

APRENDIZAJE POR REFUERZO

Giacomo Salerno – Data Science 2309





CONTENIDO

- Aprendizaje por refuerzo
- Bases e historia breve
- Diferencias entre aprendizajes
- Aplicaciones
- Conceptos básicos
- Tipos de algoritmos
- Resumen



APRENDIZAJE POR REFUERZO

¿Qué es?

Se le llama aprendizaje por refuerzo (RL por sus siglas en inglés) al proceso mediante el cual un agente aprende a tomar decisiones en base a recompensas obtenidas.

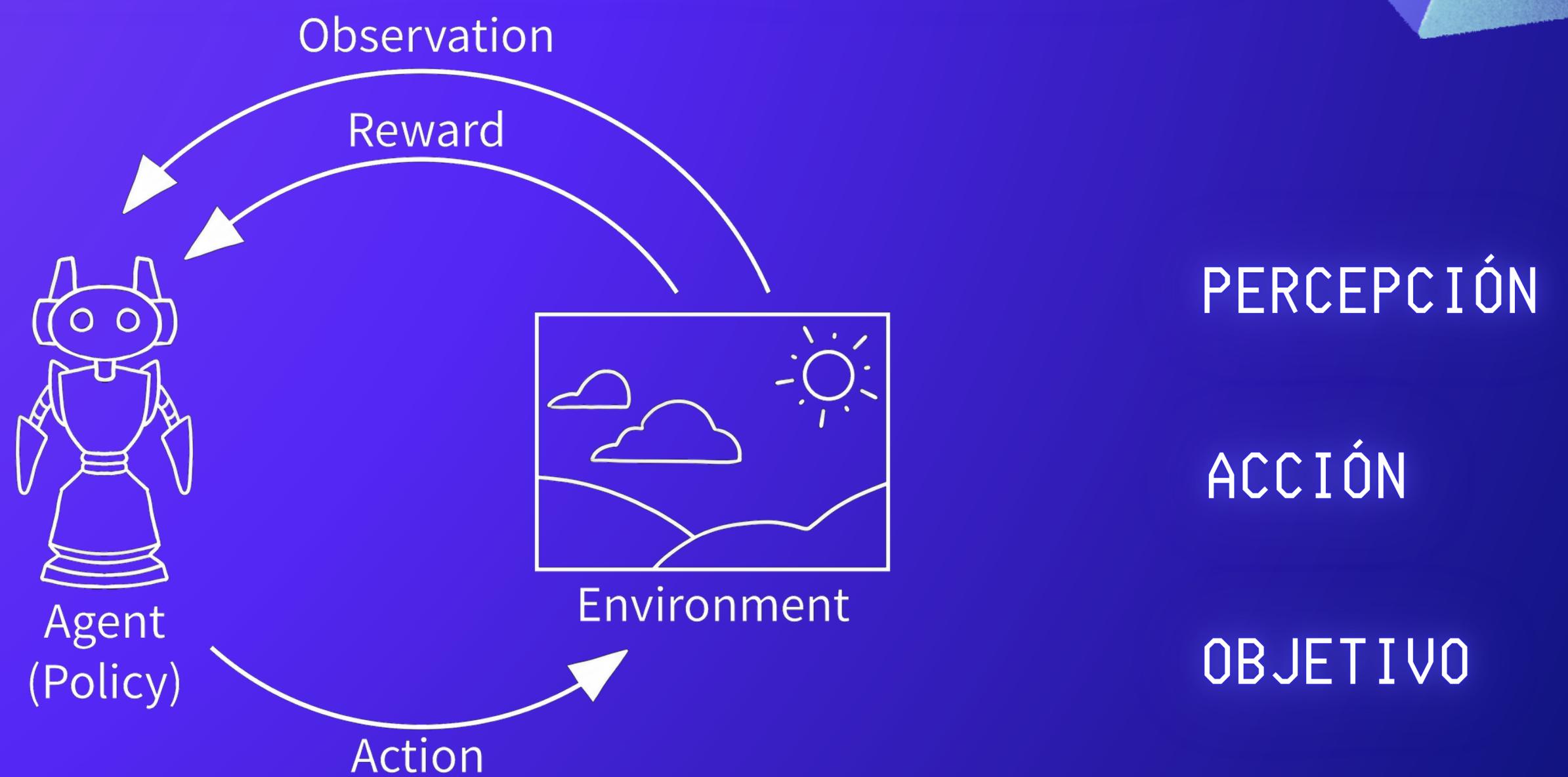
¿Qué lo caracteriza?

A diferencia de otros enfoques, el aprendizaje por refuerzo no implica recibir instrucciones directas sobre qué hacer. En su lugar, el agente explora y descubre qué acciones generan la mayor recompensa a través de la experiencia.

ENSAYO Y ERROR + RECOMPENSA



APRENDIZAJE POR REFUERZO





¿CÓMO SURGE?

¿CÓMO SE APRENDE?

Aprendemos principalmente interactuando con nuestro entorno, una idea fundamental en teorías de aprendizaje e inteligencia. Esta interacción directa nos proporciona valiosa información sobre causa y efecto, las consecuencias de nuestras acciones y cómo alcanzar metas.

1979

Empieza a ganar popularidad gracias a Sutton y Barto, cuando intentan crear un programa “hedonista” que busca maximizar recompensas, basado en la teoría del condicionamiento operante.



DIFERENCIAS ENTRE APRENDIZAJES

SUPERVISADO



DATOS ETIQUETADOS

El objetivo es extrapolar o generalizar predicciones correctamente en situaciones que no están presentes en el conjunto de entrenamiento.

En problemas interactivos, es difícil obtener ejemplos prácticos y representativos del comportamiento deseado.

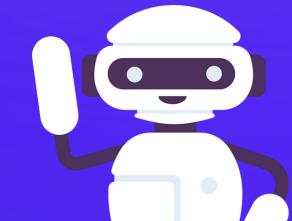
NO SUPERVISADO

DATOS NO ETIQUETADOS

El objetivo es encontrar estructuras o patrones ocultos dentro de los datos, pero no hay una "meta" concreta.

Si bien, en territorios desconocidos sería algo beneficioso, no nos interesa encontrar patrones sino explorar el entorno en profundidad

POR REFUERZO





APLICACIONES

ASPIRADORES

CHATBOTS

EXPLORADORES

SISTEMAS DE
RECOMENDACIÓN

CONDUCCIÓN
AUTOMATIZADA

BOTS EN
VIDEOJUEGOS

PSICOLOGÍA Y
NEUROCIENCIA



CONCEPTOS BÁSICOS

AGENTE

Es el “aprendiz”; ajusta su comportamiento a través de la experiencia para obtener recompensas y optimizar resultados a lo largo del tiempo.

ENTORNO

Es el contexto en el que el agente toma decisiones y realiza acciones. El entorno presenta estados, posibles acciones e interactúa en forma de recompensas, influyendo directamente sobre el agente, midiendo el tiempo en episodios.

POLÍTICA

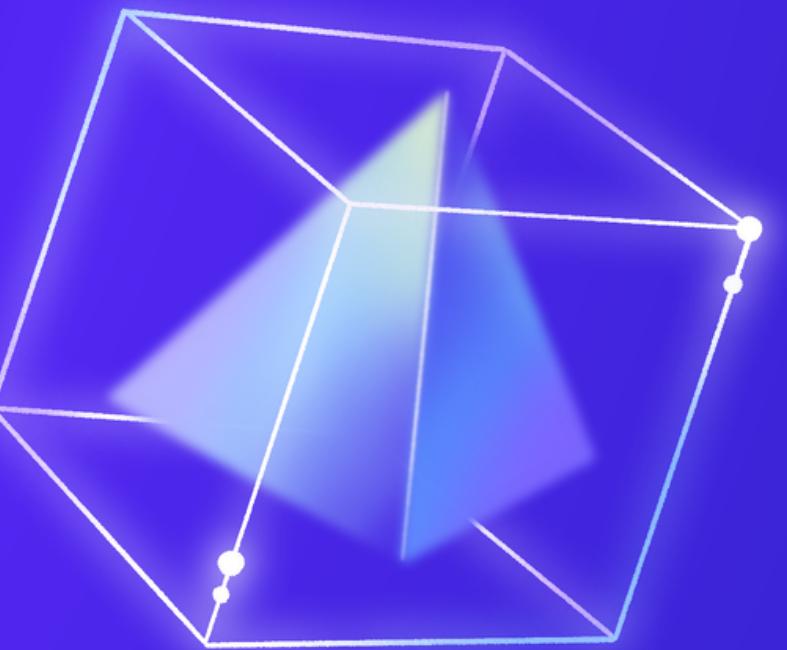
Es lo que define el comportamiento del agente: puede ser tan simple como una función o una tabla de referencia. Su naturaleza varía, pudiendo ser determinística o probabilística.



CONCEPTOS BÁSICOS

ESTADO

Es la situación actual del entorno en el que se encuentra el agente. La interacción con distintos estados es clave para maximizar el aprendizaje.



RECOMPENSA

Es la meta del aprendizaje por refuerzo, pues el único objetivo del agente es maximizar el total de recompensas a largo plazo, lo que también ayuda a definir qué es bueno o malo durante el aprendizaje en algo parecido al placer o el dolor.

FUNCIÓN DE VALOR

Es lo que especifica qué es más beneficioso a largo plazo para el agente, indicándonos su nivel de satisfacción o insatisfacción en un estado determinado.



CONCEPTOS BÁSICOS PARÁMETROS DE APRENDIZAJE

0 - 1

ALPHA

Representa la tasa de aprendizaje. Es un factor que ajusta la magnitud con la que se actualizan los valores de la función Q durante el proceso de entrenamiento.

A mayor alpha, más "confianza".

A menor alpha, más "desconfianza".

GAMMA

Representa el factor de descuento. Ajusta la importancia de las recompensas futuras, influyendo en la toma de decisiones del agente al equilibrar las recompensas inmediatas y futuras

A mayor gamma, más se conserva el valor.

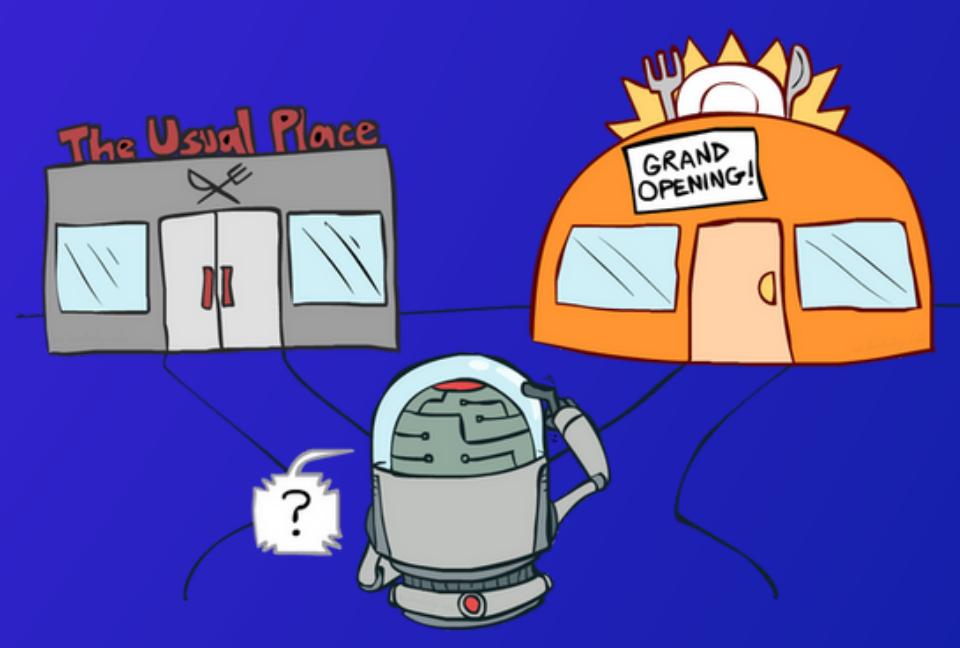
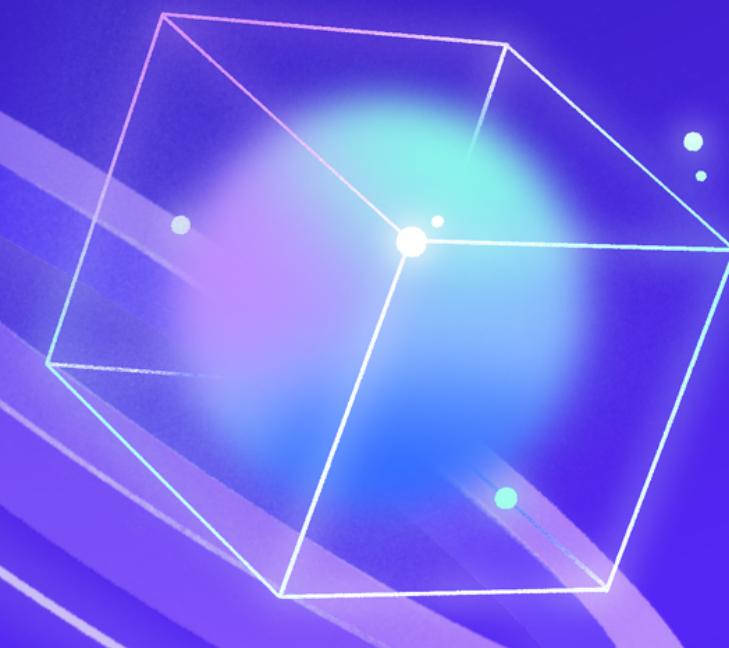
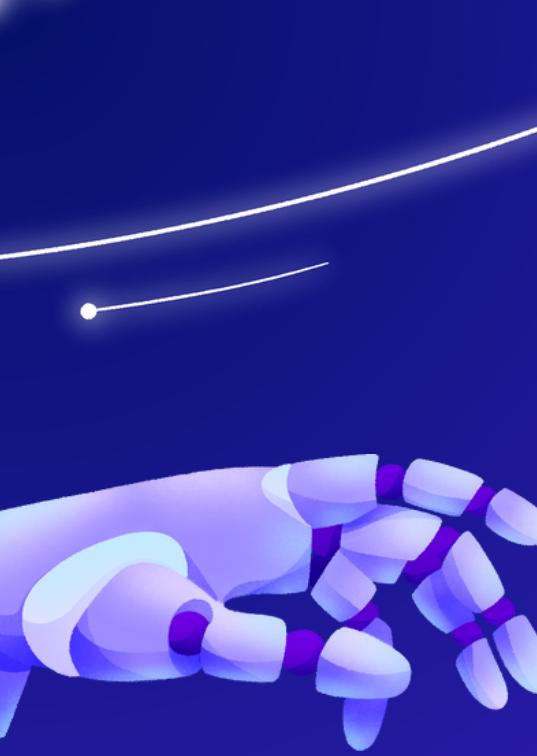
A menor gamma, más se descuenta el valor.

EPSILON

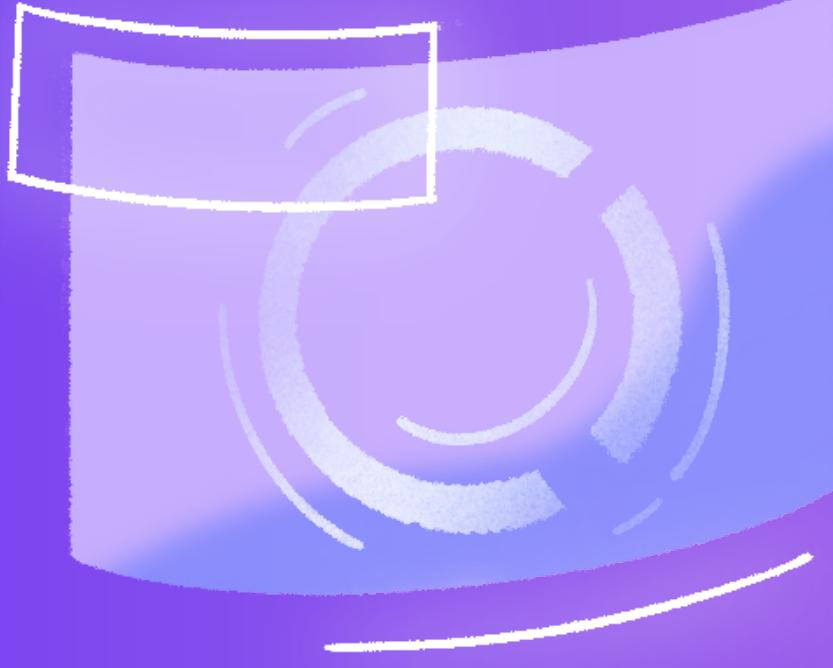
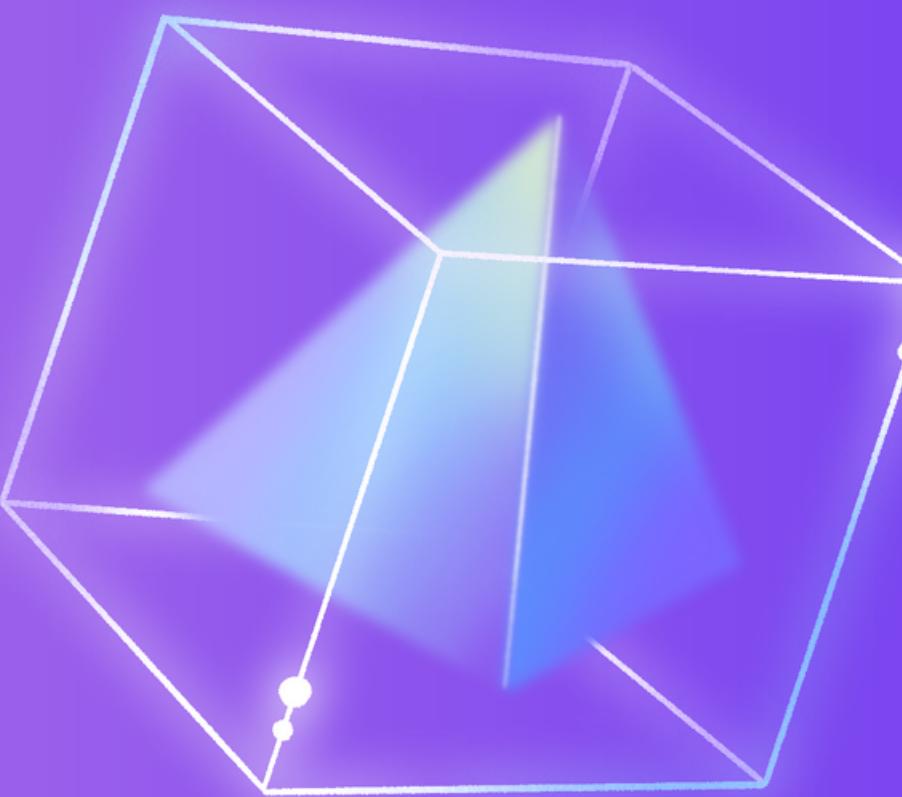
Representa la tasa de exploración. controla la probabilidad de que un agente elija una acción aleatoria en lugar de seguir su política actual.

A mayor epsilon, mayor exploración.

A menor epsilon, mayor explotación.



EXPLORACIÓN VS EXPLOTACIÓN



$$\epsilon = 0$$

Si nuestro epsilon es nulo, el agente se centra en explotar sus conocimientos....
¿pero cuáles?



$$\epsilon = 1$$

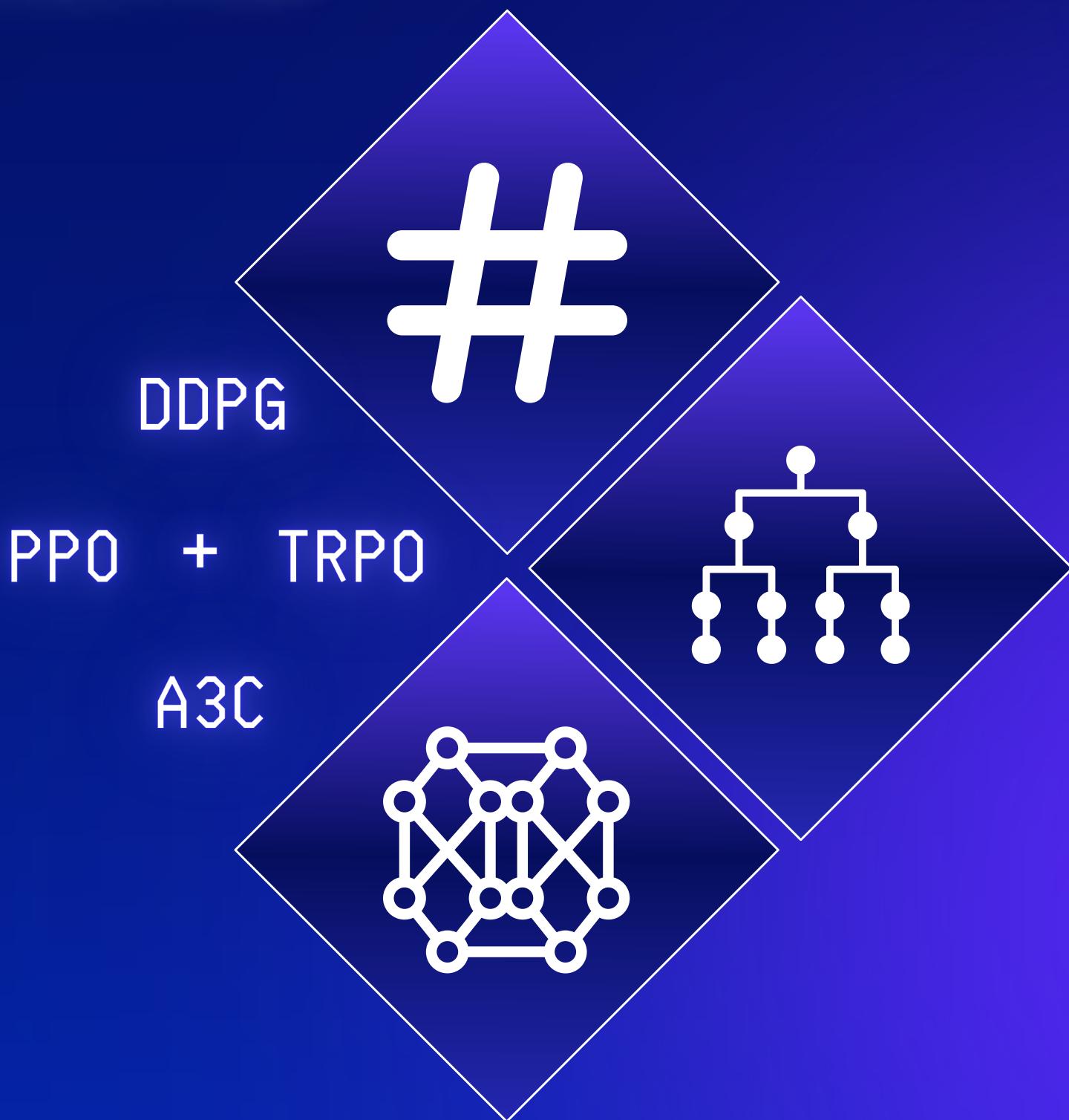
Si nuestro epsilon es total, el agente se centra en explorar el entorno sin utilizar sus conocimientos... si es que los tiene.



$$\epsilon = 0.5 ?$$

¡No!
La respuesta a “balancear” epsilon no es simplemente usar la mitad, esto en realidad haría que el agente tome decisiones completamente aleatorias entre lo que puede aprender y lo que ya sabe.

TIPOS DE ALGORITMOS



Q-LEARNING

Permite a un agente aprender a tomar decisiones óptimas mediante una Tabla Q, que evalúa la calidad de las acciones en distintos estados del entorno.

SARSA

Se enfoca en mejorar las políticas de toma de decisiones del agente evaluando y ajustando valores Q en función de las acciones tomadas (**no óptimas**) y las recompensas obtenidas durante las interacciones con el entorno

La diferencia es que se adapta de manera gradual, considerando las acciones reales para evitar riesgos innecesarios y aprender de manera más conservadora.

DEEP Q-NETWORK

Utiliza redes neuronales para aproximar la función Q, mejorando la toma de decisiones del agente en entornos complejos.



RESUMEN

- El aprendizaje por refuerzo se basa básicamente en ensayo y error.
- Es un tipo de aprendizaje diferente del supervisado y del no-supervisado.
- Sus componentes claves son agente, entorno, recompensa y función valor.
- Es fundamental mantener un buen equilibrio entre los parámetros para garantizar el comportamiento deseado.
- Existen muchísimos tipos de algoritmos en RL, sobre todo teniendo en cuenta que está en constante desarrollo; pero Q-Learning y SARSA son excelentes ejemplos para comprender la base.

MUCHAS GRACIAS

¡VAMOS A PONERLO EN PRÁCTICA!

