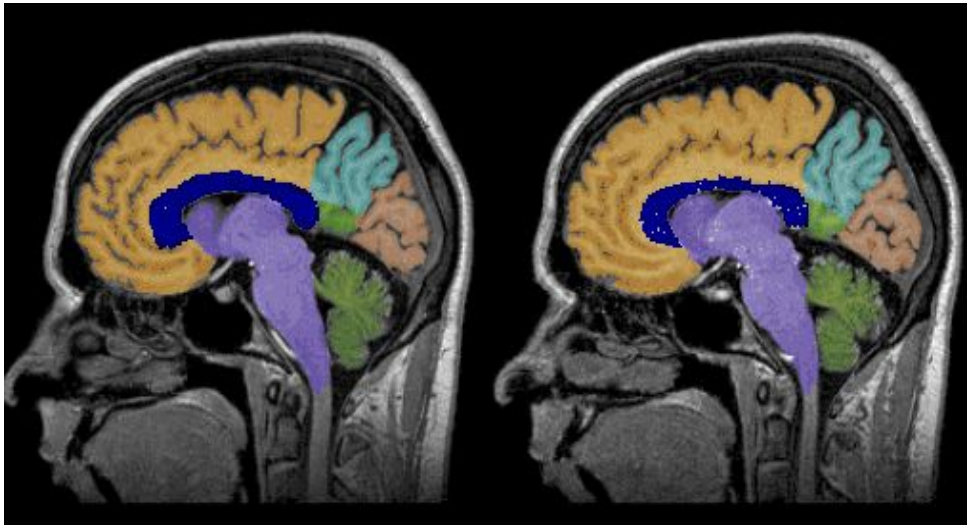
Introduction to Machine Learning

Michal Mikolajczak (mkm0796@gmail.com)

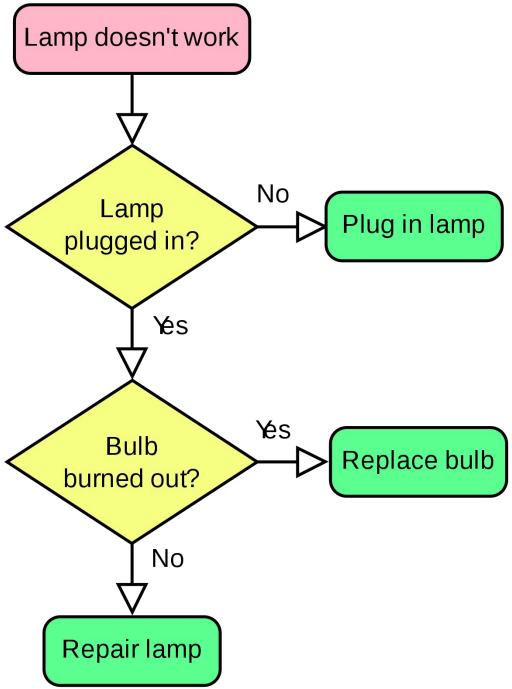
Quick introduction

Contact: mkm0796@gmail.com
or via fb/twitter (same username,photo from above)

Main fields of work: Computer Vision/Medical imaging.



Machine learning - what is it exactly?



Machine learning - what is it exactly?

Is this a dog?

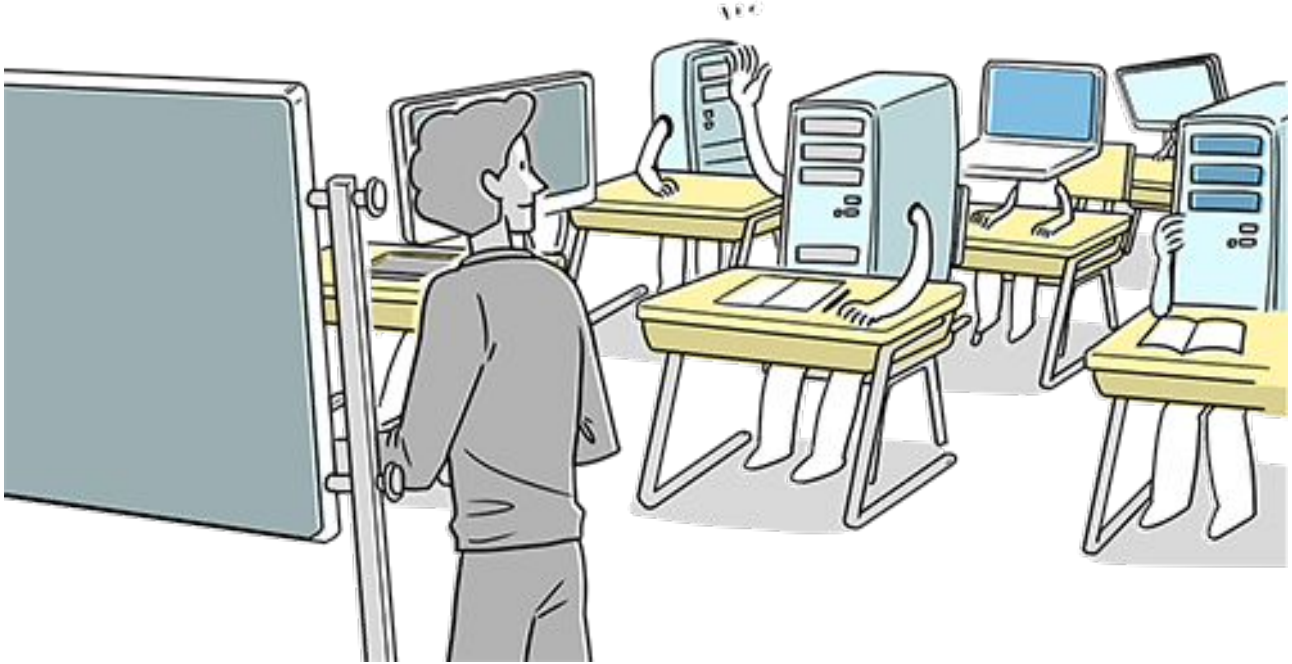
Machine learning - what is it exactly?

Is this a cat?

Machine learning - what is it exactly?

What about the computer?

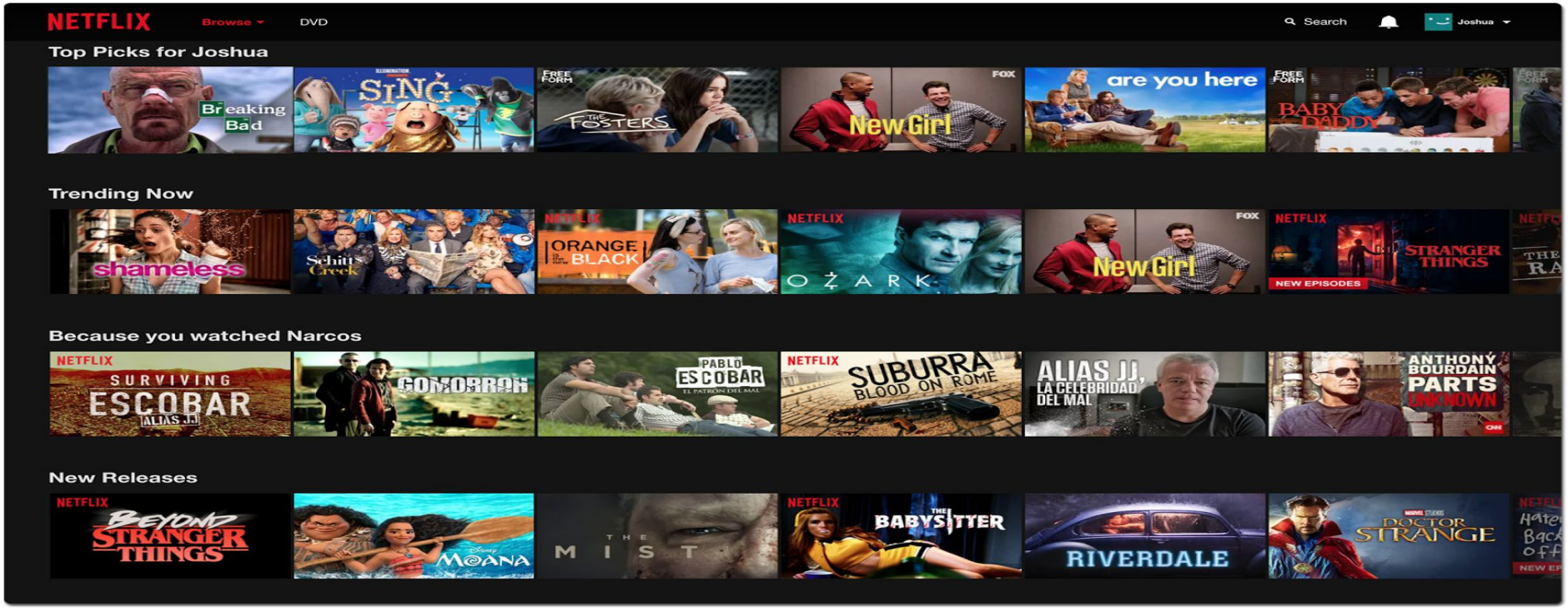
Machine learning - what is it exactly?



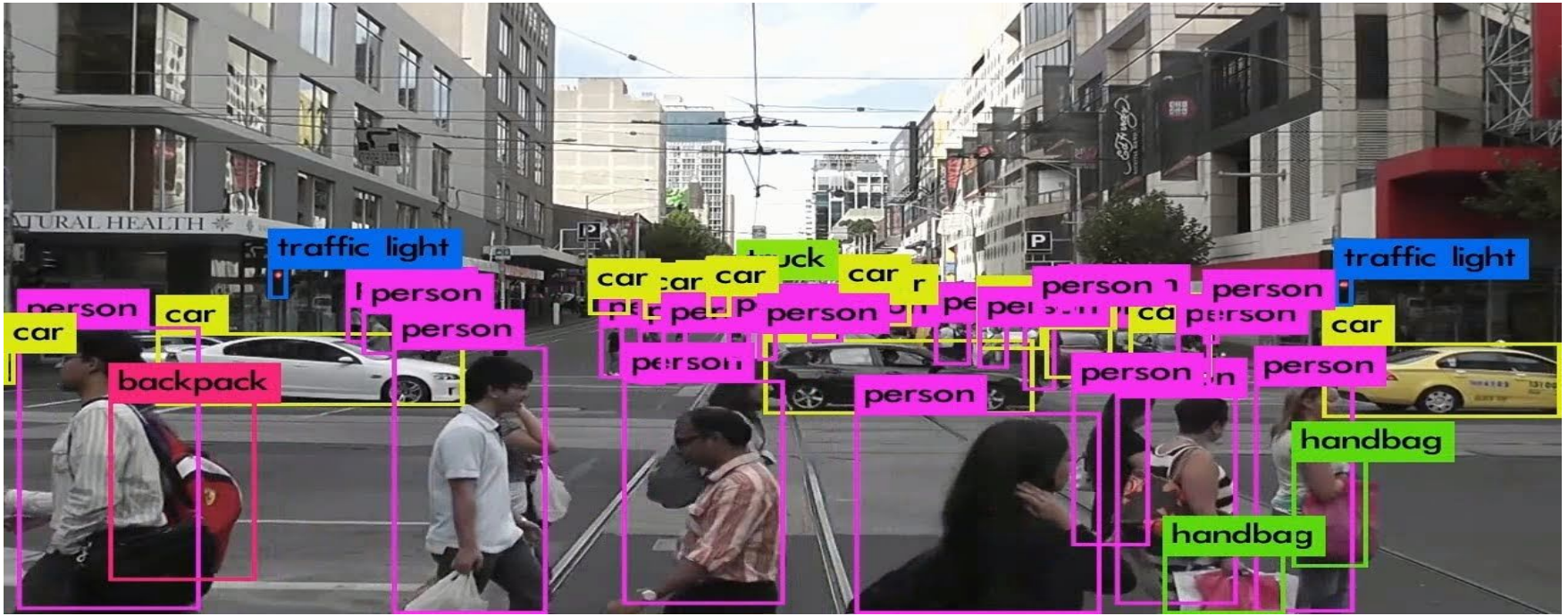
Machine learning - what is it exactly?

Machine learning is the science of programming computers so they can learn from data how to perform some specified task - without being explicitly programmed.

ML - how it can be useful?

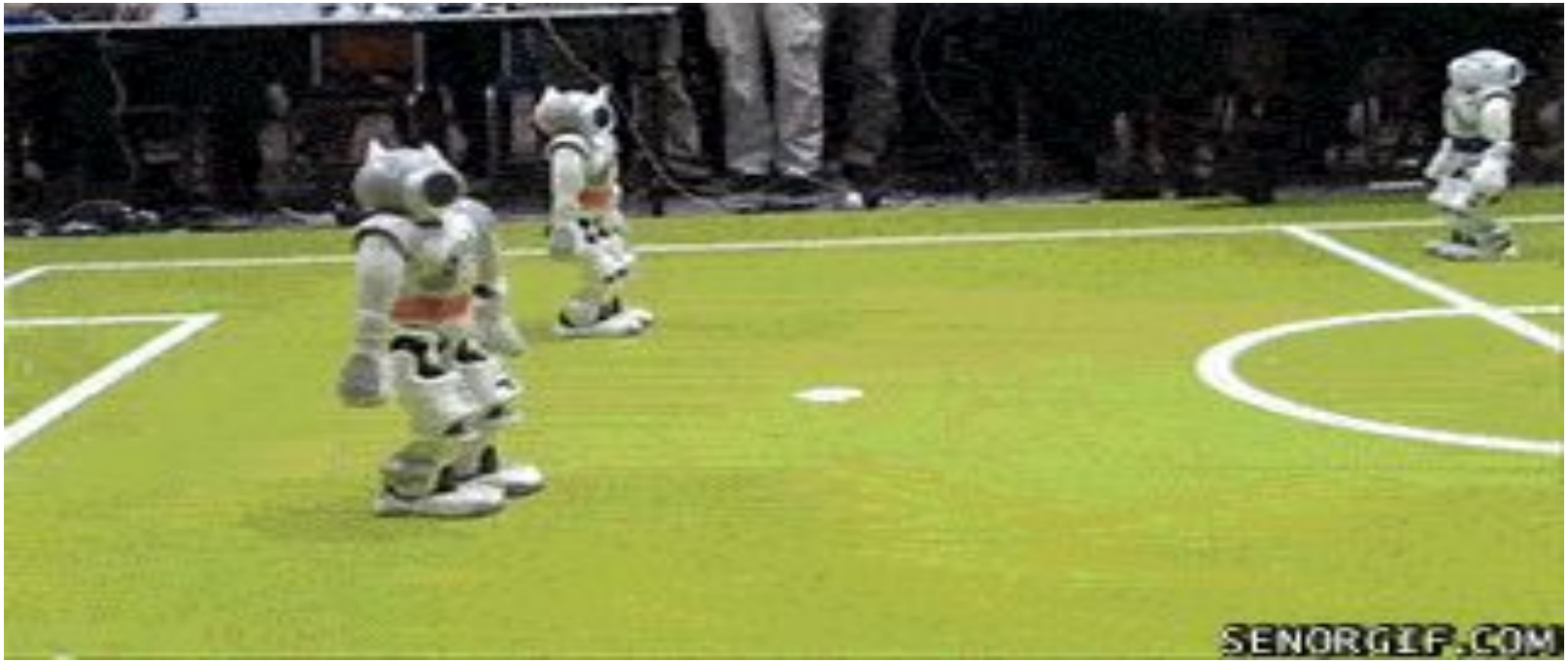


ML - how it can be useful?



ML - how it can be useful?

ML - how it can be useful?



Tasks ML is great for

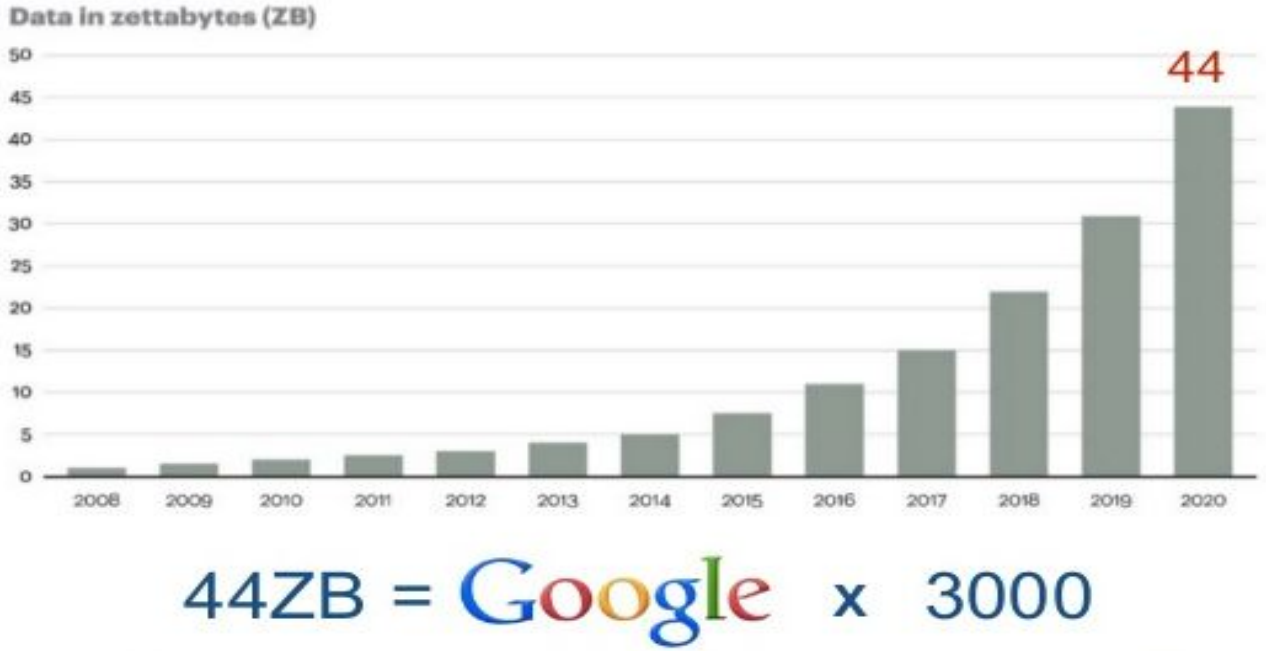
- Problems which existing solutions require a lot of manual tuning and lists of rules
- Problems with fluctuating environments, ML systems can update quite easily/automatically to new data.
- Complex problems, with no good solution at all with traditional approaches, e.g. image or audio recognition
- Reversing decision processes/finding associations/getting insights from data - so called data mining.

Sudden popularity rise

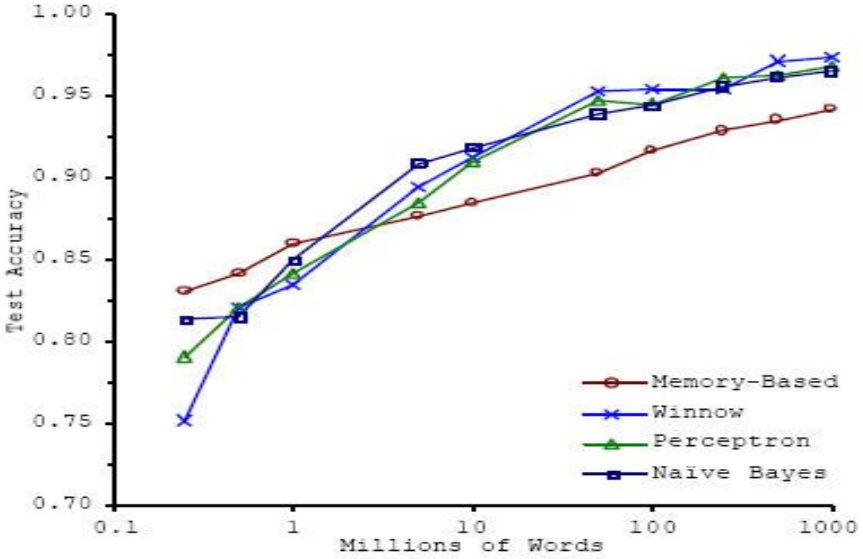
And ***why*** has ML become much more common and popular in recent years?

(besides the fact that we can do pretty cool things with it)

Exponential data growth



The unreasonable effectiveness of data



"The unreasonable effectiveness of data", Banko & Brill, 2001

- Machine learning algorithms are often called "data hungry".
- Their performance rise with the amount of training data, often outweighing selected algorithm and feature engineering.

GOOGLE DATACENTER

1,000 CPU Servers
2,000 CPUs • 16,000 cores

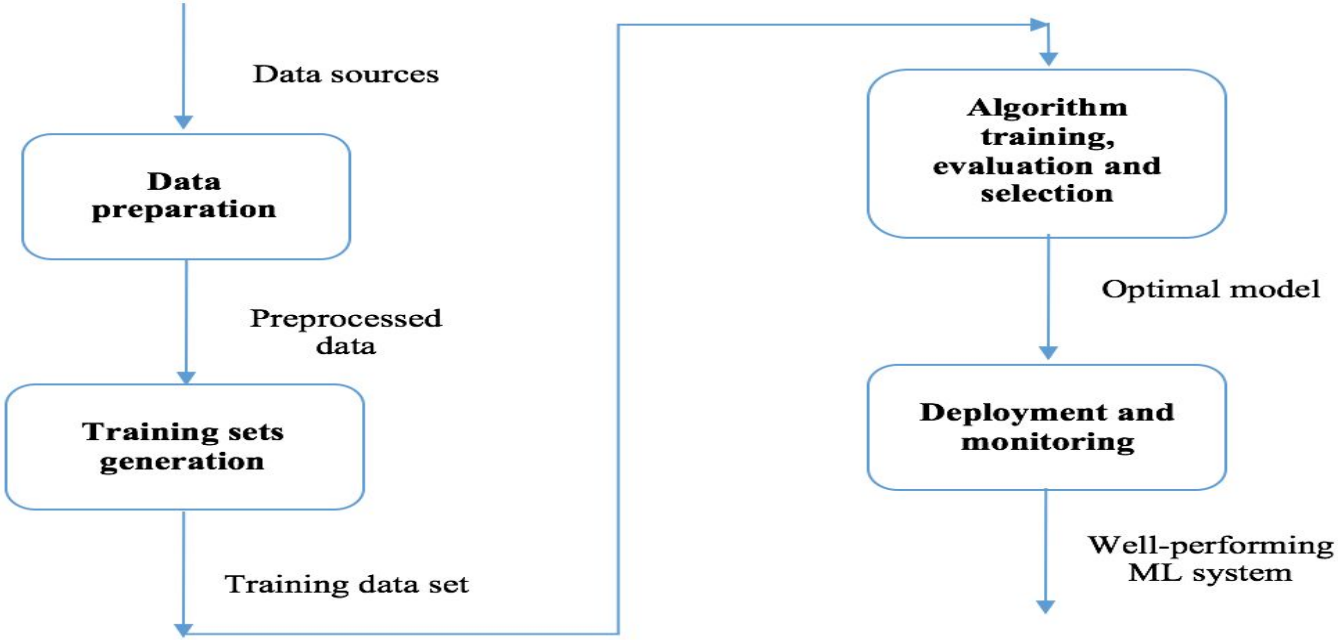
600 kWatts
$5,000,000

STANFORD AI LAB

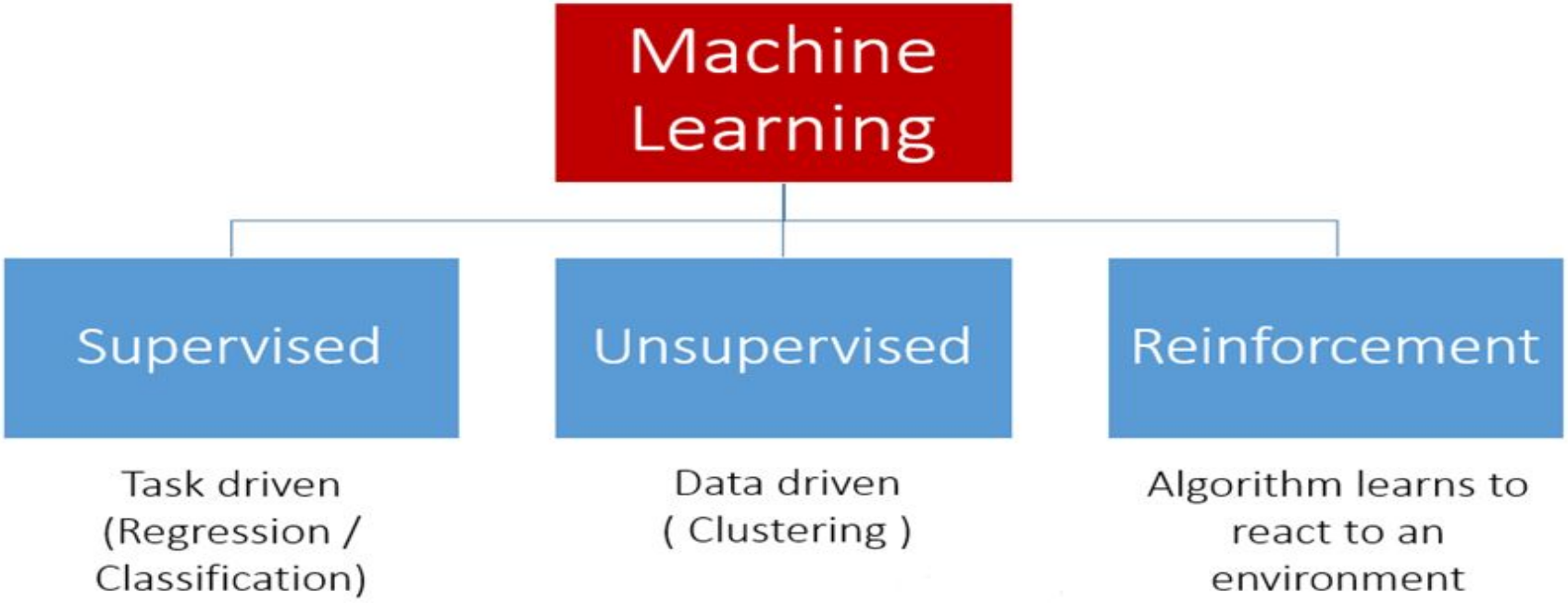
3 GPU-Accelerated Servers
12 GPUs • 18,432 cores

4 kWatts
$33,000

Machine learning workflow schema

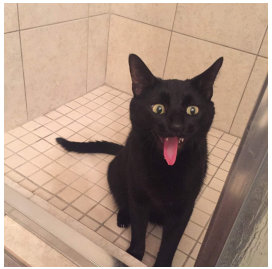


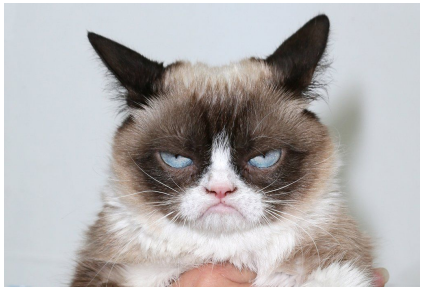
Types of machine learning



Supervised learning

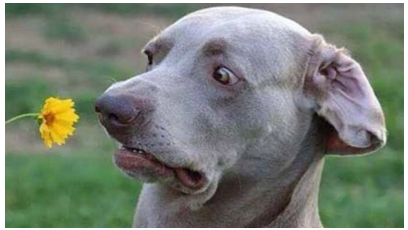
Cats



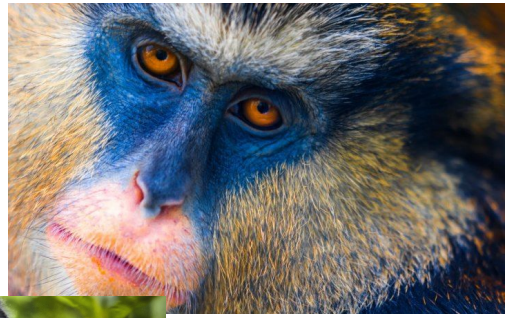
Dogs





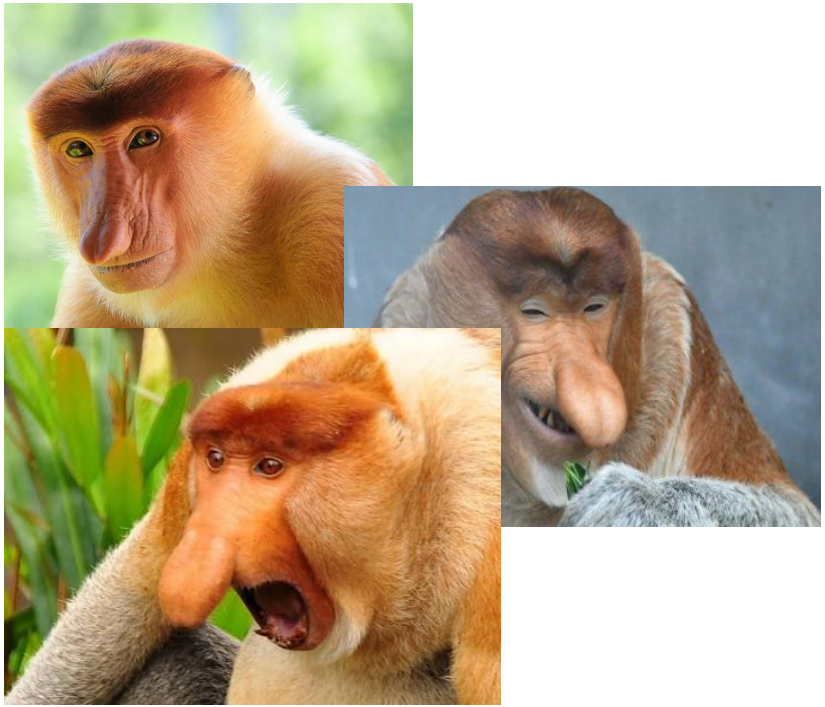

Unsupervised learning

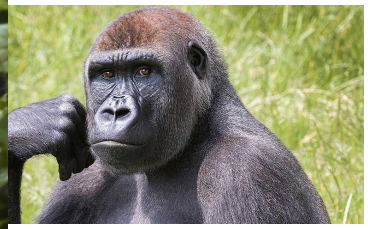




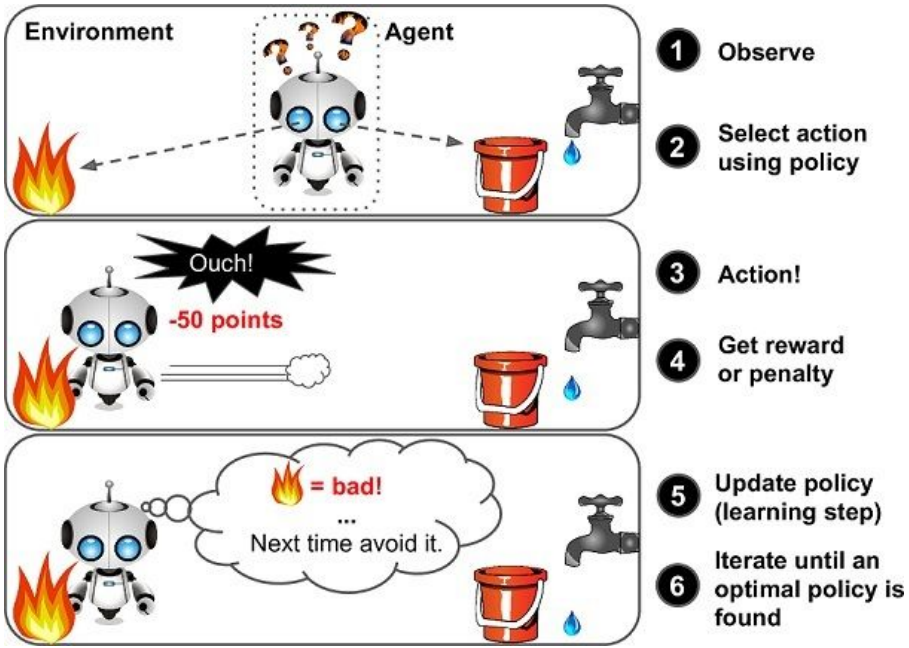


Unsupervised learning

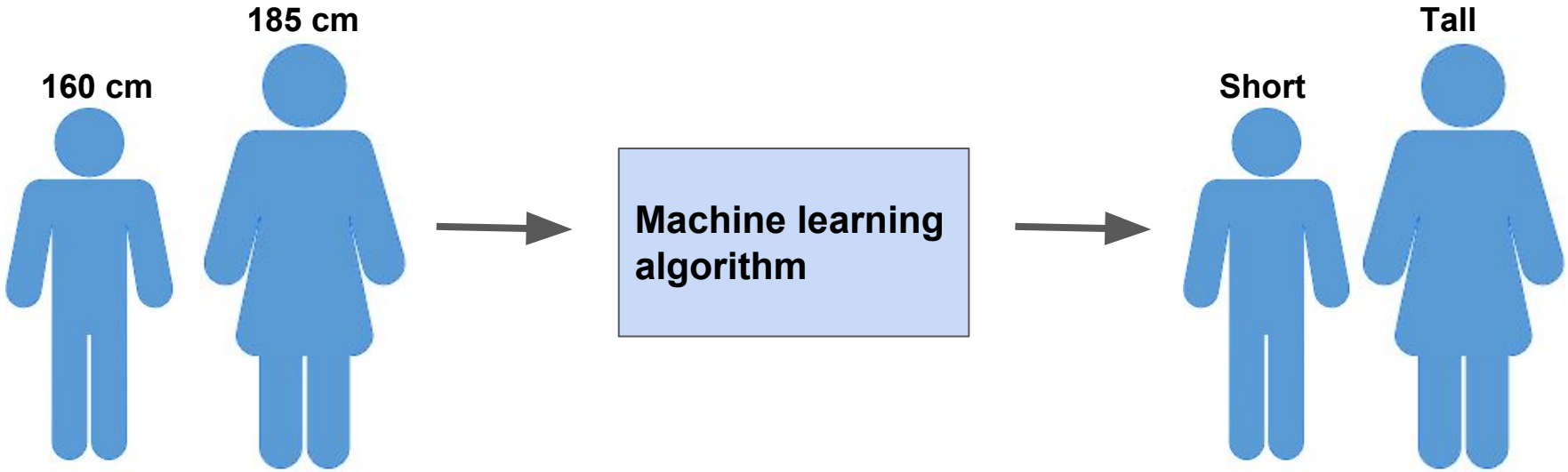




Reinforcement learning



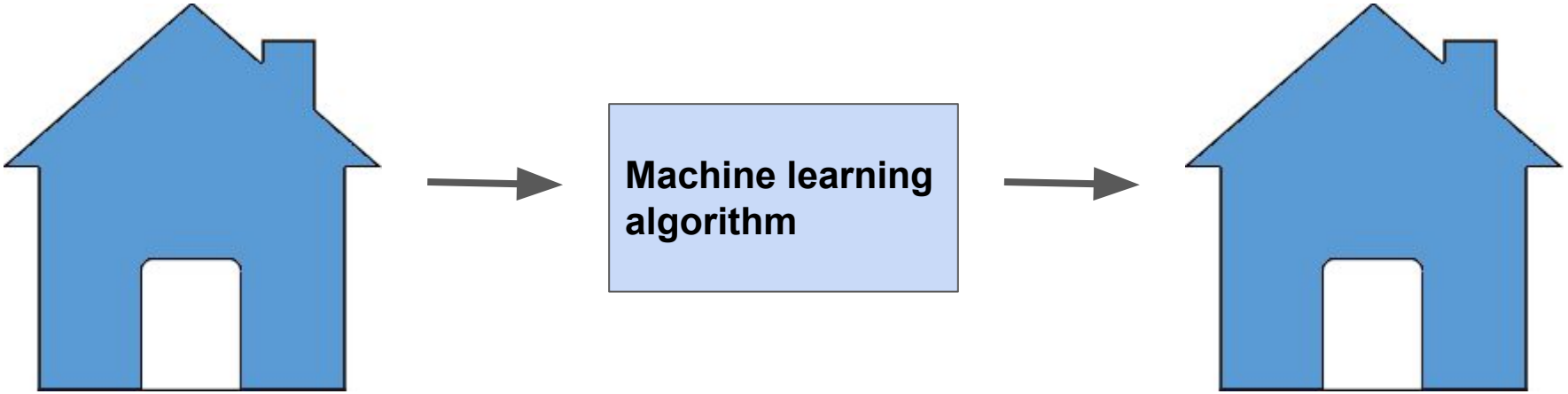
Classification problem



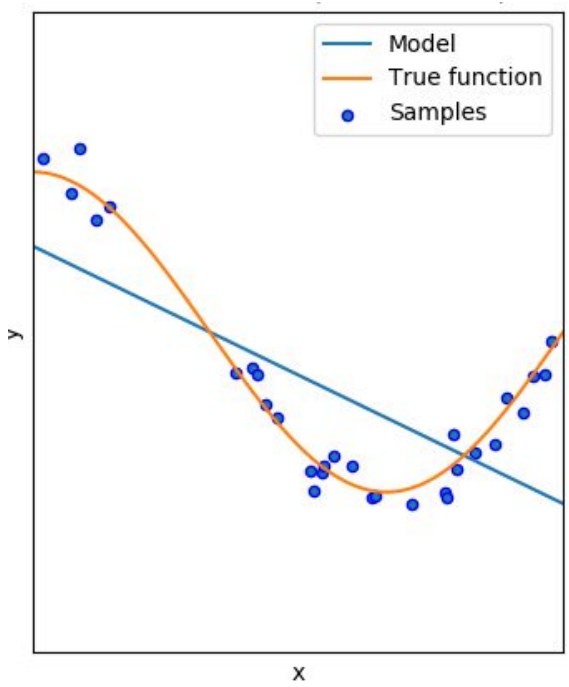
Regression problem

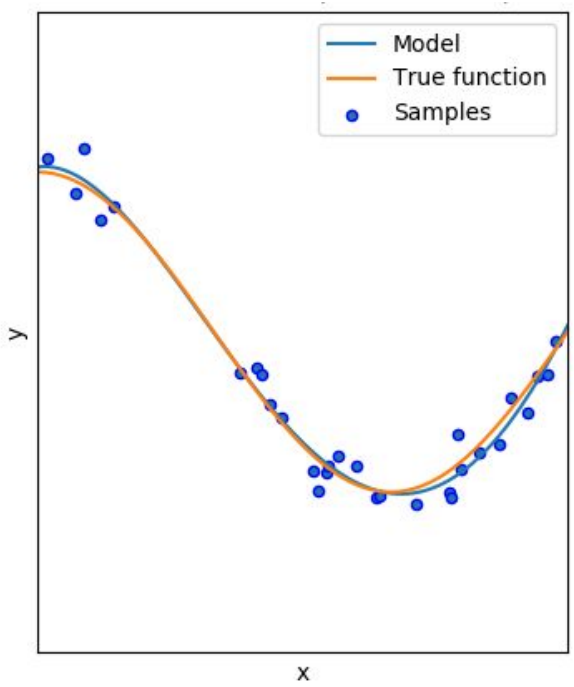
Square meters: 50
Condition: good
Location: NY

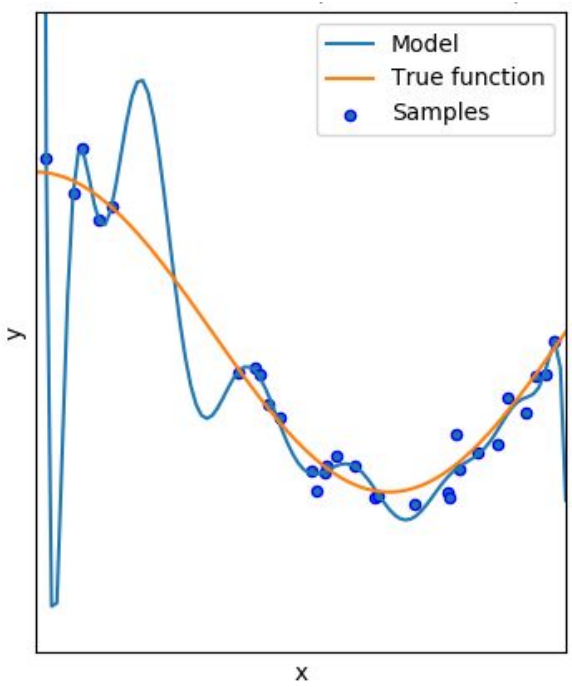
Price: 100 000$
(Continuous value)



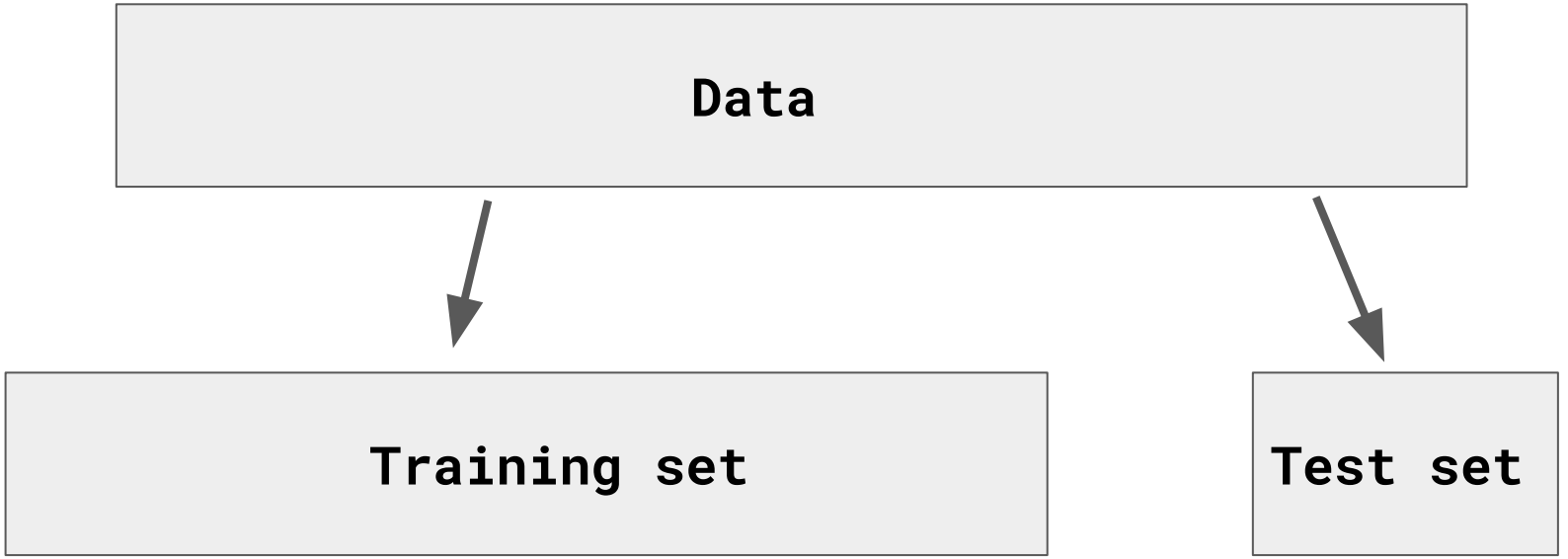
Underfitting and overfitting







Train/test splitting



Used for model fitting, feature selection, parameter tuning.

Final tests, not used in training.

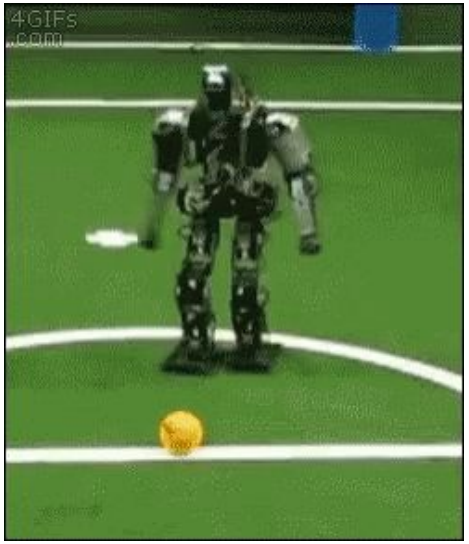
Train/test splitting

Performance on training data vs test data

But what if my machine is not enough...

- Despite what it seems like, even “casual user” machines are often good enough at the beginning of the ML adventure.
- There are some computing resources available for free, e.g. Google Colab.
- Cloud resources (Paperspace, Azure, and so on) are quite cheap (and often have some free starter/student packages).
- And at last - get used to it (as after some threshold this problem affects pretty much everyone).

Just remember - ML isn't a silver bullet...

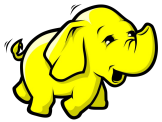


Tools and stuff





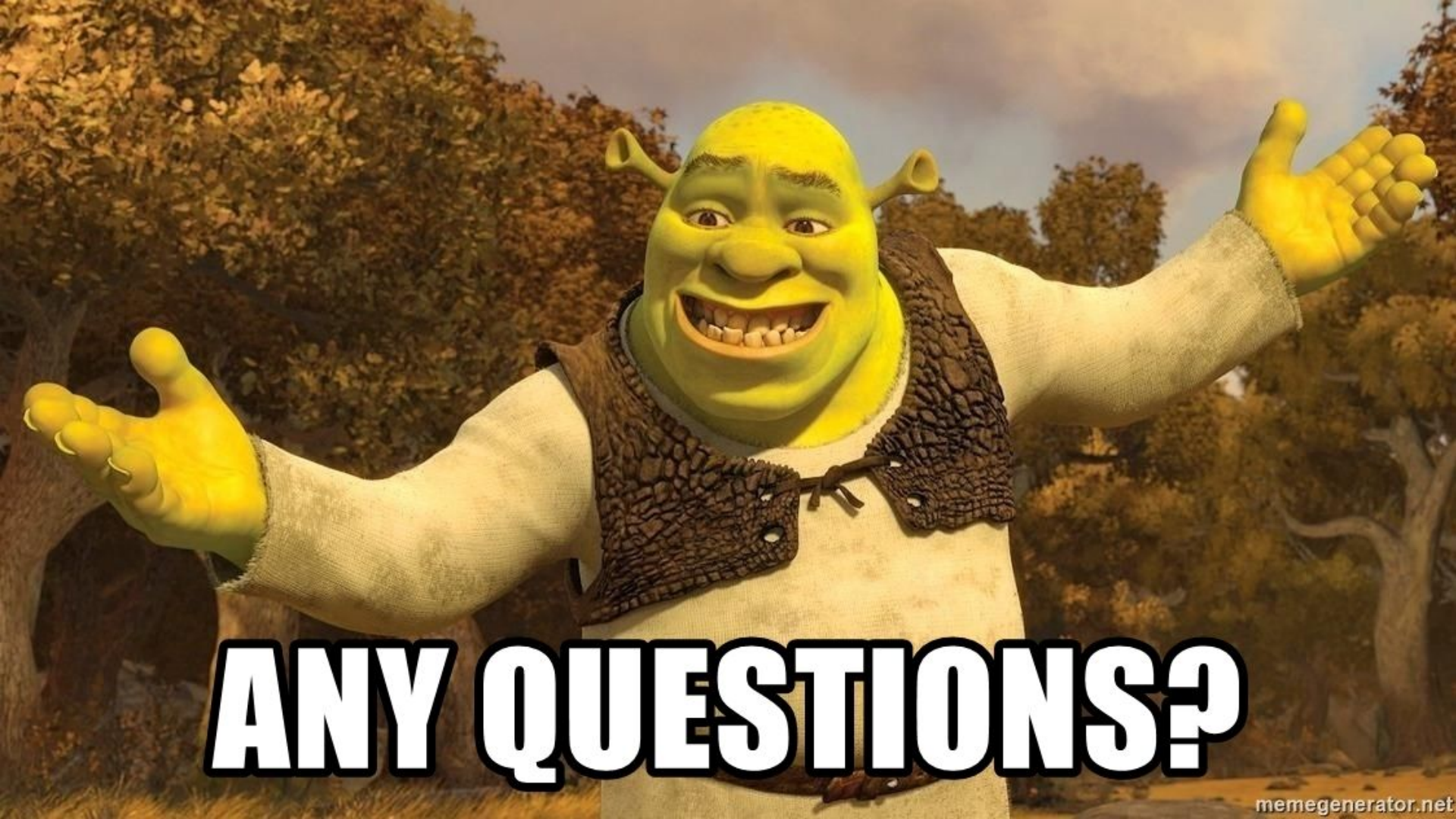









And many, many others...


ANY QUESTIONS?
memegenerator.net