

## **V0: The Pirate Ship. By: Abe & Alfredo**

- **Agente utilizado: Dyna-Q**

Dyna-Q tiene una arquitectura simple que integra las principales funciones necesarias en un agente de planificación en línea. Además, incluye todos los procesos como planificación, actuación, aprendizaje de modelos y RL directo, todos los cuales ocurren continuamente.

Para The Pirate Ship, se empleó la siguiente configuración de valores fijos para  $\text{Alpha} = 1$ ,  $\text{Gamma} = 0.95$  y  $\text{Epsilon} = 0.1$ . En el modo entrenamiento, se configuró con un total de 1500 episodios.

- **Tiempo promedio de aprendizaje**

Para lograr conseguir que el agente encuentre la solución correcta del entorno, se requiere de al menos 1500 episodios. El tiempo mínimo debería alcanzar esta cantidad, puede ser de al menos 23 o 24 minutos, aunque influyen las características de hardware del equipo en el que se corra el programa.

Como en el entorno tiene incluido 2 bombas que afectan la vida del agente, la solución del agente se verá perjudicada y tendrá hostilidad, por lo que pueden existir momentos que 1500 episodios no serán suficientes en el peor de los casos, pero es un mínimo aceptable.

- **Link del video**

<https://youtu.be/qbmOHc5kCQQ>