



UNIVERSIDADE DA CORUÑA



# **Tema 7:** Aprendizaje por refuerzo

FUNDAMENTOS DE ROBÓTICA INTELIGENTE  
Grado en Inteligencia Artificial - 3<sup>er</sup> curso

# Métodos de aprendizaje

- Por refuerzo.
  - Un agente realiza acciones en un entorno y recibe recompensa por las correctas (el objetivo es maximizar dicha recompensa acumulada).
  - Ej. algoritmos: Q-Learning, DQN, PPO...
  - Ej. en robótica: navegación, manipulación, control en general...

- Agente. Entidad que toma decisiones y aprende.
- Entorno. Mundo con el que interactúa el agente.
- Estado ( $s$ ). Representación de una situación concreta en la que puede encontrarse el agente en el entorno.
- Acción ( $a$ ). Actuación consecuencia de una decisión del agente.

- Política ( $\pi$ ). Estrategia de actuación del agente ( $s \rightarrow a$ ). Puede incorporar elementos estocásticos.
- Función de recompensa ( $R$ ). Vincula cada estado a un valor que indica cuán deseable es ese estado.
- Función de valor ( $V$ ). Recompensa acumulada esperada a partir de un estado y de la aplicación de una política dados.
- Modelo de entorno (opcional). Se utiliza para considerar posibles situaciones futuras (estados) => planificar.
- *model-free vs. model-based*

# Aprendizaje

1. El agente observa el estado actual del entorno.
2. El agente selecciona una acción basándose en el estado y la política actual.
3. El entorno transita a un nuevo estado como resultado de la acción.
4. El agente recibe una recompensa asociada con esta transición.
5. El agente actualiza su política o función de valor utilizando la información obtenida.
6. Vuelta a 1.

