# Reinforcement Learning (RL)
# (Markov Decision Process)

## Markov Decision Process (MDP):

It is the theoretical foundation of **reinforcement learning** and is the basis of any reinforcement learning process.

It is a general framework for modeling a reinforcement-learning problem.

The word **« Markov »** here refers to the **« Markov Property »**, which is a fundamental principle of the **« Markov Process »**. In the next sections, the Markov Property is explained alongside the Markov Chain (Markov Process), as well as how to convert the Markov Chain into an MDP (Markov Decision Process).

### Markov Property:

The Markov property is satisfied when the current state of the process is sufficient to predict the future state of the process.

**Example:** if we have two states (sunny or rainy) and if we are in the « sunny » state, **we do not use the sequence of previous states for future state transitions**.

In another definition. A stochastic process has the Markov property if the conditional probability distribution of the future states of the process (conditional on the past and present values) depends only on the present state.

#### Why the Markov Property?

The Markov property is important in reinforcement learning because decisions and values are assumed to be uniquely based on the current state. For these to be effective and informative.

## Markov Process (Markov Chain):

The formal definition of the Markov process is the following:

- A set of states (S) in which a system can be found.
- A transition matrix (T), with transition probabilities, which defines the dynamics of the system.

**Remark:** Because the observations form a chain or sequence of states, the Markov Process is also known as the Markov Chain.

The sequence of observations over time forms a chain of states, this sequence is called **the history**.

**Important:** To be called an MP (Markov Process), a system must satisfy the Markov Property.

## Markov Reward Process (MRP):

Now, we will need to expand our Markov process model a little bit now. For each transition, we add a value called the reward value, which is a crucial notion in reinforcement learning and can be negative or positive, as well as a value called the discount factor, which is a value between 0 and 1 and can be negative or positive (usually 0.9 or 0.99).

## The road to MDP:

Now to transform our MRP into MDP and encompass all the entities of reinforcement learning, we add a finite set of actions.

## Summary:

Markov Chain (Markov Process) + Rewards = **Markov Reward Process (MRP)**

Markov Reward Process + Actions = **MDP (Markov Decision Process)**

So:

Markov Chain (Markov Process) + Rewards + Actions

⇨ **Markov Decision Process (MDP)**.

We can also see the relationship in the following figure:


MRP