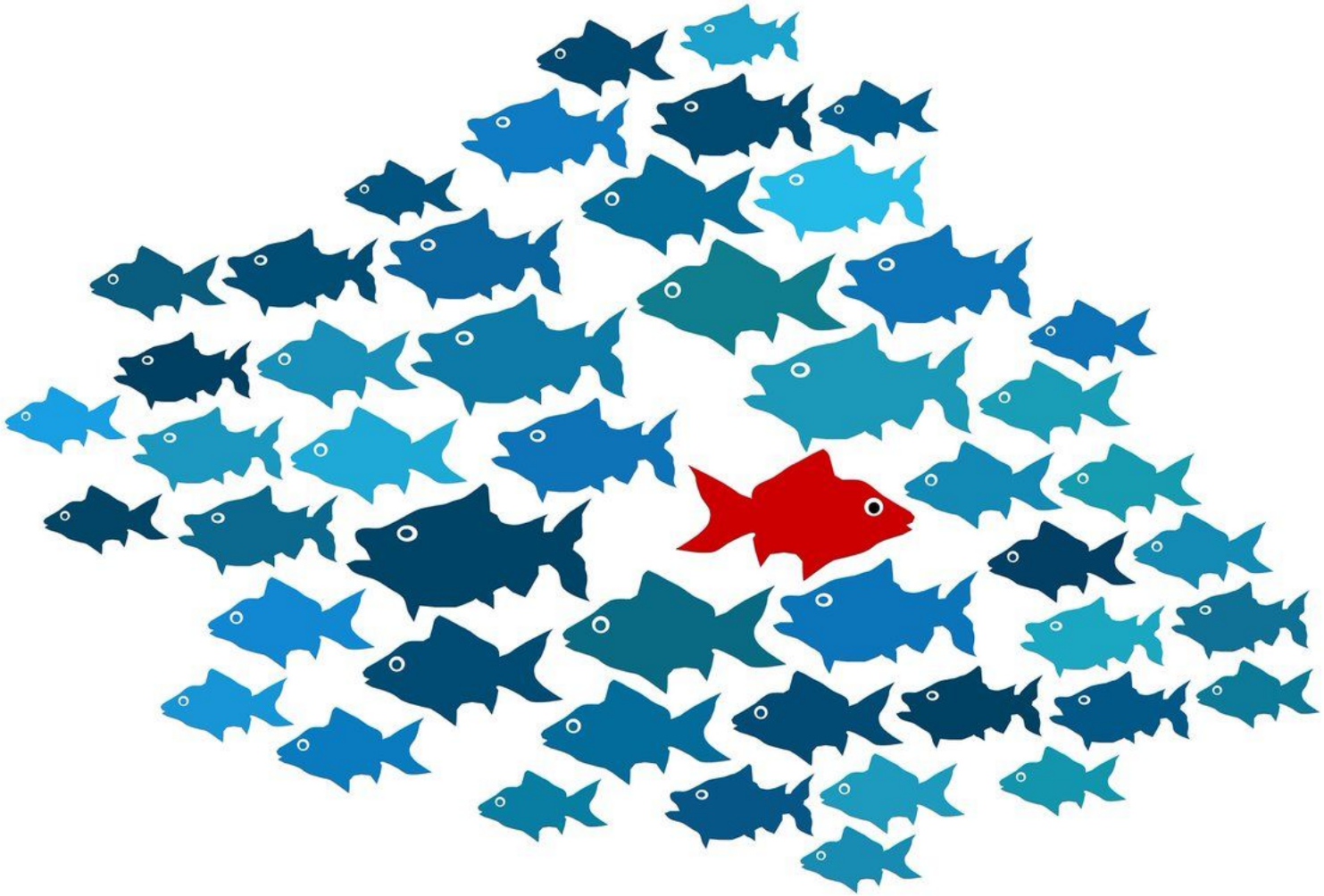
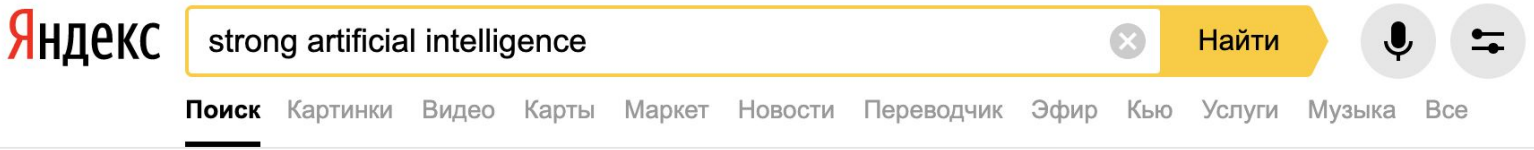



Figure 3.4 Further Examples of Clusters

# Anomaly Detection



# Ranking




**Strong** AI - Wikipedia
**en.wikipedia.org** › Strong AI
**Strong artificial intelligence** or, True AI, may refer to: **Artificial** general **intelligence**, a hypothetical machine that exhibits behavior at least as skillful and flexible as humans do, and the research program of building such an **artificial** general **i...** Читать ещё ›

Сильный и слабый искусственные **интеллекты**...
**ru.wikipedia.org** › Сильный и слабый искусственные интеллекты
Сильный и слабый искусственные **интеллекты** — гипотеза в философии искусственного **интеллекта**, согласно которой некоторые формы искусственного **интеллекта** могут...

Deep learning programming. bitcoin, blockchain.
**strongartificialintelligence.com**
**Strong Artificial Intelligence** is the born of new era for programming machines. Supercomputers need new language and different algorithms and we give the key for...
Читать ещё ›

What is **Strong** AI? | IBM
**ibm.com** › cloud/learn/strong-ai
**Strong artificial intelligence** (AI), also known as **artificial** general **intelligence** (AGI) or general AI, is a theoretical form of AI used to describe a certain mindset of AI development. If researchers are able to develop **Strong** AI, the machine would require... Читать ещё ›

**Нашлось 3 млн результатов**
21 показ в месяц
Дать объявление

# Dimensionality Reduction

# Types of Features

$D_j$ — possible values for the j-th feature

# Numerical $D_j = \mathbb{R}$

Examples:

- Area of an apartment
- Blood pressure of a patient

# Binary $D_j = \{0, 1\}$

Examples:

- Does the apartment have gas supply
- Did the patient take the pill

# Categorical $D_j = \{u_1, \ldots, u_m\}$

Examples:

- In which district the apartment is located
- Which specialist the patient visit initially?

# Categorical: Encoding

- One-hot-encoding
  - create m binary features
- Frequency Encoding
  - how often each category appear in the training dataset
- Mean Target Encoding
  - average values of a target within category

# Ordinal $D_j = \{u_1, \ldots, u_m\}$

Examples:

- Type of locality (large city, small city, village)
- Risk factor for a patient on some scale

# Learning Outcomes

After this lecture you should know:

- Course goals

- Key objects of the course
  types of ML problems, features, target variables, etc.

- Connection of course contents with applications