Data Analytics in Cyber Security
(CT115-3-M)(Version E)

# Machine Learning Methods

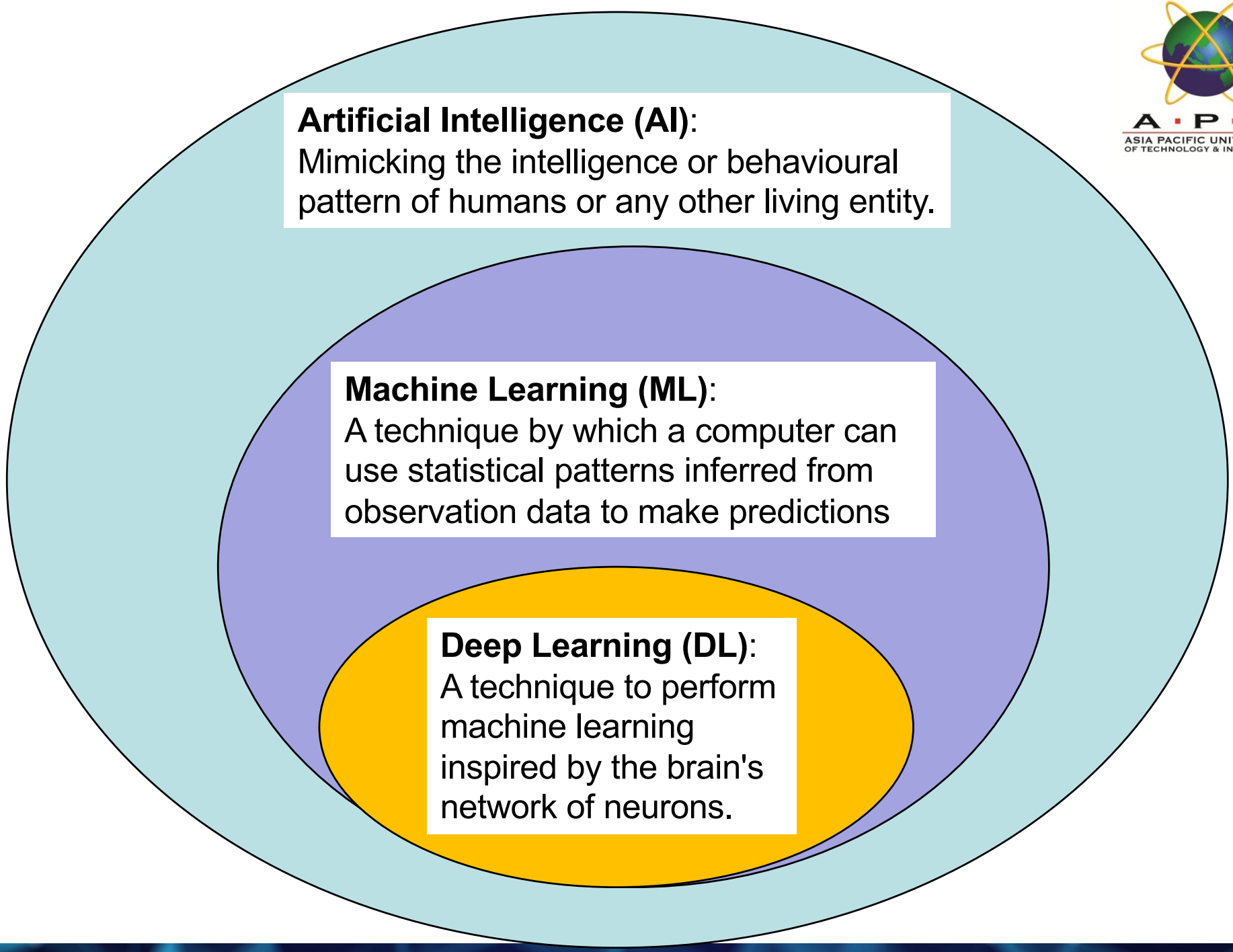
# TOPIC LEARNING OUTCOMES

At the end of this topic, you should be able to:

1. Understand the differences between Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL).
2. Understand types of Machine Learning.
3. Understand characteristics of Deep Learning.

# CONTENTS & STRUCTURE

- Artificial Intelligence
- Types of Machine Learning
- Deep Learning


**Artificial Intelligence (AI)**:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.
**Machine Learning (ML)**:
A technique by which a computer can use statistical patterns inferred from observation data to make predictions
**Deep Learning (DL)**:
A technique to perform machine learning inspired by the brain's network of neurons.

## Artificial Narrow Intelligence (ANI)

Stage-1

### Machine Learning

- Specialises in one area and solves one problem


Siri


Alexa


Cortana

## Artificial General Intelligence (AGI)

Stage-2

### Machine Intelligence

- Refers to a computer that is as smart as a human across the board

## Artificial Super Intelligence (ASI)

Stage-3

### Machine Consciousness

- An intellect that is much smarter than the best human brains in practically every field

## Artificial Intelligence

AI involves techniques that equip computers to emulate human behavior, enabling them to learn, make decisions, recognize patterns, and solve complex problems in a manner akin to human intelligence.

## Machine Learning

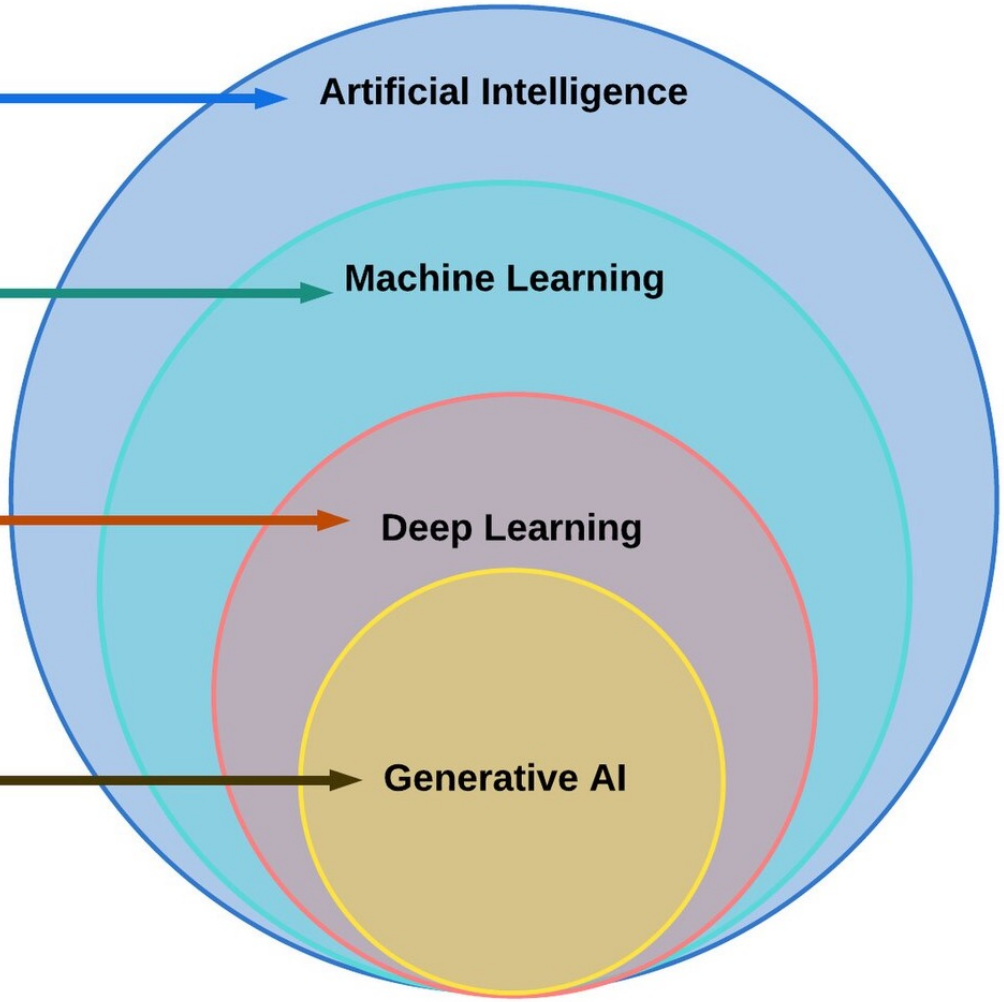
ML is a subset of AI, uses advanced algorithms to detect patterns in large data sets, allowing machines to learn and adapt. ML algorithms use supervised or unsupervised learning methods.

## Deep Learning

DL is a subset of ML which uses neural networks for in-depth data processing and analytical tasks. DL leverages multiple layers of artificial neural networks to extract high-level features from raw input data, simulating the way human brains perceive and understand the world.

## Generative AI

Generative AI is a subset of DL models that generates content like text, images, or code based on provided input. Trained on vast data sets, these models detect patterns and create outputs without explicit instruction, using a mix of supervised and unsupervised learning.




A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

**Unraveling AI Complexity - A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI.**

**(Created by Dr. Lily Popova Zhuhadar, 07, 29, 2023)**

# Artificial Intelligence (AI)

- Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.

- A science of making things smart or, in other words, human tasks performed by machines (e.g., visual recognition, understanding speech, etc.).

# Machine Learning (ML)

- An approach (one of many) to AI that uses a system that is capable of learning from experience (typically represented by data).
- It is intended not only for AI goals (e.g., copying human behavior) but it can also reduce the efforts and/or time spent for both simple and difficult tasks.
- ML is a system that can recognize patterns by using examples rather than by programming them.
- The system “learns” patterns in the data using mathematical and statistical analysis.

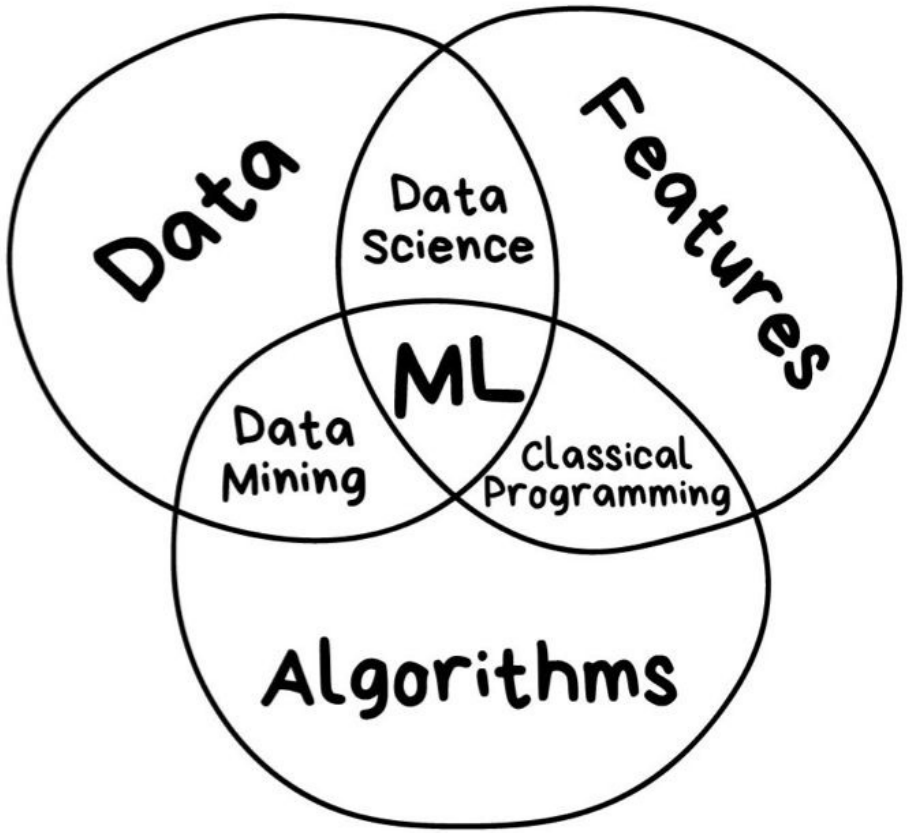
# Deep Learning (DL)

- A set of techniques for implementing machine learning that recognise **patterns of patterns**.
- For example, image recognition systems identify primary object edges, a structure, an object type, and then an object itself.
- Deep Learning is using Deep Neural Networks (DNNs), which come in many forms. DNNs are an architecture, not an algorithm.

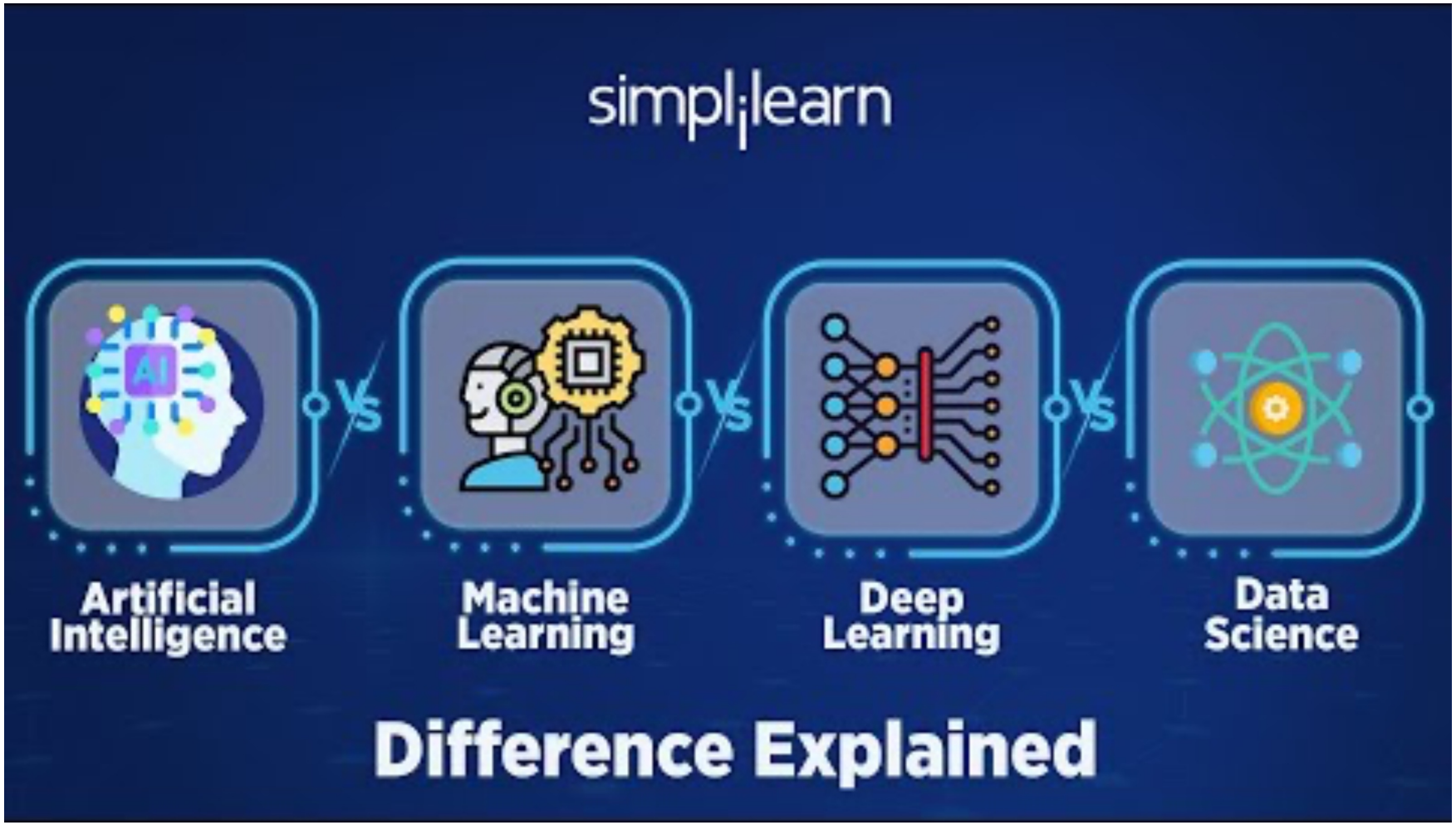
# ML Algorithms

The main mathematical tools used by all ML algorithms are a blend of:

- Linear algebra
- Analytic geometry
- Matrix decompositions
- Vector calculations
- Optimisation
- Probability/Statistics



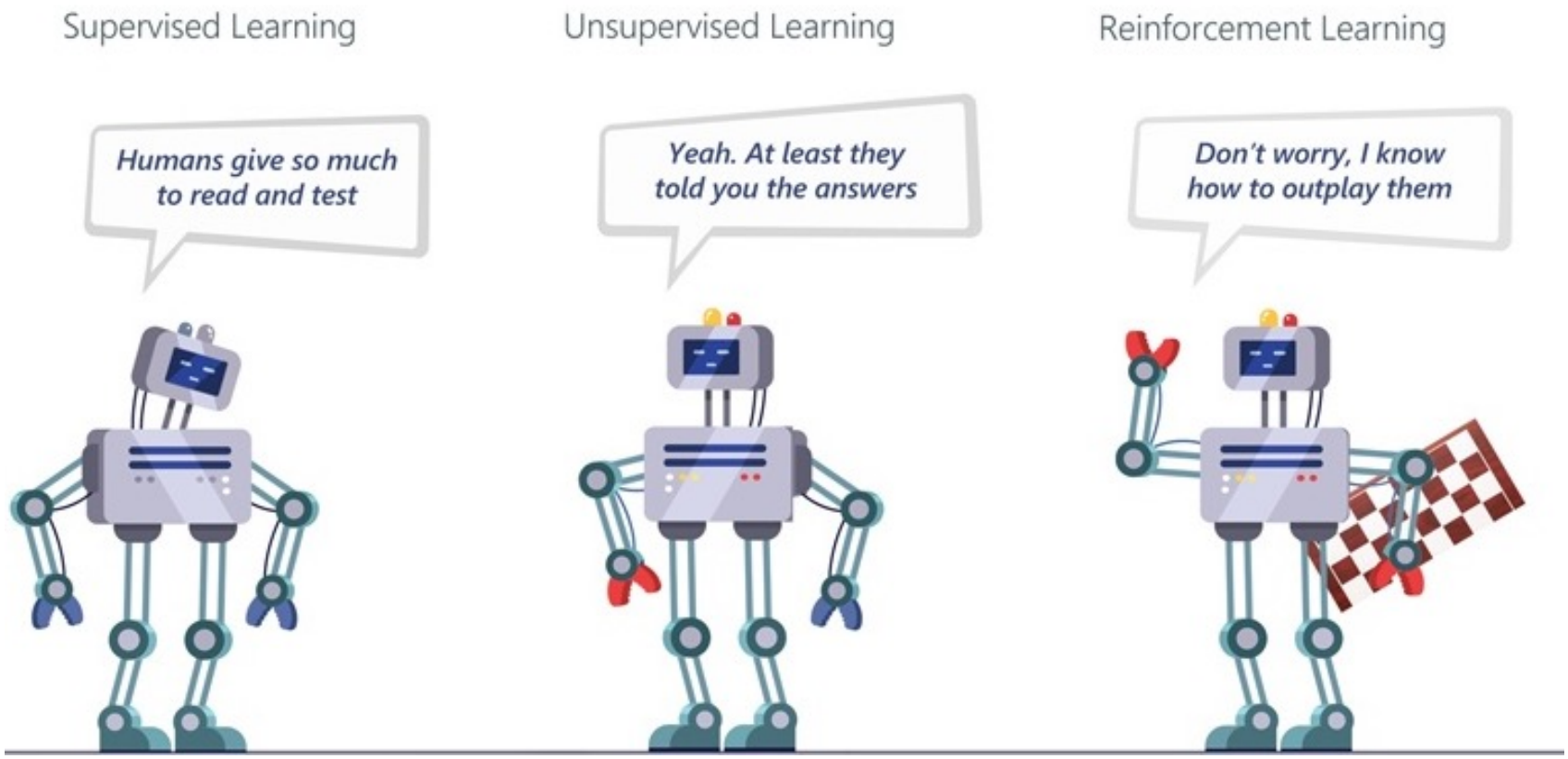
# Difference between AI, ML and DL




https://www.youtube.com/watch?v=vNc2z2u_nh0

# Types of Machine Learning

- Machine learning is the ability of a machine to improve its performance based on previous results.



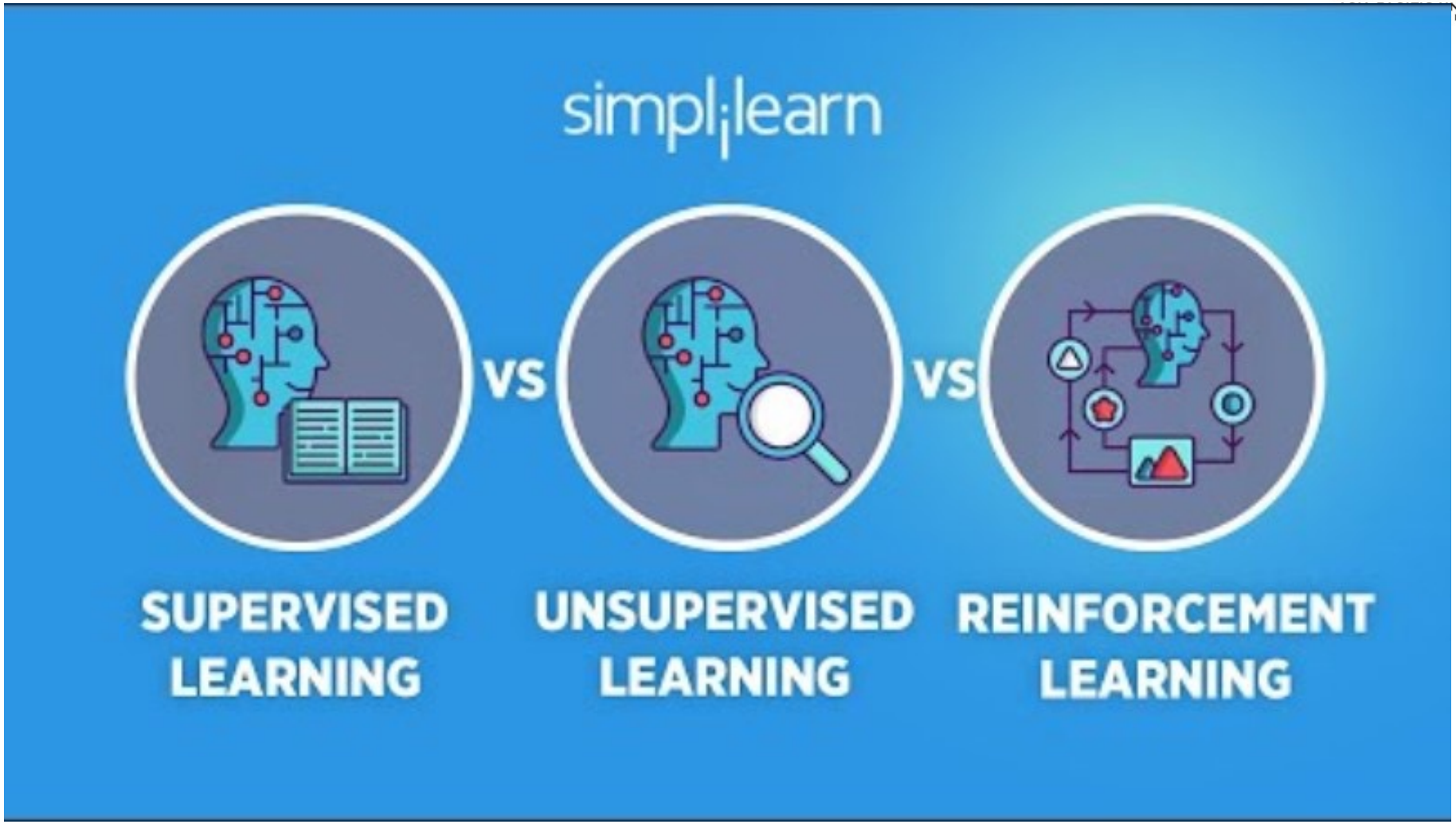
# Types of Machine Learning


- **Supervised learning:** The computer is presented with example inputs and their desired outputs, i.e., labelled data, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
  - e.g., learn to recognise cars by images of cars and non-cars.

- **Unsupervised learning:** no labels are given to the learning algorithm (i.e., unlabelled data), leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end.
  - e.g., group toys of a given colour or shape from a Lego set.

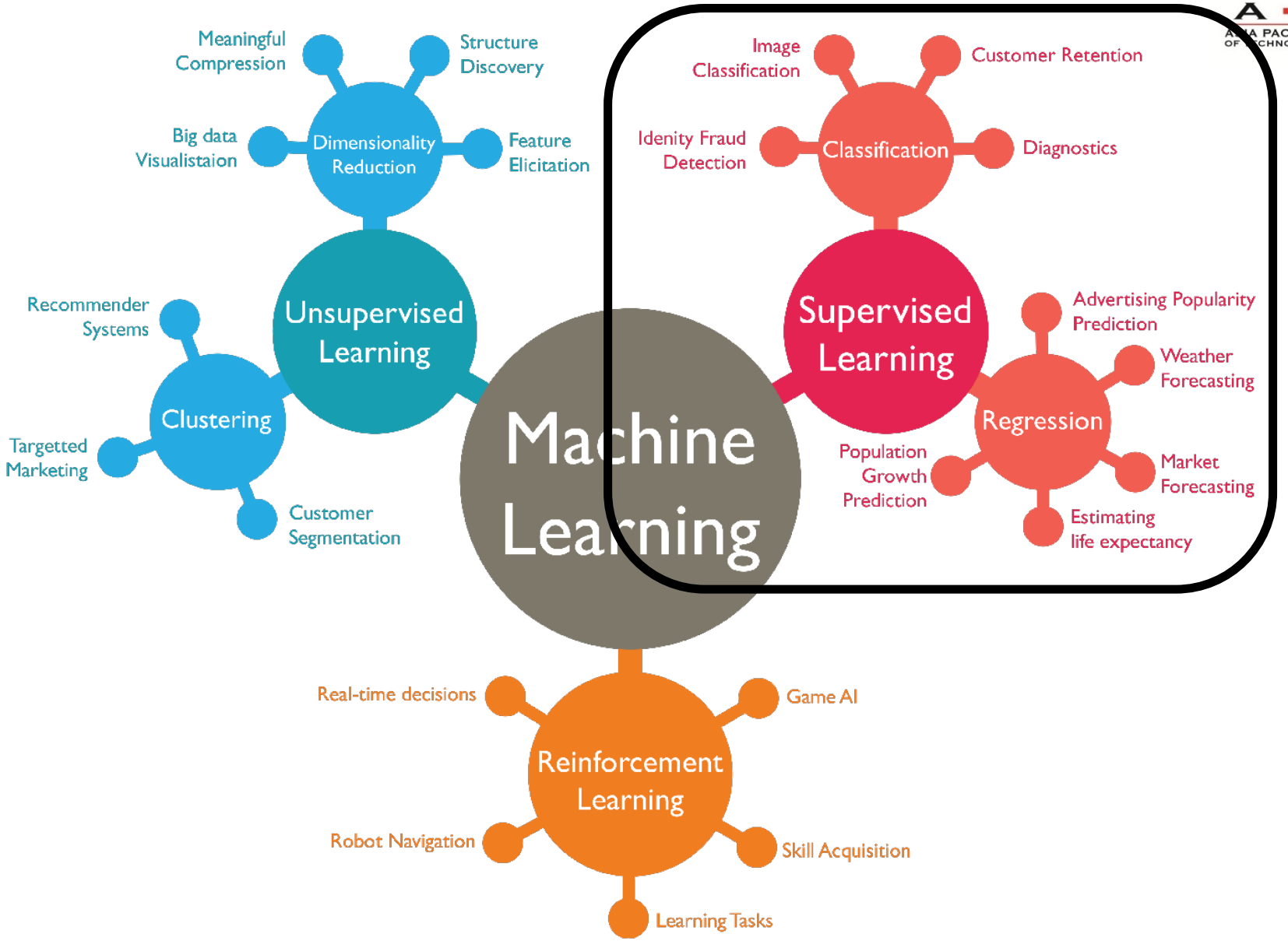
- **Reinforcement learning:** a computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle), without a teacher explicitly telling it whether it has come close to its goal or not, i.e., continuous reweighting of decision model elements after series of decisions.
  - e.g., learn to play tennis / table tennis.

# Types of Machine Learning

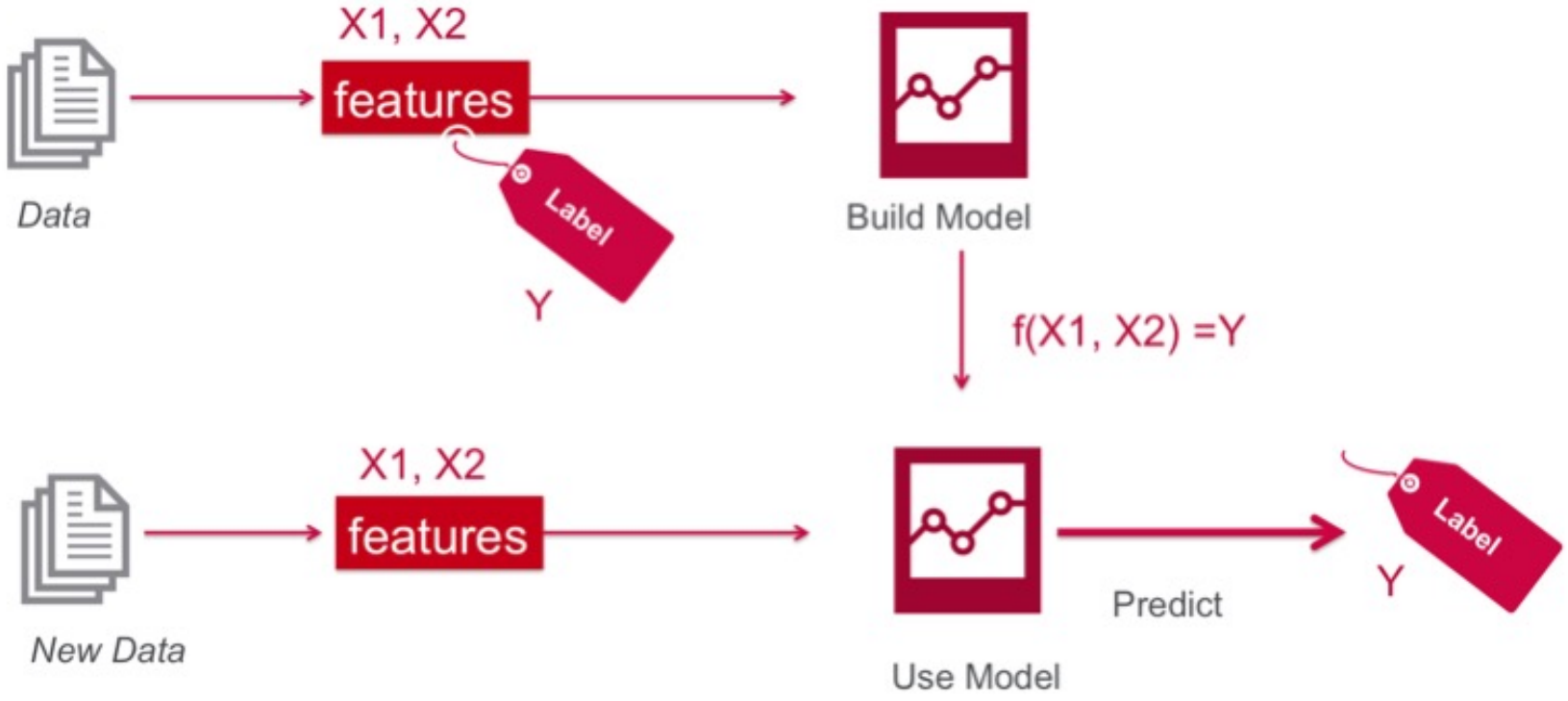



https://www.youtube.com/watch?v=1FZ0A1QCMWc&t=1s

# Machine Learning



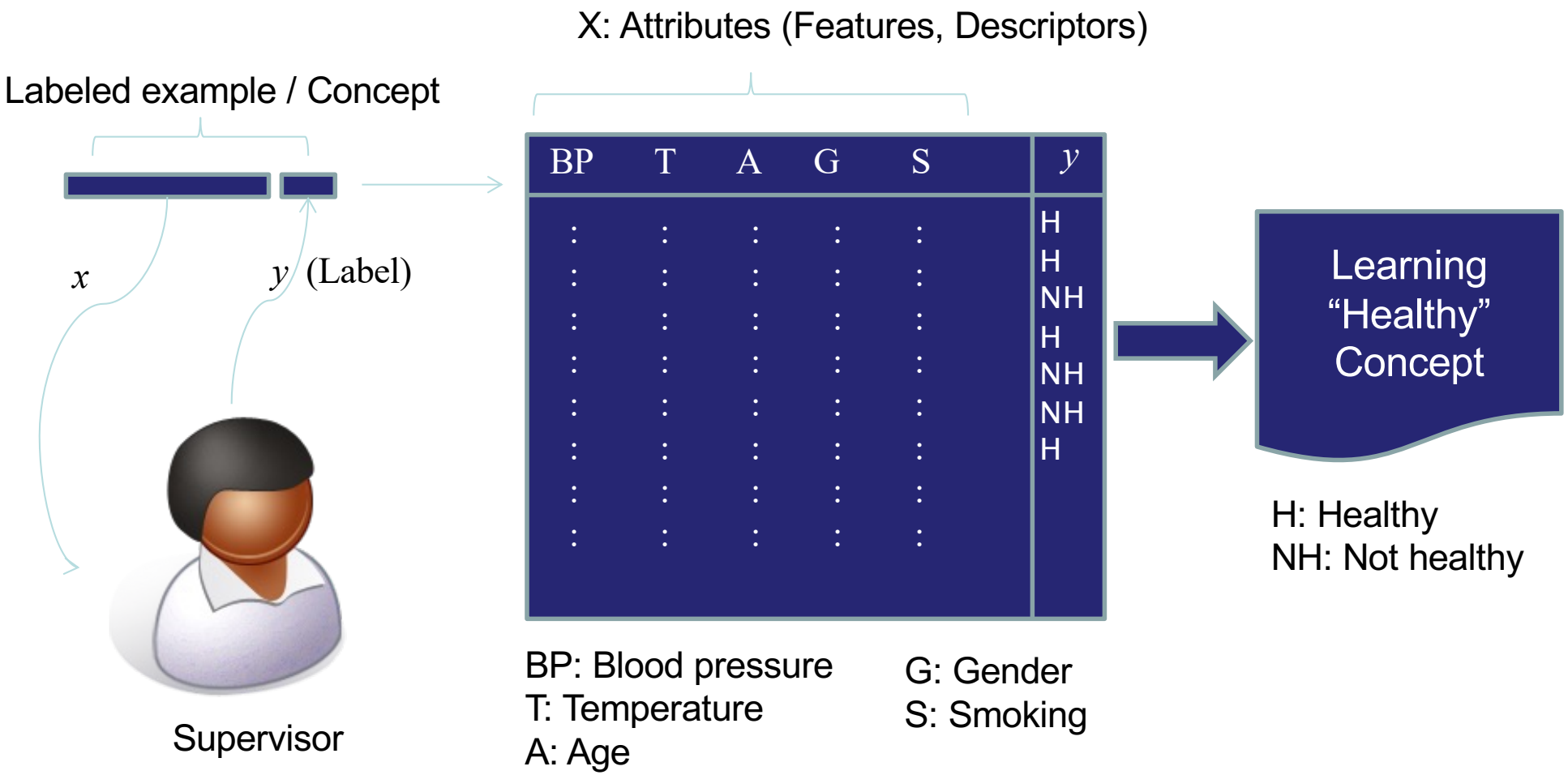
# Supervised Learning




That's nice – but where do the labels come from?

# Supervised Learning

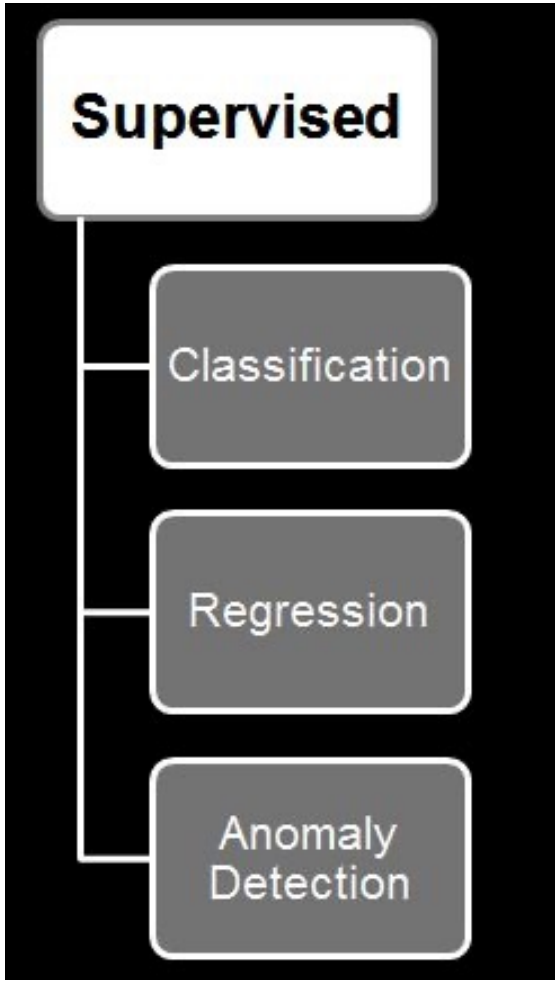
**The labels come from *human intelligence***



# Supervised Learning – Types of Problems

**Supervised Methods** (also called **Predictive**): Predict an unknown value(s) of a variable(s) from the values of some attributes

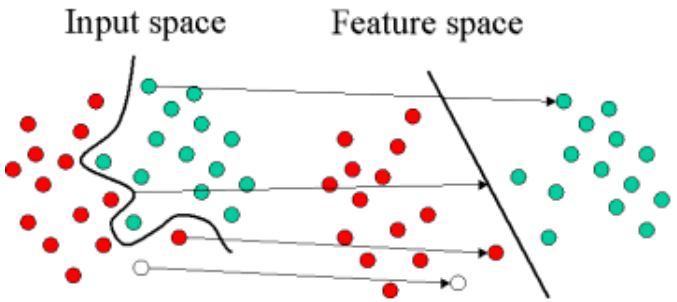
- **Classification:** predict the type/class of new cases
  - ✓ Spam Filtering, Handwriting Character Recognition, Patient Diagnosis
- **Regression:** predict a numerical value of new cases
  - ✓ Blood Pressure, Sales Amounts
- **Supervised Anomaly Detection:** identify items, events or observations deviating from expected patterns using data labeled as "normal" and "abnormal" (involves training a classifier)

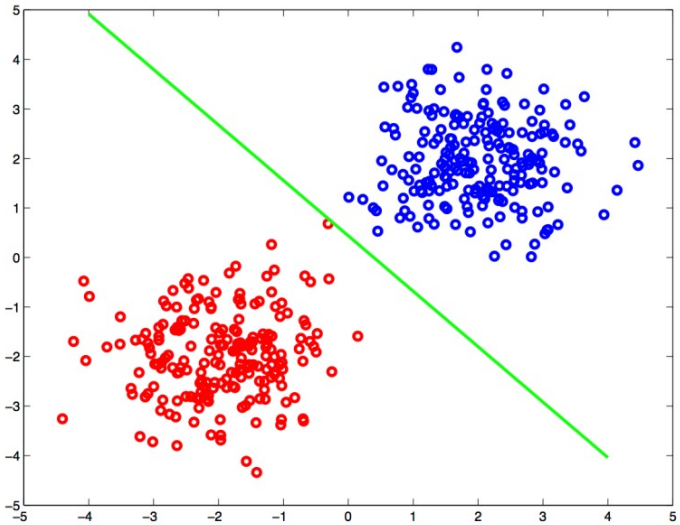


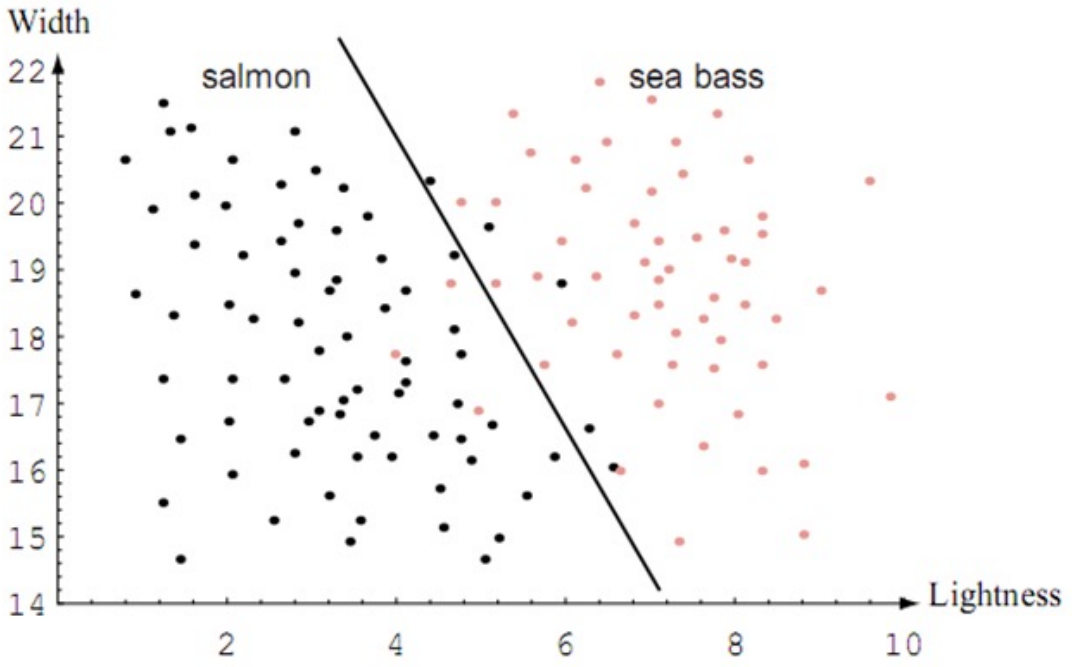

- It is common to combine different methods such as clustering and classification (Hybrid methods)

# Classification

Extracting features from given inputs allows us to separate and classify the inputs according to defined categories





# Example: Recognizing Five Letters

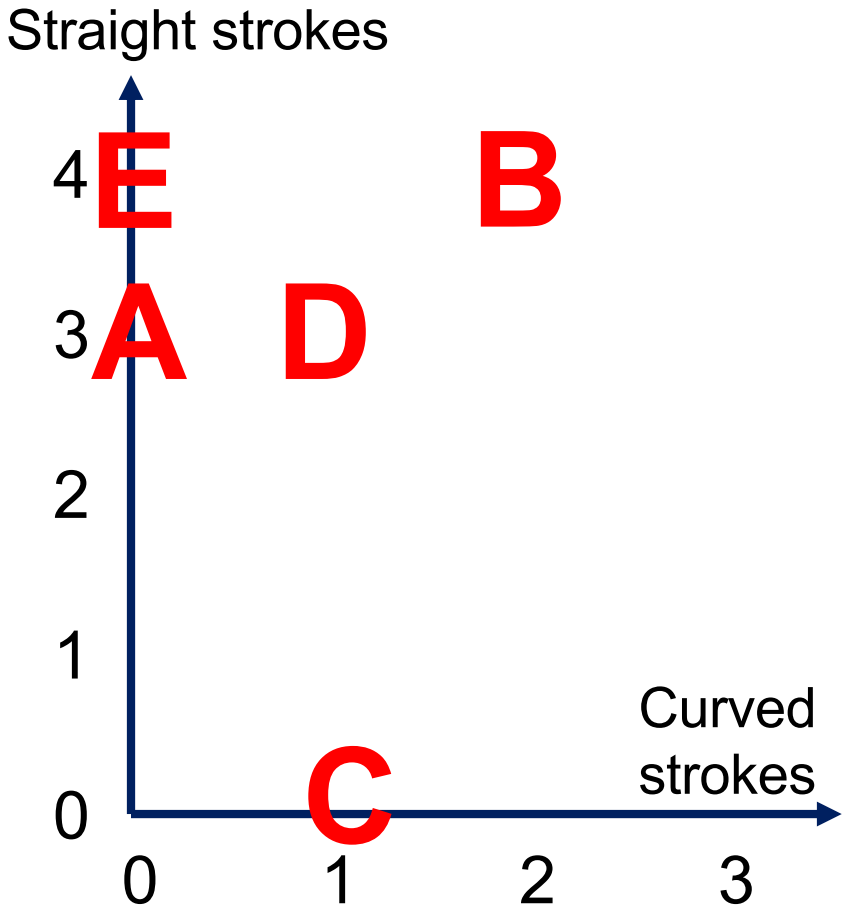
**A, B, C, D, E**

Example features:

*x*: Number of curved segments
*y*: Number of straight segments

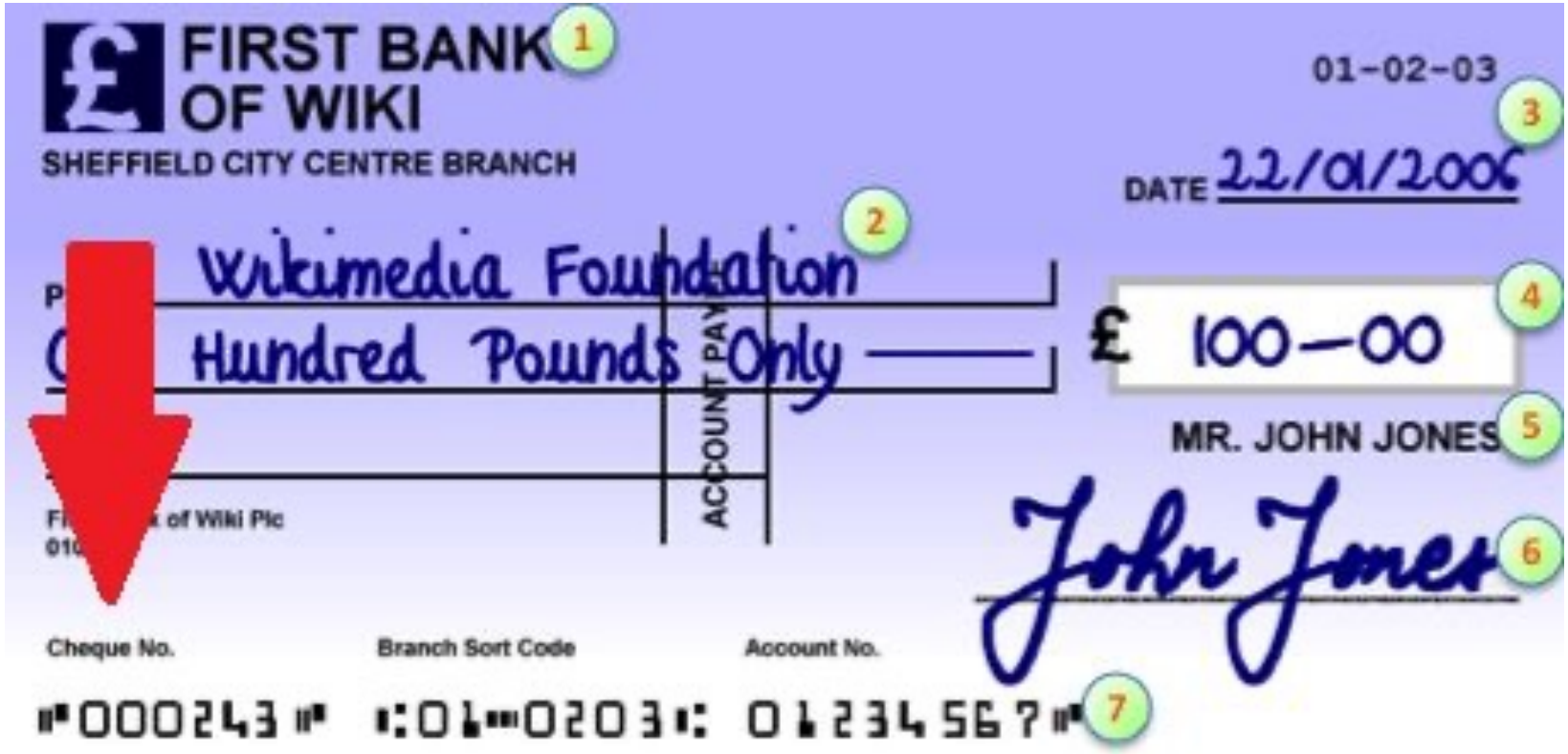
Where would “F” fall?

*Suggest an additional feature*



# Handwriting Recognition

Requires dealing with many variants of characters



# Features: Comparing Fingerprints

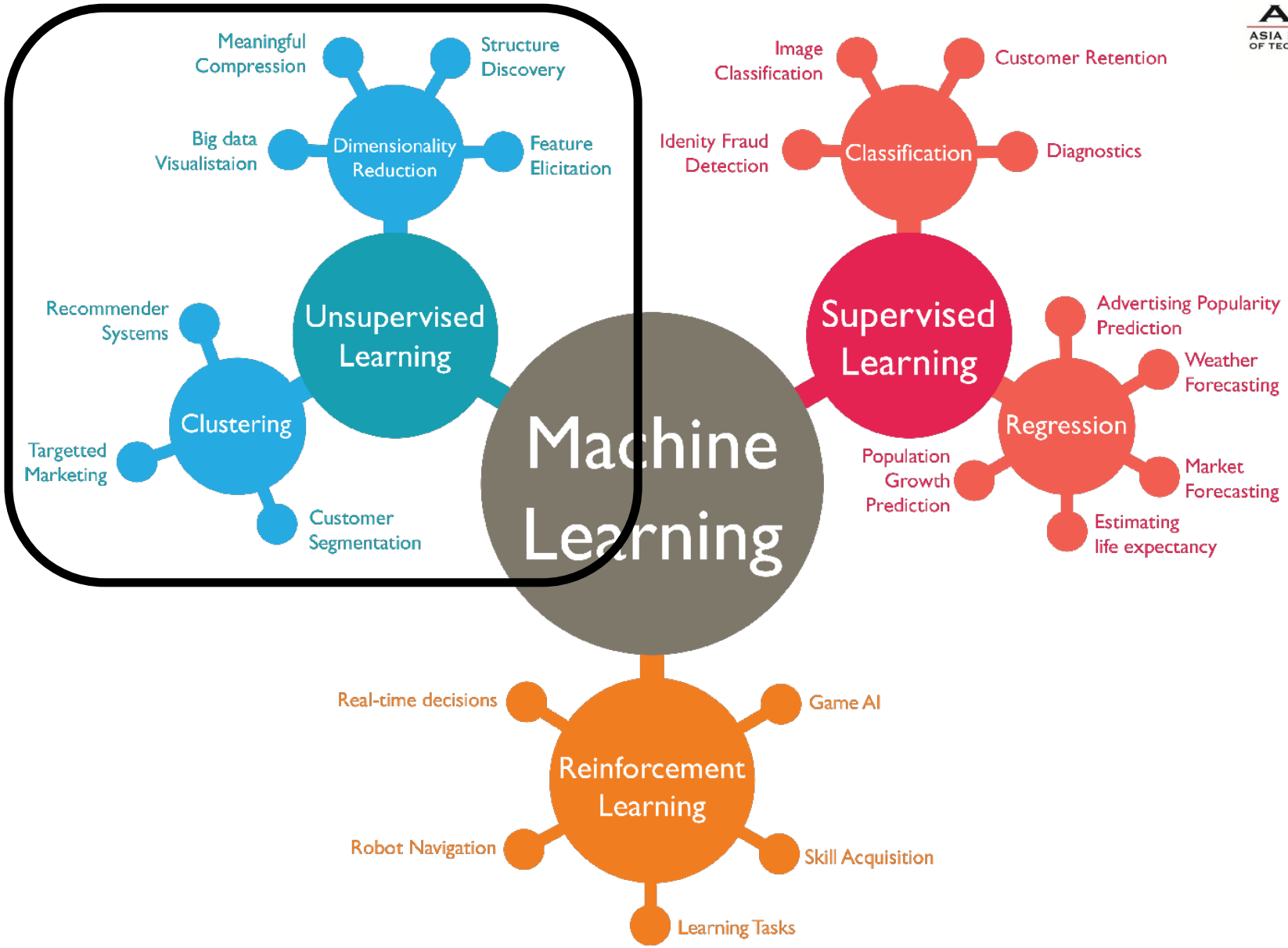
**Criminal investigations and biometric identification**
Does a fingerprint match any of the prints in a criminal database?
Does the fingerprint match one recorded for an authorized user?

Human fingerprints tend to be unique

Even identical twins have different prints

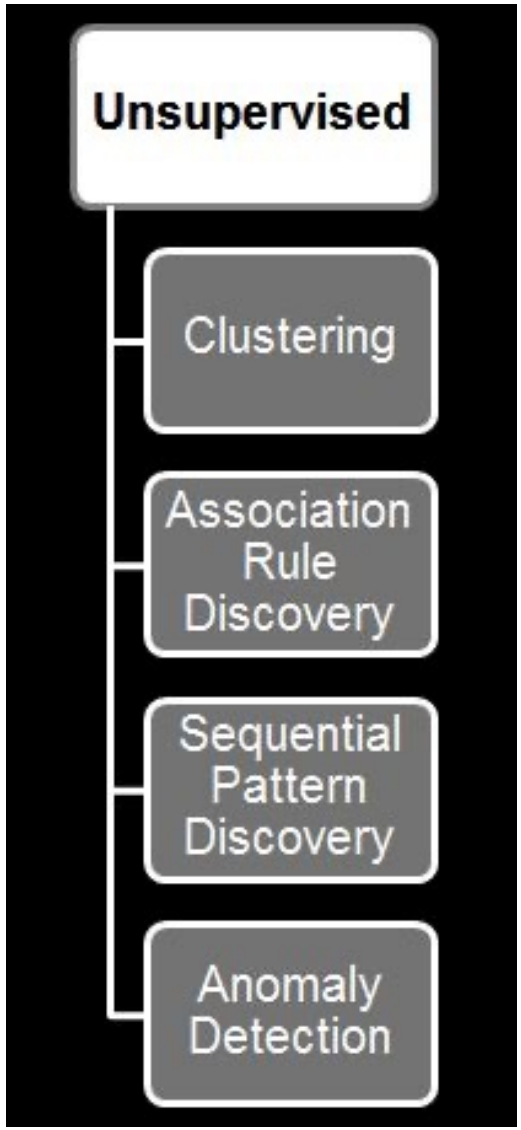

http://www.youtube.com/watch?v=IrpTqKkgygA [6min]

# Machine Learning



# Unsupervised Learning – Types of Problems

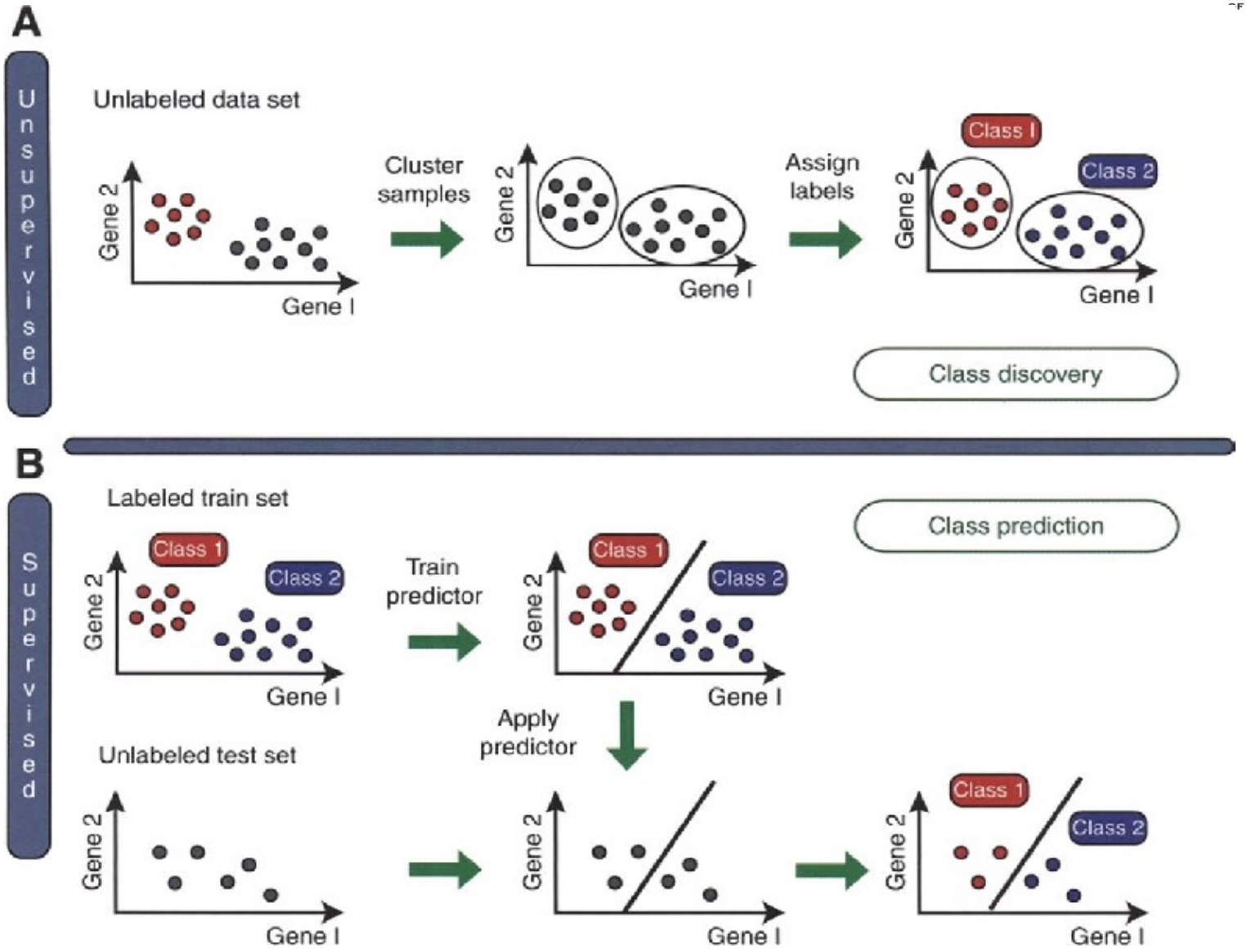
- **Unsupervised Methods** (also called **Descriptive**): Try to find meaningful patterns in the data.
  - **Clustering:** group similar data into clusters
    - ✓ Market Segmentation, Document Clustering
  - **Association Rule Discovery:** find human interpretable patterns (associations)
    - ✓ Product Recommendations, Store Shelf Management
  - **Sequential Pattern Discovery:** describe the sequential dependencies among different events
    - ✓ Buying Patterns, Gene Sequencing
  - **Unsupervised anomaly detection:** to detect anomalies in unlabeled data under the assumption that the majority of the instances are normal
    - ✓ Fraud Detection, Network Intrusion Detection



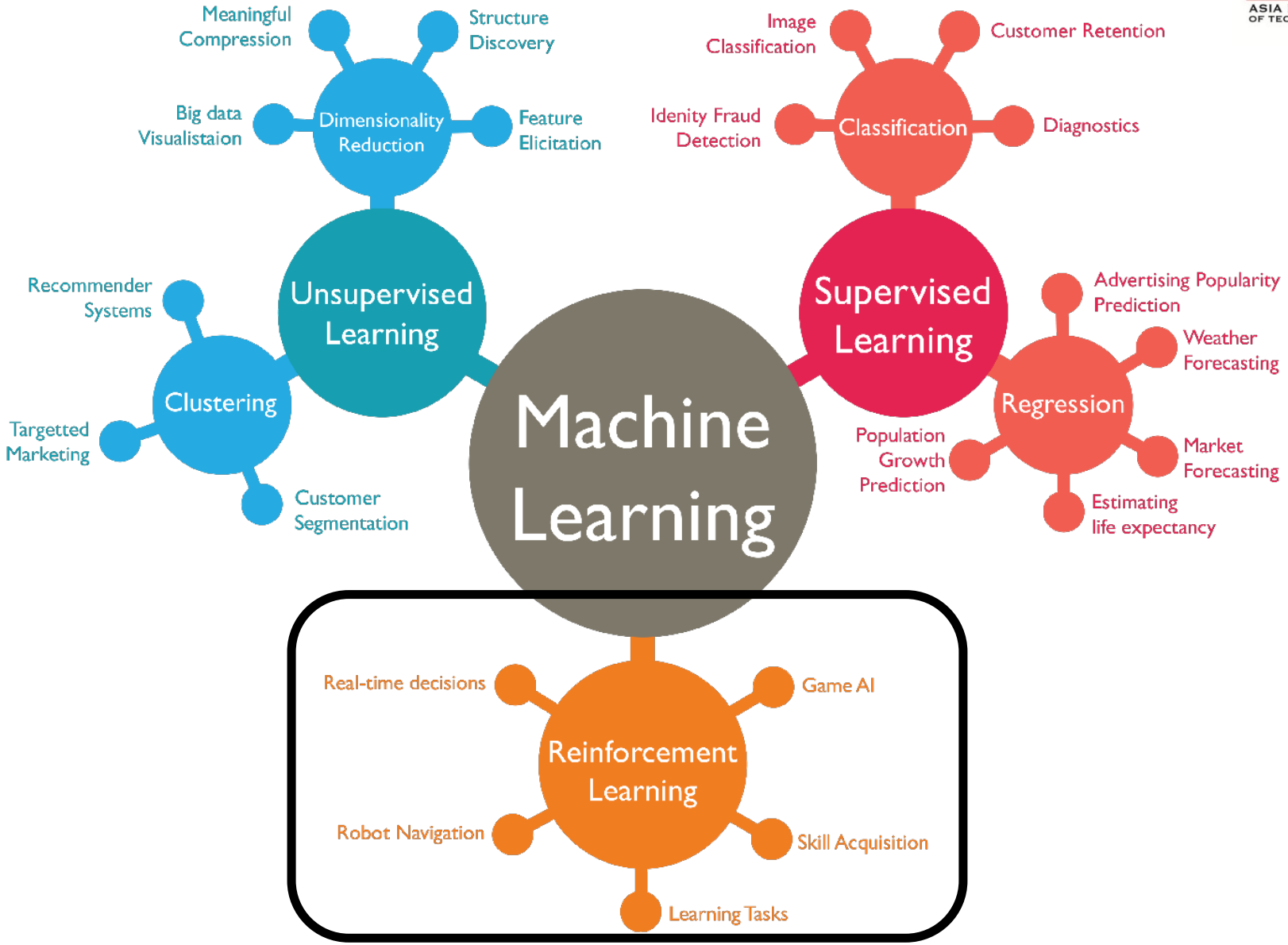
# Unsupervised: Clustering


- Clustering is similar to classification with the only but major difference. The information about the classes of the data is unknown. There is no idea whether this data can be classified.
- Usually clustering is not applied to solving a particular task in cybersecurity as it is more like one of the subtasks in a pipeline (e.g., grouping users into separate groups to adjust risk values).

# Unsupervised vs. Supervised



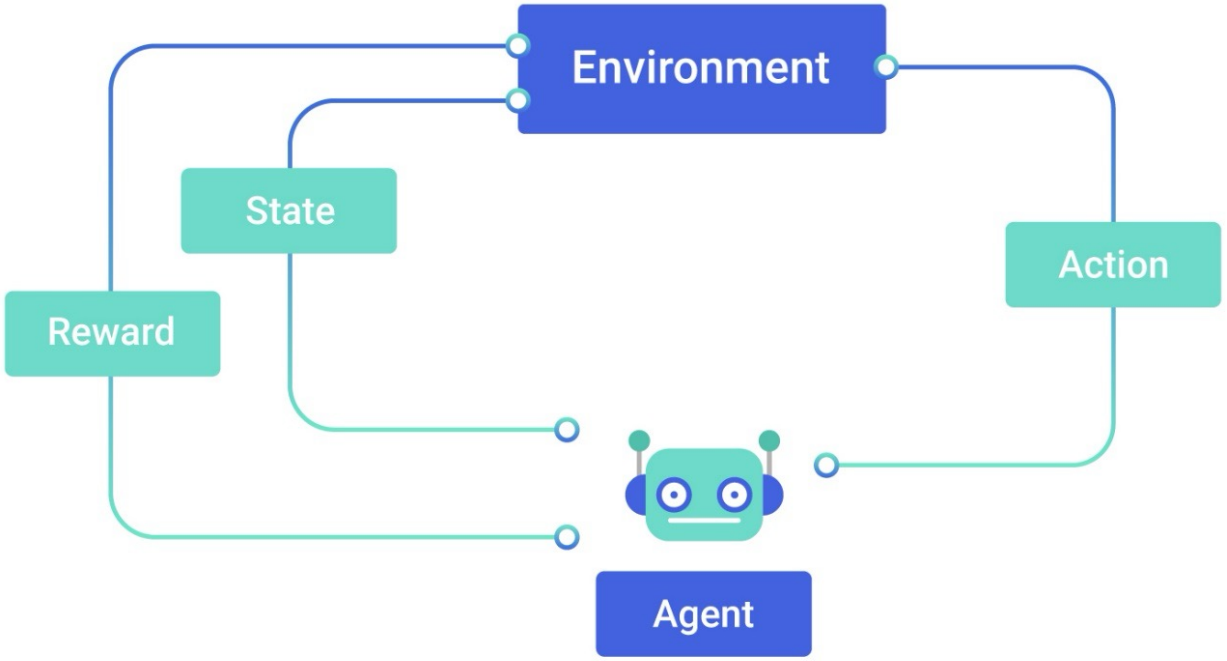
# Machine Learning



# Reinforcement learning – Types of Problems

- **Reinforcement learning** is a machine-learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

**Action (for example)**
Robot movement
Characters in games
Autonomous car driving

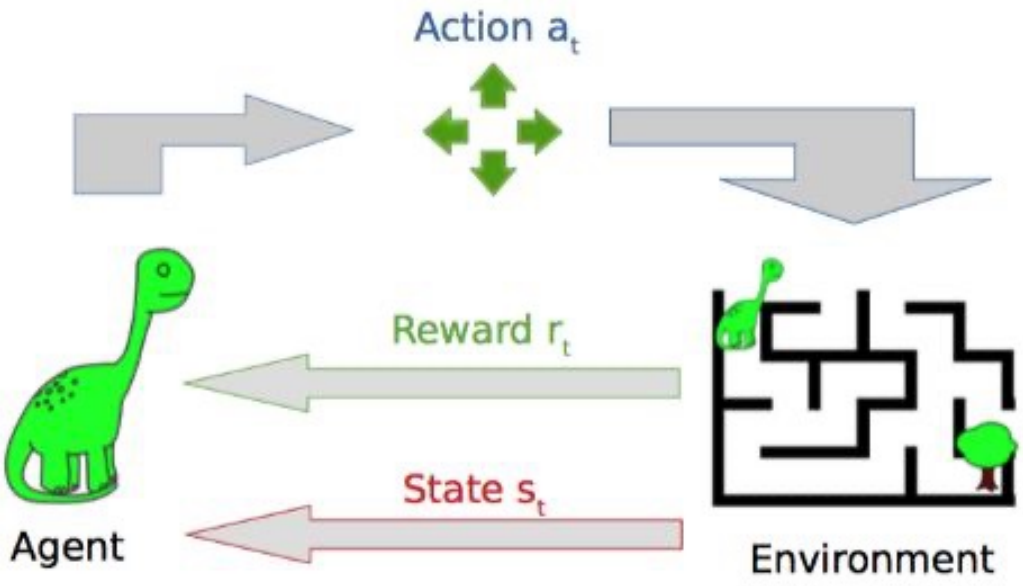


# Reinforcement learning


https://www.youtube.com/watch?v=nIgIv4IfJ6s&t=1s

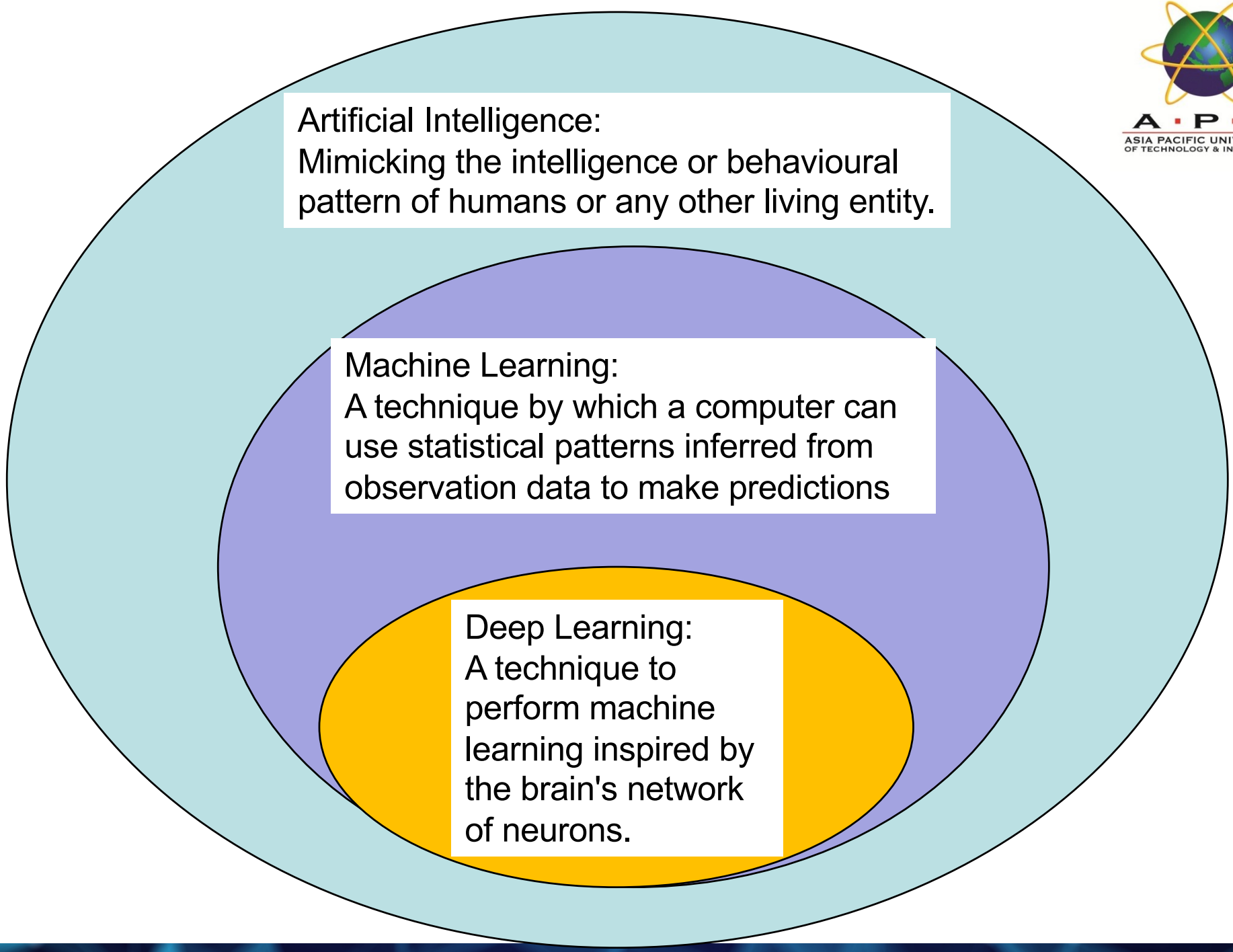
# Reinforcement learning

- https://www.theverge.com/tldr/2017/7/10/15946542/deepmind-parkour-agent-reinforcement-learning
- https://video.twimg.com/ext_tw_video/1111683489890332672/pu/vid/1200x674/WqUJEhUETw0M0gCl.mp4?tag=8



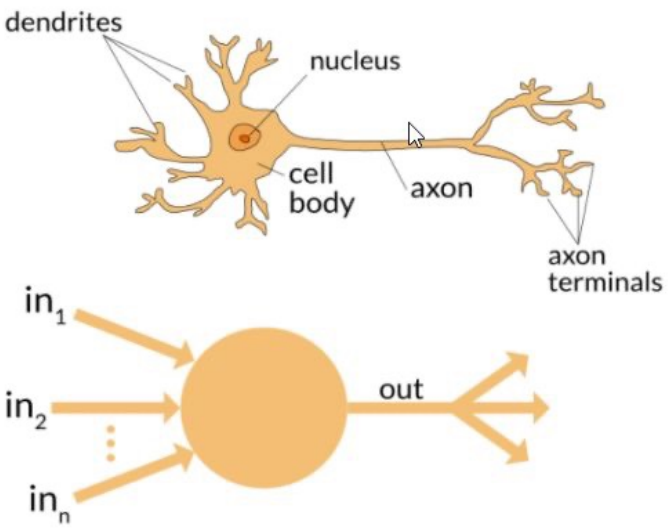
# Supervised vs Unsupervised vs Reinforcement

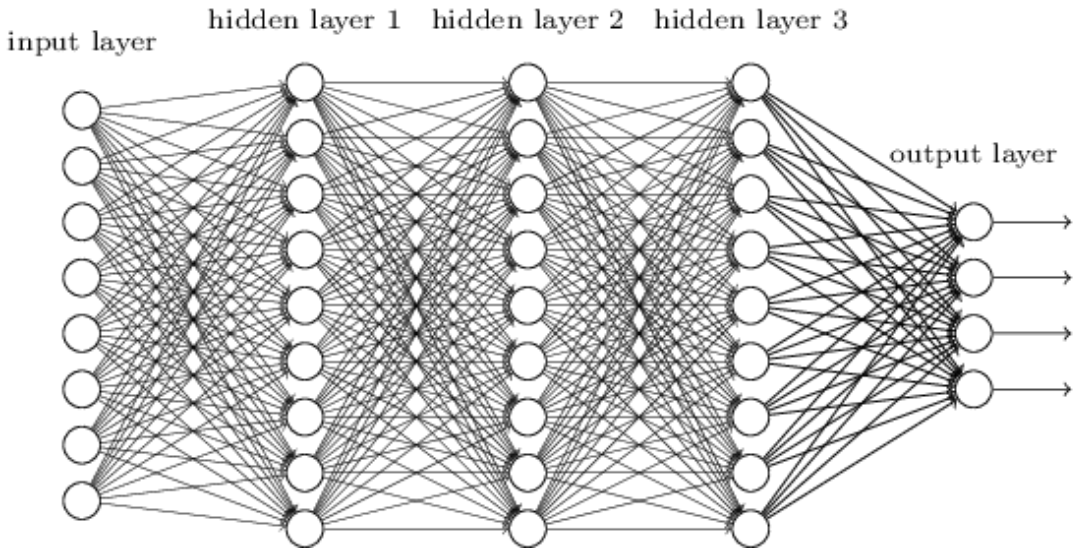

| Criteria | Supervised Learning | Unsupervised Learning | Reinforcement Learning |
|---|---|---|---|
| Definition | The machine learns by using labeled data | The machine is trained on unlabeled data without any guidance | An agent interacts with its environment by performing actions & learning from errors or rewards |
| Type of problems | Regression & classification | Association & clustering | Reward-based |
| Type of data | Labeled data | Unlabeled data | No predefined data |
| Training | External supervision | No supervision | No supervision |
| Approach | Maps the labeled inputs to the known outputs | Understands patterns & discovers the output | Follows the trial-and-error method |


Artificial Intelligence:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.
Machine Learning:
A technique by which a computer can use statistical patterns inferred from observation data to make predictions
Deep Learning:
A technique to perform machine learning inspired by the brain's network of neurons.

# Deep Learning

- **Deep learning is a system architecture, not an algorithm**
- a style of parallel computation inspired by neurons and their adaptive connections: It's a very different style from a sequential computation.

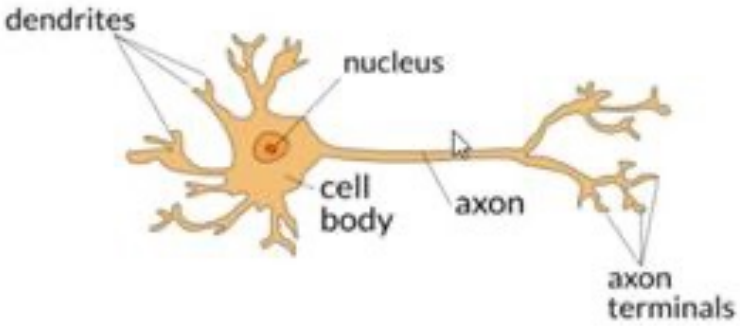


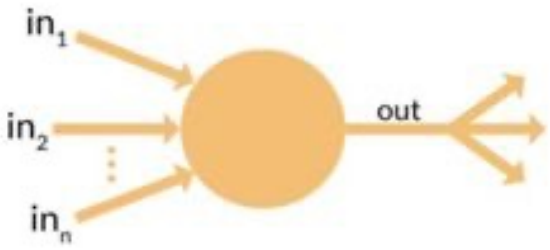



# Artificial Neural Network

ANN is a machine learning approach inspired by the way in which the brain performs a particular learning task

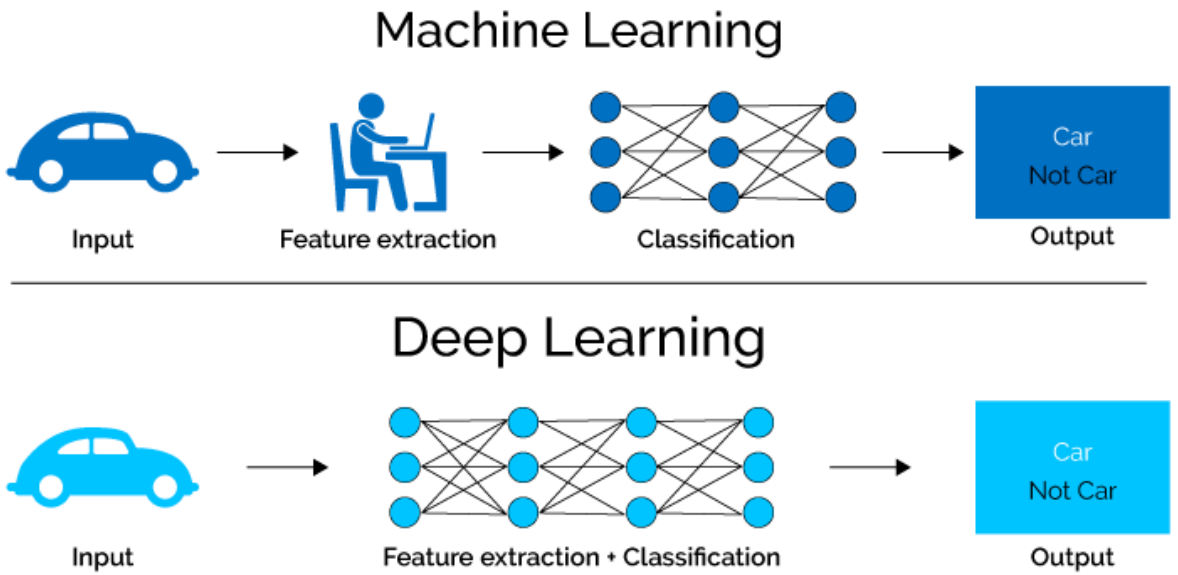
- ANN has several “neurons” interconnected (architecture)
- Knowledge about the learning task is stored as inter neuron connection strengths (weights)
- During the learning process the  weights are modified in order to model the particular learning task correctly on the training examples.



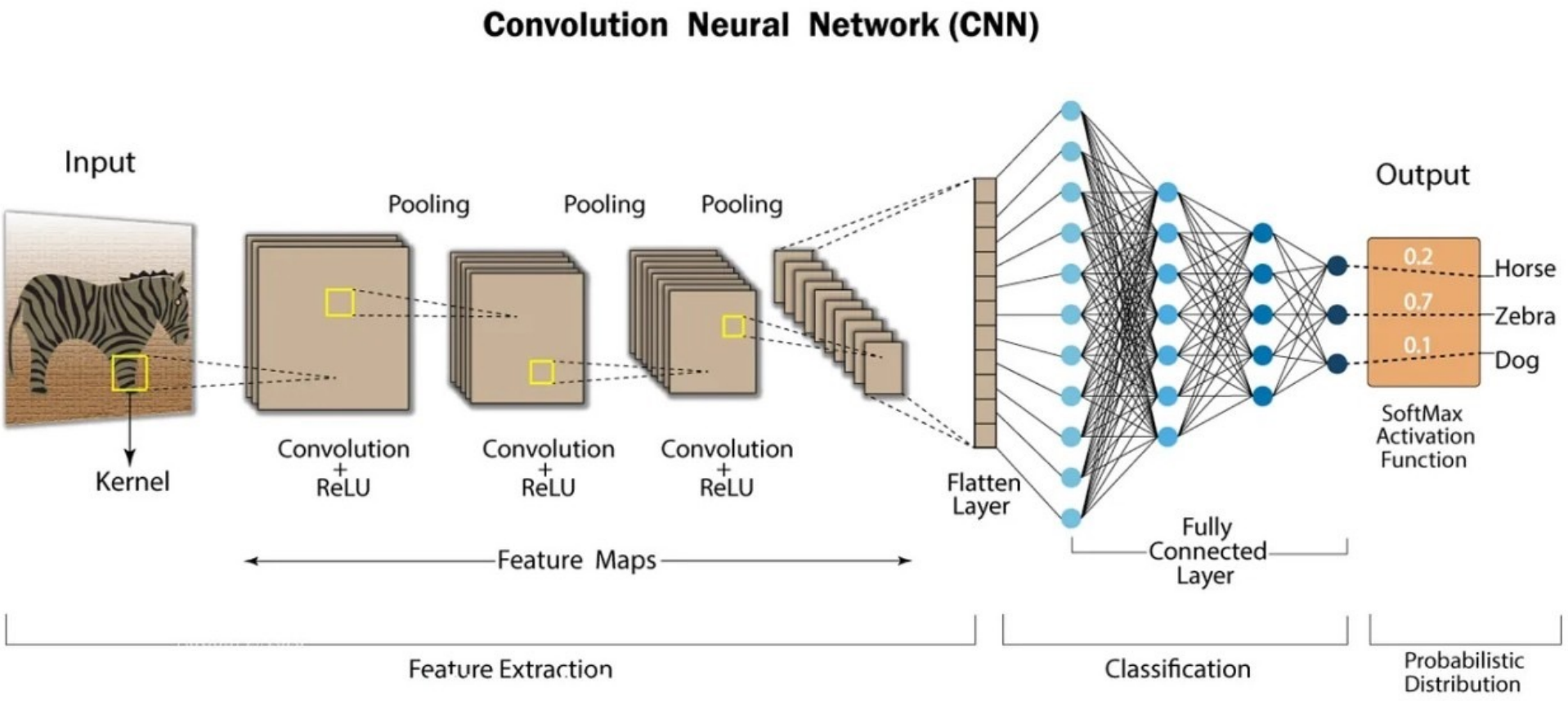



# Deep Learning

- Deep learning methods aim at learning feature hierarchies where features from higher levels of the hierarchy are formed from the lower-level features.
  - ***The Model Defines the Features***



# Deep Learning



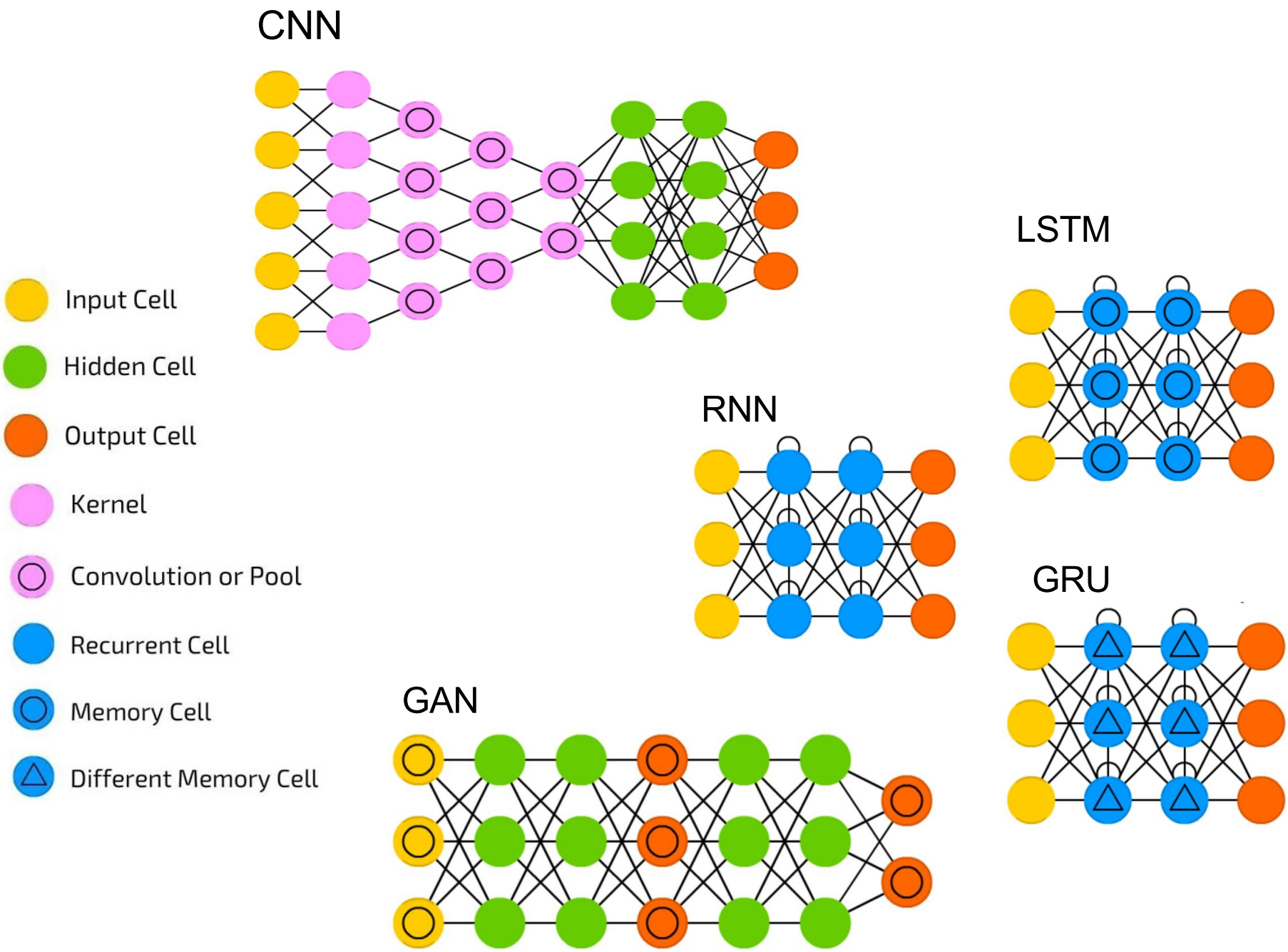
# Machine Learning v. Deep Learning

| | Machine Learning | Deep Learning |
|---|---|---|
| How it works | Uses automated algorithms that learn to predict future decisions and model functions using the data which is provided. | Interprets features in data and the relationships using neural networks which pass the data through several layers of the algorithm. |
| Intervention | Algorithms usually require human interventions to examine different variables and dataset features. | Algorithms require no human intervention for data analysis. |
| Data points | A few hundred to a few thousand. | Can be into the millions. |
| Output | A numerical value such as a score or classification. | Could possibly be anything? |
| Hardware | Requires less hardware capacity than deep learning. | Requires high- end hardware capabilities such as GPUs to perform at its best. |
| Feature extraction | Requires features to be identified. | Will look to determine features from patterns in data. |
| Training time | Training time is less. | Training time is more. |

# Current Limits

- DNNs consist of multiple layers of interconnected neurons. Each neuron and layer contributes towards the task that the network has been trained to execute.
- However, scalability and generalisation is not simple.
- The behaviour of a NN changes with the amount of data, and it also depends on methods of training and network architecture.
  - you can't just feed 10x more data to a NN and expect it to become 10x more precise.
- Must experiment with network structures and training modes for the technology to be usable.

***This has led to a LOT of variations of DNN architectures***




https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

# Hardware is Important

**Training a neural network is a compute-intensive process**

- DNNs require a large amount of training data to achieve high accuracy, meaning hundreds of thousands of input samples
- State-of-the-art DNNs can have well over one billion parameters to adjust
- Requires multiple forward passes for error detection, and multiple backward passes of adjusting the weights of millions of neurons in various layers of the network.

# Hardware is Important

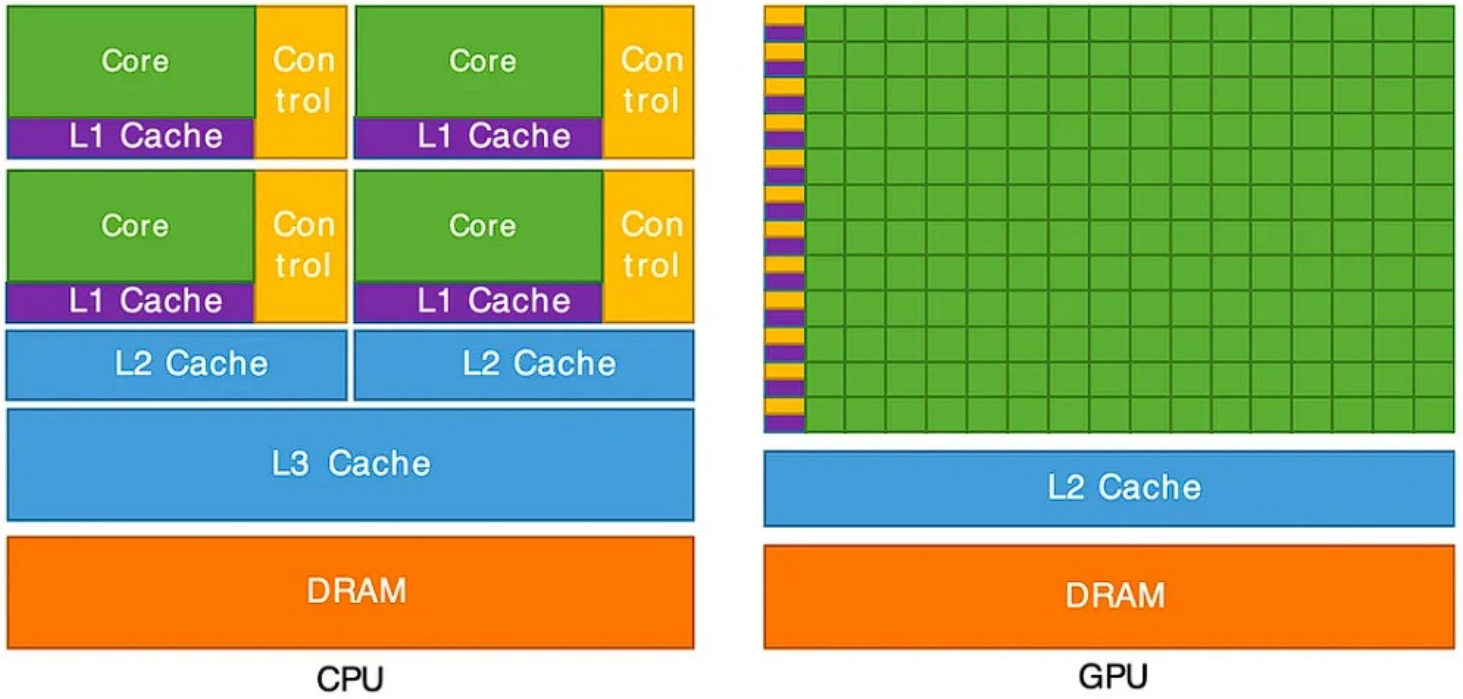

- Because neural networks are created from large numbers of identical neurons, they are highly parallel by nature.
- This parallelism maps naturally to GPUs, which provide a significant speedup over CPU-only training.
- Neural networks rely heavily on matrix math operations and require tremendous amounts of floating-point performance and bandwidth for both efficiency and speed.
- GPUs have thousands of processing cores optimized for matrix math operations

# CPU vs. GPU Architecture

Small number of powerful cores

versus

Very large number of simple stream processors




CPU, GPU Schematic [https://cvw.cac.cornell.edu/GPUarch/gpu_characteristics]

# Review Questions

1. What are the differences between AI, ML and DL?
2. What are different types of Machine Learning?
3. What are the typical characteristics of Deep Learning?

# Summary / Recap of Main Points

- The differences between AI, ML and DL.
- Types of Machine Learning.
- Characteristics of Deep Learning.