

# **Desarrollo de una exploración de los algoritmos**

**Carlos Andres Suarez Torres**

**Saira Sharid Sanabria Muñoz**

Universidad Santo Tomás

Facultad de Ingeniería Electrónica

Presentado a: DIEGO ALEJANDRO BARRAGAN VARGAS

Bogotá, Colombia

21 de noviembre de 2025



# 1. Desarrollar una exploración de los algoritmos más utilizados actualmente en la academia e industria y exponer sus características principales.

## 1.1. Algoritmos de Aprendizaje por Refuerzo Más Utilizados

Actualmente, en academia e industria, los algoritmos de aprendizaje por refuerzo más utilizados son:

### 1.1.1. Q-learning

- **Tipo:** Algoritmo basado en valor
- **Descripción:** Aprende una función Q para estimar la calidad de las acciones en cada estado
- **Ventaja:** Simple, eficiente para entornos discretos
- **Uso:** Tareas con espacios pequeños/medianos de estados y acciones

### 1.1.2. SARSA (Estado-Acción-Recompensa-Estado-Acción)

- **Tipo:** Basado en valor
- **Descripción:** Similar al Q-learning, pero actualiza su política con la acción tomada realmente, lo que lo hace más conservador
- **Uso:** Situaciones donde la ejecución óptima debe ser segura y exploratoria

### 1.1.3. Deep Q-Networks (DQN)

- **Tipo:** Basado en valor
- **Descripción:** Variante de Q-learning que utiliza redes neuronales profundas para manejar espacios de estados grandes o continuos
- **Uso:** Popular en juegos, robótica, y tareas complejas que requieren representación flexible



#### 1.1.4. REINFORCE

- **Tipo:** Algoritmo de política
- **Descripción:** Basado en gradiente de política, que aprende directamente la política para maximizar la recompensa esperada
- **Uso:** Tareas con grandes espacios de acción y políticas estocásticas

#### 1.1.5. Policy Gradient Methods (PPO, TRPO, A2C, A3C)

- **Tipo:** Métodos de política
- **Descripción:** Algoritmos avanzados basados en gradiente que optimizan políticas estocásticas y son más estables y eficientes que REINFORCE
- **Uso:** PPO (Proximal Policy Optimization) es uno de los más populares para robótica, juegos y simulaciones realistas

#### 1.1.6. DDPG (Deep Deterministic Policy Gradient)

- **Tipo:** Actor-crítico
- **Descripción:** Método actor-crítico para entornos continuos con políticas determinísticas
- **Uso:** Muy usado en control y robótica donde las acciones son continuas



## 1.2. Tabla Comparativa de Algoritmos de Aprendizaje por Refuerzo

Cuadro 1: Características principales de los algoritmos de aprendizaje por refuerzo

Algoritmo	Tipo	Características clave	Uso principal
Q-learning	Basado en valor	Tabla Q, aprendizaje por diferencia temporal, sin modelo del entorno	Entornos discretos y pequeños
SARSA	Basado en valor	Política on-policy, actualiza con acción real	Situaciones seguras y exploración
DQN	Basado en valor	Q-learning + redes neuronales profundas	Estados complejos y grandes
REINFORCE	Política	Gradiente de política, aprendizaje directo	Espacios grandes y estocásticos
PPO, TRPO, A2C, A3C	Política	Métodos avanzados, estabilidad y eficiencia	Control complejo, juegos, robótica
DDPG	Actor-crítico	Para acciones continuas, política determinística	Control, robótica con espacios continuos

## 1.3. Aplicaciones y Casos de Uso en la Industria

Estos algoritmos permiten crear agentes capaces de aprender a tomar decisiones óptimas en problemas variados de la industria y la investigación, desde videojuegos y robótica hasta finanzas y optimización de procesos [1].



### 1.3.1. Aplicaciones en la Industria

- **Videojuegos y Entretenimiento:** Los algoritmos como DQN y PPO han sido fundamentales para desarrollar agentes que superan a humanos en juegos complejos como Go, Dota 2 y StarCraft II.
- **Robótica y Automatización:** DDPG y otros algoritmos de política se utilizan para el control preciso de robots en tareas de manipulación, navegación y ensamblaje.
- **Finanzas y Trading:** Q-learning y sus variantes se aplican en sistemas de trading automatizado que aprenden a maximizar ganancias en mercados financieros volátiles.
- **Logística y Cadena de Suministro:** Algoritmos de RL optimizan rutas de entrega, gestión de inventarios y planificación de producción.
- **Salud y Medicina:** Se utilizan para personalizar tratamientos y optimizar protocolos médicos basados en la respuesta del paciente [2].

### 1.3.2. Tendencias Actuales en Investigación

- **Multi-agente:** Sistemas donde múltiples agentes aprenden a colaborar o competir en entornos compartidos.
- **Meta-aprendizaje:** Algoritmos que aprenden a aprender, adaptándose rápidamente a nuevas tareas con poca experiencia.
- **RL con Restricciones:** Desarrollo de agentes que maximizan recompensas mientras cumplen restricciones de seguridad críticas.
- **Integración con LLMs:** Combinación de aprendizaje por refuerzo con modelos de lenguaje grande para tareas de razonamiento complejo.



## 1.4. Consideraciones Prácticas para la Implementación

Al implementar estos algoritmos en proyectos reales, es importante considerar:

- **Selección del Algoritmo:** La elección depende del tipo de espacio de estados y acciones, requisitos de seguridad y recursos computacionales disponibles.
- **Exploración vs Explotación:** Balancear adecuadamente la exploración de nuevas acciones con la explotación del conocimiento actual.
- **Estabilidad del Entrenamiento:** Métodos como PPO incluyen mecanismos para mantener la estabilidad durante el entrenamiento.
- **Transfer Learning:** Utilizar conocimiento pre-entrenado para acelerar el aprendizaje en nuevas tareas relacionadas [3].

Estos algoritmos representan el estado del arte en aprendizaje por refuerzo y continúan evolucionando para abordar desafíos más complejos en inteligencia artificial.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] DataCamp. “Aprendizaje por Refuerzo: Una introducción con ejemplos en Python.” Actualizado 30 jul 2024, visitado 20 de nov. de 2024. dirección: <https://www.datacamp.com/es/tutorial/reinforcement-learning-python-introduction>.
- [2] G. B. Public. “Lo que nadie te cuenta del aprendizaje por refuerzo.” Publicado 02 jun 2025, visitado 20 de nov. de 2024. dirección: <https://gottabepublic.com/blog/aprendizaje-refuerzo>.
- [3] Coursera. “Algoritmos de aprendizaje por refuerzo y casos de uso.” Actualizado 25 de abril de 2025, visitado 20 de nov. de 2024. dirección: <https://www.coursera.org/articles/reinforcement-learning-algorithms>.