

# Introduction to Machine Learning with Applications

Liang Liang

CSC646/546

# Artificial Intelligence: General AI

<https://www.hbo.com/westworld>



[https://arrow.fandom.com/fr/wiki/Gideon\\_\(Waverider\)](https://arrow.fandom.com/fr/wiki/Gideon_(Waverider))



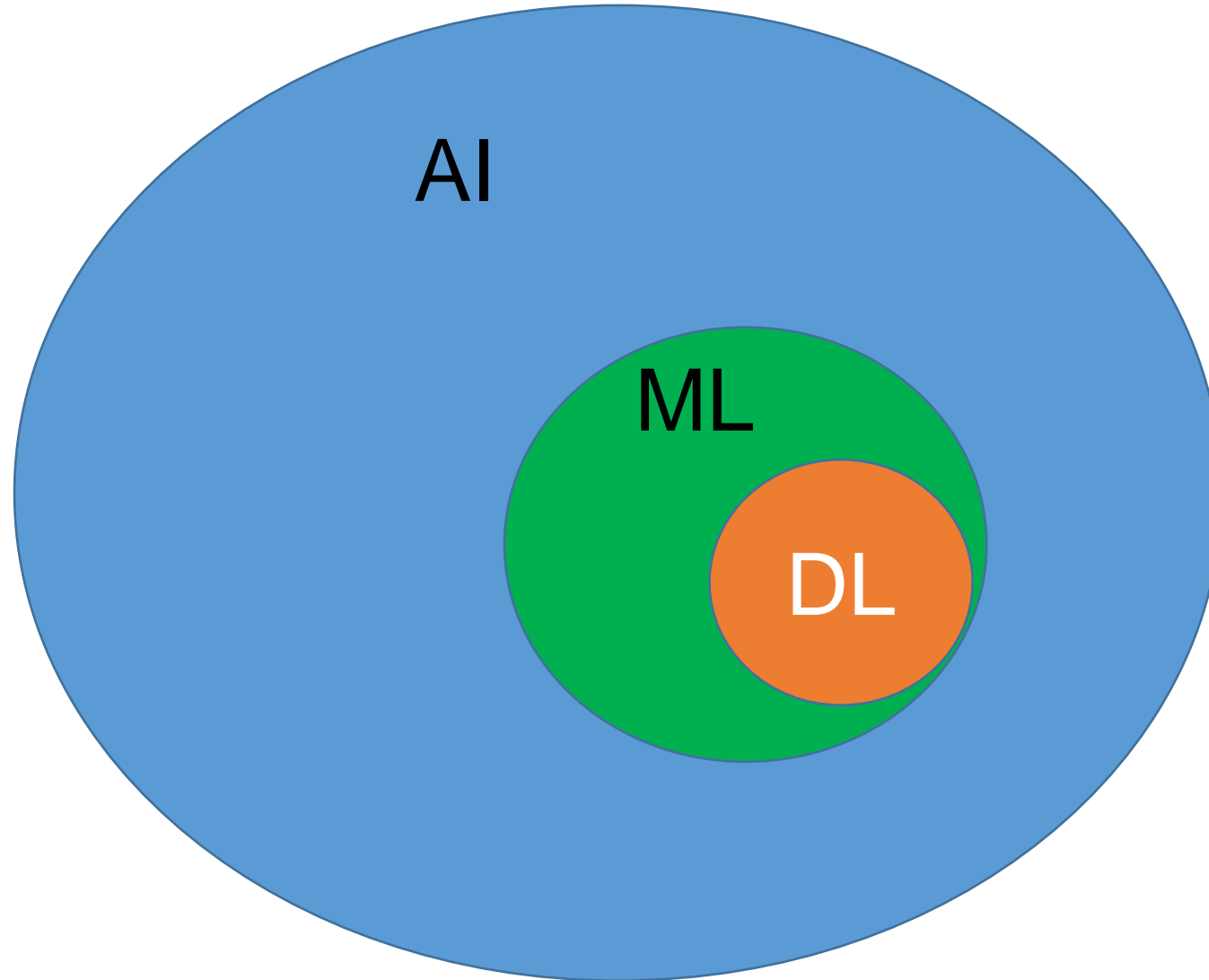
"Data" in star trek



AI: Artificial Intelligence

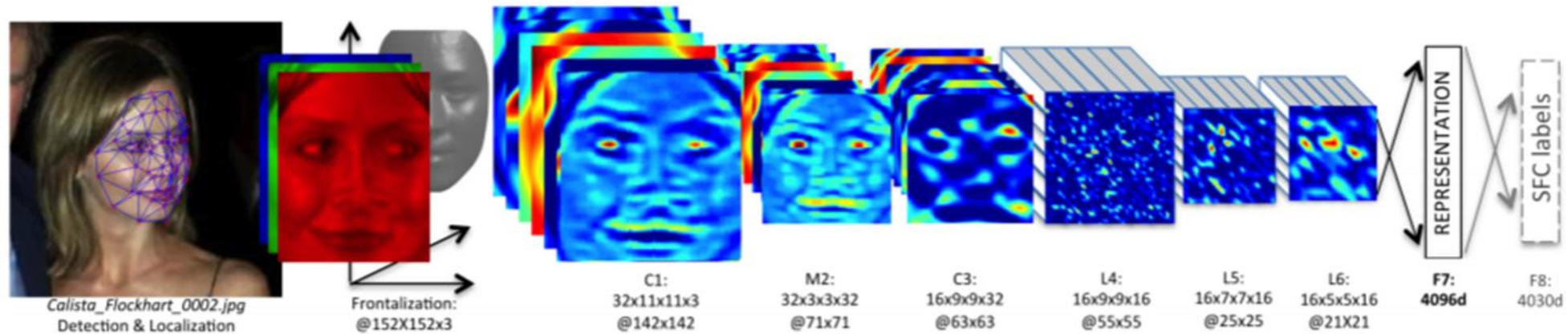
ML: Machine Learning = Specialized AI

DL: Deep (Machine) Learning = ML using Deep Neural Networks



# Machine Learning (Specialized AI)

- Vision (image recognition, semantic segmentation, etc)
  - as good as or better than humans in some applications



Facebook:

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

# Machine Learning (Specialized AI)

- Vision (image recognition, semantic segmentation, etc)
  - as good as or better than humans in some applications

computer vision system for self-driving cars

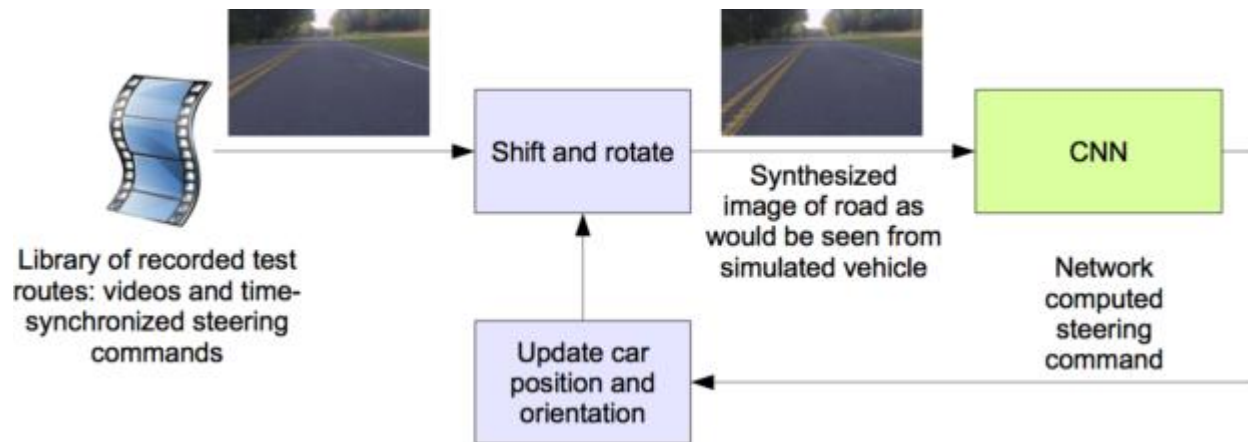


<https://www.nvidia.com/en-au/self-driving-cars/drive-px/>



# Auto-driving in 2016 by Nvidia

- <https://developer.nvidia.com/blog/deep-learning-self-driving-cars/>



<https://www.youtube.com/watch?v=NJU9ULQUwng>

# Tesla Self-Driving Test (Beta 10.12.1, May 2022)

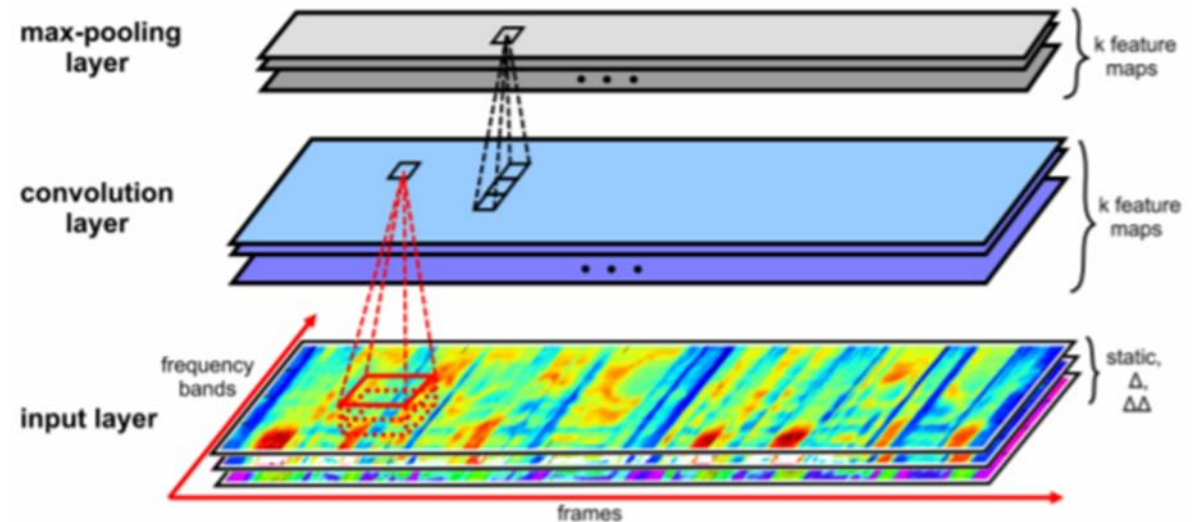
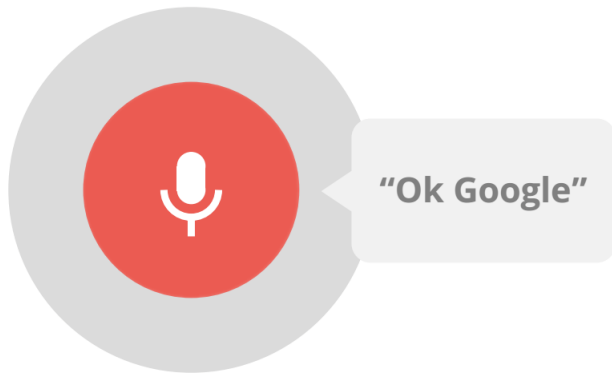
- [https://www.youtube.com/watch?v=\\_ZYEjYnmPIA](https://www.youtube.com/watch?v=_ZYEjYnmPIA)



Level 0	No automation
Level 1	Semi-automated systems, like cruise control.
<b>Level 2</b>	Semi-automated systems, like steering, speed and braking.
Level 3	Automated driving in some conditions, driver available to take over
<b>Level 4</b>	Automated driving in most conditions
Level 5	Automated driving in all conditions

# Machine Learning (Specialized AI)

- Speech (e.g. speech recognition, speaker recognition, etc)



Towards End-to-End Speech Recognition with  
Deep Convolutional Neural Networks

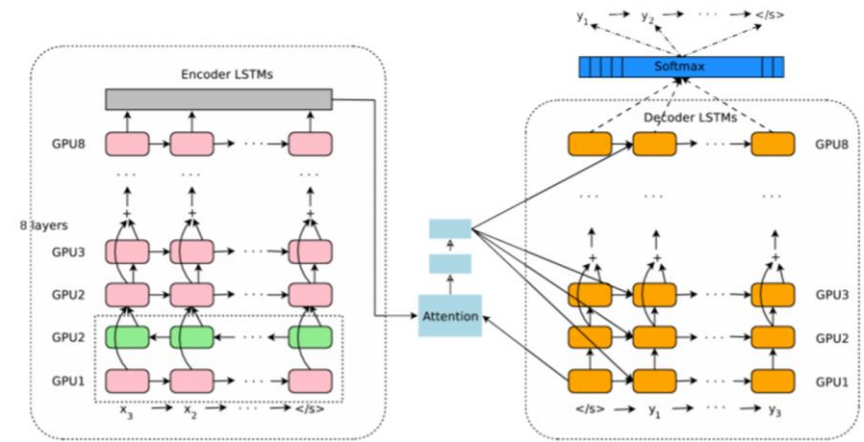
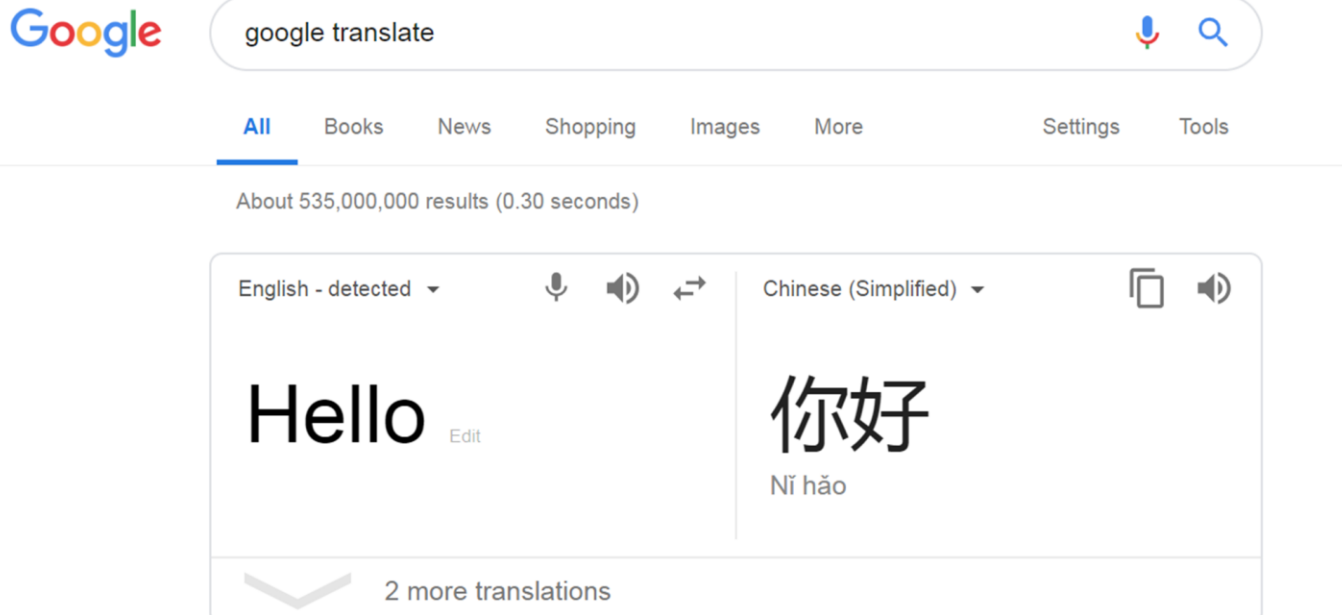
<https://arxiv.org/pdf/1701.02720.pdf>



# Machine Learning (Specialized AI)

- Text (e.g. language translation, chat-bot)

<https://arxiv.org/pdf/1609.08144.pdf>



<https://arxiv.org/pdf/1706.03762.pdf>

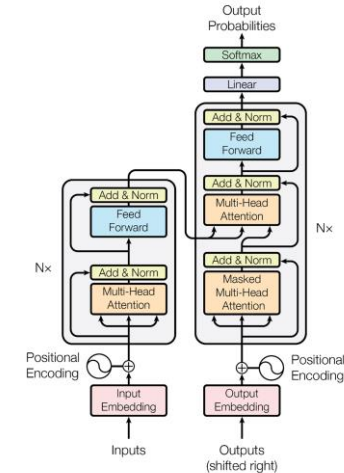


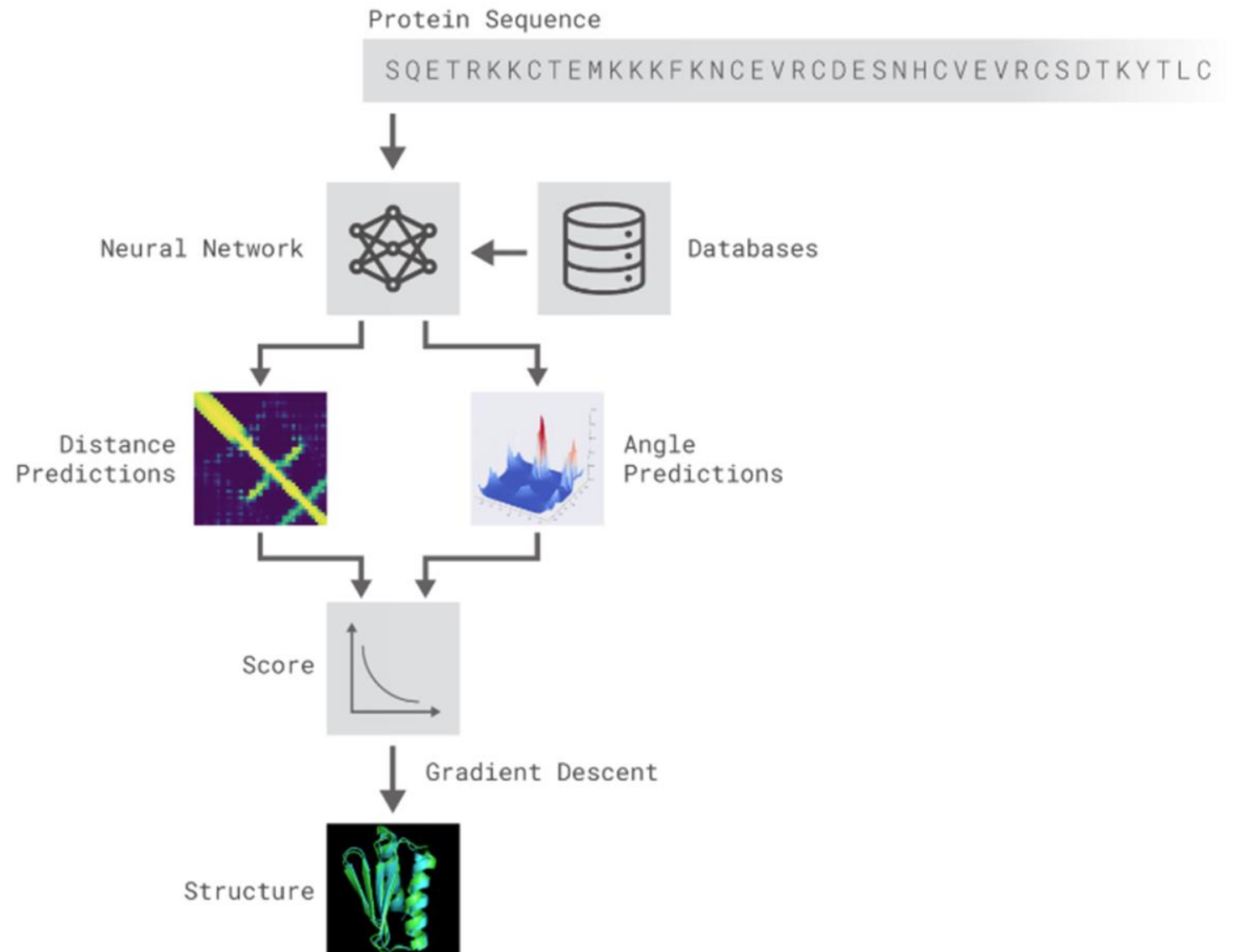
Figure 1: The Transformer - model architecture.

# Machine Learning (Specialized AI)

- Bioinformatics

<https://deepmind.com/blog/alphafold/>

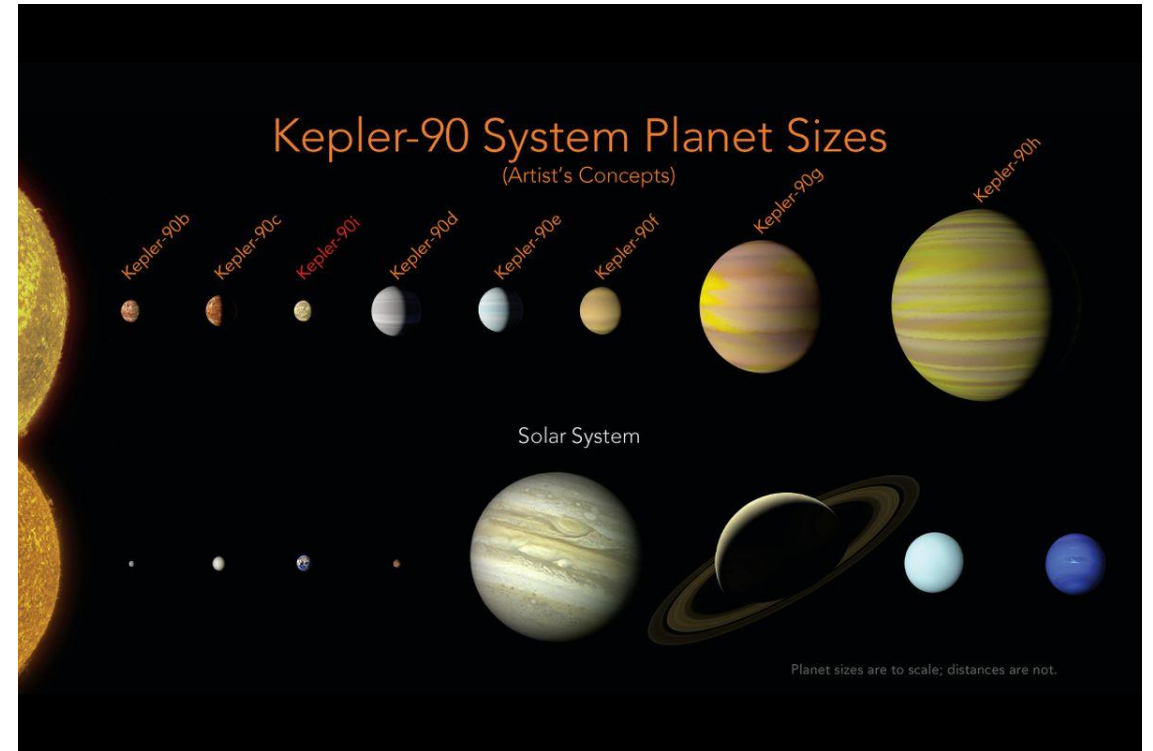
deep neural networks are trained to predict properties of the protein from its genetic sequence.



# Machine Learning (Specialized AI)

- Astronomy

<https://ai.googleblog.com/2018/03/open-sourcing-hunt-for-exoplanets.html>



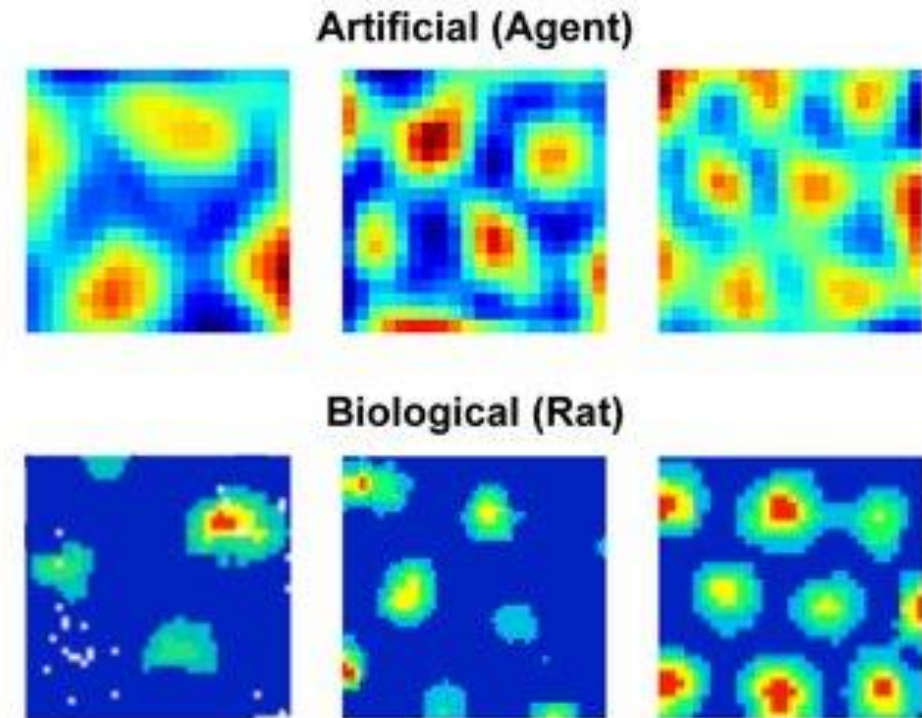
Researchers at Google in 2017 discovered two exoplanets by using ML algorithms to analyze data from NASA's Kepler space telescope and accurately identify the most promising planet signals.

# Machine Learning (Specialized AI)

- Neuroscience

Researchers at Google Deepmind in 2018 developed ML algorithms which behave like grid-cells in animal (and human) brain for navigation.

Use artificial neural networks to explain the real neural networks in brains



Our experiments with artificial agents yielded grid-like representations ("grid units") that were strikingly similar to biological grid cells in foraging mammals.

# Machine Learning (Specialized AI)

- Finance

A company named simility uses ML algorithms to detect different types of fraud activities.

- (1) Account takeover fraud
- (2) Wire Fraud: transfer money..
- (3) Money Laundering (drug dealer...)
- (4) Mobile Check Deposit Fraud  
(scan fake check using smartphone)

The algorithms take into account the following information of the user: keyboard patterns, time and location, transaction amount, frequency of transactions, etc...



PayPal acquired simility in July 2018



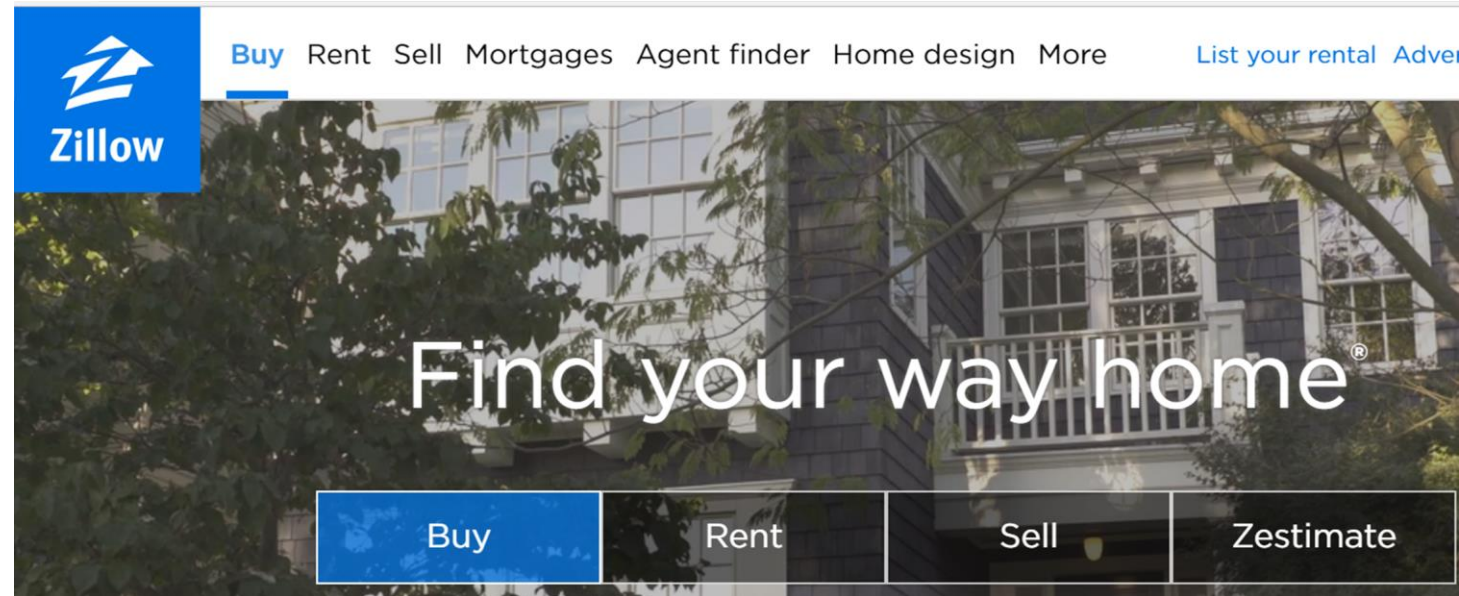
# Machine Learning (Specialized AI)

- Realestate

The company Zillow is trying to use ML-algorithms to predict future sale prices of homes.

It offered **\$1,000,000 USD** to anyone who can develop ML algorithms for price prediction in 2017

<https://www.kaggle.com/c/zillow-prize-1>



Zillow is the leading real estate and rental marketplace (online platform).

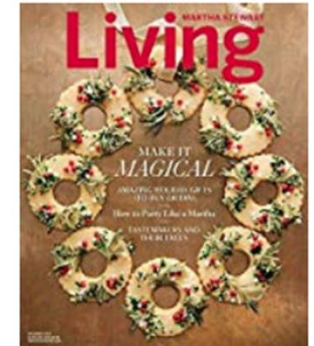
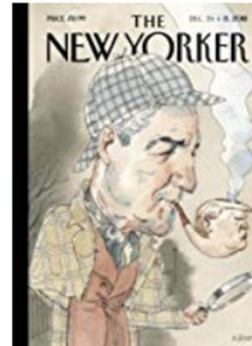
Through Zillow, people can buy, sell, and rent homes.

# Machine Learning (Specialized AI)

- Online Recommendation

Amazon makes product recommendation based on your browsing history

## Recommended for you in Magazine Subscriptions

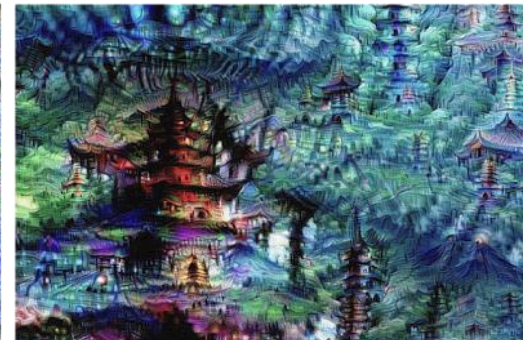
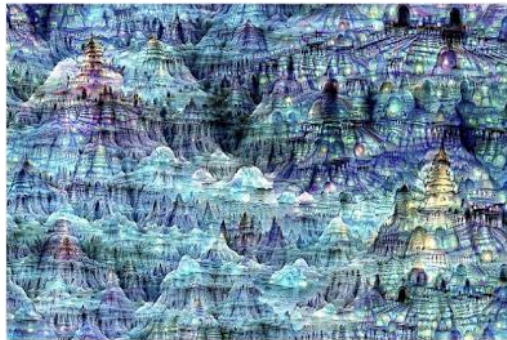
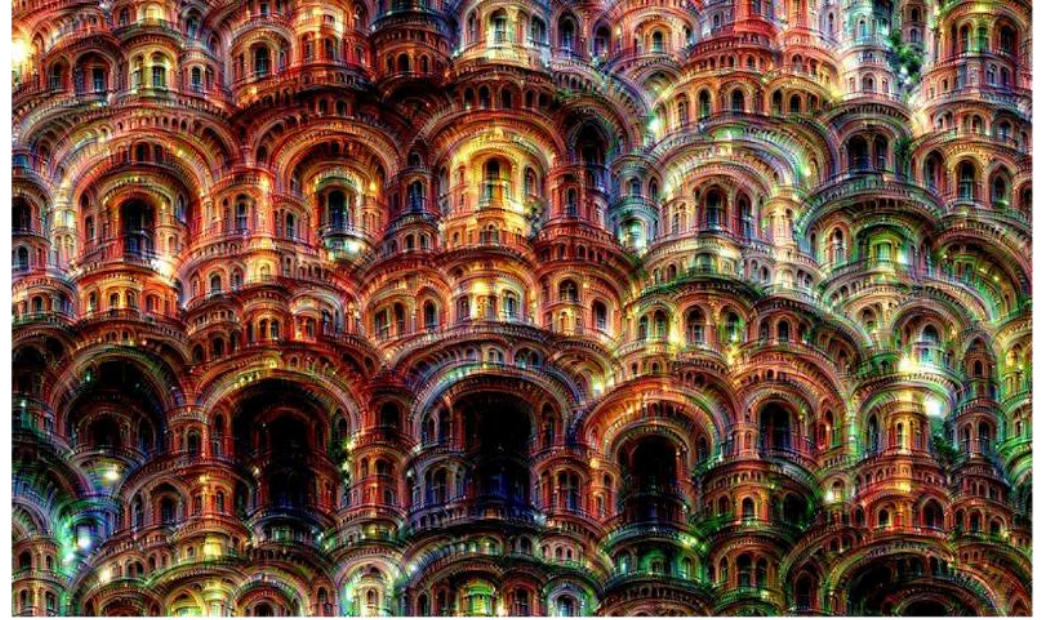




# Machine Learning (Specialized AI)

- Just for fun

deep dream (google)







Style transfer



# Machine Learning (Specialized AI)

- ML Artist (e.g. Stable Diffusion)

<https://replicate.com/stability-ai/stable-diffusion/examples>

<https://huggingface.co/spaces/stabilityai/stable-diffusion>

**a flying pig over university of miami**



<https://stablediffusionlitigation.com/>

We've filed a lawsuit challenging Stable Diffusion, a 21st-century collage tool that violates the rights of artists.

**Because AI needs to be fair & ethical for everyone.**

JANUARY 13, 2023

Hello. This is **Matthew Butterick**. I'm a writer, designer, programmer, and lawyer. In November 2022, I teamed up with the **amazingly excellent**

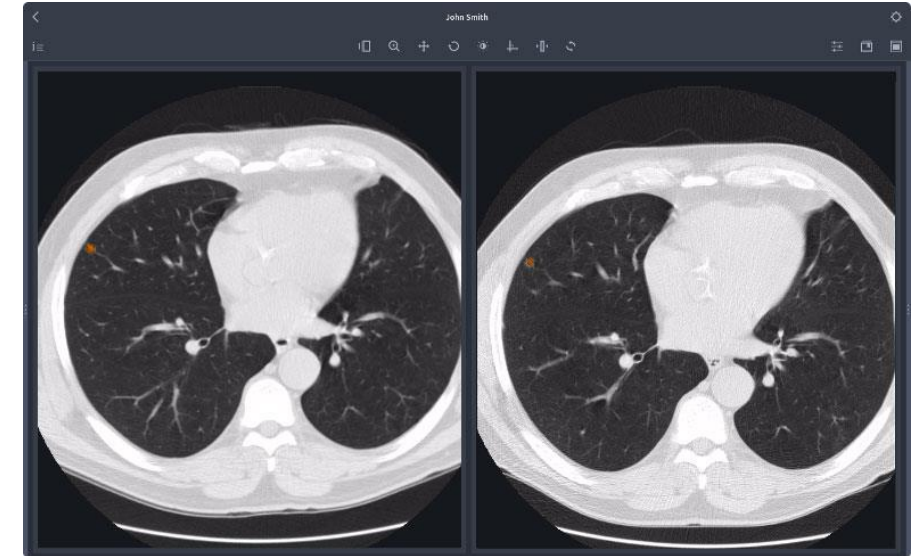




# Machine Learning (Specialized AI)

<https://www.arterys.com/lung>

- Medical Imaging and Image Analysis



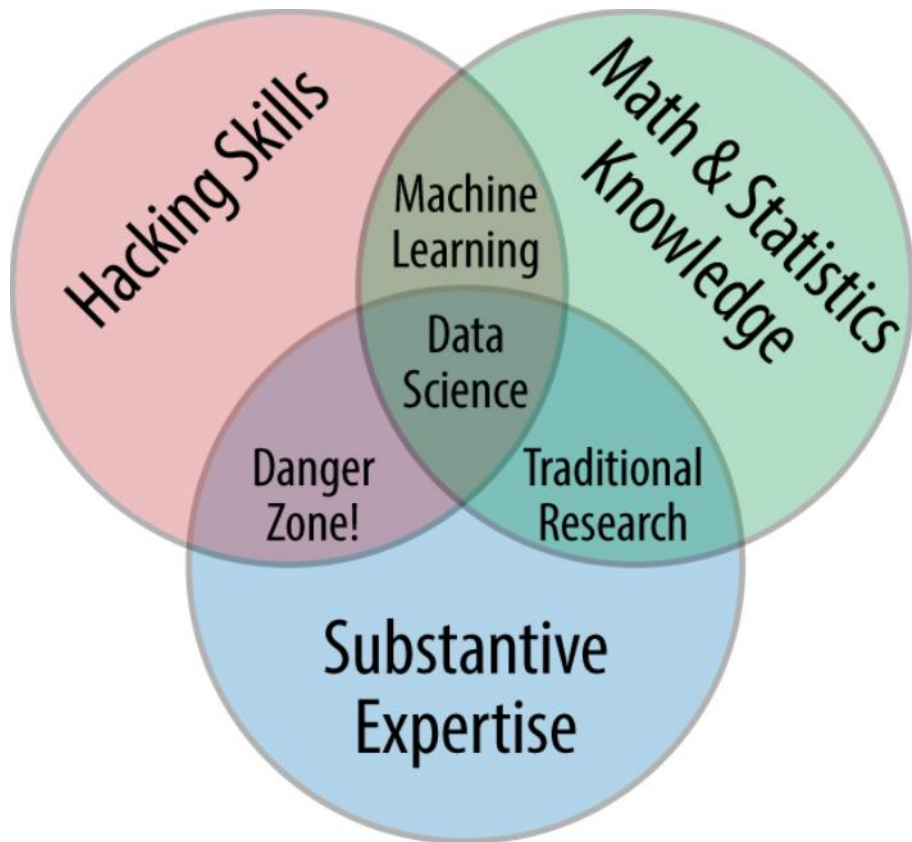
Some patients may have lung nodules.

A lung nodule is a type of lesion which could develop into cancer.

A company Arterys use ML algorithms to automatically detect lung nodules on CT images, and assess the risks. (FDA cleared)

# What is **Data Science** ?

The 'first' diagram  
to define data science



**Data science** is an interdisciplinary field that combines computer programming (hacking), math, and machine learning to solve problems in a specific domain/field.

A **data scientist** needs to have:

- (1) programming skills (hacking)
- (2) knowledge of math, especially statistics
- (3) knowledge of machine learning
- (4) domain knowledge and expertise  
e.g. biology, physics, psychology, ...

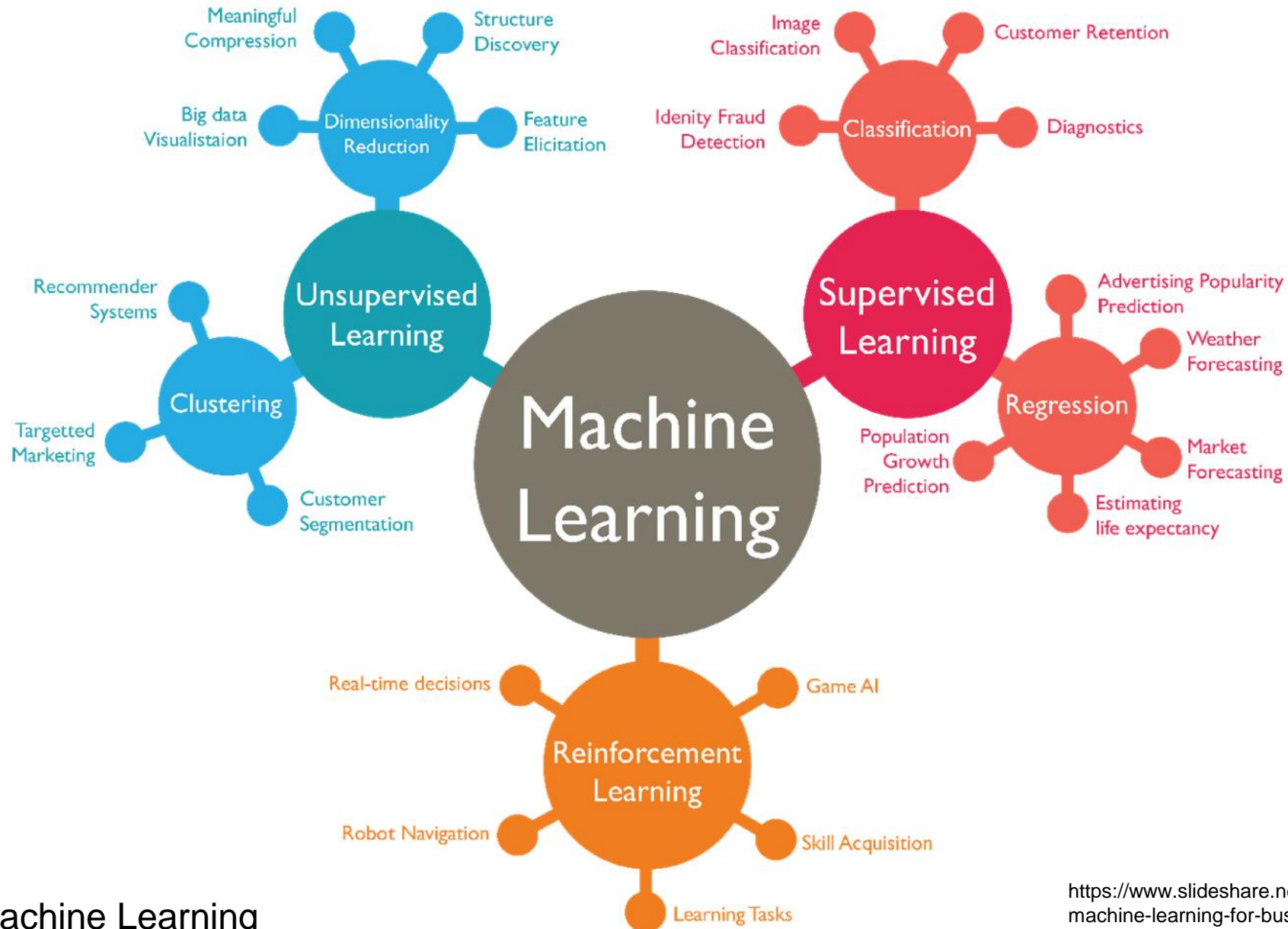
# What is **Machine Learning (ML)** ?

- Machine Learning is a sub-field of Artificial Intelligence.

It has many definitions if you google it ....

- Machine Learning is to extract patterns from data.
  - Machine Learning is to give computers the ability to learn without being explicitly learned.
  - Study of algorithms that improve their performance at some task with experience
  - Machine Learning is the study of (computer) algorithms that can learn something from data and apply the learned knowledge to perform some tasks.
- **ML** algorithms can keep improving their performance by using more data. - More Data, Better Performance.

# What is Machine Learning ?



Three types of machine Learning

<https://www.slideshare.net/awahid/big-data-and-machine-learning-for-businesses>

# Machine Learning Application

Machine Learning  
Models and Algorithms

Domain Knowledge and Data:  
Goal of Machine Learning



The Basic Models and Algorithms in Machine Learning are like **lego bricks**



Use the lego bricks to build different objects/models for different applications



# Machine Learning (ML) needs mathematics

Basics (if you want to learn ML and make applications)

- Calculus
- Linear Algebra
- Probability and Statistics

Advanced (if you want to be a ML researcher):

- Information Theory
- Numerical Method and Optimization
- Signal Processing (speech and image recognition)
- Stochastic Process (reinforcement learning)
- Control Theory (reinforcement learning)

# Machine Learning (ML) needs Python

- Python is #1 programming language for ML

Three open-source software packages for machine learning



Each package is written by using a mixture of different programming languages: C/C++ and Python.



TensorFlow

Users can use the packages through Python.

PYTORCH

It is much easier to use Python than C/C++

# Machine Learning (ML) needs Python

- Basic Python Packages for data manipulation and visualization



Numpy: store data and manipulate data



Pandas to process tabular data



Matplotlib to visualize data

# Course Syllabus

- Classical Machine Learning (50%)
- Deep (Machine) Learning (50%)
- **There are 5 assignments. Each assignment may have two parts:**
  - Math:** derive some equations
  - Programming:** use Python to complete a machine learning application

Some applications are chosen from **<https://www.kaggle.com/>**

- no exam
- The last homework assignment is the final project.



# Textbooks

(Not a single book covers everything in machine learning)

- Hastie, Tibshirani, and Friedman's The Elements of Statistical Learning  
<https://web.stanford.edu/~hastie/ElemStatLearn/>
- Machine Learning: a Probabilistic Perspective  
<https://www.cs.ubc.ca/~murphyk/MLbook/>
- Pattern Recognition and Machine Learning, Chris Bishop  
<https://www.microsoft.com/en-us/research/people/cmbishop/#!prml-book>
- Ian Goodfellow and Yoshua Bengio and Aaron Courville: Deep Learning  
<https://www.deeplearningbook.org/>

# Lecture Notes vs. Textbooks

- My lecture notes are not the replacement of textbooks
- Pick one of the textbooks and read it if you want to do some (applied or theoretical) research in machine learning

# Python (v3.x)

- Python Basics: [https://www.python-course.eu/python3\\_course.php](https://www.python-course.eu/python3_course.php)
- Python Packages: Python Data Science Handbook  
<https://jakevdp.github.io/PythonDataScienceHandbook/>
- The lectures will focus on the methods and algorithms.
- My lecture files (zip files) have example code that you can use for your homework.

# “Do I need take a Python course ?”

- If you have no programming skills in Python and its Packages, take a Python course **before** you take this ML course
- Python course: CSC315 for under-grad, DSC615 for grad
- If you are in the Data Science Track of Computer Science BS: CSC315 is a required course

# “Do I need to know the Math ?”

- Handwritten digit recognition



run the demo in MLP\_Keras.ipynb on Google Colab

## Define the model

```
1 model = Sequential()
2 model.add(Dense(units=256, activation='relu', input_shape=(784,)))
3 model.add(Dense(units=256, activation='relu'))
4 model.add(Dense(units=10, activation='softmax'))
5 model.compile(loss='categorical_crossentropy', optimizer=SGD(lr=0.01), metrics=['accuracy'])
6 model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
dense (Dense)	(None, 256)	200960
dense_1 (Dense)	(None, 256)	65792
dense_2 (Dense)	(None, 10)	2570
=====		
Total params: 269,322		
Trainable params: 269,322		
Non-trainable params: 0		

Question: Do you want to know the algorithms in the model ?  
or you just want to use it as a magic box ?



### 3.2.4.3.1. `sklearn.ensemble.RandomForestClassifier`

```
class sklearn.ensemble.RandomForestClassifier(n_estimators=100, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None, ccp_alpha=0.0, max_samples=None)
```

[\[source\]](#)

To understand the meaning of each parameter, you need to understand the algorithm

## ConvTranspose2d

```
CLASS torch.nn.ConvTranspose2d(in_channels, out_channels, kernel_size, stride=1,  
    padding=0, output_padding=0, groups=1, bias=True, dilation=1,  
    padding_mode='zeros')
```

[\[SOURCE\]](#)

To understand the meaning of each parameter,  
you need to understand the algorithm

# Objective: Introduction to Machine Learning

- The objective of this course is to give an introduction to machine learning methods and algorithms (lectures), and then, the students use ML to make some applications (homework assignments).

methods and algorithms = Math

If you want to understand ML, you need to use Math

- To further enhance your skills:
  - Try kaggle competitions: <https://www.kaggle.com/competitions>
  - Read research papers
  - Do a research project