### **MDP là gì?**

MDP là mô hình toán học tiêu chuẩn dùng để mô tả bài toán RL.

### **Cấu trúc MDP:**

* **S**: Tập các trạng thái (s ∈ S)
* **A**: Tập các hành động (a ∈ A)
* **P(s' | s, a)**: Xác suất chuyển trạng thái từ s sang s' khi thực hiện hành động a
* **R(s, a)**: Hàm phần thưởng, là phần thưởng kỳ vọng khi thực hiện hành động a tại trạng thái s
* **γ (gamma)**: Discount factor ∈ [0, 1) - đo lường tầm quan trọng của phần thưởng tương lai.

### **Tính chất Markov:**

* Mọi quyết định và kết quả chỉ phụ thuộc vào trạng thái hiện tại, không cần biết lịch sử trước đó:  
  
* 