Focus on formal definitions

Definition of MDP :

$MDP\{S; A; P^{a}_{ss'}, R^{ss'}, T, I\}$

- S: State space, a list of possible states (integer numbers or strings)
- A: Action space, a list of possible actions (integer numbers or strings)
- P: Transition function
- R: Reward function
- T: Terminal states
- I: Initial state distribution

"Dimensions" of RL (Properties):

1. **Discrete (Finite)** : State and action space finite [chess]
   Continuous (Infinite) : State and action space infinite [driving a car]

2. **Model-based** : Algorithms that find the optimal policy have access to to MDP (Stochacity (Transition function) and Rewards (R))[Optimal control]

   Model-free: Algorithms that do not have a strong prior on the MDP and try to estimate them [Google AlphaGo]

3. **Deterministic** : Doing the same action always leads to the same state [Playing a chess game against a computer with no random components]

   Stochastic : The same action could lead to different states [Chess against a human opponent]

4. **Episodic** : (slide 20) The goal is to reach a terminal state [Painting a car]

   Continuing : Reaching and keeping a state [Slackline]

5. **Markovian** : Optimal policy depends only on the actual state = transition functions are independant from the previous states [Sampling with replacement] i.e. : Consider a $MDP\{S, A, Tr, R, T, I\}$, the markov property is the following :

$$P(s_t \to s', a_t) = Tr(s_t, a_t | s_0, a_0, ...s_{t-1}, a_{t-1}) \quad (1)$$
$$= Tr(s_t, a_t | s_{t-1}, a_{t-1}) \quad (2)$$

P here is the probability of transition.
Considering a decision process Markovian or not actually depends on the modelisation of action and states. For exemple, in physics knowing the entire state of the world at time t, would allow you to infer correctly the next state at time t+1 given an action. However, if the information given in the state is not enough to allow you to infer the next state, then it is non-Markovian.
Quick example : The experience of infering the position of a ball after dropping it :

- Markovian if you know the weight, the actual speed and position of the ball
- Non-Markovian if you only have access to the position of the ball

Non-Markovian : cf. Above

6. **Observable**: The agent can always perfectly observe the state [Chess, Go : The agent can always have a look at the full board]
   Partially Observ.: The agent cannot have access to some information on the state [Poker: Other player's cards, Real-time strategy games (Google Deepmind & Starcraft)]