Topic / Unit 5: **Reinforcement Learning**

With Best wishes to readers -

Professor Jay Bhatnagar

Introduction to Artificial Intelligence(AI)

Artificial Intelligence are the development of computer systems that are capable of performing tasks that require human intelligence,

Eg: Decision making, Object detection, Solving complex problems etc.

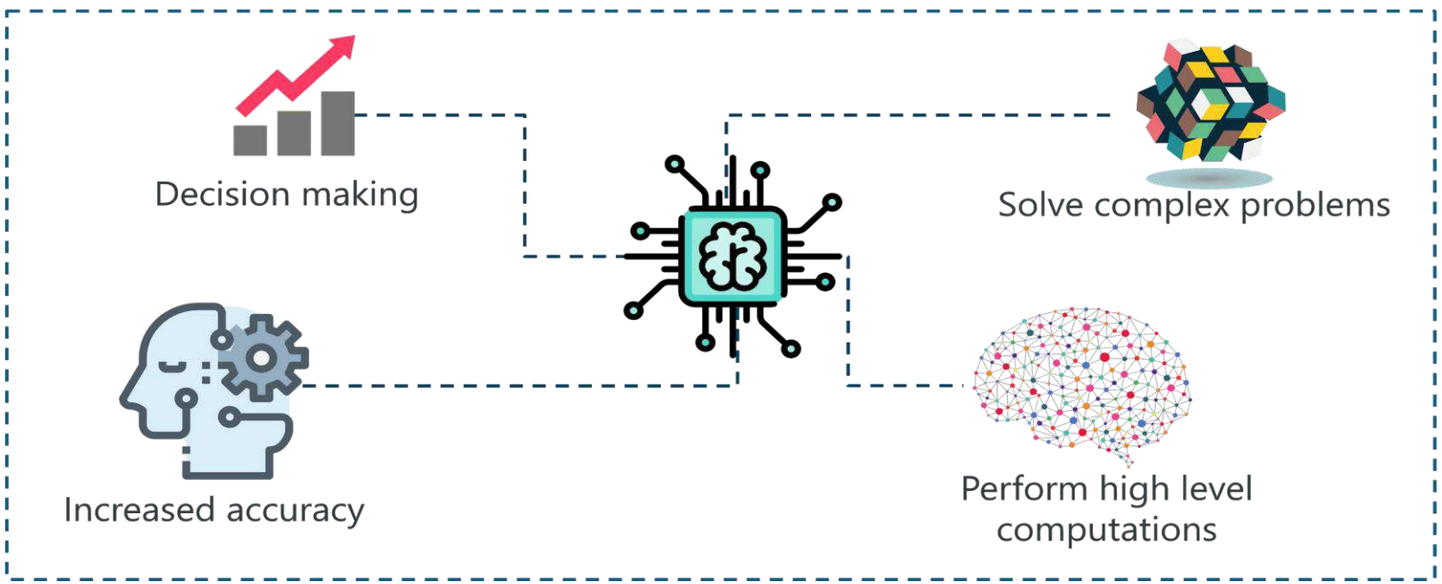


Figure Source: Edureka

Rational Decisions

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

Definition of AI short ans

- Existing definitions advocate everything from replicating human intelligence to simply solving knowledge-intensive tasks.

Examples:

"Artificial Intelligence is the design, study and construction of computer programs that behave intelligently." -- Tom Dean.

"Artificial Intelligence is the enterprise of constructing a physical symbol system that can reliably pass the Turing test." -- Matt Ginsberg.

Introduction to Artificial Intelligence

- Philosophy of AI - "*Can a machine think and behave like humans do?*"
- In Simple Words - *Artificial Intelligence is a way of* ***making a computer, a computer-controlled robot, or a software think intelligently****, in the similar manner the intelligent humans think.*
- **Artificial intelligence** (**AI**) is an area of computer science that emphasizes the creation of **intelligent** machines that work and re-act like humans.
- *AI is accomplished by studying how human brain thinks, and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.*

Rational Decisions

- We'll use the term **rational** in a very specific, technical way

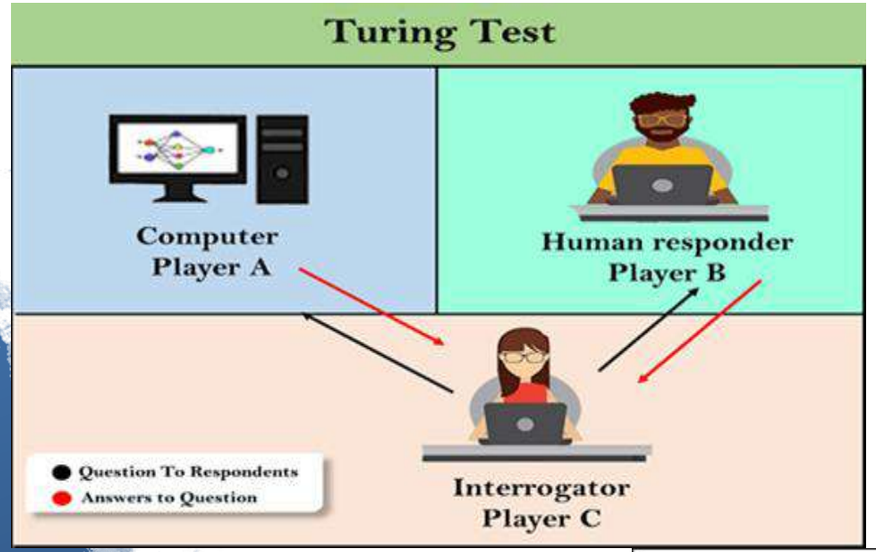
  - Rational: maximally achieving pre-defined goals
  - Rationality only concerns what decisions are made (not the thought process behind them)
  - Goals are expressed in terms of the **utility** of outcomes
  - Being rational means **maximizing your expected utility**

- **Computational Rationality**

Turing Test Approach (short ans.)

proposed by Alan Turing (1950)

THINK HUMANLY &
ACT HUMANLY

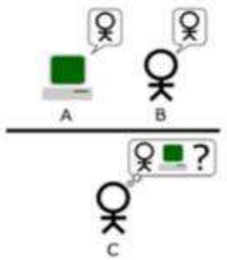


Turing Test

= Testing the machine's ability to exhibit ~~intelligence~~ human expertise.

Traditional Turing Test: a conversion.

If, during a conversation, machine can fool the interrogator into thinking that it is human, then it pass the test



Modern example, the captcha

Recognizing words is too hard for machines, but easy for people.

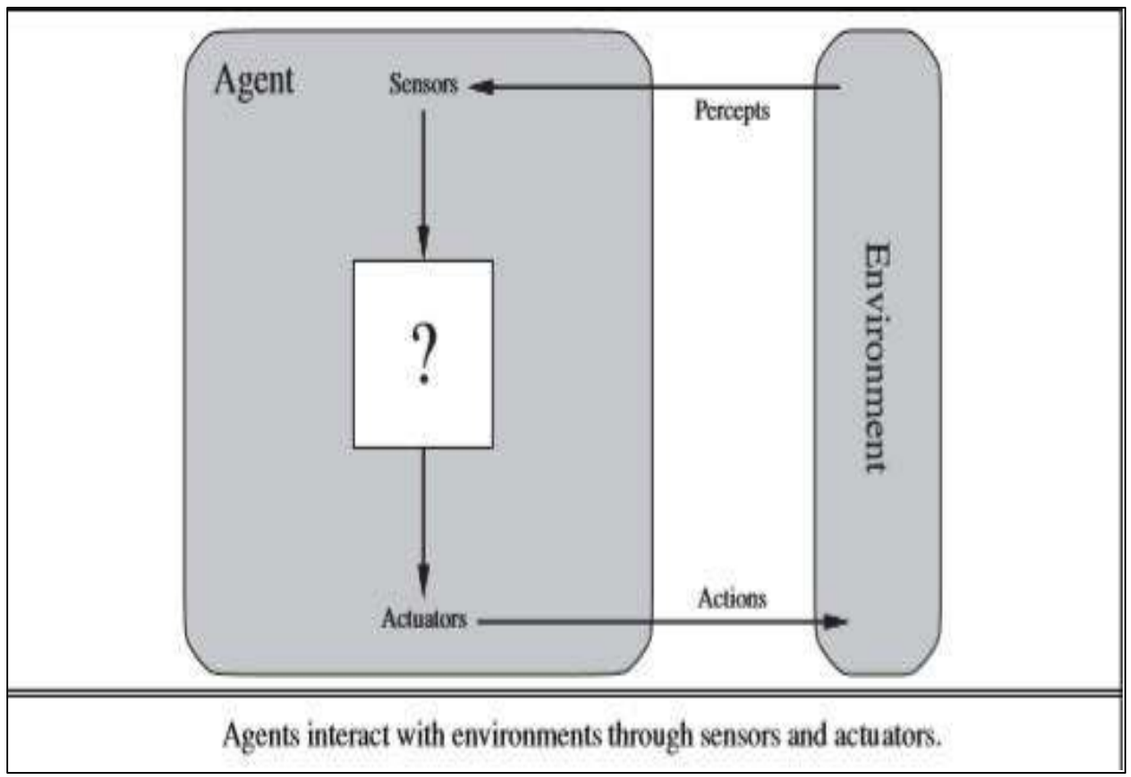
Foundations of Artificial Intelligence

Below are the disciplines that contributed **ideas**, **viewpoints** and **techniques** to AI:

1. Philosophy
2. Mathematics
3. Economics
4. Neuroscience
5. Psychology
6. Computer Engineering
7. Control theory, Linguistics

Agent & Environment

- ‘Anything’ that can gather information about its environment and take **action** based on that information.
- In AI, an intelligent agent (IA) is an autonomous entity which observes through sensors and acts upon an environment using actuators and directs its activity towards achieving goals.
- **Intelligent** agents may also learn or use knowledge to achieve their goals.
- A **rational** agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.



Agents interact with environments through sensors and actuators.

Agents
MCQ

- A human agent
  - eyes, ears, and other organs for sensors
  - hands, legs, vocal tract, and so on for actuators.
- A robotic agent
  - might have cameras and infrared range finders for sensors
  - various motors for actuators.
- A software agent
  - receives keystrokes, file contents, and network packets as sensory inputs
  - acts on the environment by displaying on the screen, writing files, and sending network packets.

Agent…

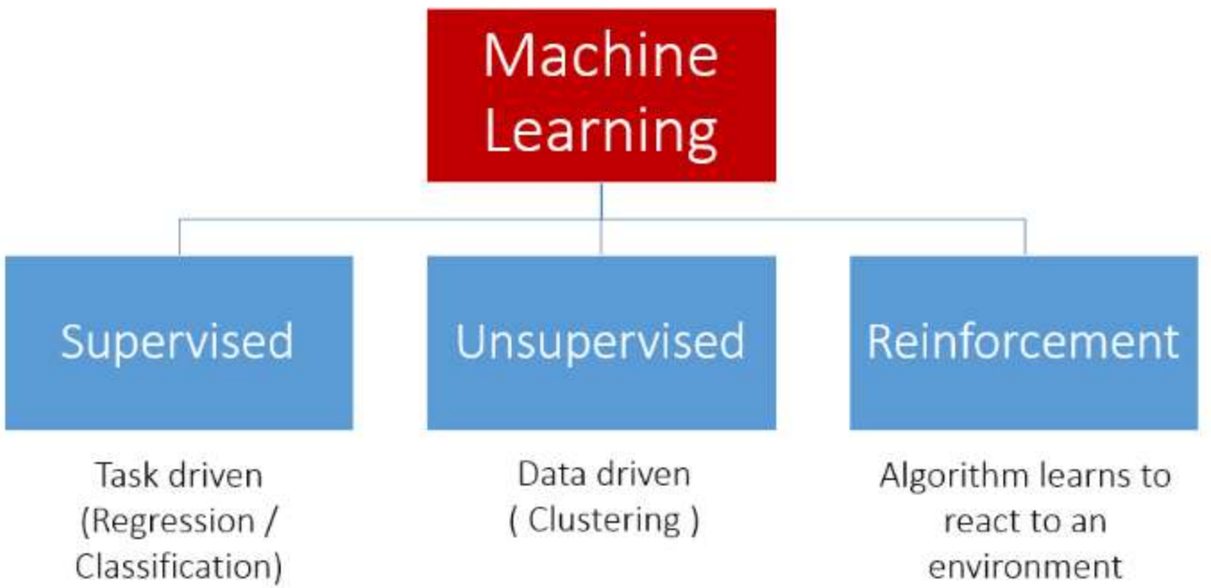
- **Percept – Concept + Hypothesis + Instances** refers to the agent's perceptual inputs at any given instant.
- An agent's **percept sequence** is the complete history of everything the agent has ever perceived.
- An agent's choice of action at any given instant can depend on the entire percept sequence observed to date, but not on anything it hasn't perceived.
- Mathematically speaking, we say that an Agent's behavior is described by the agent function that maps any given percept sequence to an action
- One could view a hand-held calculator as an agent that chooses the action of displaying "4" when given the percept sequence "2 + 2 =,"

Supervised learning is a method in which we teach the machine using labelled data

In unsupervised learning the machine is trained on unlabelled data without any guidance

In Reinforcement learning an agent interacts with its environment by producing actions & discovers errors or rewards

Types of Machine Learning



*

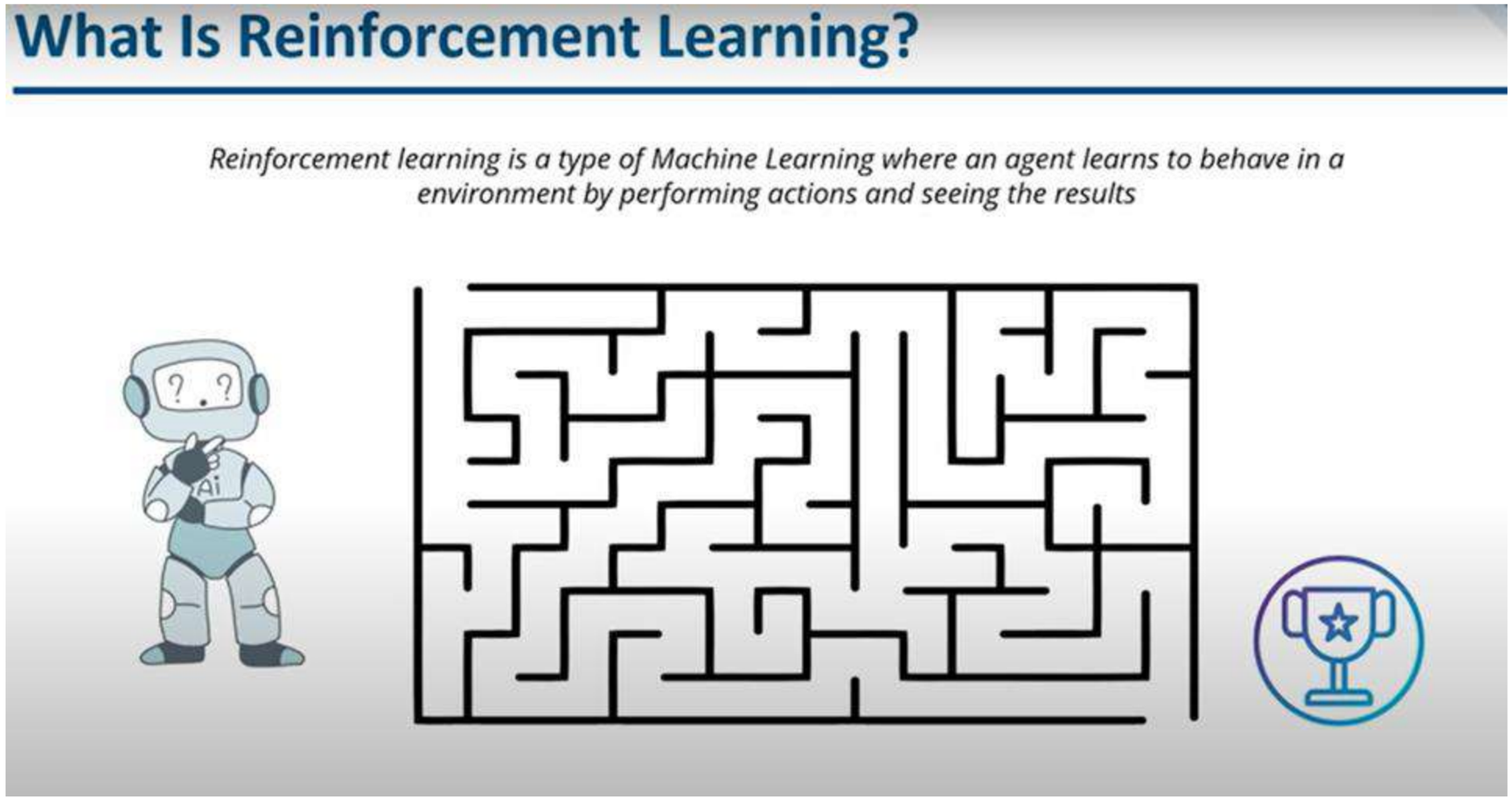
Reinforcement Learning Vs Supervised Learning (MCQ)

- The Reinforcement Learning and Supervised Learning both are the part of machine learning, but both types of learning's are far opposite to each other.
- The RL agents interact with the environment, explore it, take action, and get rewarded.
- Whereas supervised learning algorithms learn from the labeled dataset and, on the basis of the training, predict the output.
- The difference table between RL and Supervised learning is given below:

Reinforcement Learning	Supervised Learning
RL works by interacting with the environment.	Supervised learning works on the existing dataset.
The RL algorithm works like the human brain works when making some decisions.	Supervised Learning works as when a human learns things in the supervision of a guide.
There is no labeled dataset is present	The labeled dataset is present.
No previous training is provided to the learning agent.	Training is provided to the algorithm so that it can predict the output.
RL helps to take decisions sequentially.	In Supervised learning, decisions are made when input is given.

*

Re – inforcement Learning



*

What is game playing?

- The term Game means a sort of conflict in which *n* individuals or groups (known as players) participate.
- Game theory denotes games of strategy.
- Game theory allows decision-makers (players) to cope with other decision-makers (players) who have different purposes in mind.
- In other words, players determine their own strategies in terms of the strategies and goals of their opponent.

What is Reinforcement Learning? (MCQ)

- Reinforcement Learning is a ~~feedback~~ (interaction) based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions.

- For each good action, the agent gets positive ~~feedback~~ reward, and for each bad action, the agent gets negative ~~feedback~~ reward or penalty.

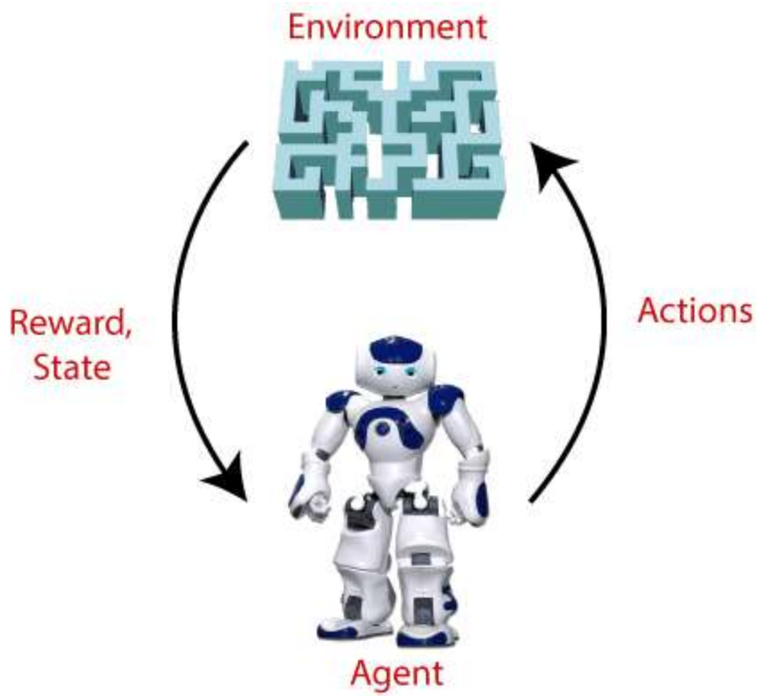
- In Reinforcement Learning, **the agent learns automatically** using feedbacks without any labeled data, unlike supervised learning.

- Since there is no labeled data, so the agent is bound to learn by its experience only.

- RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as **game-playing, robotics**, etc.

*

- The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.
- The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way.
- Hence, we can say that "Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that."
- How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.
- It is a core part of Artificial intelligence. It is a core part of Artificial intelligence, and all AI agent works on the concept of reinforcement learning.
- Here we do not need to pre-program the agent, as it learns from its own experience without any human intervention.

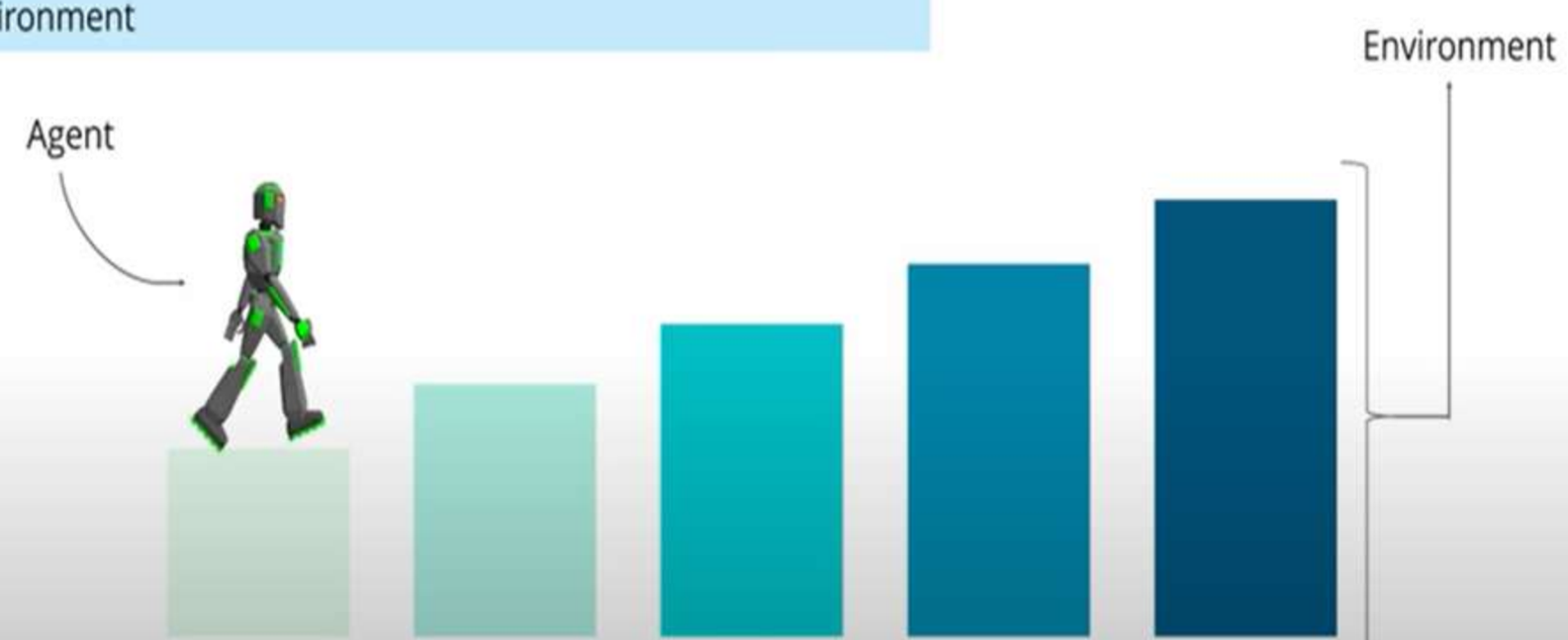


*

Reinforcement Learning Process

Reinforcement Learning system is comprised of two main components:

- Agent
- Environment

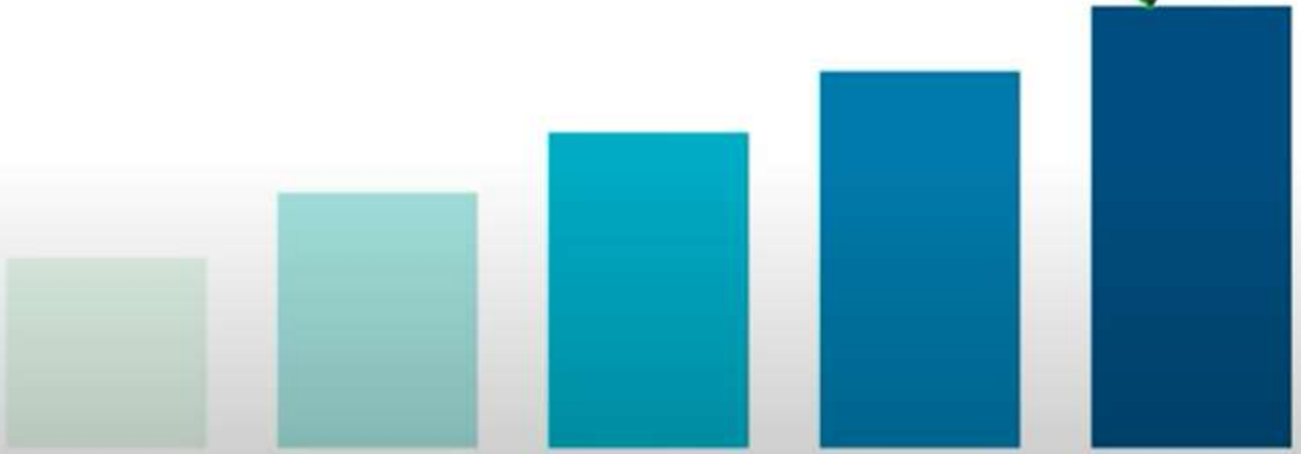


*

Reinforcement Learning Process

Reinforcement Learning system is comprised of two main components:

- Agent
- Environment



*

- Example: Suppose there is an **AI agent present within a maize environment**, and his goal is to find the diamond.
- The agent interacts with the environment by performing some actions, and based on those actions, the state of the agent gets changed, and it also receives a reward or penalty as feedback.
- The agent continues doing these three things (take action, change state/remain in the same state, and get feedback), and by doing these actions, he learns and explores the environment.
- The agent learns (**mark a path**) that **what actions lead** to **positive points** or rewards and what actions lead to penalty. As a positive reward, the agent gets a positive point, and as a penalty, it gets a negative point.

Terms used in Reinforcement Learning

Agent: The RL algorithm that learns from trial and error

Environment: The world through which the agent moves

Action (A): All the possible steps that the agent can take

State (S): Current condition returned by the environment

Terms used in Reinforcement Learning

Reward (R): An instant return from the environment to appraise the last action

Policy (π): The approach that the agent uses to determine the next action based on the current state

Value (V): The expected long-term return with discount, as opposed to the short-term reward R

Action-value (Q): This similar to Value, except, it takes an extra parameter, the current action (A)

Elements of Reinforcement Learning

There are four main elements of Reinforcement Learning, which are given below:

1. Policy
2. Reward Signal
3. Value Function
4. Model of the environment

- **Policy:** A policy can be defined as a way how an agent behaves at a given time.

- **Reward Signal:** The goal of reinforcement learning is defined by the reward signal. At each state, the environment sends an immediate signal to the learning agent, and this signal is known as a **reward signal**.

- **Value Function:** The value function gives information about how good the situation and action are and how much reward an agent can expect.

- **Model:** Which mimics the behavior of the environment. The approaches for solving the RL problems **with the help of the model** are termed as the **model-based approach**. Comparatively, an approach **without using a model** is called a **model-free approach**.

Approaches to RL model (MCQ)

There are mainly three ways to implement reinforcement-learning in ML, which are:

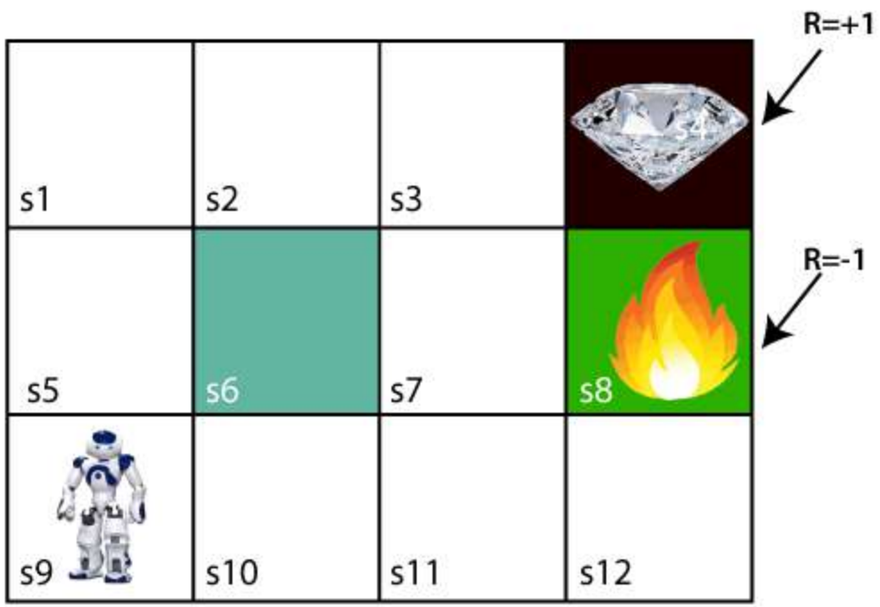
- **Value-based:**
  The value-based approach is about to find the optimal value function, which is the maximum value at a state under any policy. Therefore, the agent expects the long-term return at any state(s) under policy π.
- **Policy-based:**
  Policy-based approach is to find the optimal policy for the maximum future rewards without using the value function. In this approach, the agent tries to apply such a policy that the action performed in each step helps to maximize the future reward.
  The policy-based approach has mainly two types of policy:
  - **Deterministic:** The same action is produced by the policy (π) at any state.
  - **Stochastic:** In this policy, probability determines the produced action.
- **Model-based:** In the model-based approach, a virtual model is created for the environment, and the agent explores that environment to learn it. There is no particular solution or algorithm for this approach because the model representation is different for each environment.

How does Reinforcement Learning Work?

To understand the working process of the RL, we need to consider two main things:

- **Environment:** It can be anything such as a room, maze, football ground, etc.
- **Agent:** An intelligent agent such as AI robot.

Example of a maze environment that the agent needs to explore. Consider the below image: **OPTIMIZATION PROBLEM**



- In the image, the agent is at the very first block of the maze.
- The maze is consisting of an S_6 block, which is a **wall**, S_8 a **fire pit**, and S_4 a **diamond block.**
- The agent cannot cross the S_6 block, as it is a solid wall.
- If the agent reaches the S_4 block, then get the **+1 reward;** if it reaches the fire pit, then gets **-1 reward point**. It can take four actions**: move up, move down, move left, and move right.**
- The agent can take any path to reach to the final point, but he needs to make it in possible fewer steps. Suppose the agent considers the path **S9-S5-S1-S2-S3**, so he will get the +1-reward point.

Types of Reinforcement learning (MCQ)

There are mainly two types of reinforcement learning, which are:

- **Positive Reinforcement:**

  The positive reinforcement learning means adding something to increase the tendency that expected behavior would occur again. It impacts positively on the behavior of the agent and increases the strength of the behavior.

  This type of reinforcement can sustain the changes for a long time, but too much positive reinforcement may lead to an overload of states that can reduce the consequences.

- **Negative Reinforcement:**

  The negative reinforcement learning is opposite to the positive reinforcement as it increases the tendency that the specific behavior will occur again by avoiding the negative condition.

  It can be more effective than the positive reinforcement depending on situation and behavior, but it provides reinforcement only to meet minimum behavior.

How to represent the agent state? (MCQ)

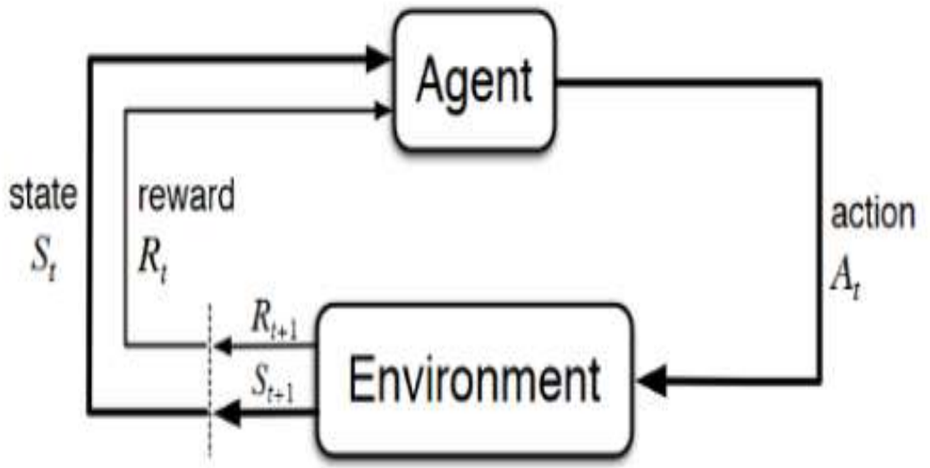
- We can represent the agent state using the **Markov State** that contains all the required information from the history.

- The State St is Markov state if it follows the given condition:

$$Pr[S_{t+1} \mid S_t] = Pr[S_{t+1} \mid S_1, \ldots\ldots, S_t]$$

- The Markov state follows the **Markov property**, which says that the future is independent of the past and can only be defined with the present.

- The RL works on fully observable environments, where the agent can observe the environment and act for the new state. The complete process is known as **Markov Decision process [MDP].**



The agent–environment interaction in a Markov decision process.

Shortest Path Problem/ Greedy algo.

- Now suppose you are at place A, the only visible path is your next destination and anything beyond that is not known.

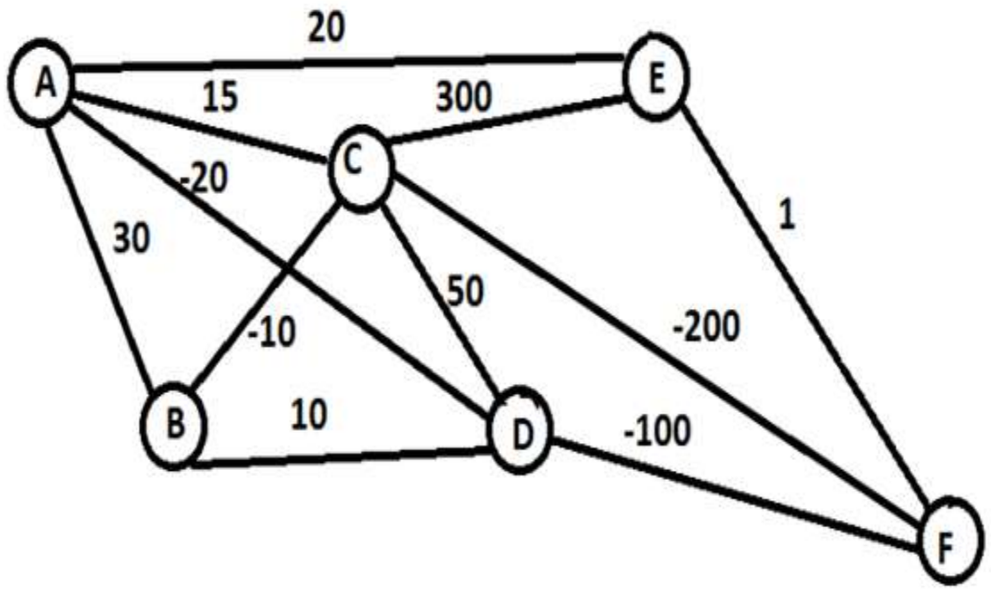
- You can take a greedy approach and take the best possible next step, which is going from {A -> D} from a subset of {A -> (B, C, D, E)}.

- Similarly now you are at place D and want to go to place F, you can choose from {D -> (B, C, F)}.

- We see that {D -> F} has the lowest cost and hence we take that path.

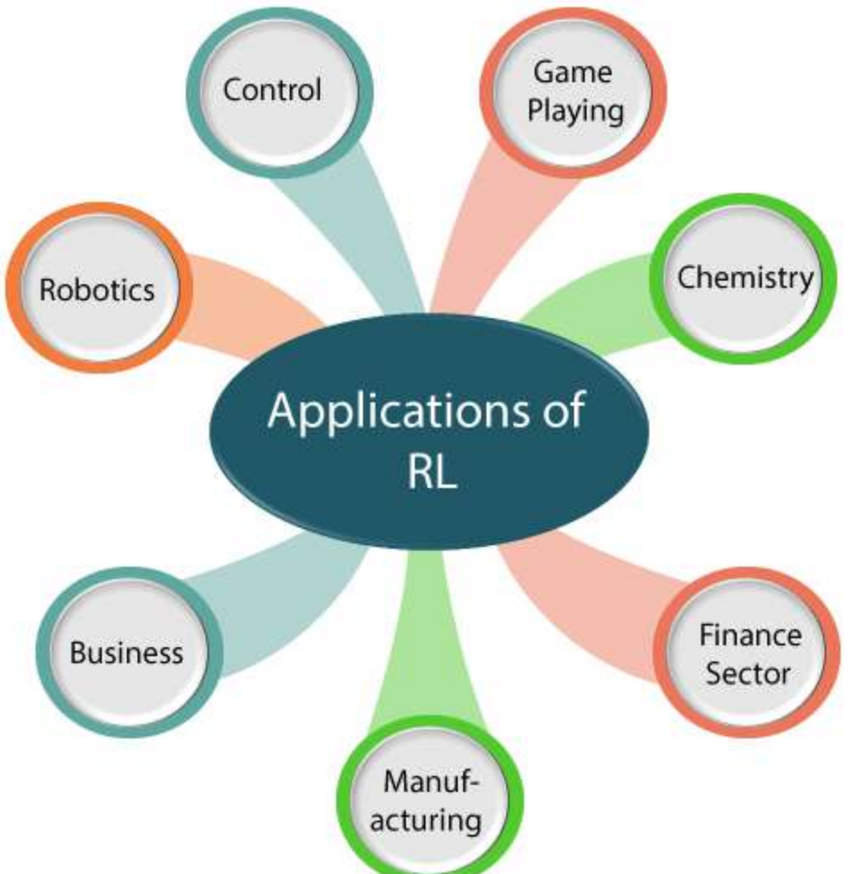
- So here, our policy was to take {A -> D -> F} and our Value is -120.

This algorithm is known as **epsilon greedy,** which is literally a greedy approach to solving the problem.



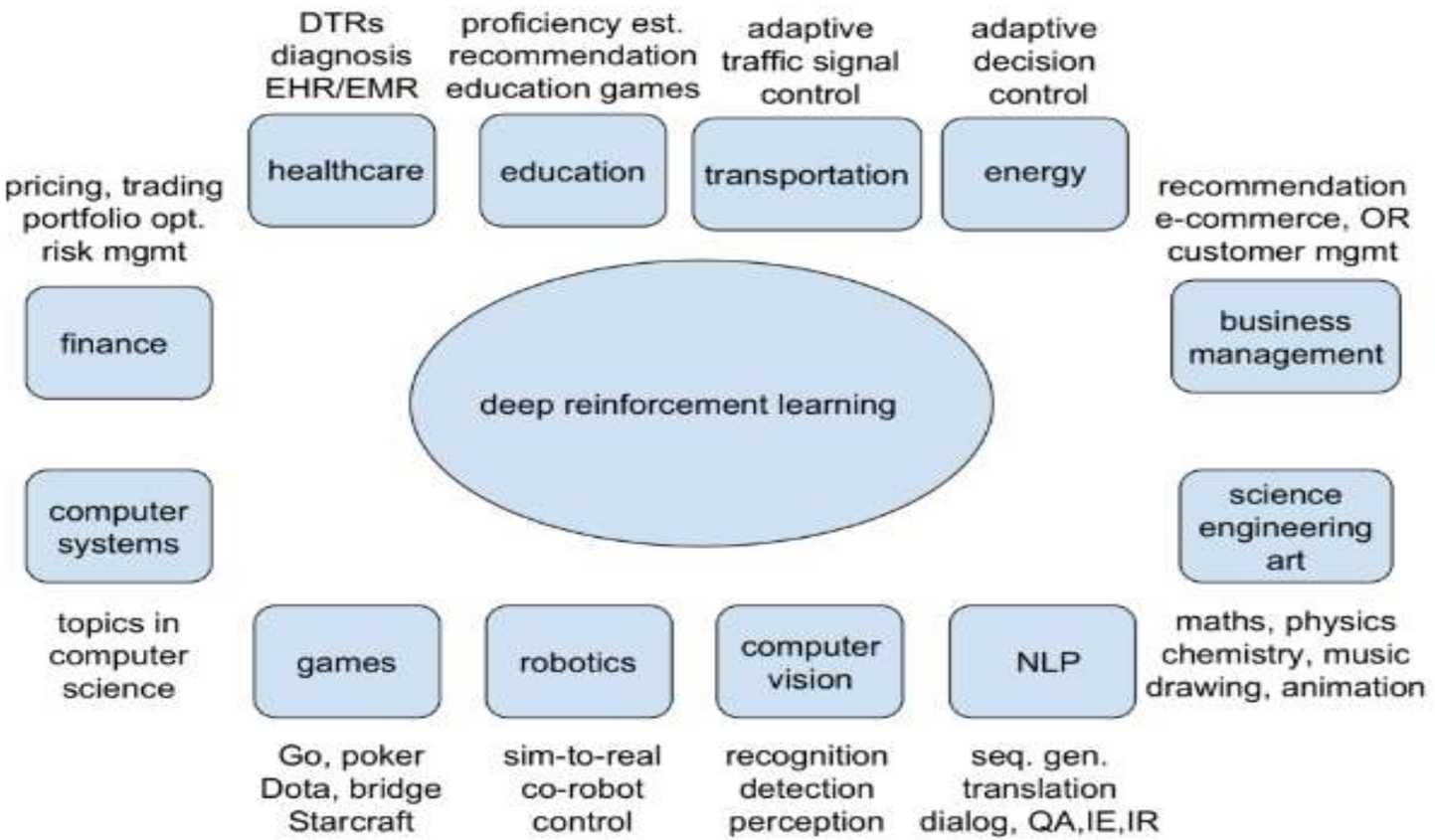
Reinforcement Learning Applications (MCQ)

- **Robotics:**
  - RL is used in **Robot navigation, Robo-soccer, walking, juggling**, etc.
- **Control:**
  - RL can be used for **adaptive control** such as Factory processes, admission control in telecommunication, and Helicopter pilot is an example of reinforcement learning.
- **Game Playing:**
  - RL can be used in **Game playing** such as tic-tac-toe, chess, etc.
- **Chemistry:**
  - RL can be used for optimizing the chemical reactions.
- **Business:**
  - RL is now used for business strategy planning.
- **Manufacturing:**
  - In various automobile manufacturing companies, the robots use deep reinforcement learning to pick goods and put them in some containers.
- **Finance Sector:**
  - The RL is currently used in the finance sector for evaluating trading strategies.



*

Applications



Yuxi Li, Deep Reinforcement Learning, arXiv, 2018

*

Game Playing [Deep Blue in Chess]

- This is the **first World Champion class chess computer** among the oldest challenges in computer science.
- When **World Chess Champion Garry Kasparov** resigned the last game of a **six-game match against IBM's Deep Blue supercomputer on 11 May 1997**, his loss marked achievement of Deep Blue to meet its goal.
- Deep Blue's 1996 debut in the first Kasparov versus Deep Blue match in Philadelphia finally eclipsed Deep Thought II.
- The 1996 version of Deep Blue used a new chess chip designed at IBM Research over the course of three years.
- A major revision of this chip participated in the historic 1997 rematch between Kasparov and Deep Blue to achieve its goal.

IBM DeepBlue beats the chess champion Garry Kasparov

https://www.youtube.com/watch?v=KF6sLCeBj0s

*

IBM Watson in Jeopardy

- **Watson is a question-answering computer system** capable of answering questions posed in natural language, developed in **IBM's DeepQA project by a research team led by principal investigator David Ferrucci**.

- **First computer to defeat TV game show Jeopardy! champions (Ken Jennings and Brad Rutter)**. Research teams are working to adapt Watson to other information-intensive fields, such as telecommunications, financial services and government.
- https://www.youtube.com/watch?v=WFR3lOm_xhE

IBM Watson computer defeats human in final

*

Google’s Deep Mind in AlphaGo

- **AlphaGo is a computer program that plays the board game Go**.
- It was developed by DeepMind Technologies which was later acquired by Google.
- **AlphaGo versus Lee Sedol, also known as the Google DeepMind Challenge Match, was a five-game Go match between 18-time world champion Lee Sedol and AlphaGo, a computer Go program developed by Google DeepMind, played in Seoul, South Korea between the 9th and 15th of March 2016**.
- It is **able to do this by using a novel form of reinforcement learning, in which AlphaGo Zero becomes its own teacher**.
- The system starts off with a neural network that knows nothing about the game of Go.
- It then plays games against itself, by combining this neural network with a powerful search algorithm.
- As it plays, the neural network is tuned and updated to predict moves, as well as the eventual winner of the games.
- Google's Deep Mind Explained! - Self Learning A.I.
- https://www.youtube.com/watch?v=mzZWPcgcRD0

*

contd …..

- Figure shows the photo for Google DeepMind AlphaGo versus Lee Sedol.
- This updated neural network is then recombined with the search algorithm to create a new, stronger version of AlphaGo Zero, and the process begins again.
- In each iteration, the performance of the system improves by a small amount, and the quality of the self-play games increases, leading to more and more accurate neural networks and ever stronger versions of AlphaGo Zero.

*

Thanks