

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación y antecedentes . . . . .	1
1.2. Descripción del problema . . . . .	2
1.3. Objetivos . . . . .	2
1.3.1. Objetivos generales . . . . .	3
1.3.2. Objetivos específicos . . . . .	3
<b>2. Marco teórico y estado del arte</b>	<b>4</b>
2.1. Procesos de decisión de Markov . . . . .	4
2.2. Aprendizaje reforzado . . . . .	5
2.2.1. Recompensas, retorno y política . . . . .	5
2.2.2. Funciones de valor . . . . .	7
2.2.3. Algoritmos de aprendizaje reforzado . . . . .	8
2.3. Aprendizaje reforzado profundo . . . . .	11
2.3.1. Funciones de aproximación y redes neuronales profundas . . . . .	11
2.3.2. DQN: Deep Q-Networks . . . . .	12
2.3.3. DDPG: Deep Deterministic Policy Gradient . . . . .	13
2.4. Navegación robótica autónoma . . . . .	15
2.4.1. Navegación robótica basada en aprendizaje reforzado profundo . . . . .	16
2.5. Aprendizaje de máquinas interactivo . . . . .	16
2.5.1. Retroalimentación correctiva humana . . . . .	17
<b>3. Formalización del problema</b>	<b>20</b>
3.1. Descripción general . . . . .	20
3.2. Espacio de observaciones . . . . .	21
3.3. Retroalimentación humana correctiva . . . . .	22
3.4. Espacio de acciones . . . . .	23
3.5. Parametrización de la política . . . . .	23
3.6. Algoritmo híbrido propuesto: D-COACH y DDPG. . . . .	24
3.7. Función de recompensa . . . . .	27
<b>4. Entrenamiento y evaluación del agente en simulaciones</b>	<b>30</b>
4.1. Ambiente de entrenamiento . . . . .	30
4.1.1. Mapa y posiciones válidas . . . . .	30
4.1.2. Condiciones episódicas del ambiente . . . . .	31
4.2. Experimentos y Resultados . . . . .	32
<b>5. Validación de las políticas del agente en simulaciones</b>	<b>42</b>

<b>6. Conclusión y trabajo futuro</b>	<b>44</b>
<b>Bibliografía</b>	<b>46</b>