



APRENDIZAJE POR REFUERZO EN ROBÓTICA COLABORATIVA: UN ESTUDIO SOBRE LA INFLUENCIA DE PARÁMETROS HUMANOS EN EL DESARROLLO DE HABILIDADES ROBÓTICAS

Tesis doctoral presentada por Asier González-Santocildes
dentro del Programa de Doctorado en Ingeniería para la Sociedad de la
Información y Desarrollo Sostenible.

Dirigida por Dr. Juan-Ignacio Vazquez y Dr. Andoni Eguiluz



Deusto

Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

APRENDIZAJE POR REFUERZO EN ROBÓTICA COLABORATIVA: UN ESTUDIO SOBRE LA INFLUENCIA DE PARÁMETROS HUMANOS EN EL DESARROLLO DE HABILIDADES ROBÓTICAS

Tesis doctoral presentada por Asier González-Santocildes
dentro del Programa de Doctorado en Ingeniería para la Sociedad de la
Información y Desarrollo Sostenible.

Dirigida por Dr. Juan-Ignacio Vazquez y Dr. Andoni Eguiluz

El doctorando

Los directores

GONZALEZ SANTOCILDES ASIER -
79005758E Digitally signed by
GONZALEZ SANTOCILDES
ASIER - 79005758E
Date: 2026.01.20
10:27:54 +01'00'

Firmado digitalmente por
VAZQUEZ GOMEZ
JUAN IGNACIO -
30669432J
Fecha: 2026.01.20
12:07:43 +01'00'

Firmado digitalmente por
ANTONIO
EGUILUZ MORAN
Fecha:
2026.01.21
07:56:09 +01'00'

Bilbao, enero de 2026

Aprendizaje por Refuerzo en Robótica Colaborativa: Un Estudio sobre la Influencia de Parámetros Humanos en el Desarrollo de Habilidades Robóticas

Autor: Asier González-Santocildes

Directores: Dr. Juan-Ignacio Vazquez y Dr. Andoni Eguiluz

Impreso en Bilbao
Primera edición, enero 2026

*Like everyone else, I know the odds are always against us, that's just
the way the universe works. I don't want guarantees, promises,
assurances or false hopes...*

*All I've ever asked for is one chance to try. To stand on my own two
feet and give it my best shot. Succeed or fail, live or die, if I know I
tried then I can accept whatever happens.*

PETER PARKER, *The Amazing Spider-Man* (Vol. 2) #34 written by
J. Michael Straczynski, 2001

Resumen

El aprendizaje por refuerzo ha demostrado ser una técnica eficaz para entrenar agentes robóticos autónomos en tareas complejas. Sin embargo, su aplicación en contextos de interacción humano-robot sigue presentando desafíos, especialmente cuando se pretende adaptar el comportamiento robótico al bienestar humano. Esta tesis se centra en el desarrollo de un sistema de entrenamiento que incorpora variables humanas en el proceso de aprendizaje, con el objetivo de generar comportamientos robóticos adaptativos que favorezcan una interacción segura, fluida y personalizada.

La primera parte del trabajo se desarrolla en entornos simulados, diseñados específicamente para exponer al agente a distintos escenarios colaborativos. En ellos se entrenan políticas de control utilizando funciones de recompensa que penalizan la invasión del espacio personal, el contacto físico con humanos u objetos, y la ineficiencia en la ejecución de la tarea. Posteriormente se aborda una fase de entrenamiento orientada al modelado del confort humano, en la que el robot aprende a ajustar su comportamiento en función del nivel de confort estimado del usuario, generando distintas estrategias de acción adaptadas a perfiles fisiológicos y contextuales.

Con el fin de validar la hipótesis en un entorno real, se diseña un experimento con participantes humanos, donde se recoge información fisiológica (frecuencia cardíaca y nivel de actividad) y se combina con cuestionarios subjetivos para evaluar la percepción del confort en distintas condiciones de interacción. Los resultados muestran que es posible identificar perfiles de usuario con diferentes umbrales de sensibilidad, y que estos perfiles pueden ser utilizados para personalizar la respuesta del robot sin requerir calibración individual ni uso de sensores invasivos.

La tesis concluye demostrando que es posible construir sistemas robóticos adaptativos que, mediante aprendizaje por refuerzo y la integración de parámetros humanos, mejoren la experiencia de interacción en escenarios colaborativos. Esta aproximación propone una línea viable para el desarrollo de soluciones robóticas personalizadas, seguras y centradas en el usuario final, con aplicaciones potenciales en entornos industriales, asistenciales o domésticos.

Abstract

Reinforcement learning has proven to be an effective technique for training autonomous robotic agents in complex tasks. However, its application in human-robot interaction contexts still presents challenges, especially when the goal is to adapt robotic behavior to human well-being. This thesis focuses on the development of a training system that incorporates human variables into the learning process, with the aim of generating adaptive robotic behaviors that enable safe, fluent, and personalized interaction.

The first part of the work takes place in simulated environments, specifically designed to expose the agent to various collaborative scenarios. In these environments, control policies are trained using reward functions that penalize personal space invasion, physical contact with humans or objects, and inefficiency in task execution. Subsequently, an advanced training phase is introduced, focused on modeling human comfort, in which the robot learns to adjust its behavior based on the user's estimated comfort level, generating different action strategies adapted to physiological and contextual profiles.

In order to validate the hypothesis in a real-world setting, an experiment with human participants is conducted. Physiological data (heart rate and activity level) is collected and combined with subjective questionnaires to assess comfort perception under different interaction conditions. The results show that it is possible to identify user profiles with varying sensitivity thresholds, which can be used to personalize the robot's behavior without requiring individual calibration or invasive sensors.

The thesis concludes by demonstrating that it is feasible to build adaptive robotic systems that, through reinforcement learning and the integration of human parameters, enhance the interaction experience in collaborative scenarios. This approach offers a viable path for the development of personalized, safe, and user-centered robotic solutions, with potential applications in industrial, assistive, or domestic environments.

Laburpena

Errefortzu bidezko ikaskuntza teknika eraginkorra izan da agente robotiko autonomoak lan konplexuetan entrenatzeko. Hala ere, gizakiaren eta roboten arteko interakzio-testuinguruetan aplikatzeak erronkak ditu oraindik ere, bereziki jokabide robotikoa giza ongizatera egokitu nahi denean. Tesi honek ikasteko prozesuan giza aldagaiak txertatzen dituen entrenamendu-sistema bat garatzen du, elkarreragin seguru, arin eta pertsonalizatua erraztuko duten portaera robotiko moldagarriak sortzeko helburuarekin.

Lanaren lehen zatia ingurune simulatueta garatzen da, agentea hainbat lankidetzagertokitan erakusteko berariaz diseinatuta daudenak. Horietan, kontrol-politikak entrenatzen dira, espazio pertsonala inbaditzea, gizakiekin edo objektuekin kontaktu fisikoa izatea eta lana eraginkortasunik gabe egitea zigortzen duten sari-funtzioak erabiliz. Ondoren, giza erosotasuna modelatzera bideratutako entrenamendu fase bati ekiten zaio. Fase horretan, robotak bere portaera erabiltzailearen erosotasun-mailaren arabera egokitzen ikasten du, profil fisiologikoetara eta testuingurukoetara egokitutako ekintza-estrategiak sortuz.

Ingurune erreala batean hipotesia baliozkotzeko, esperimentu bat diseinatu da giza parte-hartzaileekin. Esperimentu horretan, informazio fisiologikoa jasotzen da (bihotz-maiztasuna eta jardueramaila), eta galdetegi subjektiboekin konbinatu da, erosotasunaren pertzepzioa elkarrekintza-baldintza desberdinetan ebaluatzeko. Emaitez erakusten dutenez, sentsibiltate-atalase desberdinak dituzten erabiltzaile-profilak identifika daitezke, eta profil horiek robotaren erantzuna pertsonalizatzeko erabil daitezke, banakako kalibraziorik eta sentsore inbaditzailearik behar izan gabe.

Tesiaren amaieran, sistema robotiko moldagarriak eraiki daitezkeela erakusten da, errefortzuzko ikaskuntzaren eta giza parametroen integrazioaren bidez elkarlaneko agertokietan elkarrekintza esperientzia hobetuko dutenak. Hurbilketa honek irtenbide robotiko pertsonalizatuak, seguruak eta azken erabiltzailean zentratuak garatzeko ildo bideragarri bat proposatzen du, industria-, laguntza- edo etxe-inguruneetan aplikazio potentzialekin.

Agradecimientos

A mis directores, Andoni Eguiluz e Iñaki Vázquez, por su guía, confianza y apoyo constante. No podría haber tenido mejores referentes en esta etapa.

A mi familia y amigos, gracias por estar siempre, y por seguir estándolo.

Muchas gracias a todos, por todo.

Asier González Santocildes

Bilbao, enero 2026

Tabla de contenido

Lista de figuras	xiii
Lista de tablas	xvii
Acrónimos	xix
1 Introducción	1
1.1 Contexto y motivación	2
1.2 Hipótesis, objetivos y alcance	5
1.3 Metodología	7
1.4 Contribuciones	9
1.5 Estructura de la Tesis	10
2 Marco teórico y estado del arte	11
2.1 La evolución de la robótica colaborativa	12
2.2 Interacción humano-robot	15
2.3 Aprendizaje por refuerzo	20
2.4 Aprendizaje por refuerzo aplicado a la robótica	24
2.5 Aprendizaje por refuerzo centrado en el humano	28
2.6 Representación y evaluación del confort en la interacción humano-robot	32
2.7 Resumen y oportunidades de investigación	36
3 Entrenamiento simulado de comportamientos robóticos adaptativos	39
3.1 Selección y justificación del sistema y framework de entrenamiento	40
3.2 Diseño y validación de entornos simulados básicos	45
3.3 Entrenamiento en entornos avanzados con adaptación de trayectorias	54

4	Modelado del confort en la interacción humano-robot	67
4.1	Definición y representación del confort en la interacción humano-robot	69
4.2	Modelado matemático de la función de recompensa asociada al confort	72
4.3	Configuración experimental y entrenamientos	78
4.4	Resultados y discusión	81
5	Diseño de la experimentación con participantes humanos	87
5.1	Procedimiento de la experimentación	88
5.2	Configuración del entorno del experimento	91
5.3	Selección y características de los participantes del experimento . .	95
5.4	Variables y señales del experimento	96
5.5	Consideraciones éticas de la experimentación	100
5.6	Configuración del entorno complementario en VR	101
5.7	Resumen del diseño y procedimiento del experimento	105
6	Resultados del experimento y análisis de perfiles de confort	107
6.1	Resultados iniciales y observaciones exploratorias	108
6.2	Análisis de perfiles de reactividad mediante <i>clustering</i>	112
6.3	Caracterización de los perfiles de reactividad	119
6.4	Discusión e implicaciones de los resultados del experimento . . .	120
7	Conclusiones y trabajo futuro	125
7.1	Resumen del trabajo y conclusiones	125
7.2	Balance de la hipótesis y los objetivos	127
7.3	Publicaciones asociadas a la investigación	131
7.4	Líneas futuras	133
7.5	Consideraciones finales	136
A	Repositorio Comfort-RL	139
B	Repositorio Human-Robot-Comfort-Experiment	151
C	Estructura, adquisición y preprocesado de los datos del experimento	161
D	Cuestionarios de evaluación del experimento	165
	Bibliografía	171

Lista de figuras

1.1	Niveles de interacción entre humanos y robots: desde la celda aislada hasta la colaboración simultánea en una misma tarea. . . .	3
1.2	Diagrama del flujo metodológico con iteración localizada entre las fases de entrenamiento y reajuste, principal bucle de optimización del sistema.	8
2.1	Incremento porcentual en publicaciones sobre robótica colaborativa	13
2.2	Robot colaborativo UR3e de Universal Robots en entorno de laboratorio	14
2.3	Incremento porcentual en publicaciones sobre Interacción Humano-Robot	16
2.4	Distintos biosensores y su localización en el cuerpo humano (Shu et al., 2018)	18
2.5	Incremento porcentual en publicaciones sobre Aprendizaje por Refuerzo	20
2.6	Interacción agente-entorno en el aprendizaje por refuerzo (Sutton y Barto, 2018)	21
2.7	Incremento porcentual en publicaciones de RL y Robótica	25
2.8	Ejemplo de robot cuadrúpedo de Boston Dynamics aplicado a inspección industrial (Boston Dynamics, 2025).	26
2.9	Incremento porcentual en publicaciones de RL y Robótica	29
2.10	Las cinco ondas cerebrales distintas: Delta, theta, alpha, beta, and gamma.	31
2.11	Incremento porcentual en publicaciones de representaciones del confort en la interacción humano-robot	33
2.12	Mapa conceptual de las intersecciones clave entre percepción humana, entornos virtuales, aprendizaje por refuerzo y robótica colaborativa.	37
3.1	Captura de terminal con la ejecución de ML-Agents. Se observan las versiones utilizadas en las primeras iteraciones.	44

3.2	Robot UR3e importado y modificado en Unity para el desarrollo de los entrenamientos	47
3.3	Primer entorno simulado básico en Unity utilizado para entrenar una tarea de aproximación a un objetivo.	48
3.4	Gráfico de tensorboard para el entorno simple con Curriculum Learning.	50
3.5	Representación del entorno simulado avanzado en Unity, diseñado para la ejecución de tareas con múltiples puntos objetivo y trayectorias.	51
3.6	Gráfico de tensorboard para el test de entornos avanzados.	52
3.7	Gráfico de tensorboard entorno multi punto extendido. 6 millones de pasos.	53
3.8	Entorno de entrenamiento de tarea compleja. Mesa de recogida (gris) y mesas de colocación (naranja y rosa), con el agente en una posición de reposo.	55
3.9	Entorno de entrenamiento de tarea compleja, consistente en recoger y dejar una pelota de color específico, en presencia de un dummy humano.	56
3.10	Curva de entrenamiento (fases 1 y 2) sin recompensas negativas hacia el objeto. Aprendizaje más lento y menos estable.	58
3.11	Curva de entrenamiento (fases 1 y 2) con recompensa densa en fase 1 y sparse en fase 2. Convergencia más rápida gracias al uso de penalización negativa hacia el objeto.	59
3.12	Ejemplos de trayectorias generadas en presencia del humano en distintas posiciones.	61
3.13	Curva de entrenamiento (fases 1 y 2) con movimiento del humano y aplicación de curriculum learning.	62
3.14	Vista general del escenario industrial en realidad virtual diseñado para la validación y análisis de los comportamientos aprendidos por el agente	65
4.1	Gráfica de ejemplo del argumento de la tangente hiperbólica de la función de recompensa basada en distancia geométrica para dos niveles de confort del usuario. $\theta_{thr} = 40cm$ en rojo y $\theta_{thr} = 80cm$ en azul.	74
4.2	Función completa de recompensa basada en la distancia geométrica con el eje Y (reward) completo $F_{conf}(C_l, N_c)$. $\theta_{thr} = 40cm$ en rojo y $\theta_{thr} = 80cm$ en azul	75
4.3	Trayectorias obtenidas durante el entrenamiento con $\alpha = 0$ y $\beta = 0$. El agente mantiene un comportamiento eficiente y estable.	79

4.4	Evolución de la recompensa acumulada con distintos niveles de confort. Los colores indican el nivel de confort declarado, donde el rojo (nivel 1) representa estados de desconfort, y el verde (nivel 10) una situación de interacción más cómoda, natural y confortable.	81
4.5	Distancia media mantenida entre robot y humano en función del nivel de confort.	83
4.6	Número promedio de pasos empleados en función del nivel de confort.	84
4.7	Ejemplos de trayectorias generadas bajo diferentes niveles de confort. Se aprecian recorridos más directos en niveles altos y trayectorias más amplias en niveles bajos.	84
4.8	Ejemplos de trayectorias en un escenario alternativo. Las diferencias entre niveles de confort son más sutiles, aunque coherentes con la tendencia general.	85
5.1	Esquema temporal del procedimiento del experimento.	90
5.2	Vista general del laboratorio con el robot UR3e, la posición asignada al participante y las cámaras.	92
5.3	Capturas de las tres cámaras utilizadas en el experimento, mostrando de manera simultánea el rostro y la postura completa del participante.	93
5.4	Sensor fisiológico Polar Verity Sense colocado en el brazo de un participante durante el experimento.	94
5.5	Entorno complementario en VR empleado para la validación del experimento en condiciones controladas sin necesidad de robot físico.	102
5.6	Participante humano interactuando con el entorno en VR: a la izquierda en la configuración estándar y a la derecha en una versión mejorada con interacción directa con las pelotas.	104
6.1	Media de la frecuencia cardíaca de los 50 participantes bajo las condiciones lejana (azul, $\bar{d} = 71$ cm) y cercana (rojo, $\bar{d} = 49$ cm).	109
6.2	Cambio medio de la frecuencia cardíaca respecto al baseline en las condiciones lejana (azul, $\bar{d} = 71$ cm), cercana (rojo, $\bar{d} = 49$ cm) y cercana-rápida (verde, +67% de velocidad).	110
6.3	Ejemplos de capturas frontales de dos participantes en diferentes momentos de las trayectorias, con anotaciones automáticas de emociones estimadas.	111
6.4	Método del codo. Evolución de la inercia en función del número de clústeres, con un punto de inflexión en $k=3$	114

6.5	Clústeres identificados. Proyección bidimensional mediante PCA de los participantes, coloreados según su pertenencia a cada clúster ($k = 3$). La PCA se emplea únicamente con fines ilustrativos.	115
6.6	Comparación de la variación de frecuencia cardíaca (ΔHR) entre el Clúster 1 (no reactivo) y el grupo reactivo (Clúster 2 y Clúster 3).	116
6.7	Distribución de participantes por clúster de reactividad fisiológica. El Clúster 1 (Calm & Confident) agrupa al 50 % de la muestra, mientras que los Clústeres 2 (Reactive but Tolerant) y 3 (Highly Reactive & Saturable) representan el 22 % y 28 % respectivamente.	116
6.8	Índice de movimiento ponderado medio (WMI, M) por clúster de reactividad fisiológica. El Clúster 1 presenta valores bajos, el Clúster 2 valores intermedios y el Clúster 3 los valores más elevados, reflejando una mayor implicación corporal en los perfiles reactivos.	118
6.9	Ejemplo de interfaz simplificada con tres niveles de confort (verde = alto, amarillo = medio, rojo = bajo).	122
7.1	Esquema de correspondencia entre los objetivos específicos de la tesis y los capítulos en los que se desarrolla su cumplimiento.	130
7.2	Mapa de trazabilidad entre los repositorios abiertos, las publicaciones científicas y los capítulos de la tesis.	132

Lista de tablas

2.1	Principales hiperparámetros en algoritmos de aprendizaje por re- fuerzo profundo	24
3.1	Resumen comparativo de frameworks de simulación	43
3.2	Ajustes de parámetros para el algoritmo de SAC.	49
3.3	Resumen del rendimiento obtenido en los distintos entornos si- mulados tras el entrenamiento.	53
4.1	Resumen de variables consideradas para la función de recompen- sa asociada al confort.	71
4.2	Parámetros y símbolos usados en el modelado del confort.	77
5.1	Resumen simplificado de variables.	98
5.2	Síntesis estructurada de señales, captura, preprocesado y métri- cas derivadas del experimento.	99
6.1	Centroides de los clústeres para ΔHR (bpm)	115

Acrónimos

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CAD	Computer-Aided Design
CNN	Convolutional Neural Network
CPU	Central Processing Unit
EEG	Electroencephalography
GPU	Graphics Processing Unit
HCI	Human–Computer Interaction
HR	Heart Rate
HRC	Human–Robot Collaboration
HRI	Human–Robot Interaction
HRV	Heart Rate Variability
IK	Inverse Kinematics
JCR	Journal Citation Reports
k-NN	k -Nearest Neighbors
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron

PCA	Principal Component Analysis
PID	Proportional–Integral–Derivative
RL	Reinforcement Learning
ROS	Robot Operating System
SAC	Soft Actor–Critic
TD	Temporal Difference
UR	Universal Robots
URDF	Unified Robot Description Format
VR	Virtual Reality
WMI	Weighted Movement Index

If I give up now, I'll regret it!

MONKEY D. LUFFY, *One Piece*, Episode
442 (Eiichirō Oda)

Capítulo

1

Introducción

La creciente incorporación de robots en entornos laborales ha transformado la manera en que se abordan tareas industriales, logísticas y de asistencia. Esta evolución ha dado lugar al desarrollo de los llamados robots colaborativos, o *cobots*, cuya principal característica es su capacidad para trabajar de forma segura y eficiente junto a seres humanos compartiendo el mismo espacio de trabajo. A diferencia de los robots tradicionales, que operan en entornos aislados o acotados, los cobots están diseñados para detener su movimiento ante un contacto inesperado, garantizando así un entorno colaborativo seguro (Villani et al., 2018).

Esta interacción entre humanos y robots plantea nuevos retos que van más allá de la eficiencia o precisión en la ejecución de tareas. Aspectos como la seguridad, la percepción de confort o la aceptación por parte de los usuarios adquieren un papel fundamental en el diseño de estos sistemas. En particular, la proximidad física, los movimientos repentinos o impredecibles del robot y la falta de adaptación del robot a las preferencias humanas pueden generar rechazo, estrés o incomodidad, reduciendo la eficacia global de la colaboración (Sheridan, 2016).

En este contexto, surge la necesidad de proporcionar a los cobots de capacidades de adaptación que tengan en cuenta no solo el entorno físico, sino también variables humanas como la respuesta fisiológica o la percepción subjetiva de confort. Para abordar esta necesidad, el aprendizaje por refuerzo (*Reinforcement Learning*, RL) (Sutton y Barto, 2018) ofrece un marco metodológico interesante que permite entrenar agentes autónomos a través de la experiencia,

maximizando una función de recompensa definida en función del comportamiento deseado. El aprendizaje por refuerzo, además, permite generar comportamientos adaptativos que evolucionan en función del entorno y de las señales observadas, consiguiendo que el robot ajuste su comportamiento en tiempo real para cumplir la tarea asignada (Mehri Shervedani et al., 2025).

La presente tesis explora cómo integrar medidas centradas en el usuario en el proceso de aprendizaje por refuerzo para generar comportamientos robóticos adaptativos orientados al confort humano. El desarrollo de la investigación se resume en una primera fase de diseño y simulación de entornos colaborativos. Se entrenan políticas que tienen en cuenta tanto criterios de eficiencia y resolución de las tareas como parámetros de interacción segura y confortable. En una segunda fase, se valida el enfoque en un experimento con participantes humanos, recogiendo tanto señales fisiológicas como percepciones subjetivas para evaluar la calidad de la interacción (Iturrate et al., 2010; Kar et al., 2022).

El objetivo final de este trabajo es demostrar que es viable desarrollar sistemas robóticos que no solo ejecuten tareas con eficacia y sean robustos ante perturbaciones, sino que también se adapten a las necesidades individuales de los usuarios, promoviendo interacciones más seguras, eficientes y humanas. Esta línea de investigación abre nuevas posibilidades para el diseño de tecnologías centradas en las personas en sectores como la industria, la salud o la asistencia personal.

El resto de este capítulo se estructura del siguiente modo: la sección 1.1 expone el contexto y la motivación de esta investigación. A continuación, la sección 1.2 formula la hipótesis, los objetivos derivados y el alcance del trabajo. La sección 1.3 describe la metodología de investigación seguida para alcanzar dichos objetivos, y la sección 1.4 resume las contribuciones científicas y técnicas de esta tesis. El capítulo finaliza con un esquema general del documento en la sección 1.5.

1.1 Contexto y motivación

A medida que los robots colaborativos se integran en procesos productivos reales, resulta fundamental analizar cómo varía la interacción entre humanos y robots en función del grado de cooperación establecido. No todas las configuraciones colaborativas implican el mismo nivel de proximidad, sincronización o dependencia mutua, y estas diferencias condicionan tanto el diseño de los sistemas como los riesgos y oportunidades asociados. Comprender esta gradación permite identificar en qué contextos la interacción adquiere una relevancia

crítica y, por tanto, dónde resulta más urgente incorporar criterios centrados en el usuario. La Figura 1.1 ilustra estos escenarios colaborativos. El modelo Celda representa una configuración industrial tradicional con compartimentos separados para humanos y robots. El primer modelo colaborativo, denominado Coexistencia, contempla humanos y robots sin jaulas trabajando uno al lado del otro, pero en zonas diferenciadas. El modelo de interacción sincronizada implica que humanos y robots comparten el mismo espacio de trabajo, aunque no de forma simultánea, garantizando así flujos de trabajo coordinados. En el modelo de Cooperación, ambos realizan tareas al mismo tiempo en un mismo espacio, pero sobre productos o componentes distintos. Finalmente, el modelo de Colaboración ejemplifica el nivel más alto de integración, en el que humanos y robots trabajan conjuntamente sobre el mismo producto o componente (Gonzalez-Santocildes et al., 2024a; Bauer et al., 2016).

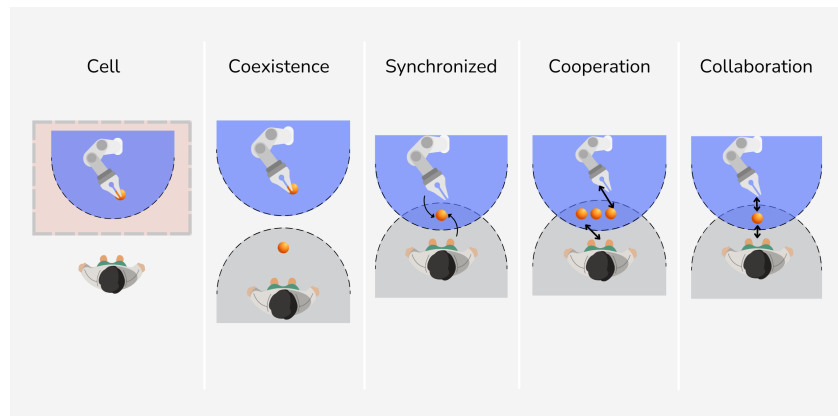


Figura 1.1: Niveles de interacción entre humanos y robots: desde la celda aislada hasta la colaboración simultánea en una misma tarea.

Es importante destacar que en los dos últimos niveles, los de Cooperación y Colaboración, la calidad de la interacción humano-robot adquiere una relevancia crítica. En estos escenarios, la cercanía física y la simultaneidad de acciones requieren no solo mecanismos de seguridad efectivos, sino también una relación fluida y adaptativa entre el operario y el sistema robótico. Una buena interacción no solo contribuye a evitar incidentes o errores en la ejecución de las tareas, sino que resulta fundamental para preservar el bienestar del trabajador, reducir la carga cognitiva y promover un entorno de trabajo más confortable, eficiente y sostenible.

Esta evolución hacia un entorno completamente colaborativo entre humanos y robots, ha motivado la búsqueda de soluciones que permitan a los robots colaborativos adaptarse dinámicamente a situaciones variables. El aprendizaje

por refuerzo ha sido una de las técnicas preferidas para dotar a los agentes robóticos de comportamientos adaptativos. Ya que a través de la interacción con su entorno y maximizar la función de recompensa, los agentes pueden descubrir estrategias eficaces para completar tareas, incluso enfrentando situaciones no previstas durante el entrenamiento (Sutton y Barto, 2018).

El aprendizaje por refuerzo ha sido ampliamente explorado en aplicaciones robóticas como la manipulación, la navegación o el ensamblaje (Kober et al., 2013; Kormushev et al., 2013). No obstante, la mayoría de estos trabajos priorizan objetivos de rendimiento como la precisión o la velocidad de ejecución, relegando a un segundo plano factores humanos como el confort subjetivo, la percepción de seguridad o la adaptabilidad del robot al estado emocional del usuario. Aunque se han propuesto modelos que introducen restricciones de seguridad o esquivan colisiones, estos enfoques rara vez consideran el bienestar del usuario como una variable explícita dentro del proceso de entrenamiento del robot (Brunke et al., 2022).

Este vacío es particularmente relevante en entornos laborales colaborativos donde los trabajadores pueden estar sometidos a altos niveles de estrés, fatiga o carga cognitiva. En tales situaciones, un comportamiento robótico mal ajustado, por ejemplo, movimientos inesperados o demasiado cercanos, puede generar incomodidad, rechazo o incluso comprometer la seguridad emocional del operario (Sheridan, 2016). Incorporar variables humanas en el diseño de sistemas autónomos no es solo un reto técnico, sino también una necesidad social para avanzar hacia tecnologías verdaderamente centradas en las personas.

Desde esta perspectiva, la presente tesis propone el desarrollo de un marco de entrenamiento basado en aprendizaje por refuerzo que permita incorporar variables humanas directamente en la función de recompensa. Para ello, se diseñan entornos simulados que posibilitan el entrenamiento y la evaluación de políticas robóticas adaptativas, así como experimentos con participantes humanos que validan la viabilidad de este enfoque en escenarios colaborativos reales. El objetivo es avanzar hacia sistemas robóticos que ajusten su comportamiento no solo en función de la tarea, sino también del estado del usuario, favoreciendo interacciones más seguras y sostenibles. Este trabajo busca responder a una necesidad creciente en entornos laborales donde la incapacidad de adaptar el ritmo o la dinámica de trabajo a las capacidades humanas puede derivar en sobrecarga cognitiva, disminución del rendimiento y reducción del bienestar general. Desarrollar robots con sensibilidad al estado del usuario representa un avance técnico clave para una colaboración más efectiva y centrada en las personas.

1.2 Hipótesis, objetivos y alcance

Teniendo en cuenta el estado actual de la colaboración humano-robot y los avances en tecnologías aplicadas, la hipótesis de la tesis es:

Hipótesis. *La incorporación de funciones de recompensa centradas en el confort humano dentro de entornos de aprendizaje por refuerzo permite desarrollar comportamientos robóticos adaptativos capaces de mejorar la calidad de la interacción colaborativa. Estos comportamientos no solo reducen el estrés percibido por el usuario, sino que también incrementan su confianza y sensación de seguridad, incluso en contextos dinámicos y no previstos. Esta adaptación personalizada puede lograrse mediante el uso de señales fisiológicas y modelos previamente entrenados, sin necesidad de calibraciones extensas ni sensores invasivos durante la fase de despliegue.*

Con el fin de validar esta hipótesis, se plantea el siguiente objetivo principal de investigación:

Objetivo principal. *Diseñar, implementar y validar un framework de entrenamiento robótico basado en aprendizaje por refuerzo que integre parámetros humanos, en particular métricas de confort fisiológico y subjetivo, con el fin de generar comportamientos personalizados y adaptativos en entornos colaborativos. Este sistema será entrenado y evaluado en simulación y experimentado posteriormente con usuarios reales, explorando la viabilidad de modelos de confort transferibles, escalables y fácilmente desplegables en escenarios reales.*

Este objetivo general se descompone en los siguientes objetivos específicos y cuantificables:

1. Diseñar e implementar entornos de simulación progresivos para distintos entrenamientos, incluyendo tanto escenarios simples sin colaboración humano-robot como escenarios con una integración completa.
2. Diseñar funciones de recompensa que integren el confort humano como criterio de optimización en el comportamiento del robot.
3. Entrenar agentes robóticos en entornos simulados mediante aprendizaje por refuerzo, con diferentes niveles de complejidad y sensibilidad al confort humano percibido.
4. Evaluar el impacto de introducir métricas de confort en el proceso de aprendizaje en comparación con controladores no adaptativos.
5. Implementar un experimento con usuarios humanos para validar la respuesta del sistema en condiciones reales.

6. Recolectar y analizar datos fisiológicos y subjetivos para evaluar el confort percibido durante la interacción humano-robot.
7. Identificar patrones de respuesta o perfiles de usuario en función de su reactividad ante diferentes comportamientos del robot.
8. Validar la aplicabilidad real del enfoque propuesto en términos de personalización, seguridad y mejora de la experiencia de usuario.

El trabajo presentado en esta tesis no aborda las siguientes condiciones:

1. Definición normativa o psicológica del confort: aunque se utilizan escalas para representar el confort percibido por el usuario, no se pretende establecer una definición formal, teórica ni psicológica del concepto de confort humano. El análisis se centra en respuestas ante diferentes niveles de confort, no en su fundamentación conceptual.
2. Despliegue en tiempo real en entornos industriales reales: la tesis no contempla la integración completa del controlador adaptativo en un sistema industrial en funcionamiento continuo. No se aborda la ejecución en tiempo real bajo condiciones de carga, latencia, seguridad o integración con software de planta.
3. Evaluación grupal o multiusuario: el estudio se limita a interacciones individuales entre un robot y una persona. No se incluyen experimentos con múltiples usuarios simultáneos ni se analiza la dinámica de confort en contextos colaborativos grupales.
4. Adaptación en tiempo real durante el despliegue final: el sistema no incluye mecanismos de aprendizaje *online* ni adaptación continua del comportamiento robótico una vez entrenado. Las políticas aprendidas se aplican tal como han sido entrenadas, sin re-entrenamiento durante el despliegue.
5. Generalización a otros dominios o tipos de robots: aunque se discuten implicaciones generales, los experimentos están acotados a un dominio concreto y no se pretende validar la solución en múltiples plataformas robóticas ni tareas industriales variadas.

A partir de esta base conceptual, en el siguiente apartado se detalla la metodología empleada para alcanzar los objetivos definidos.

1.3 Metodología

Con el fin de alcanzar los objetivos específicos propuestos en la sección 1.2 se ha establecido una estrategia metodológica basada en un enfoque iterativo y progresivo que se divide en las siguientes fases:

1. Revisión crítica del estado del arte en robótica colaborativa, aprendizaje por refuerzo e interacción humano-robot.
2. Diseño y desarrollo de entornos de simulación para el entrenamiento de agentes robóticos.
3. Definición y análisis comparativo de funciones de recompensa centradas en el confort humano.
4. Entrenamiento de agentes en simulación y evaluación de los comportamientos aprendidos.
5. Reajuste iterativo de entornos y parámetros de optimización a partir de la validación e interpretación de los resultados obtenidos.
6. Diseño e implementación del experimento físico con usuarios reales.
7. Recogida, preprocesamiento y análisis de datos fisiológicos y subjetivos del experimento físico.
8. Identificación de perfiles de usuario mediante técnicas de agrupamiento.
9. Evaluación comparativa entre la simulación y el entorno físico, y discusión crítica de los resultados.
10. Difusión científica de los resultados y validación externa a través de presentaciones y revisión por pares.

A modo de síntesis visual, la Figura 1.2 muestra el flujo general de la metodología seguida en esta tesis. El proceso se estructura de manera secuencial desde la revisión inicial hasta la difusión final, incorporando dos bucles iterativos localizados en las fases de entrenamiento y validación. Esta dinámica ciclo-iterativa durante el proceso de entrenamiento y validación constituye el núcleo de optimización del sistema, permitiendo mejorar progresivamente el comportamiento de los agentes antes de su validación experimental.

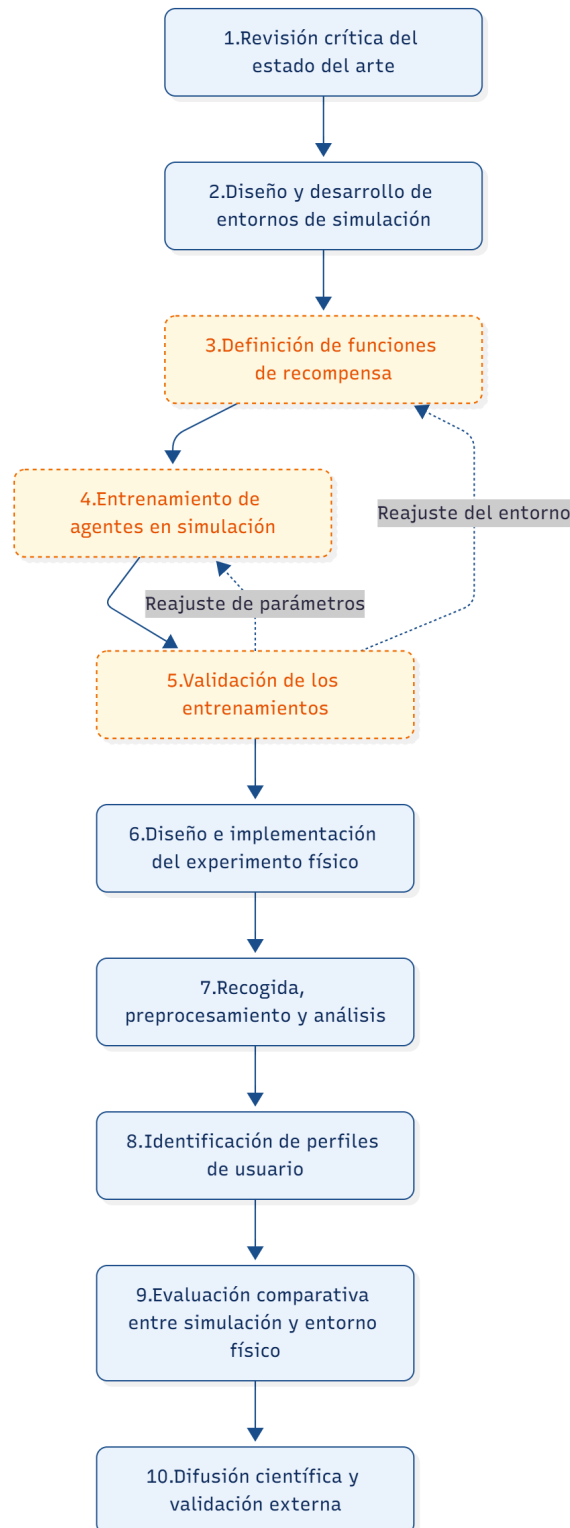


Figura 1.2: Diagrama del flujo metodológico con iteración localizada entre las fases de entrenamiento y reajuste, principal bucle de optimización del sistema.

1.4 Contribuciones

Las siguientes contribuciones científicas pueden encontrarse en esta tesis:

- **Una función de recompensa sensible al nivel de confort del usuario para la robótica colaborativa**, que integra de manera explícita las variables de distancia y velocidad de movimiento dentro del entrenamiento por aprendizaje por refuerzo. Esta contribución se introduce en el Capítulo 3 y se desarrolla en mayor detalle en los Capítulos 4 y 5.
- **Un conjunto de entornos de simulación para el aprendizaje por refuerzo basado en confort**, que abarcan desde tareas básicas de manipulación hasta escenarios avanzados de clasificación de objetos con presencia dinámica de figuras humanas. Estos entornos, descritos en el Capítulo 3 y Capítulos 4, constituyen una base para el entrenamiento y la evaluación de agentes robóticos sensibles al confort.
- **Un protocolo experimental con participantes humanos para la evaluación del confort en la interacción humano-robot**, que combina medidas fisiológicas, cuestionarios subjetivos y técnicas de visión por computador. Este protocolo se introduce en el Capítulo 5 y sus resultados se presentan en el Capítulo 6.
- **Una metodología para identificar perfiles de reactividad fisiológica**, que permite clasificar a los usuarios en función de su sensibilidad a la proximidad y la velocidad del robot. Esta metodología, descrita en el Capítulo 6, sienta las bases para la personalización del comportamiento robótico en tiempo real.

Para hacer posibles estas contribuciones científicas, también se han desarrollado las siguientes **contribuciones técnicas** disponibles en los diferentes repositorios asociados a la tesis:

- **Repositorio Comfort-RL**, disponible en Comfort-RL (Gonzalez-Santocildes, 2025a), que contiene los entornos de simulación desarrollados, las configuraciones de entrenamiento por aprendizaje por refuerzo (incluyendo distintos algoritmos y funciones de recompensa). Este repositorio contiene lo necesario para que investigadores exploren la integración del confort humano en la robótica colaborativa.
- **Repositorio Human-Robot-Comfort-Experiment**, disponible en Human Robot Comfort Experiment (Gonzalez-Santocildes, 2025b), que reúne los datos experimentales, junto con las herramientas necesarias para su procesamiento (*clustering*, análisis multimodal, cuestionarios y métricas fisiológicas). Además, incluye documentación detallada para replicar el protocolo experimental.

1.5 Estructura de la Tesis

El trabajo presentado se organiza en siete capítulos, que abarcan desde la introducción a la temática central hasta el análisis final de los resultados obtenidos y la presentación de futuras líneas de investigación.

El **Capítulo 1** introduce la motivación del trabajo, los objetivos planteados y la relevancia de abordar la interacción humano–robot desde la perspectiva del confort. Asimismo, se presenta una visión general de la metodología seguida y de las principales contribuciones científicas y técnicas de la tesis.

El **Capítulo 2** ofrece una revisión del estado del arte en robótica colaborativa, interacción humano–robot y aprendizaje por refuerzo, con especial énfasis en los enfoques centrados en el usuario y en la representación del confort. Este capítulo presenta los principios teóricos sobre las que se construye el trabajo.

El **Capítulo 3** describe los entornos de simulación desarrollados para entrenar agentes robóticos con aprendizaje por refuerzo. Se detallan tanto los escenarios básicos como las tareas avanzadas, así como los criterios de diseño empleados para reproducir condiciones realistas de interacción.

El **Capítulo 4** se centra en la formulación de modelos de entrenamiento y en la definición de funciones de recompensa sensibles al confort. Se analizan los distintos algoritmos de aprendizaje empleados y se discuten los resultados obtenidos en simulación.

El **Capítulo 5** presenta de forma diferenciada el diseño del experimento y el procedimiento seguido con participantes humanos, incluyendo la configuración del entorno, la instrumentación empleada y la recogida de datos fisiológicos y subjetivos.

El **Capítulo 6** muestra los resultados derivados del experimento con usuarios, tanto a nivel de medidas objetivas como de valoraciones subjetivas. Se introducen además metodologías de análisis de datos para identificar perfiles de reactividad fisiológica y posteriormente se discuten sus implicaciones en la personalización de la interacción.

Finalmente, el **Capítulo 7** resume las principales aportaciones de la tesis, discute sus limitaciones y plantea posibles líneas de trabajo futuro orientadas a consolidar la integración del confort humano en la robótica colaborativa.

*The more I learn, the more I realize how
much I don't know.*

ALBERT EINSTEIN

Capítulo

2

Marco teórico y estado del arte

En las últimas décadas, la robótica ha experimentado un avance significativo, dando paso a sistemas cada vez más autónomos, colaborativos, versátiles y adaptativos. Este paradigma, impulsado tanto por necesidades industriales como sociales, plantea desafíos técnicos y humanos que requieren enfoques interdisciplinarios.

Uno de los principales retos actuales en robótica colaborativa es lograr que los sistemas robóticos no solo cumplan tareas de manera eficiente, sino que además se adapten al entorno y las perturbaciones. El aprendizaje por refuerzo se ha posicionado como una estrategia prometedora al permitir a los agentes aprender los comportamientos óptimos mediante la interacción con el entorno. Sin embargo, la aplicación del aprendizaje por refuerzo en contextos físicos plantea desafíos, especialmente cuando se involucran participantes humanos en las tareas.

Este capítulo tiene como objetivo ofrecer una revisión crítica del conocimiento actual en los principales ejes que sustentan la tesis. Se aborda desde el inicio y la evolución de la robótica colaborativa, pasando por los avances en la interacción humano-robot y el aprendizaje por refuerzo aplicado a la robótica, hasta llegar a los enfoques centrados en el confort humano y su representación en sistemas de control adaptativos.

A lo largo de este capítulo, se presentan seis secciones que estructuran esta revisión. En primer lugar, en la sección 2.1, se aborda la evolución de la robótica colaborativa, desde sus orígenes industriales hasta los desarrollos actuales. A continuación, la sección 2.2 analiza los principales modelos de interacción

humano-robot. En la sección 2.4, se revisa el aprendizaje por refuerzo en el ámbito de la robótica, mientras que la sección 2.5 profundiza en los enfoques centrados en el ser humano. Posteriormente, la sección 2.6 examina cómo representar y evaluar el confort humano en contextos colaborativos. Finalmente, en la sección 2.7, se resumen las principales oportunidades de investigación que enmarcan la contribución de esta tesis.

2.1 La evolución de la robótica colaborativa

Desde los inicios de la automatización industrial, los sistemas robóticos han desempeñado un papel central en la transformación de los procesos productivos. Los primeros sistemas robóticos marcan el inicio de una etapa caracterizada por la automatización de tareas repetitivas y potencialmente peligrosas para los operarios humanos (De Santis et al., 2008).

Uno de los hitos iniciales más relevantes fue la instalación del robot *Unimate* en una planta de General Motors en 1961, considerado el primer robot industrial empleado en producción en serie (Siciliano et al., 2016).

Estos primeros robots industriales eran sistemas rígidos, de gran tamaño, pensados para operar en entornos completamente estructurados, y requerían de medidas de aislamiento físico respecto a los trabajadores debido a la poca seguridad en la interacción.

Con el paso del tiempo, aunque estos sistemas demostraron ser altamente eficaces para líneas de producción estandarizadas, también revelaron importantes limitaciones. La escasa flexibilidad, el elevado coste de reprogramación ante cambios de producción y la imposibilidad de interactuar con operarios humanos limitaban su aplicación en entornos dinámicos (Matheson et al., 2019; Bloss, 2016).

En respuesta a estas limitaciones, y en el contexto de la transformación digital de la industria conocida como *Industria 4.0*, surgieron los robots colaborativos también conocidos como *cobots* (Lasi et al., 2014). Estos nuevos sistemas están diseñados explícitamente para trabajar de forma segura junto a humanos en entornos compartidos, sin necesidad de barreras físicas. Su diseño incorpora sensores de proximidad, control de fuerzas, estructuras ligeras y estrategias de detección de colisiones que minimizan los riesgos en la interacción física (Knudsen y Kaivo-oja, 2020; Ayoubi et al., 2020).

Los robots colaborativos representan una evolución significativa en el paradigma de la automatización, al aportar una mayor flexibilidad, facilidad de

reconfiguración y una mejor adaptabilidad a tareas variadas. Gracias a su versatilidad, pueden colaborar en operaciones como carga y descarga o ensamblaje, adaptándose rápidamente a cambios en la línea de producción sin requerir una programación compleja o entornos controlados (Bibbo et al., 2025; Villagrán et al., 2020).

Grandes fabricantes como BMW o Bosch han incorporado robots colaborativos en sus líneas de montaje, donde colaboran con operarios en tareas de ensamblaje de componentes, mejorando la ergonomía y reduciendo la fatiga sin necesidad de sistemas de aislamiento físico (Wiese et al., 2019).

Este interés creciente por los robots colaborativos se refleja también en el ámbito científico, tal y como muestra la Figura 2.1. En concreto, en 2024 se registra un incremento superior al 1200 % respecto a los niveles de 2010. Además de eso la proyección de la cantidad de aportaciones científicas esperadas para 2025, basada en la tendencia observada, sugiere la continuidad de este crecimiento. Esta evidencia consolida la robótica colaborativa como una de las áreas con mayor proyección dentro de la automatización industrial.

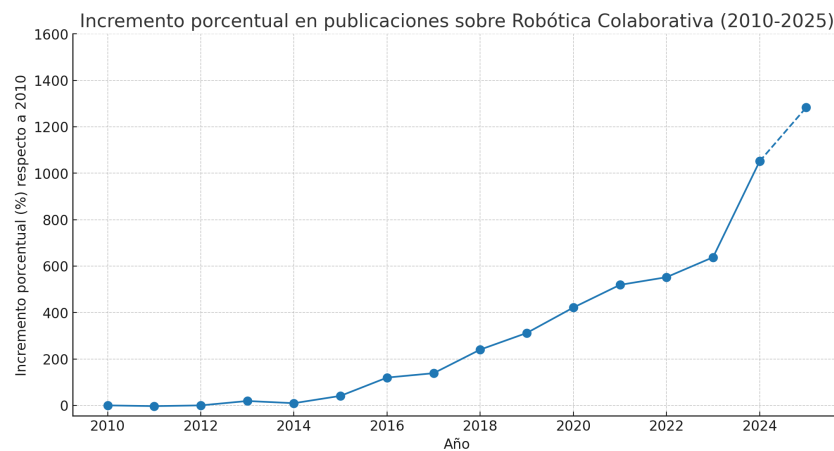


Figura 2.1: Incremento porcentual en publicaciones sobre robótica colaborativa

En la Figura 2.2 se muestra un ejemplo representativo de robot colaborativo utilizado en aplicaciones industriales y de investigación: el UR3e, desarrollado por Universal Robots. Este modelo incorpora múltiples características de seguridad y flexibilidad que permiten su uso en estaciones de trabajo compartidas con humanos.



Figura 2.2: Robot colaborativo UR3e de Universal Robots en entorno de laboratorio

A pesar de sus numerosas ventajas, la implementación de robots colaborativos no está exenta de desafíos. En particular, la seguridad en la interacción sigue siendo un aspecto crítico. La norma ISO/TS 15066 publicada por la International Organization for Standardization (ISO), establece directrices específicas para garantizar una colaboración segura, incluyendo límites cuantificables de fuerza y presión durante el contacto físico, así como requisitos sobre la forma y materiales de los robots para minimizar el daño potencial (International Organization for Standardization, 2016; Gervasi et al., 2024). Sin embargo, la mayoría de los sistemas actuales tienden a resolver esta problemática mediante estrategias reactivas, como la parada inmediata ante la detección de una colisión, lo que puede comprometer tanto la fluidez de la colaboración como la eficiencia operativa (Paulíková et al., 2021; Patalas-Maliszewska et al., 2024). Estas aproximaciones suelen centrarse únicamente en la detección y evasión del contacto, sin considerar de forma proactiva la intención humana ni adaptar el comportamiento del robot en tiempo real. Además, persisten limitaciones tecnológicas relacionadas con el procesamiento de señales y sensores, así como retos cognitivos en la interpretación del comportamiento humano y en la to-

ma de decisiones autónomas. A ello se suman aspectos normativos y éticos aún en debate, como el equilibrio entre autonomía robótica y previsibilidad para el operador humano.

En este contexto, han surgido enfoques que exploran el uso de técnicas de aprendizaje por refuerzo profundo (Kormushev et al., 2013) y señales fisiológicas como la actividad electroencefalográfica (Zhao et al., 2022), con el objetivo de dotar a los robots de una capacidad adaptativa más rica y sensible al estado del usuario. Aunque prometedores, estos enfoques se encuentran todavía en etapas iniciales de desarrollo y su integración efectiva en entornos colaborativos reales representa un reto abierto.

Esto plantea una cuestión clave en la evolución de la robótica colaborativa: ¿cómo diseñar robots que no solo respondan ante un contacto físico, sino que sean capaces de anticiparse, razonar y adaptarse dinámicamente al comportamiento humano? Esta cuestión constituye uno de los ejes centrales del presente trabajo, donde se exploran enfoques basados en aprendizaje por refuerzo y señales fisiológicas para una colaboración más inteligente y fluida.

En las siguientes secciones se revisarán las principales aproximaciones del estado del arte que abordan esta problemática, prestando atención a aquellos trabajos que integran técnicas de aprendizaje por refuerzo en contextos de interacción humano-robot. Se analizarán tanto los métodos propuestos como las limitaciones actuales, con el objetivo de identificar oportunidades concretas para avanzar hacia sistemas más adaptativos y sensibles al estado del usuario.

2.2 Interacción humano-robot

La interacción humano-robot (HRI, por sus siglas en inglés) constituye un área relevante en la robótica moderna, especialmente en aquellos contextos donde humanos y robots deben cooperar en tareas compartidas (Goodrich y Schultz, 2007; Sheridan, 2016). Por ello, se vuelve cada vez más necesario que los robots no solo sean capaces de ejecutar acciones de forma autónoma, sino también de interpretar, anticipar y adaptarse al comportamiento humano. Esta tendencia se refleja en el creciente interés científico por este área, como se muestra en la Figura 2.3. El gráfico evidencia un incremento notable a partir de 2023, con una proyección para 2025 que supera el 400% de crecimiento respecto al número de publicaciones registradas en 2010.

El estudio de la HRI proporciona los marcos conceptuales y metodológicos necesarios para diseñar interacciones seguras, eficientes y naturales entre ambos agentes, teniendo en cuenta tanto las capacidades técnicas del sistema como

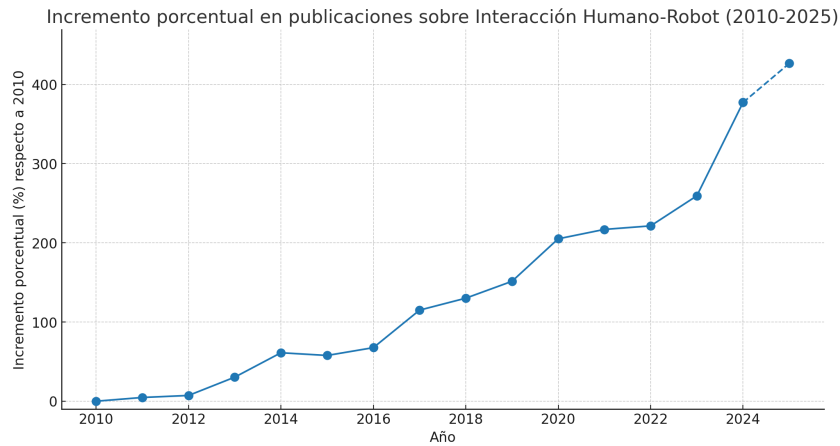


Figura 2.3: Incremento porcentual en publicaciones sobre Interacción Humano-Robot

los factores humanos implicados. A continuación, se describen las principales dimensiones que permiten caracterizar este tipo de interacción:

- **Nivel de autonomía del robot:** varía desde sistemas completamente te-
leoperados hasta robots autónomos que actúan de forma independiente,
pasando por configuraciones híbridas (Sheridan, 2011; Kim et al., 2024).
- **Tipo de control compartido:** incluye esquemas como la supervisión, el
control manual asistido, la delegación por niveles o el aprendizaje coope-
rativo, donde el humano y el robot comparten la iniciativa de forma diná-
mica (Musić y Hirche, 2017; Hoffman y Breazeal, 2004).
- **Modalidad de interacción:** puede ser explícita (por comandos, gestos o
habla) o implícita (a través de señales fisiológicas, comportamiento cor-
poral o contextos de uso) (Goodrich y Schultz, 2007).
- **Temporalidad y adaptabilidad:** la interacción puede ser puntual o con-
tinua, y más o menos adaptativa en función de la capacidad del sistema
para aprender de la experiencia o responder al estado del usuario (Dragan
et al., 2013).
- **Tipo de tarea compartida:** desde tareas físicas conjuntas (como mani-
pulación colaborativa) hasta tareas cognitivas o sociales, donde el robot
actúa como asistente, compañero o tutor (Yanco y Drury, 2002).

Estas dimensiones no solo sirven para clasificar sistemas HRI existentes, sino también para guiar el diseño de nuevas soluciones centradas en el usuario. En el caso particular de la robótica colaborativa, estas características determinan

aspectos clave como el reparto de tareas, la sincronización de movimientos o el nivel de control adaptativo requerido para una colaboración fluida y segura.

Más allá de los aspectos funcionales y técnicos de la interacción, resulta fundamental considerar los factores humanos que influyen en la calidad y efectividad. Elementos como la carga cognitiva, el confort, la confianza en el sistema o la aceptabilidad percibida condicionan en gran medida la disposición del usuario a colaborar con el robot (Hancock et al., 2011; Khavas, 2021). Por ejemplo, interacciones demasiado demandantes desde el punto de vista de la atención pueden derivar en fatiga, errores o rechazo del sistema, mientras que un diseño centrado en el usuario puede aumentar la sensación de fluidez, autonomía compartida y bienestar (explicado en detalle en sección 2.6).

No obstante, para diseñar entornos colaborativos eficaces entre humanos y robots que integren adecuadamente las dimensiones previamente descritas, resulta imprescindible complementar los canales de interacción tradicionales con tecnologías avanzadas de percepción y adaptación. Sin capacidades de percepción, la interacción se reduce a automatismos. Por ello, la percepción y la adaptabilidad son pilares fundamentales en la HRI contemporánea (Duncan et al., 2024).

En lo que respecta a la adaptación, enfoques basados en inteligencia artificial, como el aprendizaje por refuerzo (detallado en la sección 2.3) o las técnicas de lógica difusa, ofrecen mecanismos capaces de modelar la incertidumbre y ajustar dinámicamente la respuesta del sistema (Hentout et al., 2023). Por su parte, la percepción, estrechamente vinculada al concepto de HRI, requiere la incorporación de sensores fisiológicos y sistemas de percepción (Pal et al., 2021) que permitan capturar en tiempo real variables como el estrés, la carga cognitiva o el nivel de confort. Esta capacidad de interpretación del estado interno del usuario constituye un paso esencial hacia una interacción verdaderamente centrada en el ser humano.

Una dirección emergente en este ámbito de mejora de la interacción consiste en enriquecer la percepción del robot mediante señales implícitas que reflejen el estado interno del usuario. En particular, el uso de bioseñales como la actividad electroencefalográfica (EEG por sus siglas en inglés) ha mostrado potencial para estimar en tiempo real niveles de carga mental, atención o estrés (Gonzalez-Santocildes et al., 2024b). Estas señales permiten al sistema detectar variaciones que no son observables de forma directa, facilitando un tipo de interacción más sensible al contexto y al bienestar del operador (Toichoa Eyam et al., 2021). La integración de este tipo de información allana el camino hacia modelos futuros de colaboración adaptativa, donde el robot modula su comportamiento no solo

en función de la tarea, sino también de las necesidades cognitivas y emocionales del humano.

Si bien el EEG ha sido ampliamente utilizado para estimar el estado cognitivo del usuario, no constituye la única fuente de información fisiológica relevante en el diseño de sistemas adaptativos. Diversas técnicas basadas en biosensores permiten obtener una visión más amplia del estado emocional y físico del usuario, aportando señales complementarias que enriquecen la interacción humano-robot. Dispositivos como la electrooculografía (EOG), la respuesta galvánica de la piel (GSR), la frecuencia respiratoria (RSP), la electrocardiografía (ECG) o el pulso por volumen sanguíneo (BVP) permiten inferir diferentes dimensiones del estado emocional o fisiológico del usuario (Shu et al., 2018). Tal como se muestra en la Figura 2.4, estos sensores pueden situarse en diferentes partes del cuerpo y se están integrando progresivamente en el diseño de experimentos de HRI.

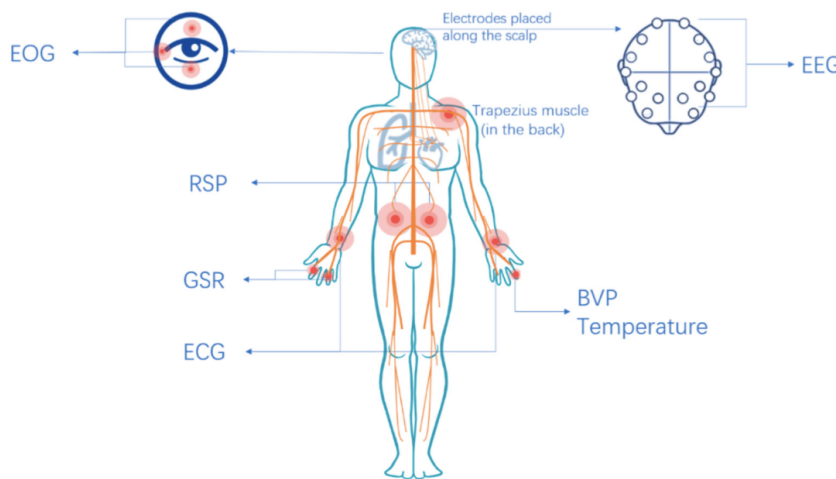


Figura 2.4: Distintos biosensores y su localización en el cuerpo humano (Shu et al., 2018)

Además de los sensores fisiológicos, técnicas como el *eye tracking* se han aplicado con éxito para inferir intención, carga visual y estrés a partir de variables como el diámetro pupilar o la densidad de fijaciones visuales (Onose et al., 2012; Onnasch et al., 2023; Mariscal et al., 2023). Este tipo de señales resulta especialmente útil en tareas colaborativas donde la atención compartida es clave para la coordinación.

Por otro lado, la visión por computador ofrece herramientas para detectar gestos, expresiones faciales o posturas, facilitando así la interpretación del comportamiento humano en tiempo real. La disciplina conocida como *Human Acti-*

vity Recognition (HAR) ha demostrado su utilidad en aplicaciones industriales, sanitarias y asistenciales, permitiendo inferencias sobre intención, fatiga o anomalías (Pérez et al., 2016; Beddiar et al., 2020; Zhu et al., 2022). No obstante, su fiabilidad puede verse comprometida por condiciones de iluminación variables o entornos complejos, y persisten desafíos vinculados al sesgo en los conjuntos de datos empleados para entrenar estos sistemas.

Complementariamente, la monitorización del ritmo cardíaco ha sido empleada como indicador de estrés o fatiga, gracias a la relación existente entre variabilidad cardíaca y estado emocional (Stamler et al., 1992; Xintarakou et al., 2022). El desarrollo de dispositivos portátiles como bandas torácicas o pulseras inteligentes ha facilitado su uso en contextos experimentales y aplicaciones en tiempo real.

Por último, los instrumentos subjetivos basados en cuestionarios siguen desempeñando un papel relevante en la evaluación del estado humano. Cuestionarios como el *Beck Depression Inventory* (BDI) o el *Perceived Stress Scale* (PSS) permiten complementar las mediciones fisiológicas con información autorreportada sobre el estado emocional y cognitivo del usuario (Jackson-Koku, 2016; Beck et al., 2011; Jumaní et al., 2022). Además, en entornos experimentales específicos, es posible diseñar cuestionarios personalizados adaptados a los objetivos de la investigación o al contexto de interacción, lo que permite capturar variables subjetivas relevantes de forma más precisa y contextualizada.

En conjunto, la combinación de sensores fisiológicos, técnicas de visión y herramientas subjetivas configura un ecosistema de medición del estado humano cada vez más sofisticado. Su integración en entornos de interacción humano-robot permite avanzar hacia sistemas capaces de responder no solo al contexto físico de la tarea, sino también al estado interno del operador, favoreciendo una interacción más empática, proactiva y efectiva (Roy et al., 2020).

Sin embargo, para que las observaciones obtenidas de los sensores se traduzcan en comportamientos robóticos realmente adaptativos, es necesario contar con mecanismos de decisión que integren dinámicamente esta información en el control del agente. En este contexto, el aprendizaje por refuerzo emerge como una técnica especialmente prometedora, ya que permite a los robots aprender políticas de acción óptimas a partir de la interacción con el entorno y la retroalimentación obtenida. En las secciones siguientes se detallan los fundamentos y aplicaciones del aprendizaje por refuerzo en robótica y los enfoques orientados a incorporar explícitamente el estado y las necesidades del usuario en el proceso de aprendizaje.

2.3 Aprendizaje por refuerzo

El aprendizaje por refuerzo constituye un marco de aprendizaje automático en el que un agente aprende a tomar decisiones mediante la interacción directa con un entorno, con el objetivo de maximizar una señal de recompensa acumulada. A diferencia de otros enfoques, como el aprendizaje supervisado, en RL el conocimiento no se proporciona en forma de ejemplos etiquetados, sino que se obtiene a través del ensayo y error. Esta característica convierte al aprendizaje por refuerzo en una herramienta especialmente adecuada para escenarios donde no se dispone de una estrategia óptima explícita.

En la Figura 2.5 se aprecia el interés creciente en los últimos años por esta técnica. Cabe destacar que la cantidad de publicaciones en el 2010 ya era bastante grande, no obstante, desde 2018 se aprecia un aumento exponencial en las aportaciones científicas, llegando a incrementos del 1600 % respecto a 2010.



Figura 2.5: Incremento porcentual en publicaciones sobre Aprendizaje por Refuerzo

En el paradigma clásico del aprendizaje por refuerzo, la interacción entre el agente y el entorno se modela como un bucle cerrado: en cada instante de tiempo t , el agente observa el estado del entorno s_t , selecciona una acción a_t según una política $\pi(a|s)$, y recibe como respuesta una recompensa r_t junto con un nuevo estado s_{t+1} . Esta secuencia de observación, acción y retroalimentación permite al agente actualizar progresivamente su política con el fin de mejorar su comportamiento futuro. La Figura 2.6, adaptada de (Sutton y Barto, 2018), ilustra este ciclo de interacción fundamental.

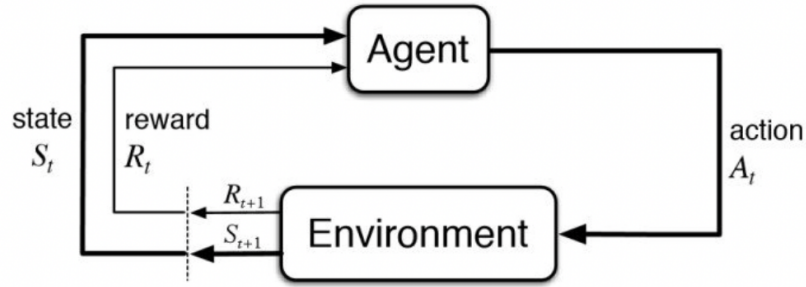


Figura 2.6: Interacción agente-entorno en el aprendizaje por refuerzo (Sutton y Barto, 2018)

Formalmente, este proceso puede representarse mediante un Proceso de Decisión de Markov (*Markov Decision Process*, MDP), definido por un conjunto de estados S , un conjunto de acciones A , una función de transición $P(s'|s, a)$, una función de recompensa $R(s, a)$ y un factor de descuento $\gamma \in [0, 1]$ que pondera la importancia relativa de las recompensas futuras. El objetivo del agente es encontrar una política óptima π^* que maximice el valor esperado de la recompensa acumulada a largo plazo, definida como $\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t]$.

Para alcanzar dicho objetivo, el RL hace uso de diversas funciones de valor. La función de valor de estado $V^\pi(s)$ estima la recompensa acumulada esperada al seguir una política π desde ese estado. La función de valor de acción, o función de valor estado acción (*state-action value function*) $Q^\pi(s, a)$, estima la recompensa esperada al ejecutar la acción a en el estado s y continuar siguiendo la política π . Estas funciones son la base de muchos algoritmos de RL, tanto de tipo *value-based* (basados en el aprendizaje de funciones de valor) como de tipo *policy-based* (centrados en el ajuste directo de la política).

Uno de los retos fundamentales en el aprendizaje por refuerzo es el equilibrio entre exploración y explotación. El agente debe decidir en cada momento si conviene ejecutar la acción que, según su conocimiento actual, proporciona la mayor recompensa esperada (explotación), o si es preferible seleccionar una acción menos conocida que podría resultar más beneficiosa a largo plazo (exploración). Una política que se base únicamente en la explotación puede conducir a óptimos locales, mientras que una política excesivamente exploratoria puede ralentizar el aprendizaje. Este compromiso es conocido como el dilema de exploración-explotación y está presente en la mayoría de algoritmos de RL (Sutton y Barto, 2018).

Diversas estrategias han sido propuestas para gestionar este equilibrio. Entre las más utilizadas se encuentran las políticas ϵ -greedy, en las que el agente selecciona la acción óptima con probabilidad $1 - \epsilon$ y una acción aleatoria con probabilidad ϵ ; los enfoques basados en softmax, que asignan una probabilidad a cada acción proporcional a su valor estimado; y los métodos que introducen regularización por entropía para fomentar la exploración en fases iniciales del aprendizaje. En algoritmos más recientes, como Proximal Policy Optimization (PPO) o Soft Actor-Critic (SAC), se integran mecanismos de exploración estocástica de forma implícita a través de técnicas de optimización de políticas con entropía máxima.

Cabe destacar que este equilibrio no solo afecta al comportamiento del agente, sino que también debe tenerse en cuenta en el diseño del entorno de entrenamiento. Una estrategia de exploración mal ajustada puede limitar la capacidad del agente para descubrir soluciones eficientes, especialmente en tareas complejas o con espacios de estado amplios. Por tanto, seleccionar o diseñar los mecanismos de exploración es un aspecto interesante del desarrollo de sistemas basados en aprendizaje por refuerzo.

El aprendizaje por refuerzo profundo (*Deep Reinforcement Learning*, DRL) surge de la combinación entre el aprendizaje por refuerzo clásico y las redes neuronales profundas, permitiendo abordar tareas con espacios de estado de alta dimensionalidad o no estructurados. Mientras que los algoritmos tradicionales de RL emplean tablas o funciones lineales para representar las políticas y funciones de valor, DRL utiliza modelos neuronales como aproximadores universales, lo que facilita la generalización en entornos complejos. Esta integración ha dado lugar a importantes avances en campos como la robótica, los videojuegos, la conducción autónoma y el control de sistemas dinámicos (Li, 2017, 2023).

Entre los algoritmos de aprendizaje por refuerzo profundo, PPO pertenece a los métodos *policy-based* y optimiza políticas estocásticas mediante un mecanismo de *clipping* que limita el cambio de política, evitando actualizaciones demasiado agresivas. Esta característica lo hace especialmente adecuado para entornos con espacios de estado continuos o de alta dimensionalidad, así como para tareas con recompensas con mucho ruido. Gracias a su equilibrio entre rendimiento y simplicidad, PPO se ha consolidado como una opción preferente en aplicaciones como el control, la robótica colaborativa o el entrenamiento de agentes en entornos virtuales complejos (Schulman et al., 2017). La flexibilidad del algoritmo contrasta con enfoques, como el de SAC o Deep Q-Networks (DQN), que presentan características específicas adaptadas a distintos tipos de

entornos y requisitos de aprendizaje (De La Fuente y Guerra, 2024).

SAC es un algoritmo *actor-critic* orientado de forma específica a entornos con espacios de acción continuos, lo que lo convierte en una de las opciones preferentes en el ámbito de la robótica. SAC introduce la maximización de la entropía de la política como parte de su objetivo, promoviendo una exploración constante que favorece la obtención de políticas más robustas y generalizables. Esta característica resulta especialmente adecuada en tareas de control donde se requiere un balance entre precisión y adaptabilidad, como la manipulación de objetos o el movimiento de brazos robóticos. Gracias a su estabilidad y capacidad de aprendizaje eficiente en entornos complejos, SAC se ha consolidado como un referente en aplicaciones robóticas modernas (Haarnoja et al., 2018).

Aunque algoritmos como PPO y SAC se han consolidado como los más utilizados en contextos con espacios de acción continuos y tareas de control, existen otras alternativas relevantes. Entre ellas, Deep Q-Networks (DQN) destaca por haber sido uno de los primeros enfoques exitosos en integrar redes neuronales con aprendizaje por refuerzo (Fan et al., 2020; Mnih et al., 2015). DQN está especialmente indicado para problemas con espacios de acción discretos y ha sido ampliamente aplicado en entornos de videojuegos y simulaciones estructuradas. No obstante, su aplicabilidad en robótica resulta más limitada, ya que muchos sistemas físicos requieren el manejo de acciones continuas, donde métodos como PPO y SAC ofrecen ventajas claras (Henderson et al., 2017; Andrychowicz et al., 2020).

La configuración de los hiperparámetros en los algoritmos de aprendizaje por refuerzo profundo tiene un papel fundamental en la eficiencia y estabilidad del proceso de aprendizaje. Factores como la tasa de aprendizaje, el tamaño de los lotes de entrenamiento o el valor del factor de descuento pueden afectar de manera significativa tanto a la velocidad de convergencia como a la calidad de las políticas obtenidas. En el caso de algoritmos como PPO o SAC, además de los parámetros comunes a otros métodos de aprendizaje automático, cobran especial relevancia ajustes específicos como el coeficiente de entropía, el parámetro de *clipping* o la actualización de las redes críticas. La selección y ajuste-fino de estos valores resulta clave para garantizar un aprendizaje estable, evitando fenómenos como la sobreexploración o la convergencia prematura a soluciones subóptimas (Henderson et al., 2017; Eimer et al., 2023). Los principales hiperparámetros utilizados en este tipo de algoritmos se resumen en la Tabla 2.1.

En definitiva, el aprendizaje por refuerzo constituye un marco versátil para el desarrollo de sistemas capaces de aprender comportamientos complejos de forma autónoma, a partir de la interacción con su entorno. La capacidad de

Tabla 2.1: Principales hiperparámetros en algoritmos de aprendizaje por refuerzo profundo

Hiperparámetro	Descripción
<i>Learning rate</i>	Controla la magnitud de las actualizaciones de la política y las redes críticas.
<i>Batch size</i>	Número de muestras utilizadas por iteración para actualizar los modelos. Afecta estabilidad y velocidad.
γ (<i>discount factor</i>)	Factor de descuento que pondera la importancia de las recompensas futuras frente a las inmediatas.
<i>Entropy coefficient</i>	Pondera el término de entropía en la función objetivo, favoreciendo la exploración.
<i>Clipping parameter</i>	En PPO, limita el cambio entre políticas consecutivas para mantener estabilidad.
<i>Target smoothing coefficient</i> (τ)	En SAC, regula la actualización suave de las redes objetivo (<i>critic networks</i>).
<i>Replay buffer size</i>	Tamaño del almacenamiento de experiencias en algoritmos off-policy, como SAC o DQN.

generar políticas adaptativas y de descubrir estrategias eficientes sin necesidad de supervisión ha dado lugar a comportamientos emergentes que, en ocasiones, superan las soluciones diseñadas manualmente. Esta característica resulta especialmente atractiva en el ámbito de la robótica, donde la diversidad de entornos y tareas plantea retos difíciles de abordar mediante enfoques tradicionales. A lo largo de las próximas secciones, se profundizará en los trabajos más recientes relacionados con la aplicación del aprendizaje por refuerzo en robótica, así como en las propuestas orientadas a incorporar factores humanos en el proceso de aprendizaje, dando lugar al aprendizaje por refuerzo centrado en el humano.

2.4 Aprendizaje por refuerzo aplicado a la robótica

En los últimos años, el aprendizaje por refuerzo se ha consolidado como una técnica de referencia dentro del ámbito de la robótica. La capacidad de los algoritmos de aprendizaje por refuerzo para generar políticas adaptativas mediante la interacción directa con el entorno resulta especialmente valiosa en escenarios robóticos, donde las soluciones tradicionales basadas en programación explícita

presentan limitaciones frente a la variabilidad y complejidad de las tareas (Tang et al., 2025).

Esta tendencia se refleja en un aumento progresivo del número de trabajos científicos y proyectos de investigación que aplican el aprendizaje por refuerzo en el ámbito de la robótica. Tal y como se observa en la Figura 2.7, entre 2010 y 2018 el crecimiento fue relativamente contenido, con incrementos acumulados inferiores al 200 % respecto a 2010. Sin embargo, a partir de 2019 se aprecia un crecimiento más acusado, alcanzando un incremento superior al 1400 % en 2024. La proyección de la cantidad de aportaciones científicas esperadas para 2025, basada en la tendencia observada, sugiere la continuidad de este crecimiento. Este patrón evidencia un interés cada vez mayor por la aplicación del aprendizaje por refuerzo en robótica, especialmente en el último lustro.

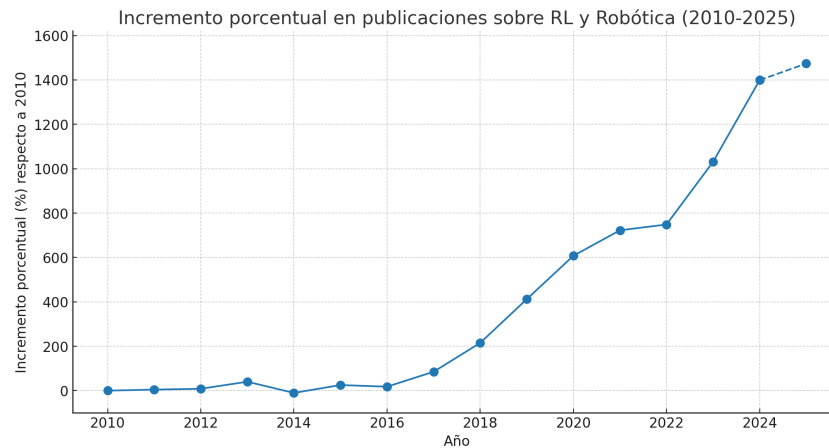


Figura 2.7: Incremento porcentual en publicaciones de RL y Robótica

A lo largo de esta sección se abordarán los hallazgos más representativos en la intersección de ambas áreas, con el objetivo de aportar una visión actualizada sobre las tendencias emergentes y los enfoques predominantes de cara al desarrollo de la presente tesis.

En el ámbito de la robótica tradicional, el aprendizaje por refuerzo ha alcanzado hitos significativos. Los avances más sólidos se han registrado en locomoción, navegación de robots terrestres y asistencia sanitaria, donde el uso de estrategias de transferencia y el diseño de recompensas densas han permitido ver progresos significativos (Zhao et al., 2020; Kumar, 2025). Estos progresos han facilitado la integración del aprendizaje por refuerzo en productos comercializados, especialmente en robots cuadrúpedos y sistemas de drones, destacando casos como el control en robots de *Boston Dynamics* (ver Figura 2.8) o el

pilotaje autónomo en carreras de drones (Kaufmann et al., 2023).

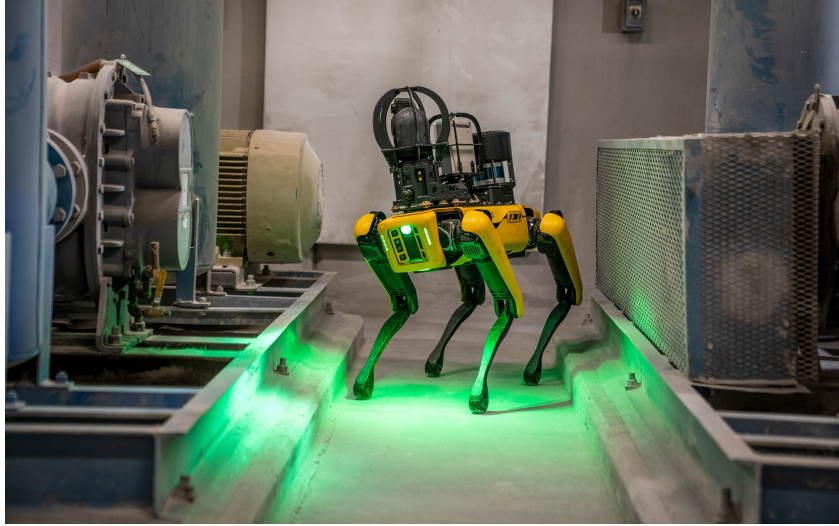


Figura 2.8: Ejemplo de robot cuadrúpedo de Boston Dynamics aplicado a inspección industrial (Boston Dynamics, 2025).

Más allá de la robótica clásica, el aprendizaje por refuerzo se ha introducido progresivamente en contextos de robótica colaborativa donde los agentes comparten espacio de trabajo. En este ámbito, los enfoques basados en aprendizaje por refuerzo multiagente han demostrado capacidad para optimizar la coordinación entre robots, especialmente en tareas de manipulación colaborativa (Tang et al., 2025).

A pesar de estos avances, la implementación del aprendizaje por refuerzo en aplicaciones robóticas industriales reales sigue siendo limitada. La mayoría de los desarrollos se encuentran en fases experimentales o en entornos controlados, con niveles de madurez tecnológica reducidos (Paduraru et al., 2021; Kormushev et al., 2013). Esta situación se debe tanto a la dificultad de transferir modelos de forma segura desde entornos simulados a contextos de producción, como a las exigencias de computación asociadas a la ejecución de modelos en tiempo real, especialmente en la toma de decisiones y planificación de trayectorias robóticas. Entre los principales retos técnicos identificados se encuentran la ineficiencia en el uso de muestras, los problemas derivados del equilibrio entre exploración y explotación, y las dificultades para garantizar la generalización y reproducibilidad de los resultados en entornos reales (Nguyen y La, 2019). En el ámbito de la exploración, las estrategias tradicionales como ϵ -greedy o la adición de ruido a las acciones presentan limitaciones importantes en entornos continuos y de alta dimensionalidad, lo que puede derivar en comportamientos

poco óptimos o inseguros. Algoritmos como PPO o SAC introducen mecanismos adicionales, como la regularización por entropía o el uso de políticas estocásticas, que permiten mejorar la estabilidad del proceso, si bien persisten desafíos relacionados con la búsqueda eficiente de nuevas soluciones sin comprometer la integridad del sistema físico.

Una cuestión crítica asociada a estos retos es la transferencia de políticas aprendidas en entornos simulados a escenarios físicos, conocida como el problema *sim-to-real* (Salvato et al., 2021). Aunque existen casos de éxito donde se han conseguido trasladar comportamientos desde simulación a robots reales, factores como las diferencias en la dinámica, la fidelidad de los sensores o el modelado de colisiones introducen discrepancias que afectan al rendimiento final. Para mitigar estas limitaciones, se han propuesto técnicas como la aleatorización de entornos de entrenamiento, la introducción de ruido sintético en el proceso de aprendizaje y el ajuste fino final en un entorno real; esto permite reducir la brecha entre simulación y realidad, aunque con un coste computacional elevado y un esfuerzo considerable en el proceso de entrenamiento. La falta de benchmarks estandarizados, además, dificulta la comparación objetiva entre algoritmos en entornos de simulación propios, y la sensibilidad a factores como las semillas aleatorias o las arquitecturas de red empleadas agrava el problema. Las investigaciones actuales se orientan precisamente a resolver estas limitaciones, explorando estrategias que permitan una integración efectiva del aprendizaje por refuerzo en entornos industriales operativos (Lou et al., 2024).

En base a las limitaciones presentadas, se han impulsado líneas de investigación centradas en la integración de sensores avanzados y arquitecturas de control híbridas, donde el aprendizaje por refuerzo se combina con técnicas supervisadas o con modelos de control tradicionales (Kober et al., 2013). Estas aproximaciones permiten abordar tareas complejas que requieren tanto precisión como adaptabilidad, como por ejemplo la manipulación de objetos deformables o el ensamblaje de piezas con tolerancias mínimas. Asimismo, han surgido propuestas orientadas a mejorar la eficiencia de entrenamiento, mediante el uso de estrategias de *curriculum learning* (Shi et al., 2025) o el aprovechamiento de experiencias previas mediante *offline reinforcement learning* (Wang et al., 2025), ampliando el rango de aplicaciones potenciales del aprendizaje por refuerzo en la robótica (Florensa et al., 2017; Levine et al., 2020).

Sin embargo, cuando el aprendizaje por refuerzo se aplica a entornos compartidos con humanos, surgen nuevas limitaciones que van más allá de los desafíos técnicos tradicionales. La mayoría de los enfoques actuales priorizan la eficiencia y, en algunos casos, la seguridad física, como evitar colisiones o respetar las zonas de seguridad. No obstante, rara vez se consideran aspectos más

sutiles pero fundamentales del estado interno del humano, como el confort, la sobrecarga cognitiva o la comprensión de la tarea (Sheridan, 2016; Giallanza et al., 2024).

En contextos colaborativos actuales, resulta insuficiente considerar al ser humano como un obstáculo o elemento externo al entorno. En lugar de ello, debe ser entendido como una parte activa y dinámica del sistema, cuyas acciones, estados y necesidades influyen directamente en el desarrollo de la tarea compartida (Nikolaidis et al., 2017). Desde esta perspectiva, el aprendizaje por refuerzo no debería centrarse únicamente en maximizar el rendimiento o la eficiencia operativa, sino también en optimizar la calidad de la interacción humano-robot, promoviendo comportamientos adaptativos que favorezcan la fluidez, la comprensión mutua y el bienestar del usuario.

Esta evolución en la concepción del entorno compartido implica la necesidad de definir nuevas métricas centradas en el usuario que complementen las tradicionales medidas de rendimiento. Indicadores como el confort, la carga cognitiva o el grado de aceptación del sistema por parte del humano se vuelven esenciales para evaluar de forma más completa la calidad de la interacción (Giallanza et al., 2024). Para capturar estas dimensiones subjetivas de manera objetiva, resulta fundamental incorporar nuevas fuentes de datos fisiológicos y de conducta. Estas señales permiten inferir el estado interno del usuario en tiempo real y guiar el comportamiento robótico en función de dichas estimaciones (Bethel et al., 2007; Pereira et al., 2025).

Las limitaciones descritas han motivado un nuevo enfoque dentro del campo, el aprendizaje por refuerzo centrado en el humano. En este nuevo enfoque se incorpora explícitamente al usuario como elemento fundamental del proceso de entrenamiento. En lugar de optimizar únicamente la tarea desde una perspectiva funcional, se opta por un aprendizaje que integre la participación del usuario en las distintas fases del ciclo de aprendizaje del agente (Jeon et al., 2020; Kormushev et al., 2013). En esta línea, el apartado 2.7 se centrará en el análisis detallado de las variables más relevantes desde el punto de vista del bienestar y el confort en la interacción humano-robot, tales como la velocidad o la distancia al robot.

2.5 Aprendizaje por refuerzo centrado en el humano

El aprendizaje por refuerzo ha demostrado ser una técnica con una relevancia importante en la investigación actual. Tal y como se ha mostrado en el apartado anterior, su aplicación en el ámbito de la robótica permite obtener comportamientos adaptativos y emergentes que superan las limitaciones de los

enfoques tradicionales. Sin embargo, la incorporación del ser humano en estos sistemas, particularmente en el contexto de la robótica colaborativa impulsada mediante aprendizaje por refuerzo, sigue representando uno de los campos con mayor potencial de desarrollo y exploración. En este apartado se van a analizar las investigaciones más recientes y relevantes que adoptan un enfoque más humano en la aplicación del aprendizaje por refuerzo, con especial atención a su integración en entornos robóticos.

El creciente interés por incorporar un enfoque más centrado en el humano dentro de las aplicaciones de aprendizaje por refuerzo queda reflejado en la Figura 2.9. A partir del año 2022 se observa un incremento con carácter exponencial en el número de publicaciones. En términos acumulados, el volumen de aportaciones científicas en esta línea ha experimentado un aumento superior al 1200% con respecto al año 2010, situando la temática como una de las áreas emergentes dentro del aprendizaje por refuerzo.

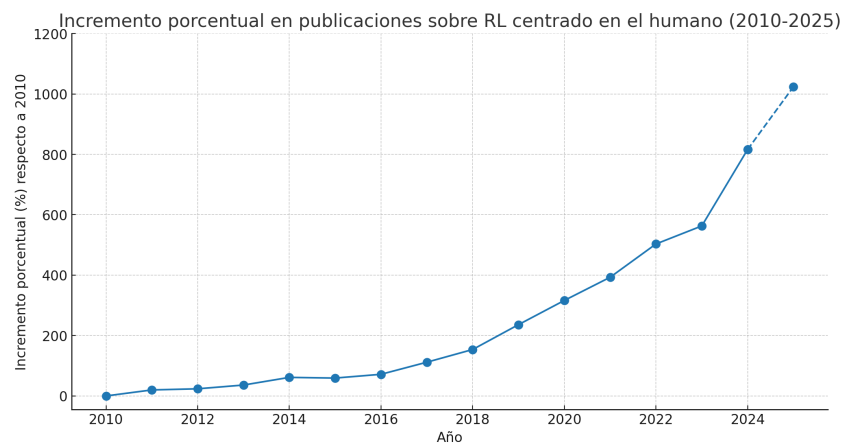


Figura 2.9: Incremento porcentual en publicaciones de RL y Robótica

El núcleo del enfoque centrado en el humano dentro del aprendizaje por refuerzo consiste en integrar de forma explícita la participación humana en el proceso de desarrollo del agente. Este concepto se conoce como *Human in the Loop* y puede materializarse en distintas fases del ciclo de vida del sistema de entrenamiento (Slade et al., 2024). Entre las distintas fases que se puede incorporar la participación del humano se encuentran, la definición del problema y los objetivos, el entrenamiento mediante retroalimentación o intervención directa, la evaluación de las políticas aprendidas y la supervisión del despliegue en entornos reales (Amershi et al., 2014). Esta visión reconoce que el conocimiento humano puede complementar de forma crítica los mecanismos de exploración del agente, permitiendo una mayor alineación con las necesidades reales del

entorno de aplicación (Retzlaff et al., 2024).

En la práctica, la participación humana puede adoptar diversas formas concretas a lo largo del proceso de entrenamiento. Una de las más relevantes es la demostración de comportamientos óptimos mediante enfoques de *imitation learning*, en los que un usuario humano resuelve una tarea con el fin de que el agente aprenda observando dichas demostraciones (Hussein et al., 2017). Otra modalidad consiste en la corrección de acciones en tiempo real durante el entrenamiento, permitiendo al usuario validar o rechazar las decisiones tomadas por el agente a medida que aprende. Asimismo, la evaluación de los resultados puede realizarse mediante señales de recompensa esporádicas. Estas señales no requieren ser necesariamente explícitas; pueden calcularse a partir de indicadores del estado del usuario, como la dirección de la mirada, la postura corporal, la actividad muscular o las respuestas neurofisiológicas. De esta manera, se favorece una integración más natural y continua de las preferencias humanas en el proceso de aprendizaje automático (Griffith et al., 2013).

En cuanto a la fase de entrenamiento, con la intención de facilitar una interacción fluida entre el usuario y el agente, una de las aproximaciones más prometedoras es la utilización de gemelos digitales humanos (*Human Digital Twins*, HDT). Esta técnica busca modelar digitalmente las características, comportamientos y estados de una persona con el fin de simular el comportamiento en las fases del entrenamiento, obteniendo modelos que al ser desplegados en entornos con interacción humana real, puedan reaccionar de manera apropiada. Los gemelos digitales humanos abren nuevas posibilidades para personalizar la experiencia del usuario, mejorar la eficiencia del entrenamiento del agente y garantizar una mayor compatibilidad entre el comportamiento del robot y las necesidades humanas (Lin et al., 2024).

Entre los distintos tipos de señales fisiológicas para recoger retroalimentación de los usuarios, la actividad cerebral ha suscitado un interés considerable debido a su estrecha relación con los estados cognitivos y emocionales. La electroencefalografía (EEG) continúa siendo una de las técnicas más empleadas para la captación de este tipo de datos, gracias a su carácter no invasivo y su alta resolución temporal (ver Figura 2.10, donde se aprecian las 5 ondas cerebrales relevantes (Toichoa Eyam et al., 2021; Iturrate et al., 2010)). No obstante, el alcance del aprendizaje por refuerzo informado por señales humanas no se limita únicamente al EEG. Señales como la respuesta galvánica de la piel (GSR), la variabilidad de la frecuencia cardíaca (HRV), la electromiografía (EMG) o incluso el análisis de expresiones faciales están siendo cada vez más exploradas como fuentes de retroalimentación implícita para agentes que operan en entor-

nos centrados en el usuario (Gonzalez-Santocildes et al., 2024b).

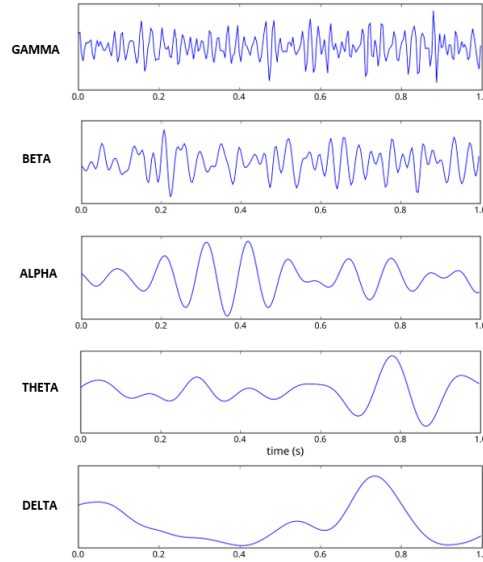


Figura 2.10: Las cinco ondas cerebrales distintas: Delta, theta, alpha, beta, and gamma.

Estas señales generadas por humanos pueden desempeñar distintos roles dentro del ciclo de aprendizaje por refuerzo. Por ejemplo, pueden utilizarse para modelar funciones de recompensa que se ajusten al estado emocional del usuario, condicionar directamente la política del agente en función del contexto fisiológico, o activar mecanismos de seguridad ante situaciones críticas. La integración de información afectiva o relacionada con la detección de errores permite que el agente no solo optimice el rendimiento de la tarea, sino que también se adapte al confort y a las expectativas humanas, algo especialmente relevante en espacios compartidos como los entornos industriales o sanitarios (Kar et al., 2022; Borboni et al., 2022).

Una línea prometedora adicional es el uso del aprendizaje por refuerzo para la personalización del comportamiento. En estos casos, los agentes no solo se entrenan para ejecutar una tarea, sino para hacerlo de manera alineada con los umbrales de estrés, las preferencias de velocidad o la carga cognitiva del usuario. Esta capacidad resulta especialmente valiosa en interacciones prolongadas, donde los modelos preentrenados pueden no generalizar adecuadamente entre distintos individuos o ante condiciones cambiantes. La naturaleza adaptativa del aprendizaje por refuerzo, combinada con la retroalimentación continua proporcionada por un gemelo digital humano, favorece el desarrollo de sistemas que

evolucionan conjuntamente con sus usuarios.

En conjunto, la integración de gemelos digitales humanos y mecanismos de retroalimentación basados en señales fisiológicas dentro de marcos de aprendizaje por refuerzo abre una nueva frontera en la robótica colaborativa. Estas técnicas no solo permiten que los agentes aprendan a resolver tareas, sino que los capacitan para ajustar su comportamiento en función de las características y necesidades del usuario humano. Este enfoque representa un cambio de paradigma respecto al aprendizaje robótico tradicional, al priorizar no únicamente la eficiencia, sino también aspectos como la empatía, la flexibilidad y el diseño centrado en el bienestar del usuario.

La combinación del aprendizaje por refuerzo, la participación humana y la robótica colaborativa representa una línea de investigación especialmente prometedora. Sin embargo, para que el comportamiento de un robot sea verdaderamente adaptativo al ser humano, no basta con incorporar señales fisiológicas o retroalimentación durante el entrenamiento, es necesario comprender en profundidad al usuario y conocer qué variables influyen realmente en su percepción de confort, seguridad y bienestar. Por ello, en el siguiente apartado se abordará la representación y evaluación del confort en la interacción humano-robot, identificando aquellas variables robóticas que más afectan a la percepción de comodidad por parte del usuario y que, por tanto, deben ser tenidas en cuenta en el diseño y ajuste de políticas de comportamiento adaptativo.

2.6 Representación y evaluación del confort en la interacción humano-robot

A medida que la robótica colaborativa avanza hacia entornos cada vez más complejos y centrados en el usuario, el confort y bienestar humano emerge como una dimensión crítica para el diseño y la evaluación de las interacciones. Lejos de ser un aspecto secundario, el confort condiciona la calidad de la experiencia compartida, la aceptación del sistema robótico y el rendimiento conjunto en tareas colaborativas. Su estudio requiere ir más allá de la eficiencia técnica, incorporando consideraciones relacionadas con la percepción subjetiva del usuario, su estado emocional y la adaptación dinámica del comportamiento robótico.

Esta creciente relevancia ha derivado en un notable incremento del número de investigaciones dedicadas a la representación y evaluación del confort en entornos de interacción humano-robot. Tal y como se muestra en la Figura 2.11, el volumen de publicaciones en esta línea ha experimentado un crecimiento cons-

tante en los últimos años, especialmente a partir de 2020. Esta tendencia refleja una transición paradigmática en el diseño de sistemas interactivos, donde el bienestar del usuario adquiere un papel central. En este apartado se abordarán las principales variables que influyen en la percepción de confort, así como las metodologías utilizadas para su evaluación objetiva y subjetiva.

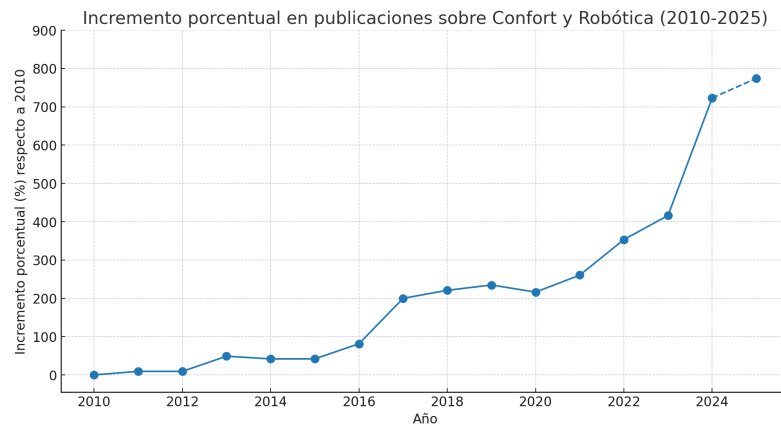


Figura 2.11: Incremento porcentual en publicaciones de representaciones del confort en la interacción humano-robot

Entre los factores que afectan al confort en entornos de interacción humano-robot se encuentran las variables cinemáticas del sistema, especialmente la velocidad, la aceleración y la suavidad de los movimientos (Rudenko et al., 2020). Estos parámetros no solo condicionan la eficiencia de la ejecución, sino que también afectan de manera directa a la experiencia del usuario. Por ejemplo, movimientos demasiado rápidos o con cambios bruscos de dirección pueden percibirse como amenazantes o impredecibles, generando desconfianza y aumentando la carga cognitiva (van den Brule et al., 2014). Por el contrario, trayectorias suaves y consistentes tienden a transmitir mayor sensación de control y seguridad, favoreciendo una colaboración fluida (Lasota et al., 2017).

La percepción de confort también está influida por la predictibilidad del movimiento, entendida como la capacidad del usuario para anticipar las acciones del robot (Dragan et al., 2013). En este sentido, perfiles cinemáticos predecibles (aceleraciones controladas y velocidad constante) permiten que el humano desarrolle ideas fiables sobre el comportamiento del sistema. Esta previsibilidad es especialmente relevante en tareas conjuntas o espacios compartidos, donde una mala interpretación del movimiento puede generar incomodidad o incluso provocar errores (Gulletta et al., 2021). Por ello, el diseño cinemático debe considerar no solo el rendimiento técnico, sino también la percepción subjetiva del

usuario en función del contexto de interacción.

La distancia física entre el robot y el humano es una de las variables espaciales más relevantes para la percepción de confort en entornos colaborativos (Pacchierotti et al., 2006). Cuando un robot invade lo que se percibe como el espacio personal del usuario puede generar incomodidad, estrés e incluso rechazo. Por el contrario, mantener una distancia adecuada y constante favorece la sensación de control, seguridad y aceptación del sistema (Mumm y Mutlu, 2011).

El nivel de confort espacial del usuario no es estático ni universal, sino que varía en función de factores contextuales como el tipo de tarea, la cultura del usuario o incluso su estado emocional. Por ello, resulta fundamental que los sistemas robóticos sean capaces de adaptar dinámicamente su comportamiento espacial en función de señales explícitas o implícitas del usuario. Algunos enfoques emplean sensores para estimar el nivel de confort en tiempo real, mientras que otros modelan esta preferencia mediante cuestionarios. La gestión inteligente de la distancia no solo mejora el confort subjetivo, sino que también optimiza la eficiencia y seguridad en la ejecución de tareas compartidas (Charalampous et al., 2017; Singamaneni et al., 2024).

Además de las características espaciales y cinemáticas, el comportamiento temporal del robot tiene un impacto directo en la percepción de confort del usuario. La sincronización adecuada entre las acciones del humano y del robot es fundamental para lograr una interacción fluida y eficiente (Jensen et al., 2017). Cuando existe desajuste temporal, por ejemplo, si el robot actúa con demasiado retraso o anticipación, puede surgir confusión y un aumento en la carga cognitiva del usuario. En cambio, una coordinación temporal precisa transmite competencia y comprensión mutua, lo que favorece la confianza y el agrado hacia el sistema robótico (Hoffman, 2019).

Otro aspecto clave cuando se habla de temporalidad es el tiempo de reacción del robot ante eventos o señales del usuario. En contextos colaborativos, los humanos tienden a esperar respuestas rápidas y coherentes por parte del sistema, especialmente cuando hay contacto visual, gestos o acciones que requieren atención conjunta. Retrasos superiores a ciertos umbrales pueden generar incomodidad o dar la impresión de que el robot no está respondiendo adecuadamente. Por ello, optimizar los tiempos de percepción, procesamiento y ejecución no solo mejora el rendimiento técnico, sino también la experiencia subjetiva del usuario y su disposición a continuar colaborando con el sistema (Leusmann et al., 2025).

Aunque las variables cinemáticas, espaciales y temporales del robot influyen de manera directa en la percepción de confort durante la interacción, no pueden entenderse de forma aislada del perfil del usuario humano. La experiencia previa con tecnología, el nivel de familiaridad con robots, el tipo de tarea, el contexto laboral y otros factores sociales condicionan cómo se interpreta y se responde a los comportamientos del robot (de Graaf et al., 2015). Por ejemplo, un mismo gesto robótico puede ser percibido como amenazante por un usuario inexperto, y como eficiente o incluso deseable por un operario con años de trabajo conjunto con sistemas automatizados (Bartneck et al., 2008).

Asimismo, variables como la edad, el nivel educativo, el perfil cultural o incluso rasgos individuales como la tolerancia al riesgo o la introversión pueden modular la respuesta emocional y fisiológica ante determinados comportamientos robóticos (Lim et al., 2020; Wang et al., 2024). Por ello, comprender el confort en la interacción humano-robot requiere tener en cuenta no solo lo que el robot hace, sino también quién es el usuario, cómo interpreta esas acciones, y en qué condiciones tiene lugar la interacción. Esta perspectiva centrada en el usuario es especialmente relevante en entornos colaborativos prolongados, donde la historia compartida, la confianza acumulada y el estado del usuario configuran una relación dinámica y sensible al contexto (Esterwood et al., 2021).

La evaluación del confort en la interacción humano-robot requiere enfoques metodológicos que combinen la objetividad de las métricas fisiológicas con la interpretación de los métodos subjetivos. Entre las herramientas más empleadas se encuentran cuestionarios estandarizados, entrevistas cualitativas, y observación directa del comportamiento. Paralelamente, se utilizan señales fisiológicas como las explicadas en la sección 2.2, que permiten estimar el estado interno del usuario en tiempo real. La combinación de ambas aproximaciones permite obtener una visión más robusta y completa del confort, facilitando su integración como variable explícita en el diseño y ajuste de políticas robóticas adaptativas.

En definitiva, queda claro que el confort en la interacción humano-robot constituye un eje fundamental en el diseño de sistemas de robótica colaborativa, especialmente en aquellos que incorporan mecanismos de aprendizaje por refuerzo para adaptar su comportamiento. Tal como se ha analizado en este apartado, las variables cinemáticas, espaciales, temporales y sociales influyen de forma directa en la percepción del usuario y deben ser tenidas en cuenta a la hora de desarrollar políticas que no solo optimicen la tarea, sino también la calidad de la experiencia compartida. La intersección de las áreas presentadas permite avanzar hacia agentes que aprendan de forma más alineada con las necesidades humanas. En la próxima sección 2.7 se detallarán las oportunidades

de investigación situadas en la intersección entre el aprendizaje por refuerzo, la robótica colaborativa y el confort humano.

2.7 Resumen y oportunidades de investigación

A lo largo de este capítulo se han explorado las principales líneas de investigación que convergen en el desarrollo de sistemas robóticos adaptativos centrados en el humano. El aprendizaje por refuerzo ofrece mecanismos potentes para generar comportamientos emergentes y personalizados; la robótica colaborativa plantea nuevos desafíos en cuanto a seguridad, fluidez e interpretación del comportamiento humano; y la noción de confort emerge como una variable clave para la aceptación y sostenibilidad de estas interacciones. Esta revisión ha permitido identificar tanto los avances recientes como las limitaciones actuales en la integración de estas tres dimensiones.

La literatura actual muestra un interés creciente hacia la incorporación del confort humano como criterio de diseño, pero su formalización dentro de marcos de aprendizaje por refuerzo sigue siendo limitada. A menudo, los agentes optimizan tareas desde una lógica puramente funcional, ignorando el impacto que su comportamiento tiene sobre la experiencia del usuario. En este contexto, la presente tesis se sitúa en la intersección de estas disciplinas, proponiendo que la integración explícita del confort humano como componente de la función de recompensa permite generar políticas más alineadas con las necesidades reales de los usuarios, especialmente en entornos colaborativos no deterministas.

Bajo esta premisa, se plantea la hipótesis (ver sección 1.2) de que es posible diseñar agentes robóticos entrenados mediante aprendizaje por refuerzo que ajusten su comportamiento a indicadores de confort humano, recogidos a través de señales fisiológicas y evaluaciones subjetivas. Esta adaptación, basada en modelos previamente entrenados, permitiría mejorar la calidad de la interacción sin necesidad de calibraciones complejas ni sensores invasivos durante el despliegue. La tesis propone validar este enfoque a través de un marco experimental que combina simulación, entrenamiento con métricas de confort y evaluación con usuarios reales.

A partir del análisis realizado, se identifican diversas oportunidades de investigación: (i) la necesidad de diseñar funciones de recompensa sensibles al estado del usuario; (ii) el desarrollo de entornos simulados progresivos que permitan incorporar el confort como señal durante el entrenamiento; (iii) la validación empírica de políticas aprendidas en interacción con humanos reales; y (iv) la exploración de mecanismos de personalización basados en perfiles fisiológicos

y subjetivos.

Como síntesis de estas oportunidades, resulta útil representar gráficamente las principales intersecciones entre las áreas tecnológicas que sustentan esta investigación. La Figura 2.12 muestra un mapa conceptual en el que convergen los cuatro ámbitos, sensores humanos, realidad virtual, aprendizaje por refuerzo y robótica colaborativa. La intersección de estas áreas definen el espacio de innovación hacia sistemas adaptativos centrados en el confort humano.

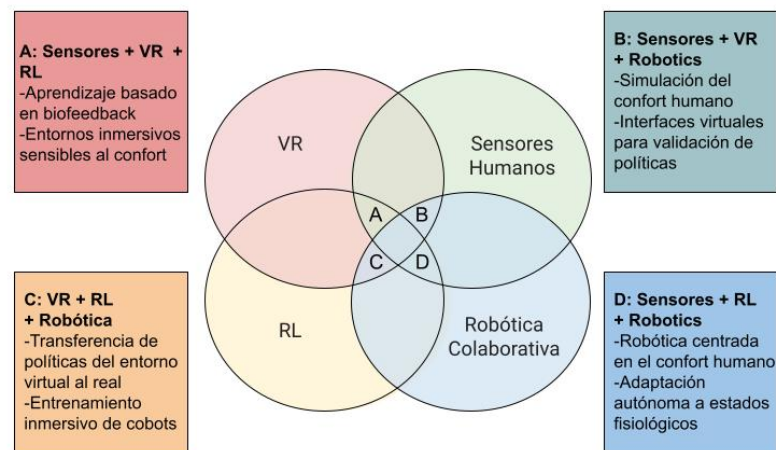


Figura 2.12: Mapa conceptual de las intersecciones clave entre percepción humana, entornos virtuales, aprendizaje por refuerzo y robótica colaborativa.

Las intersecciones del mapa recogen cuatro oportunidades principales. El área **A** combina sensores humanos, realidad virtual y aprendizaje por refuerzo para crear entornos inmersivos que respondan al estado del usuario. El área **B** une sensores humanos, realidad virtual y robótica, permitiendo evaluar el confort en tareas colaborativas simuladas. El área **C** integra realidad virtual, aprendizaje por refuerzo y robótica, facilitando la transferencia de políticas desde la simulación al entorno real. Por último, el área **D** conecta sensores humanos, aprendizaje por refuerzo y robótica, orientándose a robots capaces de adaptar su comportamiento según el confort humano.

En este contexto, la tesis se sitúa en la intersección de las cuatro áreas, proponiendo un enfoque que combina la percepción fisiológica del usuario, la simulación inmersiva, el aprendizaje por refuerzo y la robótica colaborativa. Esta

convergencia permite abordar el confort humano no solo como un indicador pasivo, sino como un componente activo en el aprendizaje y la adaptación del comportamiento robótico.

En los siguientes capítulos se describe el diseño experimental desarrollado para dar respuesta a estos retos. A través de una serie de experimentos progresivos, se evaluará la viabilidad de integrar el confort humano como variable en entornos de aprendizaje por refuerzo robóticos, tanto en simulación como en interacción directa con usuarios.

*If you are curious, you'll find the puzzles
around you. If you are determined, you
will solve them.*

ERNÓ RUBIK

Capítulo

3

Entrenamiento simulado de comportamientos robóticos adaptativos

Con el objetivo de avanzar en el diseño de comportamientos robóticos adaptativos, seguros y centrados en el ser humano, y tomando como referencia el estado del arte presentado en aprendizaje por refuerzo e interacción humano-robot, se planteó el desarrollo de una serie de entornos de simulación que permitan entrenar, validar y comparar diferentes algoritmos y estrategias. Estos entornos virtuales permiten replicar condiciones específicas de colaboración, explorar diversas configuraciones de tareas compartidas, y evaluar el impacto de distintos parámetros tanto del agente robótico como del comportamiento humano simulado. A través de esta aproximación, es posible generar políticas adaptativas que luego podrán ser transferidas con mayor robustez a escenarios reales, minimizando los riesgos y costes asociados a la experimentación física directa.

Este capítulo se centra en describir las decisiones adoptadas durante el diseño y puesta en marcha del proceso de entrenamiento simulado, haciendo especial hincapié en la selección del framework empleado, la construcción progresiva de los entornos y la justificación de su estructura. En una primera fase, se desarrollaron escenarios robóticos de baja complejidad, sin la presencia de agentes humanos, con el fin de facilitar el ajuste de hiperparámetros y el testeo inicial de los algoritmos sin interferencias externas. Posteriormente, se introdujeron agentes humanos simulados para modelar dinámicas de colaboración más realistas, sin embargo, las cuestiones relacionadas con la percepción del

confort humano y la interpretación de su estado se abordarán en capítulos posteriores. Esta secuencia gradual responde a la necesidad de construir una base experimental sólida antes de incorporar componentes más sensibles y complejos vinculados a la interacción humano-robot.

El capítulo se estructura en cuatro secciones principales. En primer lugar, se presenta la justificación de la elección del framework de entrenamiento y las herramientas empleadas, considerando tanto criterios de compatibilidad como de escalabilidad y reproducibilidad. A continuación, se describe el diseño de los entornos simulados básicos utilizados en las fases iniciales, orientados a la validación funcional de los algoritmos y herramientas sin presencia humana. En el tercer apartado, se aborda el desarrollo de escenarios más complejos que incorporan la adaptación de trayectorias en contextos colaborativos, incluyendo las primeras versiones de agentes humanos simulados. Finalmente, se expone un análisis detallado de los resultados obtenidos durante el proceso de entrenamiento, evaluando la eficacia de las estrategias implementadas y la evolución del comportamiento del agente en función de los distintos niveles de complejidad.

3.1 Selección y justificación del sistema y framework de entrenamiento

La aplicación práctica del aprendizaje por refuerzo requiere del uso de herramientas y tecnologías específicas que permitan definir entornos, agentes, recompensas y políticas de aprendizaje de manera integrada y escalable. En este contexto, existen numerosos frameworks de entrenamiento que ofrecen distintas capacidades en términos de simulación física, visualización, control de agentes y personalización del flujo de entrenamiento. La elección del entorno de entrenamiento adecuado resulta, por tanto, un aspecto crítico que condiciona no solo la eficiencia del proceso de aprendizaje, sino también la capacidad de transferir posteriormente los modelos a escenarios reales. En esta sección se detallan los criterios considerados para la selección del framework empleado en esta tesis, así como las principales ventajas y limitaciones observadas en relación con los objetivos del proyecto.

Uno de los frameworks más consolidados en el ecosistema de Python es *Stable Baselines3* (SB3), una biblioteca que proporciona implementaciones estables y mantenidas de los algoritmos de aprendizaje por refuerzo más utilizados, como PPO, DDPG, TD3, SAC o A2C (Raffin et al., 2021). SB3 destaca por su facilidad de uso, su integración nativa con la interfaz *Gymnasium* (antes OpenAI Gym), y su compatibilidad con entornos personalizados, lo que lo convierte en

una opción de referencia tanto en contextos académicos como en entornos de experimentación aplicada (Towers et al., 2024).

A pesar de su robustez y versatilidad para el entrenamiento de agentes, SB3 no incluye herramientas específicas para la creación de entornos simulados complejos. Aunque permite definir entornos simples directamente en Python, la simulación de dinámicas físicas realistas, interacciones visuales avanzadas o la representación de agentes humanos requiere el uso de motores externos. Por este motivo, el diseño del entorno de entrenamiento se delega habitualmente a otras plataformas más especializadas, que se sincronizan con SB3 mediante interfaces intermedias o *wrappers* de *Gymnasium*. Esta arquitectura modular permite aprovechar las capacidades de simulación avanzada de herramientas como Unity (Cohen et al., 2022), manteniendo la flexibilidad y eficiencia del entrenamiento proporcionado por SB3 (Raffin et al., 2021; Towers et al., 2024).

Para la creación de entornos simulados físicamente realistas, existen diversas plataformas y motores físicos que permiten modelar cuerpos rígidos, colisiones, sensores y otros aspectos clave del comportamiento robótico. Entre las opciones más extendidas se encuentran *MuJoCo*, *PyBullet*, *Unity ML-Agents* e *Isaac Gym*, cada una con características particulares en términos de precisión física, rendimiento computacional, facilidad de integración y soporte para tareas complejas. En los párrafos se analizan estas alternativas de forma individual, destacando sus principales ventajas y limitaciones, así como su idoneidad en función de los objetivos específicos de la presente tesis.

MuJoCo (Multi-Joint dynamics with Contact) es un motor de simulación física frecuentemente utilizado en investigación debido a su alta precisión en la modelación de dinámicas articuladas y contactos. Desarrollado inicialmente por Emo Todorov (Todorov et al., 2012) y actualmente mantenido por DeepMind, *MuJoCo* destaca por su capacidad para simular sistemas mecánicos complejos con un alto nivel de realismo, lo que lo convierte en una herramienta de referencia en estudios de control fino, locomoción y manipulación robótica. No obstante, su uso está más orientado a entornos abstractos y de bajo nivel gráfico, lo que limita su aplicabilidad en experimentos donde la representación visual o la interacción con humanos simulados resulta relevante.

Unity ML-Agents es un framework desarrollado por *Unity Technologies* que combina un motor gráfico avanzado propio con un conjunto de herramientas específicas para el entrenamiento de agentes mediante aprendizaje por refuerzo. Su principal fortaleza reside en la capacidad de construir entornos tridimensionales, visuales, altamente configurables y con simulación física integrada, lo que permite modelar tareas complejas con un elevado grado de realismo. Además,

proporciona soporte nativo para entornos multiagente y permite la ejecución de simulaciones en entornos de realidad virtual, facilitando así el estudio de escenarios inmersivos e interactivos. El entrenamiento de agentes puede realizarse a través de su propio framework o mediante la integración con otras librerías externas como Stable Baselines3. Estas características lo convierten en una opción versátil para investigaciones que requieren evaluar el comportamiento de los agentes en contextos visuales, dinámicos y orientados a la interacción (Cohen et al., 2022).

Isaac Gym es un framework desarrollado por NVIDIA que combina simulación física con entrenamiento acelerado mediante GPU, lo que lo convierte en una opción interesante para tareas de RL que requieren una gran cantidad de capacidad de computación. Su arquitectura permite la ejecución simultánea de simulaciones paralelas en una única GPU, lo que reduce significativamente el tiempo de entrenamiento frente a otros motores más convencionales. Además, integra un motor de físicas propio y ofrece soporte directo para controladores de robots reales, facilitando así la transferencia de los entrenamientos a entornos reales (NVIDIA, 2025). No obstante, su uso requiere un conocimiento técnico avanzado, hardware compatible y una inversión inicial considerable en la configuración del entorno. Además, su enfoque está orientado principalmente al entrenamiento de tareas físicas, por lo que no ofrece soporte nativo para la simulación de agentes humanos, la integración con dispositivos de realidad virtual, ni la interacción mediante mandos de control externo, lo que limita su aplicabilidad en estudios centrados en la interacción humano-robot o en la evaluación participativa del comportamiento robótico.

PyBullet es un motor de simulación física en tiempo real que destaca por su ligereza y facilidad de uso. Utilizado ampliamente en investigación y docencia, permite modelar cuerpos rígidos, sensores y robots articulados con un nivel de precisión aceptable en escenarios simplificados. Su integración con Python y su compatibilidad con entornos *Gymnasium* y *Stable Baselines3* lo convierten en una opción accesible para pruebas rápidas y prototipado de agentes (Coumans y Bai, 2016–2021). Sin embargo, la capacidad gráfica es limitada, y aunque permite una visualización básica en 3D, no está diseñado para generar entornos visualmente complejos ni para integrar agentes humanos o experiencias inmersivas. Estas limitaciones lo hacen menos adecuado para investigaciones centradas en la interacción humano-robot o el análisis del comportamiento desde una perspectiva perceptiva y contextualizada.

Gazebo es un entorno de simulación física de código abierto ampliamente adoptado en robótica, especialmente en contextos académicos e industriales vinculados al ecosistema ROS. Ofrece un motor de físicas robusto y un rende-

rizado tridimensional suficiente para representar interacciones realistas entre robots y entornos complejos, incluyendo colisiones. Su principal fortaleza radica en la integración nativa con ROS, lo que permite ejecutar el mismo código de control tanto en simulación como en robots reales, facilitando significativamente la transferencia de políticas de aprendizaje o control. Además, soporta múltiples motores de físicas y dispone de una amplia biblioteca de modelos de robots y escenarios. No obstante, Gazebo presenta limitaciones en términos de rendimiento gráfico y eficiencia computacional frente a motores más modernos como Unity o Isaac Sim, especialmente cuando se busca entrenar agentes mediante aprendizaje por refuerzo a gran escala. Asimismo, la configuración inicial y la sincronización precisa con ROS pueden resultar complejas, lo que demanda una curva de aprendizaje considerable para lograr un entorno de simulación estable y reproducible. Finalmente, aunque ofrece un soporte visual aceptable, carece de capacidades nativas para entornos inmersivos, y su diseño no está orientado a la simulación de factores humanos o sociales, lo que complica su uso en estudios de interacción humano-robot donde la percepción y el confort del usuario juegan un papel relevante (Koenig y Howard, 2004).

Con el fin de sintetizar las principales características de cada framework considerado para el desarrollo de los entornos relacionados con la experimentación de la tesis, la Tabla 3.1 presenta una comparativa en términos de precisión física, capacidad gráfica, soporte para agentes humanos y realidad virtual, facilidad de integración con otras librerías y rendimiento computacional. Esta visión general permite valorar de forma conjunta los puntos fuertes y las limitaciones de cada opción en relación con los objetivos específicos de esta tesis.

Tabla 3.1: Resumen comparativo de frameworks de simulación

Framework	Física	Gráficos	Hum./VR	Integrac.	Rendimiento
<i>MuJoCo</i>	Alta	Baja	No	Media	Media
<i>ML-Agents</i>	Muy Alta	Muy alta	Sí	Alta	Media
<i>Isaac Gym</i>	Muy alta	Baja	No	Media	Muy alta
<i>PyBullet</i>	Media	Baja	No	Alta	Media
<i>Gazebo</i>	Alta	Media	No	Muy alta	Media

Se ha decidido utilizar *Unity ML-Agents* como entorno de simulación principal para el desarrollo de esta tesis debido a su versatilidad para modelar agentes complejos y entornos interactivos. Una de sus principales fortalezas radica en la posibilidad de crear y representar gemelos digitales humanos, permitiendo incorporar agentes humanos simulados con comportamientos parametrizables y observar cómo estos interactúan con el robot en distintas condiciones. Asimismo, su compatibilidad con dispositivos de realidad virtual facilita el test inner-

sivo de interacciones humano-robot, aspecto fundamental para evaluar aspectos como la percepción de confort o la adaptabilidad del sistema en contextos colaborativos. A nivel técnico, ML-Agents ofrece un motor de físicas integrado suficientemente preciso para tareas de manipulación y locomoción, así como una arquitectura abierta que permite su integración fluida con librerías externas de entrenamiento, como *Stable Baselines3*, mediante wrappers estandarizados. Además, dada la naturaleza modular y el formato exportable de los entornos desarrollados en Unity, el paso a pruebas en sistemas físicos o simuladores ROS es viable sin necesidad de una reestructuración completa del sistema.

Para el desarrollo de los entornos simulados se utilizó la versión *2021.3.18f1* de Unity, correspondiente a una de las ediciones *LTS (Long Term Support)*, que garantizan mayor estabilidad y soporte a largo plazo. En cuanto a *ML-Agents*, se trabajó con el **Release 20**, una distribución oficial publicada por Unity que incluye las versiones *ml-agents 0.30.0*, *ml-agents-envs 0.30.0*, *gym-unity 0.30.0*, *Communicator 1.5.0*, y los paquetes de C# *com.unity.ml-agents 2.3.0-exp.3* y *com.unity.ml-agents.extensions 0.6.1-preview*. Esta combinación proporciona un conjunto consolidado de herramientas para el entrenamiento de agentes, soporte para entornos multiagente y sensores personalizados. Aunque existen versiones más recientes, se ha verificado que el funcionamiento de los entornos y scripts desarrollados es compatible con versiones actuales y estables, como la **Release 22**, sin presentar incompatibilidades relevantes. Tanto Unity como ML-Agents disponen de documentación oficial detallada, donde se listan todas las versiones disponibles, sus características y las instrucciones de instalación (Unity Technologies, 2019, 2025). Para ilustrar este aspecto, en la Figura 3.1 se muestra una captura de la consola de ejecución de ML-Agents, donde se indican las versiones utilizadas durante el desarrollo, incluyendo el paquete principal, la API de comunicación y la versión de PyTorch.



Figura 3.1: Captura de terminal con la ejecución de ML-Agents. Se observan las versiones utilizadas en las primeras iteraciones.

Con el fin de garantizar la replicabilidad de los experimentos y proporcionar un contexto adecuado para la interpretación de los tiempos de entrenamiento que se muestren secciones y capítulos siguientes, es importante especificar las características del sistema empleado para los entrenamientos. Todos los entrenamientos fueron realizados en un equipo con sistema operativo *Windows 11*, procesador *Intel Core i7-12700H*, *32 GB de memoria RAM* y una tarjeta gráfica *NVIDIA GeForce RTX 3060*. Esta configuración permitió ejecutar simulaciones complejas en tiempo razonable, aprovechar la aceleración por GPU durante las fases de entrenamiento y mantener una experiencia interactiva fluida durante el diseño de los entornos en Unity.

A partir de las decisiones descritas, el entorno principal de desarrollo y simulación queda definido por la combinación de *Unity* para la construcción de escenarios tridimensionales interactivos, y *ML-Agents* como framework de entrenamiento. Esta integración permite una configuración flexible, visual y personalizable, adecuada para el modelado de comportamientos adaptativos en contextos robóticos colaborativos. Todos los experimentos se llevaron a cabo en la máquina presentada anteriormente, cuyas características garantizan un entorno de ejecución estable y reproducible. Con esta base técnica establecida, la siguiente sección describe el primer proceso de pruebas y entrenamiento realizado, incluyendo el diseño de los entornos iniciales y los resultados preliminares obtenidos.

3.2 Diseño y validación de entornos simulados básicos

Con el objetivo de validar la viabilidad técnica del framework adoptado y establecer una base experimental sólida, esta sección presenta el diseño y la evaluación de los primeros entornos robóticos simulados implementados en Unity. Se trata de escenarios de baja complejidad en los que se prueba el comportamiento de agentes robóticos en condiciones controladas, sin la presencia de elementos humanos ni dinámicas colaborativas. Estos entornos permiten ajustar los parámetros fundamentales del sistema, comprobar la estabilidad de la simulación, verificar la correcta configuración de las recompensas simples y evaluar el rendimiento inicial de los algoritmos de aprendizaje. Además, facilitan la detección temprana de posibles errores de integración entre el entorno visual y el motor de entrenamiento. A partir de esta fase preliminar, se establecen las condiciones necesarias para la construcción de escenarios más avanzados en capítulos posteriores.

Un aspecto fundamental en esta fase es la correcta representación física del agente robótico dentro del entorno simulado. Para ello, se ha utilizado la librería *URDF Importer* (Unity Technologies, 2023), una herramienta desarrollada

por Unity que permite importar modelos robóticos a partir de archivos URDF (*Unified Robot Description Format*), ampliamente utilizados en plataformas como ROS. Esta librería facilita la transferencia de las características estructurales, dinámicas y cinemáticas del robot real al entorno virtual, permitiendo así una simulación más precisa y realista. En este trabajo, se ha empleado un modelo URDF extraído directamente del robot físico que será utilizado en fases posteriores, garantizando así la coherencia entre los entrenamientos simulados y las pruebas experimentales reales.

Durante el proceso de integración del modelo URDF con ML-Agents, surgieron diversas dificultades relacionadas con el comportamiento físico del robot, especialmente al intentar realizar pruebas de movimiento manual mediante mando o scripts interactivos. Estos problemas estuvieron principalmente asociados a la configuración de las mallas de colisión y los componentes físicos importados automáticamente, los cuales no siempre se ajustan de forma óptima a los requisitos del motor de físicas de Unity. Para solventarlos, fue necesario realizar ajustes manuales en los colliders y en los parámetros dinámicos del modelo. Las versiones corregidas y funcionales de estos elementos han sido documentadas y se encuentran disponibles en el repositorio asociado.

En este caso, es importante destacar que la librería *URDF Importer* controla el movimiento del robot mediante la aplicación de fuerzas sobre las articulaciones, replicando un comportamiento dinámico realista. Sin embargo, para los objetivos concretos de este proyecto, se optó por una aproximación basada en el control por posición, en la que cada articulación es desplazada hacia un valor angular objetivo (por ejemplo, de x radianes a $x + \delta$). Esta elección responde a la necesidad de facilitar la transferencia de los comportamientos aprendidos a un robot físico, cuya arquitectura admite el uso de planificadores propios que gestionan internamente la trayectoria entre puntos definidos. Por tanto, en lugar de replicar la dinámica completa, se diseñó el sistema de control simulado para calcular directamente los puntos de destino de las articulaciones, lo que permite compatibilizar el entrenamiento virtual con las capacidades nativas del robot real.

En la Figura 3.2 se muestra el modelo del robot *UR3e* una vez importado en el entorno de Unity. Este será el modelo principal utilizado en todos los entrenamientos simulados descritos a lo largo de la tesis.

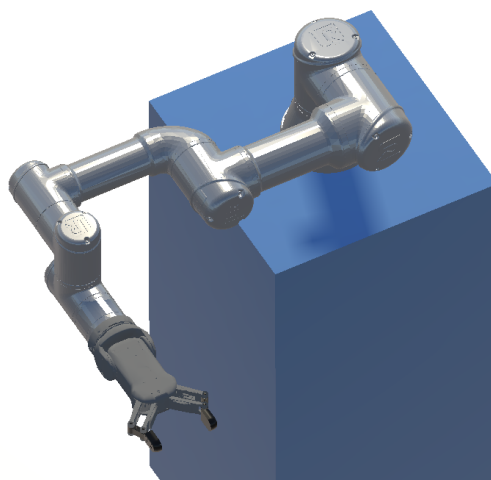


Figura 3.2: Robot UR3e importado y modificado en Unity para el desarrollo de los entrenamientos

Aunque las observaciones utilizadas durante el entrenamiento varían según la configuración y el objetivo de cada entorno simulado, las acciones se mantuvieron constantes a lo largo de todos los escenarios. Estas se definen como un vector continuo de seis valores, cada uno correspondiente a una articulación del brazo robótico UR3e. Estos valores representan incrementos angulares relativos que el agente de RL propone en cada paso de simulación, y se encuentran limitados al rango de $[-2, 2]$ grados para evitar movimientos bruscos o inestables. Dichas acciones son aplicadas directamente sobre el modelo simulado mediante funciones específicas del controlador del robot personalizado, ajustando progresivamente la posición de cada articulación.

El primero de los entornos desarrollados corresponde a un escenario robótico extremadamente simple, diseñado con el objetivo de verificar la viabilidad básica del proceso de entrenamiento mediante aprendizaje por refuerzo en el marco implementado. En este entorno, el agente robótico debe aproximar su efector final a un punto fijo en el espacio (una pelota naranja), con un criterio de precisión muy bajo (aproximadamente 20 cm de distancia). La posición objetivo no varía a lo largo del episodio, y no se requiere una orientación concreta ni una trayectoria específica. Este entorno inicial se empleó como prueba de concepto para comprobar si el agente es capaz de aprender una política mínima que le permita aproximarse al objetivo, validando así la correcta integración entre la simulación, el entorno de entrenamiento y el sistema de recompensas más elemental. La Figura 3.3 muestra una vista general de este entorno básico.

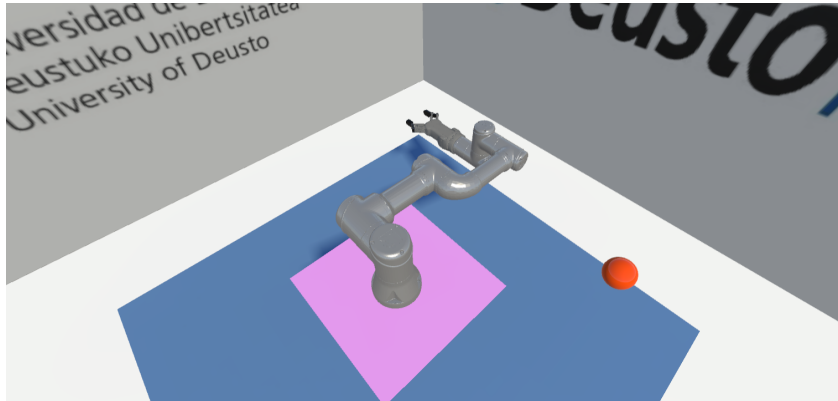


Figura 3.3: Primer entorno simulado básico en Unity utilizado para entrenar una tarea de aproximación a un objetivo.

En cuanto a las observaciones proporcionadas al agente, estas consisten en la posición de los motores del robot, así como, en este caso particular, la posición relativa del efector final respecto al objetivo. Esta información resulta suficiente para que el agente pueda deducir su progreso hacia la meta y ajustar sus acciones en consecuencia. Durante las pruebas iniciales se evaluaron dos funciones de recompensa distintas, una función *sparse*, que otorgaba una recompensa binaria únicamente cuando el efector final se encontraba dentro del radio de precisión especificado, y una función *densa* basada en la distancia euclídea entre el efector y el objetivo. Aunque ambas variantes permitieron al agente aprender una política válida, la función densa proporcionó una guía más informativa y consistente durante el entrenamiento, favoreciendo una convergencia más rápida y estable. Por este motivo, la recompensa densa fue empleada en los experimentos realizados con este primer entorno. No obstante, se prevé que en los siguientes entornos simulados ambas funciones resulten útiles dependiendo del nivel de dificultad y los objetivos específicos de cada tarea.

En este caso se han evaluado dos de los algoritmos más consolidados según el estado del arte del aprendizaje por refuerzo SAC y PPO. Aunque ambos algoritmos fueron capaces de aprender una política válida en este entorno simple, SAC mostró un rendimiento notablemente superior, especialmente debido a la naturaleza continua del espacio de acciones utilizado en este proyecto. En particular, SAC logró tiempos de convergencia aproximadamente un 20% más rápidos que PPO, lo que lo convierte en una opción más eficiente para escenarios donde la rapidez en el aprendizaje es relevante. Por este motivo, SAC fue seleccionado como el algoritmo principal y se utilizó de manera consistente a lo largo de todo el proyecto. A lo largo del proceso de entrenamiento se llevaron a cabo diversas iteraciones con ajustes progresivos en los hiperparámetros, con el objetivo de mejorar la estabilidad y la velocidad de convergencia del modelo.

Tras estos reajustes, se identificaron un conjunto de valores que ofrecieron un equilibrio adecuado entre rendimiento y robustez en el entorno experimental empleado. Los parámetros empleados finalmente se pueden ver en la tabla 3.2.

Parámetros	Valor
Learning Rate	0.0003
Batch Size	1028
Buffer Size	1000000
Tau	0.005
Hidden Units	256
Number of Layers	2
Normalize	true
Gamma	0.99
Entropy Strength	1.0

Tabla 3.2: Ajustes de parámetros para el algoritmo de SAC.

Adicionalmente, se han implementado técnicas más avanzadas, como el *curriculum learning*, con el objetivo de evaluar su viabilidad y su posible automatización dentro del entorno de simulación desarrollado en Unity. Esta estrategia se ha aplicado para investigar si es posible alcanzar niveles de precisión significativamente más altos, incluso en tareas que presentan una complejidad creciente. Concretamente, el criterio de progresión se basaba en reducir gradualmente el umbral de precisión cada vez que el agente alcanzaba un número determinado de episodios con recompensas consecutivas. Este enfoque parte de la premisa de que resulta más sencillo aprender a alcanzar un objetivo dentro de un margen amplio (por ejemplo, 20 cm) que hacerlo desde el inicio con una tolerancia estricta (1 cm).

La Figura 3.4 muestra el comportamiento del agente SAC en el entorno de precisión simple cuando se aplica la estrategia de *curriculum learning*. Se observa un incremento inicial estable en la recompensa acumulada, seguido de una caída brusca en el rendimiento coincidiendo con una transición en el nivel de dificultad, lo que refleja el ajuste automático del umbral de precisión a medida que el agente mejora. Este patrón evidencia que el agente requiere un periodo de adaptación tras cada reducción del margen de error, lo cual es coherente con los principios del aprendizaje gradual. No obstante, tras esta fase de adaptación, se aprecia una recuperación progresiva y una posterior estabilización del

rendimiento en niveles cercanos al máximo. Estos resultados sugieren que el *curriculum learning* no solo es viable en entornos simulados, sino que permite alcanzar comportamientos precisos de forma más estructurada y eficiente.

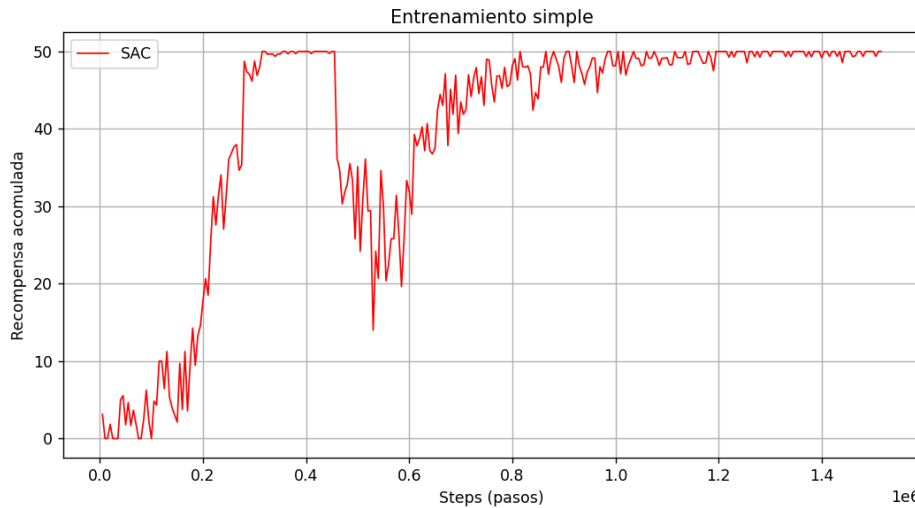


Figura 3.4: Gráfico de tensorboard para el entorno simple con Curriculum Learning.

Los resultados obtenidos en el entorno simple permiten concluir que *ML-Agents* es una herramienta viable para el entrenamiento de agentes en simulaciones robóticas. No obstante, este tipo de entornos presenta una estructura altamente determinista y una tarea constante, lo que limita su capacidad para representar escenarios más cercanos a aplicaciones reales. Por ello, se plantea la necesidad de desarrollar y evaluar entornos de mayor complejidad, en los que tanto la configuración del entorno como los objetivos del agente varíen a lo largo del episodio. En particular, se han diseñado dos entornos adicionales: el primero exige que el agente alcance un punto aleatorio en el espacio al inicio de cada episodio, introduciendo así una fuente de aleatoriedad controlada; el segundo consiste en una versión extendida de este entorno, en la que el agente debe desplazarse desde un punto inicial A hasta un punto final B, lo que implica una planificación secuencial de movimientos. Estos nuevos escenarios permitirán analizar el comportamiento del agente en contextos menos estructurados y evaluar su capacidad para generalizar políticas ante condiciones dinámicas.

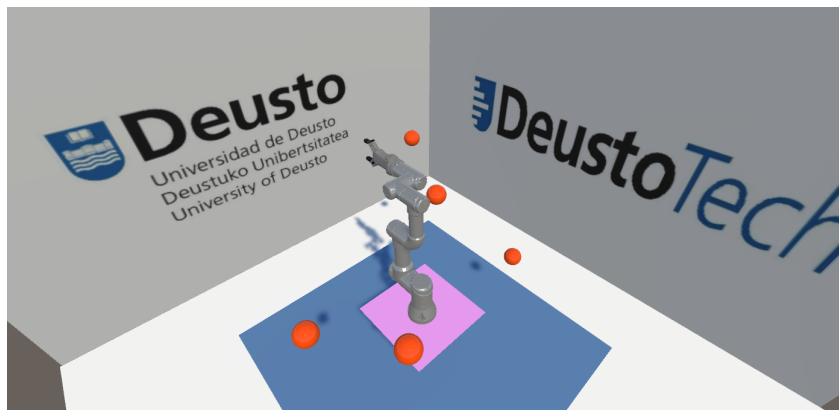


Figura 3.5: Representación del entorno simulado avanzado en Unity, diseñado para la ejecución de tareas con múltiples puntos objetivo y trayectorias.

La Figura 3.5 muestra una representación del entorno avanzado, en el que se han introducido múltiples puntos de destino para facilitar la visualización y comprensión de la tarea. Como primer paso, se llevó a cabo un entrenamiento preliminar de aproximadamente 800.000 pasos con el objetivo de valorar la viabilidad del entorno y comprobar si ambas tareas propuestas (alcanzar un punto aleatorio y recorrer una secuencia de puntos) eran resolubles por parte del agente. Los resultados de esta evaluación se presentan en la Figura 3.6, donde la recompensa máxima se ha establecido en 50, correspondiente a la finalización satisfactoria de la tarea. Aunque no se observa una convergencia totalmente estable (en el sentido de mantener una recompensa constante a lo largo del tiempo), ambos entornos muestran un comportamiento prometedor, ya que el agente consigue alcanzar el objetivo de forma repetida en múltiples ocasiones consecutivas. Esto indica que el entorno es resoluble y puede ser empleado como base para experimentos más prolongados y exigentes.

Es importante destacar que, en comparación con los entornos simples, estas tareas más complejas introducen múltiples factores de variabilidad que dificultan una convergencia clara del entrenamiento. Entre ellos se incluyen la presencia de distintos cuadrantes de trabajo, la aleatoriedad en el punto de aparición del robot y la variabilidad en la dificultad asociada a cada destino, no es equivalente alcanzar un objetivo situado frente al robot que uno ubicado detrás y en una posición baja. Esta falta de uniformidad en las condiciones iniciales y en la complejidad de los episodios hace que la política aprendida no alcance una convergencia estable en el sentido clásico. Sin embargo, como se observa en la Figura 3.7, incluso tras seis millones de pasos, el agente logra completar con éxito la tarea en la mayoría de los episodios, manteniendo un rendimiento elevado aunque variable. Este comportamiento pone de manifiesto la necesidad

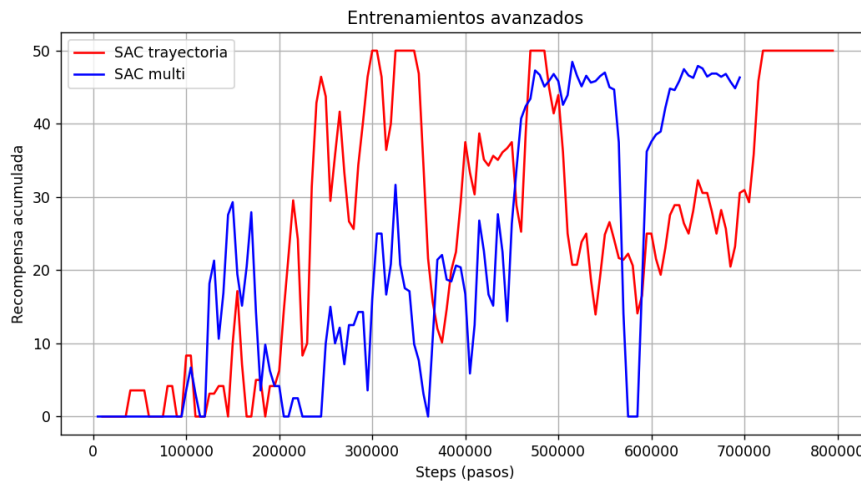


Figura 3.6: Gráfico de tensorboard para el test de entornos avanzados.

de emplear estrategias avanzadas como el *curriculum learning*, entrenamientos prolongados y aproximaciones más sofisticadas en los escenarios que incluirán interacción con humanos, los cuales se abordarán en los apartados siguientes.

Los entrenamientos realizados en los tres entornos simulados permiten confirmar la viabilidad del uso del aprendizaje por refuerzo en escenarios robóticos implementados en Unity. No obstante, los resultados también evidencian que el rendimiento del agente está fuertemente condicionado por diversos factores, como la complejidad de la tarea, la configuración del entorno, la variabilidad introducida y el diseño de la función de recompensa. Por tanto, el éxito del entrenamiento no solo depende del algoritmo utilizado, sino también de un diseño cuidadoso del entorno y de un ajuste progresivo de la dificultad. Como resumen general, en la Tabla 3.3 se recoge el entrenamiento más largo realizado en cada uno de los tres entornos, junto con una métrica de éxito basada en la proporción de episodios completados correctamente en diez ejecuciones aleatorias tras el entrenamiento.

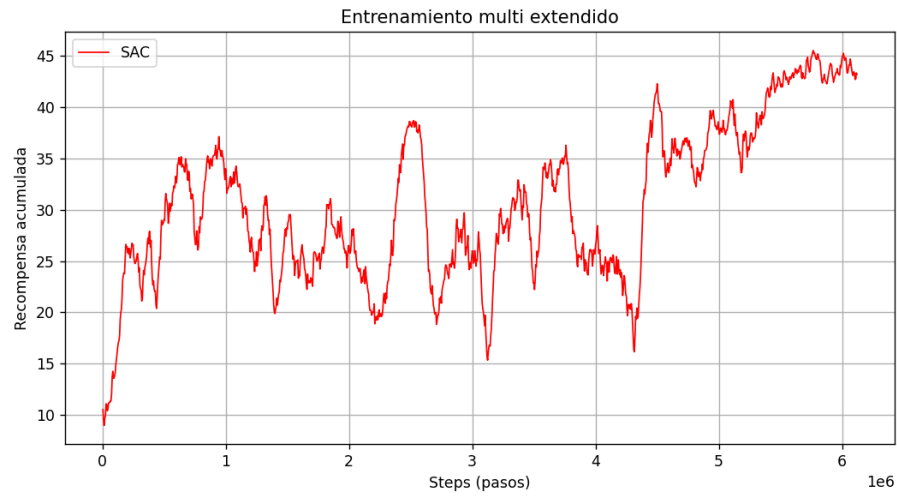


Figura 3.7: Gráfico de tensorboard entorno multi punto extendido. 6 millones de pasos.

Tipo de entorno	Algoritmo	Nº de pasos	Tasa de éxito
Punto fijo	SAC	1M	100 %
Punto aleatorio	SAC	7M	90 %
Trayectorias	SAC	8M	80 %

Tabla 3.3: Resumen del rendimiento obtenido en los distintos entornos simulados tras el entrenamiento.

En conjunto, esta fase experimental confirma que es posible aplicar aprendizaje por refuerzo en simulaciones robóticas desarrolladas con Unity, siempre que se preste atención al diseño del entorno, a la complejidad de las tareas y al ajuste progresivo de los parámetros. A partir del siguiente apartado, se incorporará la participación de gemelos digitales humanos en escenarios con trayectorias complejas, lo que permitirá avanzar hacia formas de interacción más realistas y empezar a evaluar el impacto de señales humanas en la planificación de movimientos colaborativos.

3.3 Entrenamiento en entornos avanzados con adaptación de trayectorias

Con el objetivo de avanzar hacia escenarios más representativos de la interacción humano-robot, esta sección presenta el diseño y la evaluación de entornos simulados que incorporan la presencia de un humano en el espacio de trabajo. A partir de la validación preliminar realizada en entornos simples, se han explorado configuraciones avanzadas en las que la trayectoria del agente debe adaptarse dinámicamente a la posición y el movimiento del humano. Estos escenarios permiten analizar la capacidad del sistema para generar comportamientos flexibles y adaptativos en tiempo real, y constituyen un paso clave hacia la implementación de estrategias colaborativas sensibles al entorno.

Para esta fase y las posteriores, resulta fundamental definir una tarea compleja que permita evaluar no solo la precisión del agente, sino también su capacidad de adaptación a situaciones dinámicas. Con este objetivo, se ha diseñado un escenario en el que el agente parte desde una posición fija inicial y debe completar una secuencia de acciones encadenadas. En primer lugar, debe desplazarse hasta una mesa para recoger una pelota. La recogida se ha implementado de forma sintética: al alcanzar una proximidad determinada con respecto al objeto (umbral de distancia), se considera que la pelota ha sido recogida. A continuación, el agente debe identificar el color de la pelota y trasladarla a una de las dos mesas destino disponibles. La elección correcta depende del color observado: si la pelota es naranja, debe depositarse en una mesa específica (de color naranja); si es rosa, en la otra (de color rosa). La acción de colocación también se considera completada al alcanzar un umbral mínimo de distancia con respecto a la mesa destino.

Finalmente, una vez entregada la pelota en el lugar correspondiente, el agente debe regresar a su posición inicial, cerrando así el ciclo completo de la tarea. Esta estructura de subtarefas de detección del color, planificación de la trayectoria, ejecución de la entrega y retorno, ofrece un contexto adecuado para analizar el rendimiento del sistema en términos de toma de decisiones secuenciales y adaptación ante posibles perturbaciones. La representación visual del entorno puede observarse en la Figura 3.8, donde se aprecian las mesas de recogida y destino junto con el agente robótico en su posición inicial. Además, esta tarea permite estudiar cómo el aprendizaje por refuerzo puede ser utilizado para generar comportamientos adaptativos en entornos colaborativos, especialmente cuando se introduce la presencia de un humano que modifica dinámicamente el entorno de trabajo.

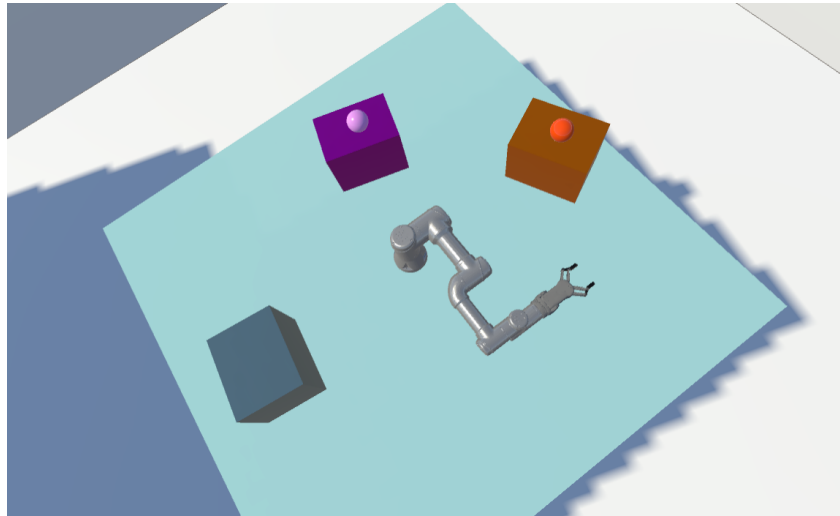


Figura 3.8: Entorno de entrenamiento de tarea compleja. Mesa de recogida (gris) y mesas de colocación (naranja y rosa), con el agente en una posición de reposo.

Con el objetivo de analizar el comportamiento del agente en presencia de humanos, se ha introducido en el entorno de entrenamiento un dummy humano que actúa como gemelo digital. Este modelo permite representar de forma realista la presencia física de una persona en el espacio de trabajo, ofreciendo la posibilidad de evaluar la capacidad del agente para ajustar sus trayectorias en función del comportamiento humano. El dummy se puede situar de forma estática en posiciones aleatorias del entorno, lo que introduce variabilidad espacial y obliga al agente a modificar sus movimientos planificados.

Para ampliar el realismo de la simulación, se han explorado diferentes formas de dotar de movimiento al dummy, con la intención de imitar patrones naturales de desplazamiento humano. Tras valorar distintas alternativas, se ha optado por un modelo de movimiento aleatorio con dirección predominante (por ejemplo, hacia adelante o hacia atrás), combinado con ligeras perturbaciones que emulan los ajustes posturales o cambios de trayectoria propios de una persona en movimiento. Esta estrategia, que incluye además la posibilidad de permanecer estático durante ciertos periodos, permite generar una dinámica fluida y creíble sin necesidad de controles complejos. Así, el entorno resultante se asemeja a un escenario de colaboración real, donde los humanos comparten el espacio de trabajo con el robot y no siguen trayectorias perfectamente predecibles (ver entorno con el dummy en la Figura 3.9).

Dado que este entorno se va a utilizar en múltiples experimentos a lo largo de la presente tesis, se ha diseñado para ser completamente configurable a través

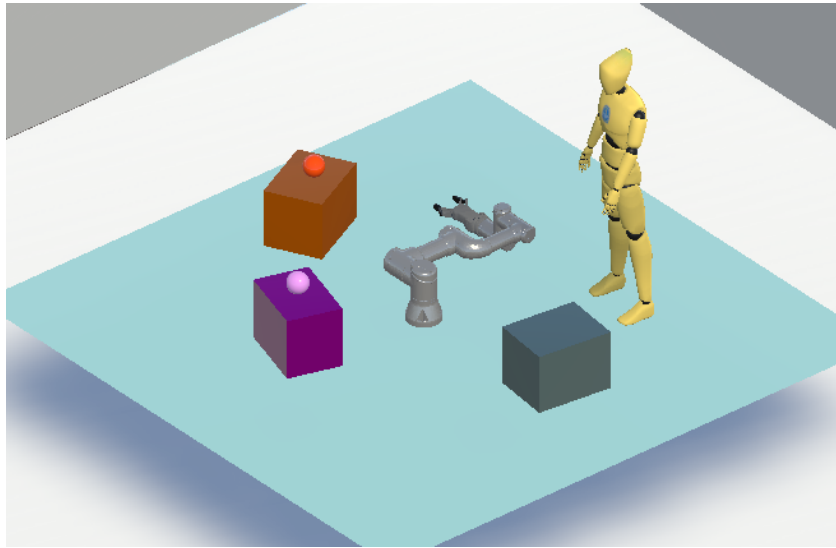


Figura 3.9: Entorno de entrenamiento de tarea compleja, consistente en recoger y dejar una pelota de color específico, en presencia de un dummy humano.

de un archivo externo en formato `.json`. Esta decisión permite adaptar fácilmente las condiciones del entorno sin necesidad de modificar el código fuente, lo que facilita tanto el ajuste de parámetros durante el proceso experimental como su futura reutilización o extensión por parte de otros investigadores. La inclusión de esta capacidad de configuración también responde al objetivo de ofrecer el entorno como recurso abierto a la comunidad científica, promoviendo así la reproducibilidad de los resultados y la exploración de variantes experimentales con un bajo coste de implementación.

Toda la información relativa a la estructura y uso del archivo de configuración se encuentra disponible en el repositorio asociado a esta tesis. No obstante, cabe destacar algunas de las opciones más relevantes que afectan directamente al diseño del entorno. Entre ellas se incluye la activación condicional de fases específicas del comportamiento del agente, así como el control total de la interacción con el humano digital, permitiendo activar o desactivar su presencia, configurar sus posiciones de aparición (*spawn points*) y definir su comportamiento (estático, movimiento aleatorio o teletransportación). Asimismo, el archivo permite establecer con precisión la ubicación de los elementos clave del entorno, como las mesas de recogida y colocación o la posición inicial de la pelota. Estas funcionalidades facilitan la personalización y replicación del entorno de entrenamiento bajo condiciones experimentales diversas. La configuración admite también opciones como el registro de trayectorias, la selección del número de entornos paralelos, el uso de múltiples cámaras o la integración con

sistemas de realidad virtual, dotando al sistema de una gran versatilidad sin necesidad de modificar el código base.

Cuando se hace referencia a la activación de fases específicas del comportamiento del agente dentro del archivo de configuración, se alude a la posibilidad de activar de manera conjunta o individual las tres fases diferenciadas de la tarea: recogida de la pelota, entrega en la mesa correspondiente y retorno a la posición inicial. Esta división modular permite implementar distintas estrategias de entrenamiento según las necesidades del experimento. Por ejemplo, es posible llevar a cabo un entrenamiento conjunto mediante una única red neuronal que abarque todo el ciclo de la tarea, aplicar técnicas de *curriculum learning* en las que se entrena primero una fase y luego se incorporan progresivamente las siguientes, o bien entrenar cada fase de manera independiente y gestionarlas mediante un *switcher*, que selecciona la red adecuada en función de la fase activa. Además, se contempla la posibilidad de utilizar un modelo de decisión más avanzado encargado de determinar dinámicamente qué red debe tomar el control del agente en cada momento, optimizando así la eficiencia y especialización del comportamiento aprendido.

En lo que respecta al diseño de funciones de recompensa, referidas por ahora únicamente a la resolución de la tarea en ausencia de un humano, el entorno configurable permite optar por recompensas de tipo *dense* o *sparse*, según la fase y el tipo de aprendizaje deseado. En la segunda fase de la tarea (la colocación de la pelota en la mesa correspondiente), se ha optado por una recompensa de tipo *sparse*, ya que el objetivo debe cumplirse completamente para recibir señal positiva y resulta conveniente mantener una formulación no guiada, de modo que sea posible observar comportamientos emergentes en futuros entrenamientos. No obstante, el agente dispone de información observacional relativa a su distancia respecto a ambas mesas de destino, y debe aprender a interpretar esta información en función del color de la pelota, desarrollando así un comportamiento condicional.

En cambio, para las fases 1 (recogida) y 3 (retorno a la posición de origen), se ha evaluado el impacto de distintas formulaciones de recompensa densa. En las primeras pruebas se compararon variantes lineales y logarítmicas negativas, observándose que las segundas ofrecían una convergencia más estable y trayectorias más suaves, especialmente en los últimos tramos de aproximación. Finalmente, se ha adoptado la recompensa logarítmica negativa como configuración definitiva, al proporcionar una señal más informativa y favorecer movimientos progresivos y eficientes, particularmente en tareas de precisión como la recogida de objetos.

Con el objetivo de validar la viabilidad del entorno y la tarea propuesta antes de introducir elementos humanos, se han realizado los primeros entrenamientos en condiciones controladas, sin presencia de obstáculos dinámicos. Inicialmente se experimentó con el entrenamiento conjunto de las tres fases (recogida, entrega y retorno), pero se comprobó que esta aproximación presentaba limitaciones: la fase de retorno, a pesar de ser la más sencilla, raramente llegaba a ejecutarse debido a la longitud total de la secuencia de tareas, lo que dificultaba su aprendizaje efectivo. En cambio, una estrategia más eficiente ha consistido en entrenar de forma conjunta las fases 1 (recogida) y 2 (entrega), y abordar la fase 3 (retorno a la posición inicial) como un proceso independiente. En este caso, se simuló el retorno desde posiciones aleatorias próximas a la mesa de entrega, logrando una tasa de éxito del 100 % en aproximadamente 9 millones de pasos. Estos resultados demuestran que la tarea es resoluble de forma eficiente mediante un sistema modular, en el que un *switcher* cambia la red activa al finalizar la entrega, permitiendo una especialización por fases que mejora la estabilidad y la velocidad del aprendizaje.



Figura 3.10: Curva de entrenamiento (fases 1 y 2) sin recompensas negativas hacia el objeto. Aprendizaje más lento y menos estable.

Las Figuras 3.11 y 3.10 ilustran la evolución del aprendizaje durante el entrenamiento conjunto de las fases más complejas y críticas del proceso que son la recogida y entrega (fases 1 y 2) de la pelota. Estas fases son especialmente relevantes porque van a representar los momentos de mayor proximidad al humano en escenarios colaborativos. En ambos casos, no se ha introducido todavía la presencia del humano, pero se han evaluado dos configuraciones distintas del sistema de recompensas introducidas anteriormente. En la Figura 3.10, el en-

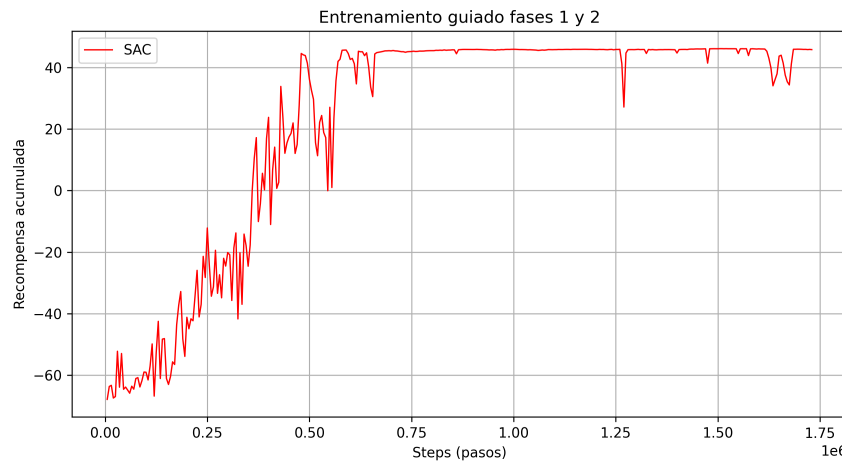


Figura 3.11: Curva de entrenamiento (fases 1 y 2) con recompensa densa en fase 1 y sparse en fase 2. Convergencia más rápida gracias al uso de penalización negativa hacia el objeto.

trenamiento se ha realizado sin incorporar penalizaciones negativas asociadas a la distancia al objeto, dando lugar a una curva de aprendizaje más lenta y menos estable, aunque útil para observar comportamientos emergentes no dirigidos. Por el contrario, la Figura 3.11 muestra la misma tarea con un esquema de recompensas guiado, en el que se penaliza la distancia durante la fase de recogida y se optimiza la entrega para reducir movimientos innecesarios. Este enfoque conduce a una convergencia más rápida y estable, pero tiende a limitar la diversidad de trayectorias, haciendo que el agente repita siempre la ruta más eficiente. En conjunto, ambos resultados validan tanto el diseño del entorno como la eficacia del sistema de recompensas propuesto.

Cabe destacar que las métricas mostradas en ambas gráficas no son directamente comparables en cuanto al valor final alcanzado, ya que los sistemas de recompensas penalizan aspectos distintos, como el número de pasos totales o la distancia al objetivo. Esta diferencia influye en los valores absolutos representados en el eje vertical, pero no afecta a la interpretación general del proceso de aprendizaje. En este caso, el propósito principal es evidenciar que el agente es capaz de resolver la tarea de forma eficiente en un entorno prácticamente determinista (a excepción de la elección de mesa, que depende del color de la pelota), logrando una política convergente en torno a los 2 millones de pasos bajo ambas configuraciones de recompensas.

En cuanto a la fase 3, correspondiente al retorno a la posición inicial, se ha comprobado que puede resolverse de manera muy eficiente cuando se entrena de forma independiente. Aunque no se presenta una figura específica para esta fase, los resultados cuantitativos obtenidos demuestran una tasa de éxito del 100 % en menos de 1 millón de pasos. Estos resultados son concluyentes y suficientes para validar su viabilidad. Este comportamiento también refuerza la idoneidad del enfoque modular planteado.

En conjunto, los entrenamientos realizados en ausencia de presencia humana han demostrado que la tarea es resoluble de forma eficiente mediante una estrategia modular. Dividiendo el proceso en dos fases consecutivas, una inicial que agrupa recogida y entrega, y una posterior dedicada al retorno, se ha logrado obtener un comportamiento exitoso y optimizado en menos de 3 millones de pasos combinados. Esta aproximación fraccionada no solo reduce significativamente el tiempo de entrenamiento en comparación con el enfoque conjunto, sino que además permite especializar redes por subtarea, lo que mejora la estabilidad del aprendizaje. Además, el análisis previo ha permitido identificar que las fases 1 y 2 son especialmente críticas en términos de proximidad al humano, por lo que resultan prioritarias en el estudio de escenarios colaborativos.

Una vez validada la configuración y realizados los primeros entrenamientos en un entorno sin interferencias humanas, se ha procedido a analizar cómo varían los comportamientos ante la introducción del humano digital en posiciones aleatorias del entorno. Este cambio impone una adaptación adicional al agente, que ya no solo debe optimizar la ejecución de las subtarefas definidas, sino también modificar dinámicamente sus trayectorias en función de la localización del humano, manteniendo siempre la integridad de la tarea principal. Como puede observarse en la Figura 3.12, la política aprendida muestra una capacidad notable para redirigir sus movimientos y evitar al humano, incluso en condiciones donde su presencia bloquea parcialmente el camino más directo hacia la mesa objetivo. Estas trayectorias adaptativas confirman la utilidad del esquema de recompensas empleado y evidencian la robustez del comportamiento aprendido.

Cabe destacar que, durante el entrenamiento, se ha observado que penalizar de forma excesiva el contacto con el humano mediante una recompensa fuertemente negativa inducía un comportamiento demasiado conservador por parte del agente, que desarrollaba un rechazo al entorno y evitaba incluso zonas seguras próximas al humano. Por ello, se ha optado por una estrategia alternativa: reiniciar el episodio en caso de colisión sin dar ninguna penalización, lo que impide al agente alcanzar el objetivo y, por tanto, obtener la recompensa positiva. Esta formulación indirecta refuerza que el agente quiera evitar al humano sin inducir bloqueos ni trayectorias artificialmente alejadas; además, ha

demostrado ser eficaz para lograr comportamientos eficientes y sin colisiones.

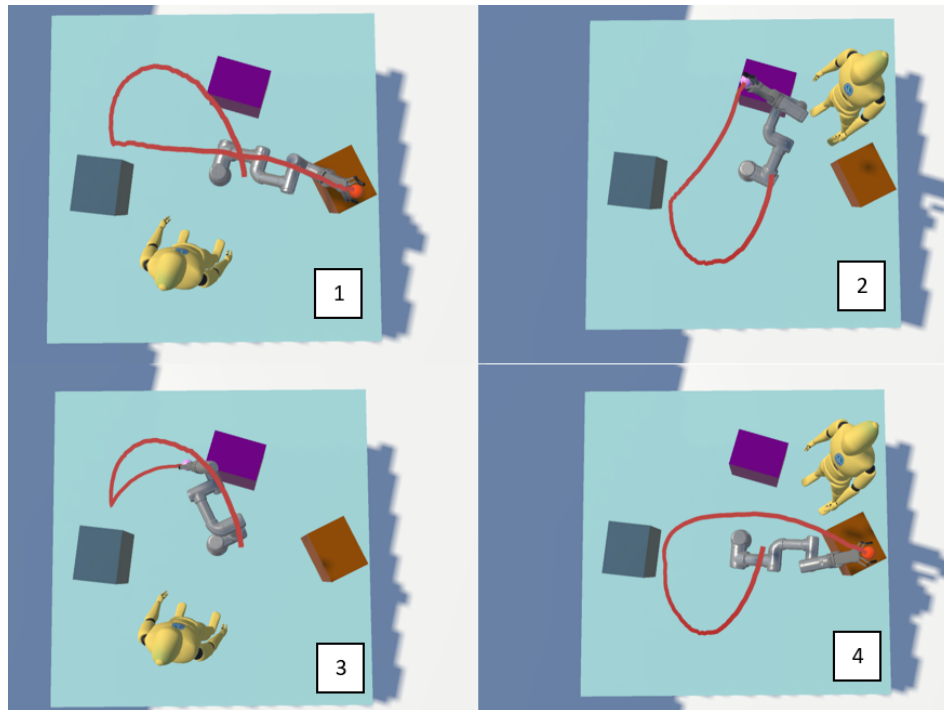


Figura 3.12: Ejemplos de trayectorias generadas en presencia del humano en distintas posiciones.

En los distintos ejemplos representados se aprecia cómo el agente selecciona rutas alternativas, priorizando evitar al humano sin comprometer el éxito de la tarea. Las curvas más pronunciadas o desviaciones laterales respecto a la trayectoria óptima en vacío reflejan la influencia directa de la posición del humano sobre la planificación del movimiento. Esta adaptación se realiza en tiempo real, demostrando la capacidad del modelo para generalizar ante condiciones no vistas durante el entrenamiento inicial.

En cuanto al proceso de entrenamiento con presencia activa del humano, se ha mantenido el enfoque por fases, centrando los experimentos en las etapas 1 y 2. En este caso, la convergencia ha requerido un mayor número de pasos, debido tanto a la complejidad añadida por el movimiento del humano como a la introducción de ajustes específicos en la función de recompensa. Concretamente, se han reajustado los pesos relativos de las recompensas terminales y se ha seleccionado una recompensa densa en la primera fase y *sparse* en la segunda fase, lo que permite priorizar maniobras de evasión sin recurrir a penalizaciones explícitas que podrían inducir comportamientos excesivamente conservadores,

precisamente en la fase de mayor proximidad potencial al humano. Estas modificaciones se implementaron tras un proceso iterativo de pruebas y validación de los entrenamientos, en el que se evaluaron diferentes combinaciones de recompensas y configuraciones del entorno.

Además, se ha empleado una estrategia de *curriculum learning* centrada en dos ejes: por un lado, la precisión requerida para las acciones de recogida y entrega; y por otro, la cantidad y velocidad del movimiento del humano. Esta progresión curricular ha permitido observar periodos de mejora seguidos de pequeñas caídas temporales en el rendimiento, reflejadas en la Figura 3.13, que corresponden con incrementos puntuales en la dificultad del entorno. El límite de recompensa en este escenario se ha fijado en un máximo de 150 puntos, manteniendo la coherencia con los experimentos anteriores.

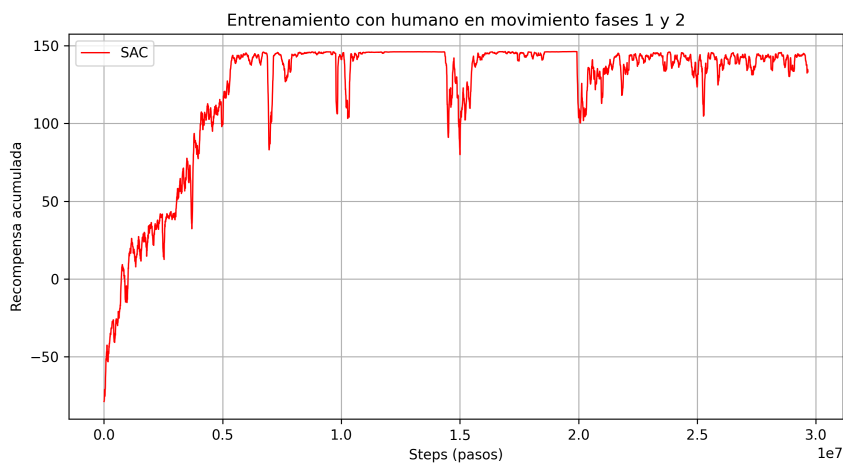


Figura 3.13: Curva de entrenamiento (fases 1 y 2) con movimiento del humano y aplicación de curriculum learning.

Los resultados obtenidos en esta fase permiten concluir que el agente es capaz de mantener un comportamiento eficiente incluso en presencia de un humano en movimiento y sin guía densa durante la fase de entrega. A pesar del incremento de dificultad asociado a la incertidumbre en la posición y desplazamiento del humano, el sistema ha mostrado una notable capacidad de adaptación, resolviendo la tarea de forma consistente tras un periodo de entrenamiento más prolongado. La supresión de la señal de recompensa continua en la segunda fase no solo ha incrementado el desafío de la tarea, sino que ha favorecido el desarrollo de comportamientos más autónomos y generalizables, donde el agente debe inferir la mesa de destino adecuada y planificar trayectorias se-

guras sin depender de una orientación directa. Las trayectorias aprendidas se caracterizan por desviaciones estratégicas que minimizan el riesgo de colisión sin comprometer el éxito de la tarea, evidenciando la robustez, estabilidad y coherencia del comportamiento adquirido.

En este contexto, el sistema de recompensas ha desempeñado un papel fundamental en la estabilidad y eficiencia del aprendizaje. Se ha optado por un esquema basado en reforzadores positivos acumulativos por fase completada, combinados con recompensas y penalizaciones continuas proporcionales a la distancia en las fases 1 y 3. Además el utilizar la penalización por paso, utilizada de forma controlada, ha incentivado la optimización temporal sin inducir bloqueos. Asimismo, se ha evitado el uso de castigos explícitos por interacción con el humano, en favor de mecanismos de refuerzo indirecto como el reinicio del episodio, lo que ha permitido mantener un equilibrio entre exploración, seguridad y rendimiento.

Cabe destacar que las funciones de recompensa y penalizaciones asociadas se pueden cambiar desde el archivo de configuración del entorno, no obstante los resultados obtenidos se han llevado a cabo con las siguientes funciones de recompensa finales implementadas en el entorno de entrenamiento:

En la primera fase (recogida), el agente recibe una recompensa negativa continua que disminuye de forma logarítmica a medida que se aproxima al objeto. Este diseño favorece movimientos más estables y progresivos, ya que la penalización es menor cuando la distancia al objetivo es pequeña. La finalización de la fase se produce cuando la distancia entre el efector final y la pelota (d_{catch}) es menor o igual que la precisión establecida ($d_{\text{precision}}$), cumpliéndose así la condición de éxito definida en la Ecuación 3.1.

$$F_{\text{task1}} = \begin{cases} +50, \text{ acabar fase 1} & \text{si } d_{\text{catch}} \leq d_{\text{precision}}, \\ -\log(d_{\text{catch}} + 1)/10, & \text{en otro caso.} \end{cases} \quad (3.1)$$

En la segunda fase (entrega), la señal de recompensa se define como *sparse*: el agente solo recibe una recompensa positiva al depositar correctamente la pelota en la mesa correspondiente según su color. La condición de éxito se cumple cuando la distancia entre el efector final y la posición objetivo de la mesa (d_{place}) es menor o igual que la precisión establecida ($d_{\text{precision}}$), tal y como se muestra en la Ecuación 3.2. En caso de error o de no completar la acción, no se obtiene recompensa.

$$F_{\text{task2}} = \begin{cases} +100, \text{ acabar fase 2} & \text{si } d_{\text{place}} \leq d_{\text{presition}}, \\ 0, & \text{en otro caso.} \end{cases} \quad (3.2)$$

Por último, en la fase de retorno, se emplea una formulación logarítmica análoga a la utilizada en la primera fase. La fase concluye cuando la distancia entre el efector final y la posición inicial del robot (d_{home}) es menor o igual que la precisión establecida ($d_{\text{presition}}$), cumpliéndose así la condición de finalización definida en la Ecuación 3.3.

$$F_{\text{task3}} = \begin{cases} +150, \text{ acabar fase 3} & \text{si } d_{\text{home}} \leq d_{\text{presition}}, \\ -\log(d_{\text{home}} + 1)/10, & \text{en otro caso.} \end{cases} \quad (3.3)$$

La Ecuación 3.4 resume la formulación general de la función de recompensa total empleada en el entorno final, donde cada término $F_{\text{task}i}$ representa la contribución de las tres fases principales de la tarea. Los parámetros $k_1, k_2, k_3 \in \{0, 1\}$ permiten activar o desactivar cada componente según el escenario y fase de entrenamiento. Adicionalmente, se incluyen las penalizaciones R_{step} y $R_{\text{collision}}$, que pueden configurarse para ajustar la eficiencia y la robustez del aprendizaje. En el caso concreto de esta investigación, la penalización por colisión se ha establecido en cero. Por su parte, la penalización temporal R_{step} puede adoptar un valor reducido (por ejemplo, 0.1 por paso) con el fin de agilizar la convergencia del entrenamiento y promover trayectorias más eficientes.

$$R_{\text{total}} = k_1 \cdot F_{\text{task1}} + k_2 \cdot F_{\text{task2}} + k_3 \cdot F_{\text{task3}} - R_{\text{step}} - R_{\text{collision}} \quad (3.4)$$

Para favorecer una progresión estable del aprendizaje, se aplica un esquema de *curriculum learning* que ajusta gradualmente la dificultad del entorno. En las primeras etapas se utilizan precisiones mas altas y movimientos del humano menos bruscos.

En síntesis, los experimentos realizados han validado la viabilidad de un entorno avanzado en el que el agente completa tareas complejas de forma eficiente, incluso bajo la presencia de un humano en movimiento. El enfoque modular planteado ha facilitado la especialización de políticas en subtareas diferenciadas, mejorando tanto la estabilidad como la velocidad de convergencia. Asimismo, se ha confirmado que la inclusión de un modelo humano, ya sea estático o dinámico, no compromete la ejecución de la tarea siempre que el diseño de

recompensas esté cuidadosamente formulado.

Con el fin de validar los comportamientos aprendidos en condiciones más cercanas a la interacción real, se ha generado un entorno de realidad virtual (VR) que permite a investigadores y usuarios introducir de manera directa las redes entrenadas y observar su rendimiento en situaciones simuladas pero inmersivas. Este escenario ofrece la posibilidad de reproducir movimientos humanos y gestos naturales, evaluando la capacidad del agente para adaptarse a dinámicas más complejas y variadas que las consideradas en los entornos de simulación tradicional.

El uso de VR no solo facilita la visualización y análisis de las trayectorias robóticas, sino que también habilita la interacción directa del usuario con el entorno. De este modo, los participantes pueden experimentar en primera persona la colaboración con el agente, aportando observaciones cualitativas y valoraciones inmediatas sobre aspectos como la fluidez de los movimientos, la seguridad percibida o el nivel de naturalidad en la interacción. Para ilustrar este planteamiento, se han incluido capturas representativas del entorno virtual (ver Figura 3.14).

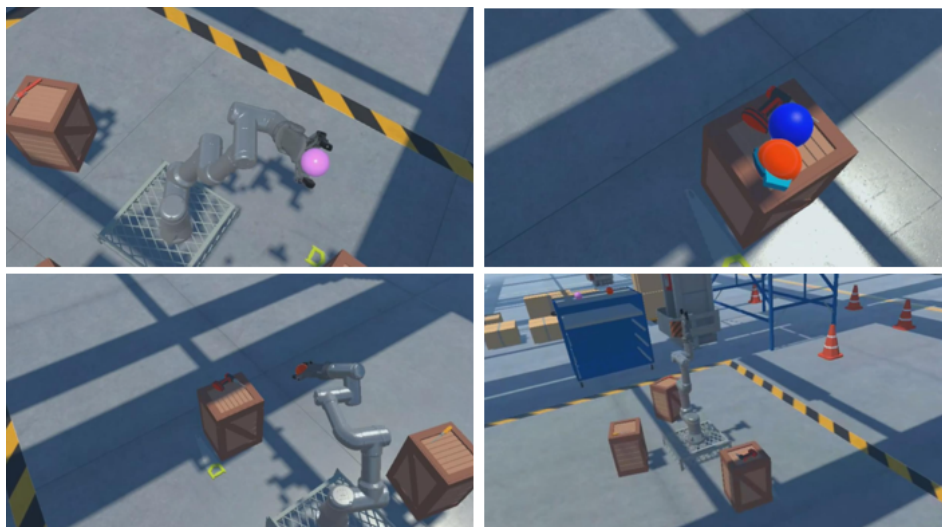


Figura 3.14: Vista general del escenario industrial en realidad virtual diseñado para la validación y análisis de los comportamientos aprendidos por el agente

El entorno de realidad virtual desarrollado no solo se ha planteado como un complemento ilustrativo, sino que ha sido utilizado para la detección temprana de limitaciones y la validación de los modelos entrenados. Gracias a la flexibilidad y seguridad que aporta, el entorno de VR ha permitido evaluar el compor-

tamiento de los agentes en escenarios colaborativos con movimiento humano realista, confirmando la solidez de los comportamientos aprendidos e identificando al mismo tiempo ajustes necesarios antes de seguir modificando el entorno y las funciones de recompensa previas al despliegue. De este modo, se consolida como una etapa intermedia clave entre la simulación computacional y el despliegue de los modelos en robots reales.

A pesar de estos resultados positivos, también se han identificado limitaciones relevantes. Hasta este punto, la interacción humano-robot se ha planteado únicamente como un problema de seguridad física, en el que el objetivo principal del agente consiste en evitar el contacto con el humano. Dicho enfoque, aunque necesario como requisito básico de cualquier escenario colaborativo, resulta insuficiente para representar una interacción real, donde influyen factores adicionales como la percepción subjetiva de confort, la naturalidad de los movimientos o la capacidad del robot para anticiparse a las intenciones del usuario.

Esta limitación pone de manifiesto la necesidad de avanzar hacia un modelo más sofisticado de interacción, en el que el confort humano se considere de forma explícita dentro del proceso de aprendizaje. En esta dirección, se ha planteado la incorporación de métricas fisiológicas y valoraciones subjetivas que permitan formalizar el confort como variable cuantificable y operativa. De este modo, el agente no solo ejecuta la tarea de manera segura, sino que también adapta sus comportamientos para optimizar la experiencia de colaboración desde la perspectiva del usuario humano.

El entorno configurable desarrollado en los capítulos anteriores se ha planteado como base para las siguientes fases de la investigación, integrando funciones de recompensa que contemplen dichas dimensiones adicionales de la interacción. De esta forma, la infraestructura presentada servirá como punto de partida para abordar el modelado del confort humano y su incorporación efectiva en el entrenamiento del agente, cuestión que se detalla en el próximo Capítulo 4.

La lentitud no es falta de ritmo; es dominio del ritmo.

RENÉ LAVAND

Capítulo

4

Modelado del confort en la interacción humano-robot

En el capítulo anterior se ha validado la capacidad del agente para interactuar en entornos avanzados con presencia humana, garantizando la seguridad básica evitando colisiones con el dummy humano. Este planteamiento, aunque imprescindible como condición mínima en cualquier escenario colaborativo contemporáneo, resulta insuficiente para reflejar la complejidad que representa una interacción humano-robot actual (Sheridan, 2016; Villani et al., 2018). La colaboración efectiva no puede limitarse únicamente a la ausencia de contacto físico, sino que debe integrar dimensiones adicionales relacionadas con la percepción y el estado psicológico del usuario (Hancock et al., 2011; Khavas, 2021). En este contexto, adquiere especial relevancia la noción de confort, entendida como un componente esencial para evaluar la calidad de la interacción y orientar el diseño de comportamientos robóticos más adaptativos y aceptables (Bibbo et al., 2025).

La literatura en interacción humano-robot ha señalado que factores como la distancia interpersonal, la velocidad de los movimientos o la fluidez de las trayectorias influyen de manera directa en la percepción de confort del usuario (Pacchierotti et al., 2006; Mumm y Mutlu, 2011). Estos parámetros condicionan tanto la aceptación del sistema como la confianza y disposición del humano a colaborar con el robot (Hancock et al., 2011; van den Brule et al., 2014). El confort, entendido como una experiencia que combina dimensiones físicas y psicológicas, se establece así como un requisito fundamental para garantizar interacciones sostenibles y eficaces, especialmente en escenarios donde la coope-

ración se mantiene en el tiempo o implica un contacto frecuente con el robot (Howell et al., 2023).

En este capítulo se plantea formalizar el confort como una variable explícita dentro de un entorno de aprendizaje por refuerzo, trasladando este concepto subjetivo a parámetros observables que puedan integrarse en el proceso de entrenamiento (Retzlaff et al., 2024; Slade et al., 2024). Para ello, se ha extendido el entorno configurable descrito en capítulos anteriores, incorporando nuevas funciones de recompensa que permiten equilibrar la resolución de la tarea con el mantenimiento de condiciones de interacción acordes al nivel de confort definido. Este planteamiento posibilita analizar cómo el agente adapta sus trayectorias y comportamientos en función de distintos umbrales de confort, ofreciendo así un marco operativo para estudiar la influencia de esta dimensión en escenarios colaborativos. Los aspectos relativos a la definición del confort, su modelado matemático, la configuración experimental y el análisis comparativo de los distintos niveles se desarrollan en detalle en las secciones posteriores.

El objetivo principal de este capítulo es doble. En primer lugar, se ha buscado definir y modelar el confort como una variable operativa que pueda incorporarse de forma explícita en el proceso de entrenamiento del agente. En segundo lugar, se ha pretendido demostrar cómo la introducción de esta dimensión modifica el comportamiento aprendido, comparando ejecuciones sin consideración del confort, con el confort activado y con diferentes niveles de exigencia asociados a dicha variable. De esta forma, resulta posible evaluar no solo la viabilidad de integrar el confort en el marco del aprendizaje por refuerzo, sino también la relevancia práctica de su incorporación en la calidad y aceptación de la interacción humano-robot.

En este sentido, el trabajo desarrollado en este capítulo sienta las bases para fases posteriores de la investigación en las que se incorporarán métricas adicionales y validaciones con usuarios reales. La formalización del confort en un entorno simulado constituye un paso intermedio necesario, que permite consolidar un marco metodológico robusto antes de abordar escenarios reales de mayor complejidad. De este modo, los avances aquí presentados no solo muestran la viabilidad de considerar el confort como variable de entrenamiento, sino que también preparan el terreno para su posterior aplicación en contextos experimentales presenciales, donde la percepción del usuario humano adquiere un papel central.

4.1 Definición y representación del confort en la interacción humano-robot

El confort se ha consolidado en los últimos años como una dimensión clave en la interacción humano-robot, especialmente en escenarios de colaboración en los que el robot comparte espacio de trabajo con una persona (Villani et al., 2018; International Organization for Standardization, 2016). A diferencia de la seguridad, que se centra en evitar daños físicos derivados de la interacción (Sheridan, 2016), el confort abarca aspectos más amplios vinculados a la percepción subjetiva del usuario y a la aceptación de la tecnología (Goodrich y Schultz, 2007; Hancock et al., 2011). La consideración explícita de este concepto resulta fundamental para garantizar que la cooperación entre humano y robot no solo sea físicamente segura y funcional, sino también percibida como natural y satisfactoria, condición indispensable para favorecer la confianza y la disposición a colaborar de manera sostenida en el tiempo Redondo et al. (2023).

El confort en la interacción humano-robot puede entenderse como una experiencia de varias dimensiones que combina factores físicos, cognitivos y emocionales. Desde una perspectiva física, aspectos como la distancia interpersonal o la velocidad de los movimientos condicionan directamente la sensación de comodidad percibida por el usuario (Pacchierotti et al., 2006; Mumm y Mutlu, 2011). En el plano cognitivo, variables como la inteligibilidad (*predictability* en inglés) de las trayectorias del robot influyen en el grado de naturalidad atribuido a la interacción (Dragan et al., 2013). Finalmente, en la dimensión emocional, estudios recientes han señalado que la confianza y la percepción de seguridad subjetiva desempeñan un papel decisivo en la aceptación de la colaboración a largo plazo (Hancock et al., 2011; Khavas, 2021). Esta visión integradora permite concebir el confort como un componente esencial para evaluar la calidad de la interacción, complementando métricas clásicas de seguridad y eficiencia (Gonzalez-Santocildes et al., 2024b) y ha comenzado a explorarse también mediante modelos cuantitativos que predicen o estiman el confort en función de parámetros observables (Yan et al., 2023; Howell et al., 2023).

Diversos estudios han identificado un conjunto de factores que influyen de manera recurrente en la percepción de confort durante la interacción con robots colaborativos Yan y Jia (2022). Entre los más citados se encuentran la distancia interpersonal, la velocidad y fluidez de los movimientos, así como la capacidad del robot para anticipar las acciones humanas (Pacchierotti et al., 2006; Mumm y Mutlu, 2011). Estos elementos, si bien han sido analizados de forma independiente en distintos trabajos, conforman un marco común que permite comprender cómo las características del comportamiento robótico inciden en la experiencia subjetiva del usuario. En consecuencia, el confort puede abordarse

no como una variable aislada, sino como el resultado de la interacción dinámica entre múltiples dimensiones físicas y cognitivas, cuya adecuada integración resulta esencial para el diseño de sistemas colaborativos aceptables y eficaces.

La distancia interpersonal se ha seleccionado en este trabajo como una de las variables principales para modelar el confort, al ser uno de los factores más mencionados en la literatura y, al mismo tiempo, por su relevancia práctica en la tarea planteada. Esta variable puede medirse de forma explícita a través de las trayectorias generadas por el agente, observando la distancia media mantenida entre el robot y el humano. En el enfoque propuesto, este valor se ajustará dinámicamente en función del nivel de confort declarado por el usuario, de modo que el robot incremente o reduzca su proximidad para favorecer la percepción de comodidad del usuario.

La velocidad de movimiento se ha seleccionado también como variable principal en el modelado del confort, al tratarse de un factor destacado en el estado del arte y directamente vinculable a la percepción de comodidad y seguridad en la interacción. Su control puede implementarse de manera explícita mediante la regulación del incremento angular máximo permitido en cada paso de simulación, lo que permite modular la fluidez y rapidez de los desplazamientos del robot. Al igual que la distancia, la velocidad se adaptará en función del nivel de confort reportado, de forma que la interacción mantenga un ritmo coherente con las necesidades del usuario.

Aunque en la literatura se han propuesto otras métricas relevantes para evaluar el confort en la interacción humano-robot, como la inteligibilidad de las trayectorias o la generación de movimientos con características human-like (Gulletta et al., 2021), estas no se han incorporado en el presente trabajo. La razón principal es que dichas métricas no se ajustan directamente al diseño experimental planteado. En este caso, el objetivo es disponer de variables cuantificables y fácilmente integrables en la función de recompensa, lo que justifica la elección de parámetros físicos como la distancia y la velocidad. Cabe señalar, no obstante, que la robustez inherente al entrenamiento mediante aprendizaje por refuerzo permite que el agente desarrolle comportamientos adaptativos frente a estados no previstos, contribuyendo indirectamente a mejorar la previsibilidad y naturalidad de sus movimientos.

En consecuencia, se determina que las variables con mayor influencia sobre la percepción de confort en el contexto experimental considerado son la distancia y la velocidad del robot con respecto al usuario. Ambas ofrecen un marco de control explícito, cuantificable y coherente con la tarea diseñada, y permiten analizar de forma sistemática cómo diferentes umbrales de proximidad y

dinámica de movimiento impactan en la experiencia del colaborador humano. Estas dos dimensiones constituyen, por tanto, la base sobre la que se construirá el modelado matemático del confort presentado en el siguiente apartado.

Tabla 4.1: Resumen de variables consideradas para la función de recompensa asociada al confort.

Variable	Estado	Justificación
Distancia	Seleccionada	Factor ampliamente mencionado en la literatura y relevante para la tarea; control explícito mediante trayectorias.
Velocidad	Seleccionada	Relacionada con la percepción de comodidad y seguridad; controlable mediante incremento angular máximo por paso.
Inteligibilidad	No seleccionada	Relevante en otros contextos, aunque no modelada de forma explícita. Puede emerger indirectamente gracias a la robustez del entrenamiento en RL y a la consistencia de las trayectorias aprendidas.
Movimientos human-like	No seleccionada	Contribuyen a la naturalidad de la interacción, aunque no se han incorporado como variable específica. El entrenamiento en RL favorece comportamientos adaptativos que pueden aproximarse a este efecto.

Como se resume en la Tabla 4.1, la distancia y la velocidad se consideran no solo como parámetros de control del agente, sino como variables directamente vinculadas al confort humano. Su análisis conjunto permitirá evaluar, en el contexto experimental planteado, si la adaptación del robot a diferentes niveles de confort contribuye a generar interacciones más seguras y percibidas como naturales.

En el marco del aprendizaje por refuerzo, estas variables seleccionadas deben formalizarse dentro de la función de recompensa, que constituye el mecanismo principal a través del cual el agente aprende a ajustar su comportamiento. En el enfoque propuesto, el nivel de confort reportado por el usuario actuará como señal de entrada, de modo que la distancia y la velocidad se adapten dinámicamente para mantener la interacción en condiciones acordes a las necesida-

des del colaborador humano. De este modo, el confort deja de ser únicamente una medida descriptiva para convertirse en un componente operativo del entrenamiento, capaz de influir directamente en las políticas aprendidas por el robot.

En síntesis, el planteamiento desarrollado en este apartado parte de la consideración del confort como un estado dinámico que puede ser reportado por el usuario en una escala cuantitativa. La intención es que dicho nivel de confort actúe como señal de entrada para el sistema, permitiendo que el agente ajuste de manera adaptativa parámetros como la distancia y la velocidad en función de las necesidades actuales del colaborador humano. Esta aproximación abre la puerta a un marco de interacción más flexible, donde el comportamiento del robot no solo persigue la resolución eficiente de la tarea, sino también la optimización continua de la experiencia del usuario. En el siguiente apartado se presenta el modelado matemático de la función de recompensa, que integra de manera explícita estas variables con el objetivo de formalizar cómo diferentes niveles de confort pueden influir en el comportamiento aprendido por el agente.

4.2 Modelado matemático de la función de recompensa asociada al confort

El aprendizaje por refuerzo se fundamenta en la definición de una función de recompensa que orienta la optimización del comportamiento del agente. En el contexto de la interacción humano-robot, esta función no puede limitarse únicamente al cumplimiento de la tarea encomendada, sino que debe integrar dimensiones adicionales que condicionan la aceptación y la calidad de la colaboración. Entre estas dimensiones, el confort humano se reconoce como un factor decisivo, lo que hace necesario formalizarlo matemáticamente para que pueda incorporarse de manera explícita en el proceso de entrenamiento. En este apartado se presenta el modelado de la función de recompensa asociada al confort, considerando como variables principales la distancia y la velocidad del robot con respecto al usuario. Con este enfoque, el confort deja de ser un concepto meramente descriptivo y se convierte en un componente operativo del entrenamiento, capaz de influir en la adaptación del comportamiento del agente en función del estado del colaborador humano.

La estructura general de la función de recompensa se muestra en la Ecuación 4.1 y sigue el planteamiento de la Ecuación 3.4. En ella se combinan, por un lado, los términos asociados a las tareas principales que debe realizar el agente y, por otro, un término vinculado al confort. De este modo, el agente recibe una recompensa positiva por completar las subtarear planteadas (por ejemplo, la

recogida de la pelota, su colocación en la mesa correspondiente o el retorno al punto inicial), mientras que la función de confort actúa reduciendo la recompensa total en caso de que el comportamiento del robot no se ajuste al nivel de confort esperado por el usuario.

$$R = k_1 \cdot F_{\text{task1}} + k_2 \cdot F_{\text{task2}} + k_3 \cdot F_{\text{task3}} - R_{\text{step}} - R_{\text{collision}} - F_{\text{confort}}(C_l, N_c) \quad (4.1)$$

En esta expresión, $k_1, k_2, k_3 \in \{0, 1\}$ son parámetros de activación que permiten decidir si cada una de las funciones de tarea contribuye a la recompensa total en función del escenario de entrenamiento. El término $F_{\text{confort}}(C_l, N_c)$ representa la función de confort, que depende tanto del número total de niveles de confort definidos por el usuario (N_c) como del nivel concreto en el que este se encuentre en un momento dado (C_l). La función de confort $F_{\text{confort}}(C_l, N_c)$ se ha definido como la *suma ponderada* de dos componentes principales, F_{conf1} y F_{conf2} , que se aprecia en la Ecuación 4.2 y representan los distintos aspectos del confort percibido por el usuario.

$$F_{\text{confort}}(C_l, N_c) = F_{\text{conf1}}(C_l, N_c) + F_{\text{conf2}}(C_l, N_c) \quad (4.2)$$

El nivel de confort C_l es un valor entero que indica el grado de tolerancia del usuario frente a interacciones cercanas o movimientos rápidos del robot. A menor valor de C_l , el confort reportado es más bajo, lo que implica una mayor sensibilidad del usuario a la proximidad o a la velocidad; por el contrario, un valor elevado refleja una mayor tolerancia. En la práctica, C_l actúa como una señal de entrada al sistema, lo que permite al robot adaptar dinámicamente su comportamiento en función del estado de confort estimado o proporcionado por el usuario.

El modelado del confort se ha realizado a partir de dos variables seleccionadas en el apartado anterior, la distancia entre el robot y el humano, y la velocidad de los movimientos del robot. Cada una de estas dimensiones se asocia a una función matemática específica, de forma que la combinación de ambas genera la función total de confort. En primer lugar, se define el componente relacionado con la distancia (Ecuación 4.3). Esta función se apoya en una formulación basada en la tangente hiperbólica, seleccionada tras evaluar y comparar su rendimiento frente a funciones lineales, exponenciales y logarítmicas. El uso de la tangente hiperbólica proporciona una transición más suave entre la penalización y la recompensa que las alternativas probadas, evitando saltos abruptos y

gestionando de manera eficaz los valores umbral.

$$F_{\text{conf}}(C_l, N_c) = -0,1 + \tanh\left(\frac{d - \theta_{\text{thr}}(C_l, N_c)}{\text{máx}(d, \varepsilon)} \cdot (\text{máx}(d_{\text{máx}} - d, \varepsilon))^{-1}\right) \quad (4.3)$$

En esta expresión, d representa la distancia entre el robot y el humano, mientras que $d_{\text{máx}}$ define la distancia máxima considerada óptima. El término $\theta_{\text{thr}}(C_l, N_c)$ corresponde al umbral que determina cuándo un valor de distancia debe considerarse cómodo o incómodo, calculado dinámicamente en función del nivel de confort del usuario y del número de niveles definidos. Finalmente, se añaden factores de estabilización numérica ($\varepsilon = 10^{-10}$) únicamente para evitar divisiones por cero o multiplicaciones indeterminadas. La constante $-0,1$ del inicio introduce una penalización base que el agente debe compensar manteniendo distancias adecuadas, lo que favorece trayectorias más seguras y coherentes con las preferencias del usuario.

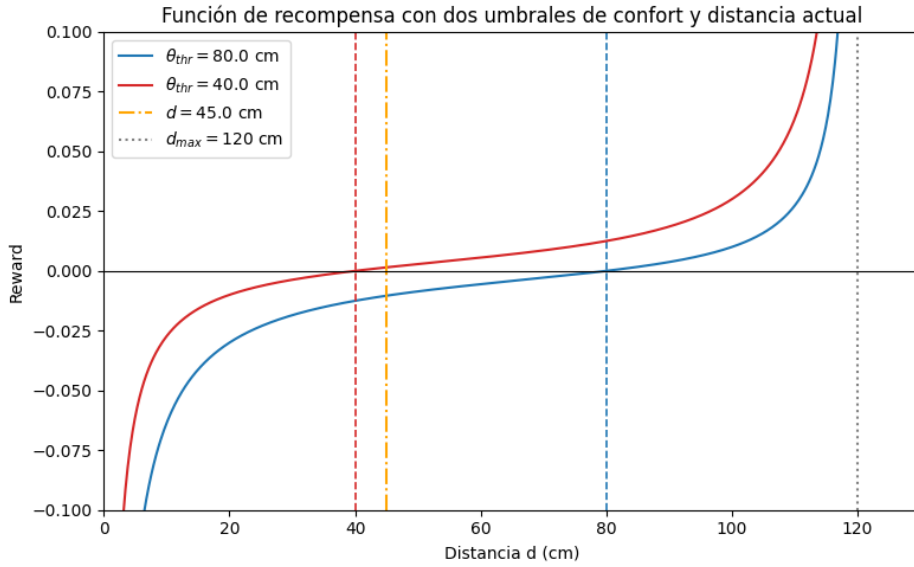


Figura 4.1: Gráfica de ejemplo del argumento de la tangente hiperbólica de la función de recompensa basada en distancia geométrica para dos niveles de confort del usuario. $\theta_{\text{thr}} = 40\text{cm}$ en rojo y $\theta_{\text{thr}} = 80\text{cm}$ en azul

La Figura 4.1 ilustra el comportamiento del argumento de la tangente hiperbólica dentro de la función de recompensa $F_{\text{conf}}(C_l, N_c)$ para dos niveles de confort representativos dentro de un rango definido entre 0 y 120 cm. En este

caso, un nivel de confort bajo ($C_l = 4$) se asocia a un umbral de confort más amplio ($\theta_{thr} = 80$ cm), lo que implica que el usuario prefiere mantener una mayor distancia con el robot. Como consecuencia, a una distancia de 45 cm el valor de la función es negativo, penalizando la proximidad y promoviendo que el robot se aleje. Por el contrario, un nivel de confort alto ($C_l = 7$) corresponde a un umbral más reducido ($\theta_{thr} = 40$ cm), reflejando que el usuario tolera o incluso prefiere interacciones más cercanas. En este caso, la misma distancia de 45 cm produce una recompensa cercana a cero o positiva, permitiendo que el robot opere en una zona más próxima sin penalización. Además, puede observarse que cuando la distancia tiende a 0 o se aproxima al valor máximo definido, la función converge hacia -1 o 1 respectivamente. A partir del límite máximo, el valor de la tangente hiperbólica se mantiene constante en un valor de 1. Cabe destacar que el desplazamiento de $-0,1$ incluido en la formulación original tiene como finalidad optimizar la estabilidad del proceso de aprendizaje, pero no se representa en la Figura 4.1 para mantener la claridad de la figura, la función completa $F_{conf2}(C_l, N_c)$ con el umbral se aprecia en la Figura 4.2.

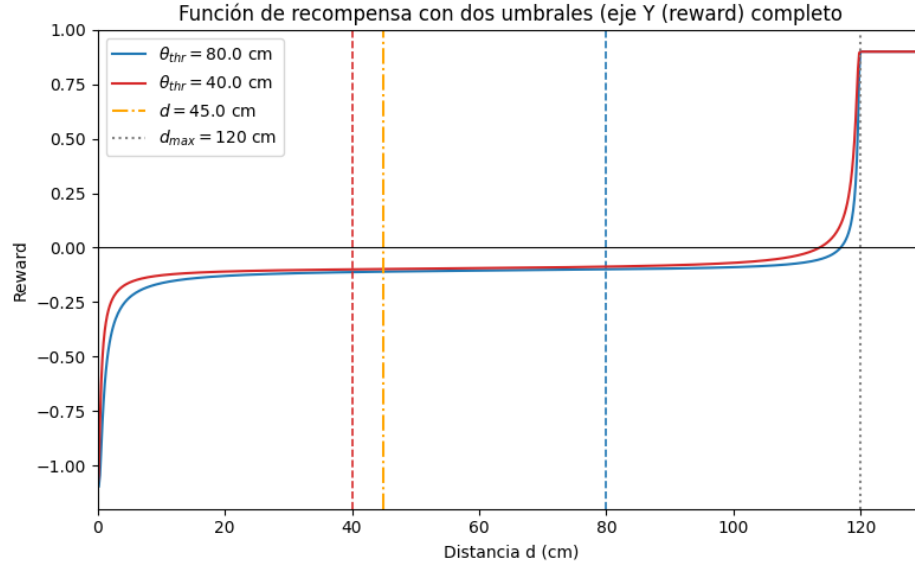


Figura 4.2: Función completa de recompensa basada en la distancia geométrica con el eje Y (reward) completo $F_{conf}(C_l, N_c)$. $\theta_{thr} = 40cm$ en rojo y $\theta_{thr} = 80cm$ en azul

El segundo componente del confort, $F_{conf2}(C_l, N_c)$, se centra en la velocidad de los movimientos del robot, representada por el incremento angular máximo de las articulaciones en cada paso de simulación. Su formulación se detalla en la Ecuación 4.4, donde se establece una penalización cuando el incremento an-

gular supera el umbral definido en función del nivel de confort.

$$F_{\text{conf2}}(C_l, N_c) = \begin{cases} -0,1 & \text{si } \Delta\alpha > \Delta\alpha_{\text{thr}}(C_l, N_c) \\ 0 & \text{en otro caso} \end{cases} \quad (4.4)$$

En esta fórmula, $\Delta\alpha$ indica el incremento angular realizado en un paso, mientras que $\Delta\alpha_{\text{thr}}(C_l, N_c)$ corresponde al umbral máximo permitido para ese nivel de confort. De este modo, en niveles bajos de confort se aplican restricciones más severas a la rapidez de los movimientos, penalizando aquellos que resulten demasiado bruscos. Por el contrario, en niveles altos de confort el umbral se amplía, permitiendo movimientos más veloces sin que ello genere incomodidad en el usuario.

Un ejemplo práctico ayuda a entender la dinámica de esta función. Supóngase que el robot realiza un movimiento con un incremento angular de $0,85^\circ$ en un escenario con $N_c = 10$. Si el nivel de confort del usuario es bajo ($C_l = 4$), el umbral será reducido y la función devolverá un valor negativo, indicando que el movimiento ha sido demasiado rápido. Si, en cambio, el nivel de confort es más alto ($C_l = 6$), el umbral es mayor y la función devolverá 0, permitiendo ese mismo movimiento sin penalización. De este modo, el robot puede adaptar la fluidez de sus trayectorias a las preferencias del usuario.

Ambos componentes se integran finalmente en una única función de confort, combinando la influencia de la distancia y la velocidad mediante factores de ponderación. La Ecuación 4.5 muestra la formulación general:

$$F_{\text{comfort}}(C_l, N_c) = \alpha \cdot F_{\text{conf1}}(C_l, N_c) + \beta \cdot F_{\text{conf2}}(C_l, N_c) \quad (4.5)$$

Los coeficientes α y β permiten ajustar la relevancia relativa de cada componente, otorgando mayor peso a la distancia o a la velocidad según las características de la tarea o las necesidades específicas del experimento. En la práctica, esta flexibilidad facilita el ajuste fino de la función de recompensa y abre la posibilidad de adaptar el sistema a diferentes contextos de colaboración.

Con el fin de facilitar la comprensión de las expresiones presentadas, en la Tabla 4.1 se recopilan los principales parámetros y símbolos utilizados en el modelado de la función de recompensa asociada al confort. La tabla resume para cada variable su tipo, unidades y significado dentro del esquema propuesto, ofreciendo una referencia clara que permite interpretar de manera más directa

cómo se formaliza el impacto de la distancia y la velocidad, junto con otros factores, en la adaptación del comportamiento del agente.

Tabla 4.2: Parámetros y símbolos usados en el modelado del confort.

Símbolo / Parámetro	Tipo / Unidades	Descripción
d	Distancia (m)	Distancia robot–usuario. Entra en F_{conf1} y puede compararse con umbrales dependientes de C_l .
$d_{\text{máx}}$	Distancia (m)	Distancia a partir de la cual se considera “confort máximo” por separación; acota la parte positiva de F_{conf1} .
$\theta_{\text{thr}}(C_l, N_c)$	Distancia (m)	Umbral (función del nivel de confort C_l y de los niveles totales N_c) que marca el cambio penalización/recompensa en F_{conf1} .
$\Delta\alpha$	Ángulo (deg/step)	Incremento articular máximo por paso de simulación. Entra en F_{conf2} .
$\Delta\alpha_{\text{thr}}(C_l, N_c)$	Ángulo (deg/step)	Umbral de incremento articular por paso, dependiente de C_l y N_c , que limita la velocidad percibida.
C_l	Nivel (sin unidad)	Nivel de confort reportado/estimado (entrada del sistema).
N_c	Entero (sin unidad)	Número total de niveles de confort definidos.
α, β	Peso [0,1]	Ponderaciones de F_{conf1} (distancia) y F_{conf2} (velocidad) en F_{comfort} .
k_1, k_2, k_3	Activación $\in \{0, 1\}$	Activación de los términos de tarea (p. ej., recoger, entregar, retorno) en la recompensa total R .
F_{conf1}	Recompensa (sin unidad)	Término de confort basado en d (suavizado con tanh), dependiente de C_l, N_c .
F_{conf2}	Recompensa (sin unidad)	Término de confort basado en $\Delta\alpha$ con umbral por C_l, N_c .
F_{comfort}	Recompensa (sin unidad)	Combinación ponderada: $F_{\text{comfort}} = \alpha F_{\text{conf1}} + \beta F_{\text{conf2}}$.
R	Recompensa (sin unidad)	Recompensa total: $R = \sum_k k F_{\text{task}k} - F_{\text{comfort}}$.

En conjunto, la formulación presentada transforma el confort de un concepto subjetivo a una variable cuantificable que puede ser incorporada directamente en el entrenamiento por refuerzo. Al recibir como entrada el nivel de confort del usuario, el agente dispone de un mecanismo explícito para ajustar sus trayectorias y velocidades en tiempo real, equilibrando el éxito de la tarea con la experiencia subjetiva de su colaborador. Esta integración constituye un paso decisivo hacia la creación de comportamientos robóticos adaptativos, capaces de responder no solo a los requisitos funcionales de la tarea, sino también a las necesidades emocionales y perceptivas de la persona. En el próximo apartado se describen las condiciones experimentales empleadas para validar este modelo, así como los resultados obtenidos en diferentes escenarios con y sin la incorporación explícita del confort.

4.3 Configuración experimental y entrenamientos

En este apartado se describe la configuración experimental utilizada para validar el modelado de la función de recompensa asociada al confort. Partiendo del entorno previamente desarrollado, se han introducido diversas modificaciones que amplían sus capacidades, incluyendo la posibilidad de parametrizar la función de recompensa a través de un archivo de configuración externo. Este archivo permite ajustar de manera flexible los coeficientes de entrada relacionados con el confort, así como definir el número de niveles y los umbrales asociados a cada uno de ellos. Todas las implementaciones realizadas han sido probadas y verificadas, y se encuentran disponibles en el repositorio asociado a esta tesis, de modo que puedan ser reutilizadas por otros investigadores o adaptadas a nuevos escenarios. El objetivo de esta sección es detallar dichas adaptaciones y mostrar ejemplos iniciales de entrenamiento, antes de abordar en profundidad los resultados comparativos en los apartados posteriores.

Además de las adaptaciones teóricas descritas en el apartado anterior, se han llevado a cabo múltiples implementaciones en el código del entorno con el fin de integrar de forma completa el modelado del confort en el proceso de entrenamiento. Entre las principales mejoras introducidas se encuentran: la incorporación de un archivo de configuración externo en formato `.json`, desde el que es posible definir los parámetros de entrada de la función de recompensa (valores de α , β , número de niveles de confort N_c , umbrales asociados, etc.); la extensión de las observaciones del agente para incluir variables de estado relacionadas con la proximidad y la dinámica de movimiento; y la parametrización de las recompensas de cada subfase de la tarea para su activación o desactivación condicional. Estas implementaciones no solo permiten replicar los experimentos de forma transparente, sino que también facilitan la exploración de

variantes con bajo coste de configuración.

Como primer caso de prueba, se ha ejecutado un entrenamiento con los parámetros $\alpha = 0$ y $\beta = 0$, lo que anula la contribución de la función de confort en la recompensa total. En este escenario, el nivel de confort definido por el usuario no influye en la optimización, de acuerdo con la formulación matemática presentada en el apartado anterior. El algoritmo y los parámetros de entrenamiento empleados han seguido la configuración utilizada en experimentos previos de la investigación. Los resultados muestran un desempeño muy similar, e incluso ligeramente superior, al reportado en entornos anteriores sin modelado de confort, lo que confirma tanto la correcta integración de las nuevas funciones en el entorno como la optimización de su implementación. Este entrenamiento sirve así como línea base para contrastar posteriormente los efectos de la activación de la función de confort con distintos niveles de exigencia. En la Figura 4.3 se muestra una comparativa entre 4 de las trayectorias obtenidas, siendo un comportamiento similar a los de la Figura 3.12 debido al valor de los parámetros α y β . Tanto en los escenarios 1 y 2, como en los escenarios 3 y 4, el humano mantiene una posición fija mientras el agente deposita las pelotas de distinto color.

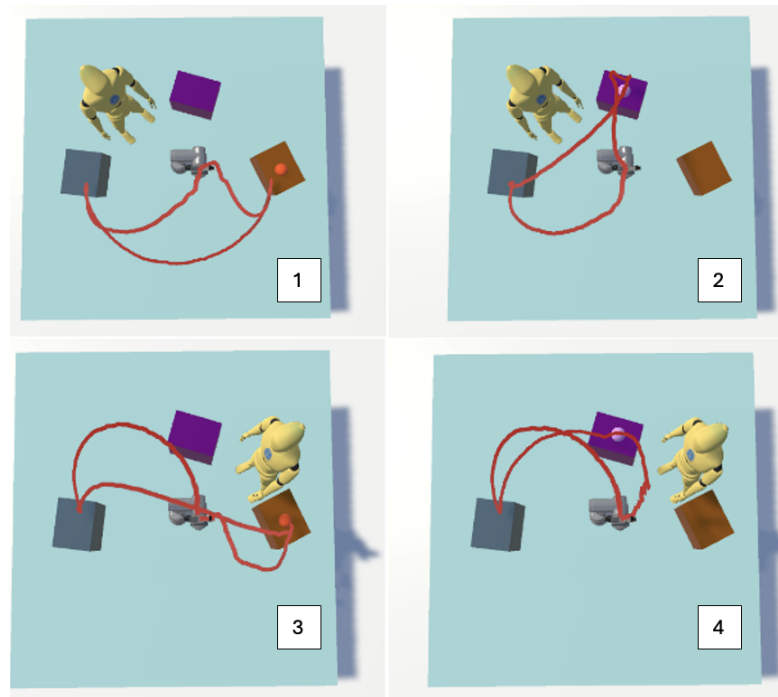


Figura 4.3: Trayectorias obtenidas durante el entrenamiento con $\alpha = 0$ y $\beta = 0$. El agente mantiene un comportamiento eficiente y estable.

Para la siguiente fase experimental se ha activado la función de confort utilizando una configuración inicial con $\alpha = 1$ y $\beta = 0$. De este modo, la recompensa se encuentra condicionada únicamente por la distancia, lo que permite observar de forma directa la correlación entre este parámetro y la percepción de confort. El objetivo de este experimento es analizar métricas globales como la distancia media mantenida, el número de pasos empleados y las variaciones derivadas de los distintos niveles de confort definidos en el sistema. Con este planteamiento se busca no solo demostrar la correcta integración de la función de confort en el entrenamiento, sino también evaluar en qué medida la modulación progresiva de los niveles (del 1 al 10) influye en el comportamiento del agente y en la dinámica de la interacción.

Los parámetros de entrenamiento empleados se han mantenido consistentes con los utilizados en las fases previas, con el fin de facilitar la replicación y garantizar la comparabilidad de resultados. La precisión en la tarea se ha fijado en 30 cm, de manera que se reduzca el tiempo de convergencia en un contexto en el que la complejidad aumenta debido a la consideración explícita del confort. En concreto, los entrenamientos se han realizado con un entorno configurado con movimiento humano aleatorio, cinco entornos simultáneos en paralelo y tareas activas de recogida y entrega. Asimismo, se ha definido un total de 10 niveles de confort, que se han explorado de manera sistemática a lo largo de distintos entrenamientos individuales, con una duración aproximada de 8 millones de pasos por nivel (alrededor de 1,1 días de cómputo por sesión). Este diseño experimental permite obtener resultados comparables y observar diferencias claras entre escenarios más restrictivos (niveles bajos) y más permisivos (niveles altos).

Cabe señalar que, en pruebas preliminares con precisiones más exigentes (del orden de 10 cm), el número de pasos requeridos para alcanzar una convergencia estable ascendía a unos 50 millones, lo que hace recomendable la utilización de estrategias de *curriculum learning* para progresar de niveles de precisión más amplios hacia otros más estrictos. En el presente trabajo, la reducción de la precisión a 30 cm ha permitido centrarse en la validación funcional de la recompensa asociada al confort, dejando para fases posteriores el análisis exhaustivo en entornos con mayores restricciones de precisión.

La Figura 4.4 muestra la evolución de la recompensa acumulada durante los entrenamientos realizados con los distintos niveles de confort definidos en el sistema. Se observa que las fases iniciales de aprendizaje presentan pendientes más pronunciadas en los niveles bajos de confort, donde el agente opera bajo mayores restricciones y la penalización por proximidad es más determinante. En cambio, en los niveles altos, caracterizados por condiciones más permisivas,

la recompensa tiende a estabilizarse antes, alcanzando un techo de rendimiento de forma más rápida. En conjunto, todos los entrenamientos convergen de manera estable, lo que confirma que la función de recompensa permite al agente adaptarse adecuadamente a diferentes umbrales de confort sin comprometer su capacidad para completar la tarea.

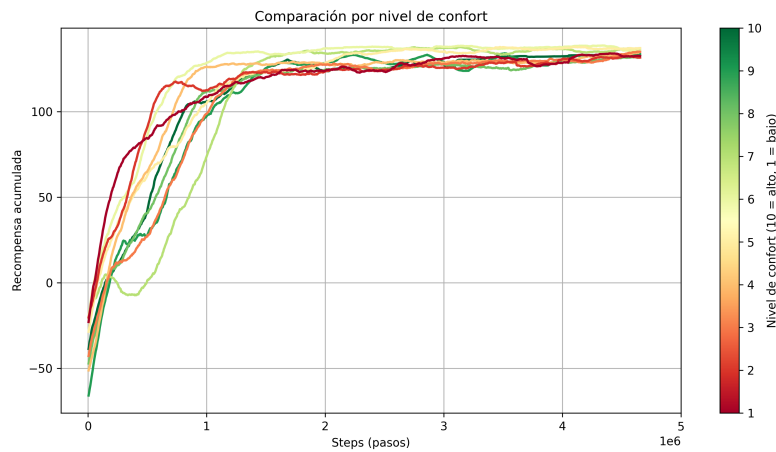


Figura 4.4: Evolución de la recompensa acumulada con distintos niveles de confort. Los colores indican el nivel de confort declarado, donde el rojo (nivel 1) representa estados de disconfort, y el verde (nivel 10) una situación de interacción más cómoda, natural y confortable.

En conjunto, los experimentos realizados en este apartado permiten validar la integración de la función de confort en el proceso de entrenamiento. Se ha demostrado que, independientemente del nivel de confort definido, el agente converge de manera estable, manteniendo un desempeño comparable en la tarea. Todos los experimentos descritos han sido implementados y se encuentran disponibles en el repositorio de GitHub asociado a esta tesis, lo que garantiza su replicabilidad y facilita su extensión en futuros trabajos. En el siguiente apartado se presentan los resultados detallados, donde se analizan métricas como la distancia media, el número de pasos empleados y las trayectorias generadas en distintos niveles de confort, con el fin de evaluar de forma más exhaustiva el impacto de esta variable en la interacción humano-robot.

4.4 Resultados y discusión

En este apartado se presentan los resultados derivados de los entrenamientos realizados con distintos niveles de confort (entendido como la tolerancia

del usuario a la cercanía física y a la velocidad de movimiento del robot), una vez validada en el apartado anterior la integración de la función de recompensa en el proceso de aprendizaje por refuerzo. El objetivo ahora es analizar de forma detallada el impacto que tiene la activación de la función de confort sobre el comportamiento del agente, comparando métricas cuantitativas y cualitativas en diferentes configuraciones. En particular, se examinan la distancia media mantenida entre el robot y el humano, el número de pasos requeridos para completar la tarea y las trayectorias generadas bajo distintos niveles de confort. Estas métricas permiten evaluar en qué medida la adaptación del agente a los umbrales de confort definidos contribuye a modificar la dinámica de la interacción y a generar comportamientos más acordes con la percepción subjetiva del colaborador humano.

Antes de abordar métricas concretas, resulta relevante analizar la evolución global de los entrenamientos con distintos niveles de confort. En todos los casos, el agente alcanza una convergencia estable (ver Figura 4.4), lo que confirma la integración correcta de la función de recompensa en el proceso de aprendizaje. No obstante, se aprecian diferencias en las fases iniciales. Los entrenamientos con niveles de confort altos presentan una pendiente de mejora más pronunciada, mientras que en niveles bajos del usuario (menos tolerancia a distancias cercanas) el progreso inicial resulta más lento debido a la necesidad de mantener mayores márgenes de separación con el humano. Estas variaciones reflejan el impacto directo que ejerce la variable de confort sobre la dinámica de aprendizaje y justifican el análisis detallado de métricas específicas, como la distancia media y el número de pasos, que se presentan a continuación.

Esta diferencia se explica por el grado de libertad que el agente tiene en cada configuración. En niveles altos de confort del usuario, donde las restricciones de proximidad son mínimas, el agente dispone de un espacio de acción más amplio y puede aproximarse directamente al objetivo sin penalización significativa, lo que favorece trayectorias más eficientes desde fases tempranas del entrenamiento. En cambio, en niveles bajos, la necesidad de mantener una mayor distancia impone condiciones más estrictas sobre las rutas posibles, lo que prolonga ligeramente el tiempo requerido para encontrar una política óptima. No obstante, estas diferencias en la velocidad de convergencia son relativamente pequeñas y no afectan a la estabilidad global del aprendizaje.

La Figura 4.5 muestra la distancia media mantenida entre el robot y el humano durante los entrenamientos realizados con diferentes niveles de confort. Los resultados reflejan una tendencia clara: a medida que se define un mayor nivel de confort (mayor tolerancia a la proximidad), el agente tiende a acercarse más al usuario, reduciendo la separación media. En niveles bajos de confort,

el robot mantiene trayectorias más amplias y conservadoras, priorizando la seguridad frente a la proximidad. Por el contrario, en niveles altos, las trayectorias se vuelven más directas y cercanas, coherentes con la mayor tolerancia del colaborador humano a interacciones próximas. Este resultado confirma que el parámetro de distancia actúa como modulador efectivo de la interacción y que el sistema es capaz de ajustar su comportamiento en función del nivel de confort definido.

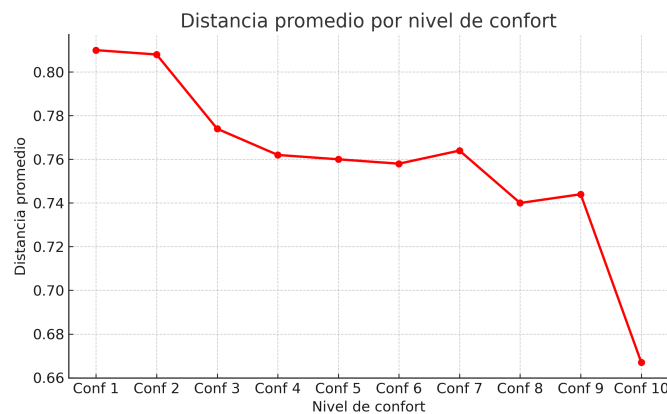


Figura 4.5: Distancia media mantenida entre robot y humano en función del nivel de confort.

La segunda métrica analizada corresponde al número promedio de pasos requeridos para completar la tarea (Figura 4.6). De forma consistente con el comportamiento anterior, se observa que en niveles bajos de confort, donde el agente se ve obligado a mantener mayor separación, el recorrido es más largo y el número de pasos aumenta. En cambio, en niveles altos de confort, la posibilidad de acercarse al usuario permite generar trayectorias más cortas y directas, reduciendo el coste en pasos necesarios para finalizar la tarea. Estos resultados ponen de manifiesto que la introducción de la variable de confort no solo impacta en la percepción subjetiva de seguridad, sino también en la eficiencia de la ejecución, logrando un equilibrio más favorable entre proximidad y optimización del movimiento.

Los resultados obtenidos permiten comprobar que el robot adapta su comportamiento en función de los niveles de confort definidos por el usuario, modificando la distancia mantenida y la forma de sus trayectorias de acuerdo con los parámetros establecidos. Las diferencias entre niveles son, en general, claras, aunque en algunos casos consecutivos la proximidad de los resultados puede dificultar su distinción visual. Sin embargo, con una configuración adecuada

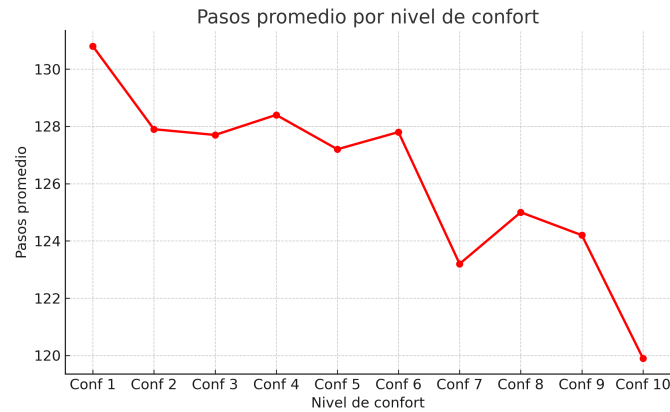


Figura 4.6: Número promedio de pasos empleados en función del nivel de confort.

de parámetros del entorno, niveles de confort y funciones de recompensa, esta separación puede hacerse más evidente. En la Figura 4.7 se presentan ejemplos de trayectorias correspondientes a cuatro niveles distintos de confort (alto, medio-alto, medio-bajo y bajo), lo que permite apreciar cómo el robot adapta su recorrido en un mismo escenario de interacción.

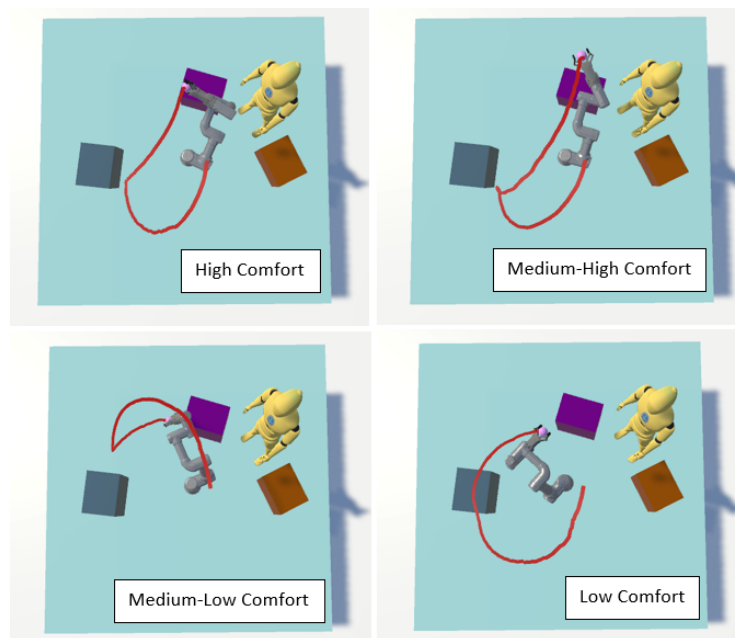


Figura 4.7: Ejemplos de trayectorias generadas bajo diferentes niveles de confort. Se aprecian recorridos más directos en niveles altos y trayectorias más amplias en niveles bajos.

Como se observa en la Figura 4.7, a medida que el nivel de confort disminuye, el robot tiende a incrementar la distancia mantenida respecto al humano, lo que se traduce en trayectorias más largas y menos directas. En el nivel de confort alto, el robot sigue un recorrido corto y eficiente, mientras que en niveles medio-bajos o bajos, la trayectoria se expande y describe curvas más amplias. Este comportamiento confirma, tanto a nivel numérico como visual, el efecto modulador de la variable de confort en la interacción.

Con el objetivo de comprobar la consistencia de estos resultados en situaciones alternativas, se ha realizado un segundo conjunto de experimentos en los que varía tanto la posición inicial del humano como la tarea concreta del robot. En la Figura 4.8 se muestran ejemplos adicionales de trayectorias bajo distintos niveles de confort. En este caso, las diferencias entre niveles son más sutiles. Mientras que en configuraciones de confort alto y medio-alto se aprecia una separación clara en la proximidad mantenida, en niveles consecutivos más bajos las trayectorias resultan visualmente más similares.

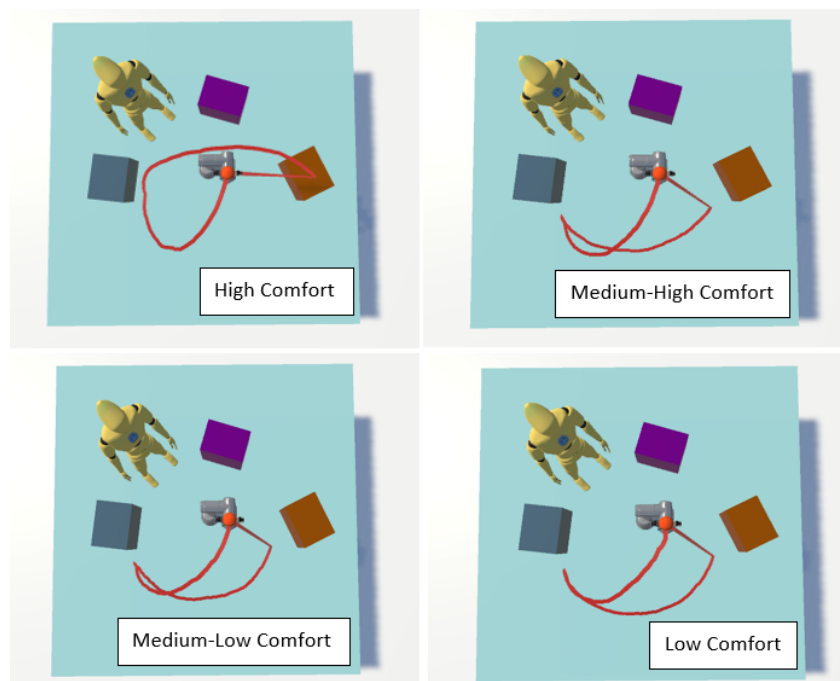


Figura 4.8: Ejemplos de trayectorias en un escenario alternativo. Las diferencias entre niveles de confort son más sutiles, aunque coherentes con la tendencia general.

La comparación de ambos conjuntos de resultados pone de manifiesto que la influencia del confort en la trayectoria depende no solo del nivel definido,

sino también de la situación del entorno y de la posición relativa del humano. En escenarios más directos, la separación entre niveles es clara y fácilmente identificable, mientras que en otros casos puede resultar más tenue. Esta variabilidad, lejos de ser un inconveniente, evidencia la flexibilidad del modelo y su capacidad para adaptarse a diferentes configuraciones contextuales, manteniendo siempre la coherencia con los niveles de confort especificados.

En conjunto, los resultados muestran que la introducción explícita del confort como variable en la función de recompensa permite generar comportamientos diferenciados y adaptativos en función de las necesidades del usuario. Tanto la distancia como la velocidad, integradas como parámetros moduladores, actúan de manera coherente con los niveles definidos, confirmando que es posible ajustar dinámicamente el comportamiento del robot sin comprometer la estabilidad del aprendizaje ni la correcta ejecución de la tarea. De este modo, se evidencia que el éxito del modelo no reside únicamente en optimizar la eficiencia de la tarea, sino en alcanzar un equilibrio entre rendimiento técnico y confort percibido, condición indispensable para trasladar este enfoque a escenarios reales de interacción. Este equilibrio entre eficiencia y percepción subjetiva abre nuevas oportunidades para diseñar sistemas colaborativos que no solo sean técnicamente robustos, sino también percibidos como seguros, comprensibles y cómodos por parte de quienes colaboran con ellos.

Más allá de los resultados obtenidos en simulación, el enfoque propuesto sienta las bases para líneas de trabajo futuras en las que se incorporen nuevas dimensiones del confort, como la inteligibilidad de los movimientos o la predictibilidad de las trayectorias. Cabe señalar, además, que la velocidad de ejecución constituye un parámetro fácilmente manipulable, al poder ajustarse de manera directa en la configuración de las trayectorias, lo que permitirá explorar su efecto de manera más específica en fases posteriores.

En los próximos capítulos se abordará el diseño de un experimento físico con usuarios reales, cuyo objetivo es analizar hasta qué punto las diferencias observadas en las trayectorias y la velocidad se traducen en una percepción diferenciada por parte del humano. De este modo, se busca validar si los patrones identificados en simulación tienen una influencia tangible en la interacción y explorar la posibilidad de definir un nivel de confort personalizado a partir de la combinación de métricas objetivas y subjetivas. Este paso constituye la transición necesaria hacia un marco metodológico que permita asignar dinámicamente comportamientos robóticos en función de la experiencia real del usuario.

*You must understand that there is more
than one path to the top of the mountain.*

MIYAMOTO MUSASHI

Capítulo

5

Diseño de la experimentación con participantes humanos

El objetivo central de este diseño del experimento es comprobar si los comportamientos generados a partir de distintos niveles de confort producen efectos diferenciados en los participantes. En otras palabras, se busca analizar si trayectorias más cercanas o más lejanas, así como movimientos más rápidos o más suaves, son percibidos de forma distinta por los usuarios en términos de seguridad, comodidad y confianza. Este planteamiento permitirá validar hasta qué punto la variable de confort, introducida durante el entrenamiento, se traduce en un impacto real en la experiencia del colaborador humano.

Para ello, se expondrá a los participantes a una serie de trayectorias parametrizadas en función del nivel de confort, registrando tanto indicadores fisiológicos (frecuencia cardíaca, variabilidad, movimientos corporales) como respuestas subjetivas obtenidas mediante cuestionarios. La combinación de estas fuentes de información permitirá determinar si los patrones de comportamiento del robot inducen reacciones diferenciadas entre usuarios y, en consecuencia, si es posible asociar perfiles específicos de reactividad a determinados niveles de confort. De esta manera, se podrán identificar grupos de usuarios más sensibles a la proximidad o a la velocidad, frente a otros que toleran interacciones más cercanas y dinámicas.

La verificación de estas diferencias resulta clave para validar la hipótesis de partida: que los niveles de confort definidos en simulación no solo modifican la trayectoria del robot, sino que también generan efectos observables en los hu-

manos que interactúan con él. Si se confirma esta relación, los entrenamientos podrán considerarse exitosos, al demostrar que las políticas aprendidas son capaces de inducir comportamientos coherentes con las expectativas y reacciones de los usuarios. Este hallazgo consolidaría la propuesta de integrar métricas de confort en el diseño de funciones de recompensa, abriendo la puerta a la personalización dinámica de la interacción humano-robot.

Con el fin de presentar de forma clara el desarrollo del experimento, este capítulo se organiza en seis secciones principales. En primer lugar, se expone el procedimiento seguido con participantes humanos, detallando la secuencia de fases de cada sesión y el esquema de contrabalanceo aplicado para controlar posibles efectos de orden. A continuación, se describe la configuración del entorno del experimento, incluyendo el laboratorio, la disposición del robot UR3e, las cámaras y los sensores empleados para la adquisición de datos. Posteriormente, se caracteriza la muestra de participantes, explicando los criterios de selección, el número de voluntarios involucrados en el estudio. Seguidamente, se introducen las variables y señales adquiridas durante las sesiones, describiendo los parámetros manipulados y las métricas de evaluación utilizadas, tanto objetivas como subjetivas. A continuación, se detallan las consideraciones éticas aplicables al diseño del experimento. Finalmente, se describe el entorno complementario de realidad virtual empleado para las primeras pruebas de concepto. Esta estructura permite ofrecer una visión integral y sistemática del diseño y procedimiento de la experimentación, sentando la base metodológica para el análisis de resultados que se desarrollará en el capítulo siguiente.

5.1 Procedimiento de la experimentación

El procedimiento del experimento se ha diseñado con el objetivo de garantizar la consistencia entre sesiones y facilitar la comparabilidad de los resultados entre participantes. En este apartado se describen las distintas fases que han compuesto cada sesión, desde la preparación inicial del voluntario hasta la finalización del experimento, así como el orden en el que se han presentado las trayectorias del robot. Asimismo, se ha contemplado un esquema de contrabalanceo para minimizar posibles efectos de aprendizaje o fatiga, asegurando que las diferencias observadas puedan atribuirse a las condiciones del experimento y no a factores externos.

En primer lugar, a cada participante se le ha solicitado la firma del consentimiento informado aprobado por el Comité de Ética de la Investigación de la Universidad de Deusto, garantizando el cumplimiento de la normativa en materia de protección de datos y la posibilidad de abandonar el estudio en cualquier

momento sin penalización. Tras esta formalidad, se les ha ofrecido una breve explicación del procedimiento, indicando que debían situarse en las huellas marcadas en el suelo, observar el extremo del robot y prestar atención a los movimientos realizados. Una vez finalizada la explicación inicial, se coloca un sensor de frecuencia cardíaca en el brazo del participante y se activan las cámaras de grabación, quedando así lista la fase de preparación.

Cada sesión se ha estructurado en dos fases. Al término de cada fase, los participantes han completado un cuestionario diseñado para recoger sus valoraciones subjetivas en relación con la interacción (estos cuestionarios se detallan en la sección 5.4 de variables y señales del experimento). Antes de comenzar las trayectorias, se ha incluido una fase inicial de reposo de dos minutos, acompañada de música neutra, con el fin de que los participantes alcanzaran un estado de calma y que las pulsaciones sirvieran como referencia de base para posteriores comparaciones con las medidas registradas durante la interacción con el robot.

El diseño ha contemplado la existencia de dos grupos de control, A y B, que han recibido las trayectorias en distinto orden durante la primera fase. El grupo A ha visualizado en primer lugar la trayectoria próxima (nivel de confort alto, con una distancia mínima de 22 cm y distancia media de 49 cm) seguida de la trayectoria lejana (nivel de confort bajo, con una distancia mínima de 45 cm y distancia media de 72 cm), mientras que el grupo B lo ha hecho en orden inverso. En la segunda fase, ambos grupos han sido expuestos a una misma condición: una trayectoria próxima ejecutada a mayor velocidad (Un 67% más rápida que la trayectoria base), con el objetivo de valorar cómo influye este factor en la percepción global de confort. Entre trayectorias y fases se ha introducido un minuto de reposo con música, con el fin de restablecer el estado basal de los participantes y evitar efectos acumulativos.

De esta forma, el procedimiento se ha estructurado en fases claramente diferenciadas: preparación, reposo inicial, ejecución de trayectorias, cumplimentación de cuestionarios y descansos intermedios. Este esquema ha permitido controlar el orden de exposición, garantizar la homogeneidad entre sesiones y obtener tanto datos fisiológicos como percepciones subjetivas en condiciones comparables.

Con el fin de ilustrar de manera esquemática la secuencia de fases descrita, en la Figura 5.1 se incluye un diagrama que resume el procedimiento seguido en cada sesión. En él se representa el flujo desde la preparación inicial hasta la finalización del experimento, diferenciando las trayectorias mostradas en la primera fase según el grupo asignado (A o B) y destacando los momentos de

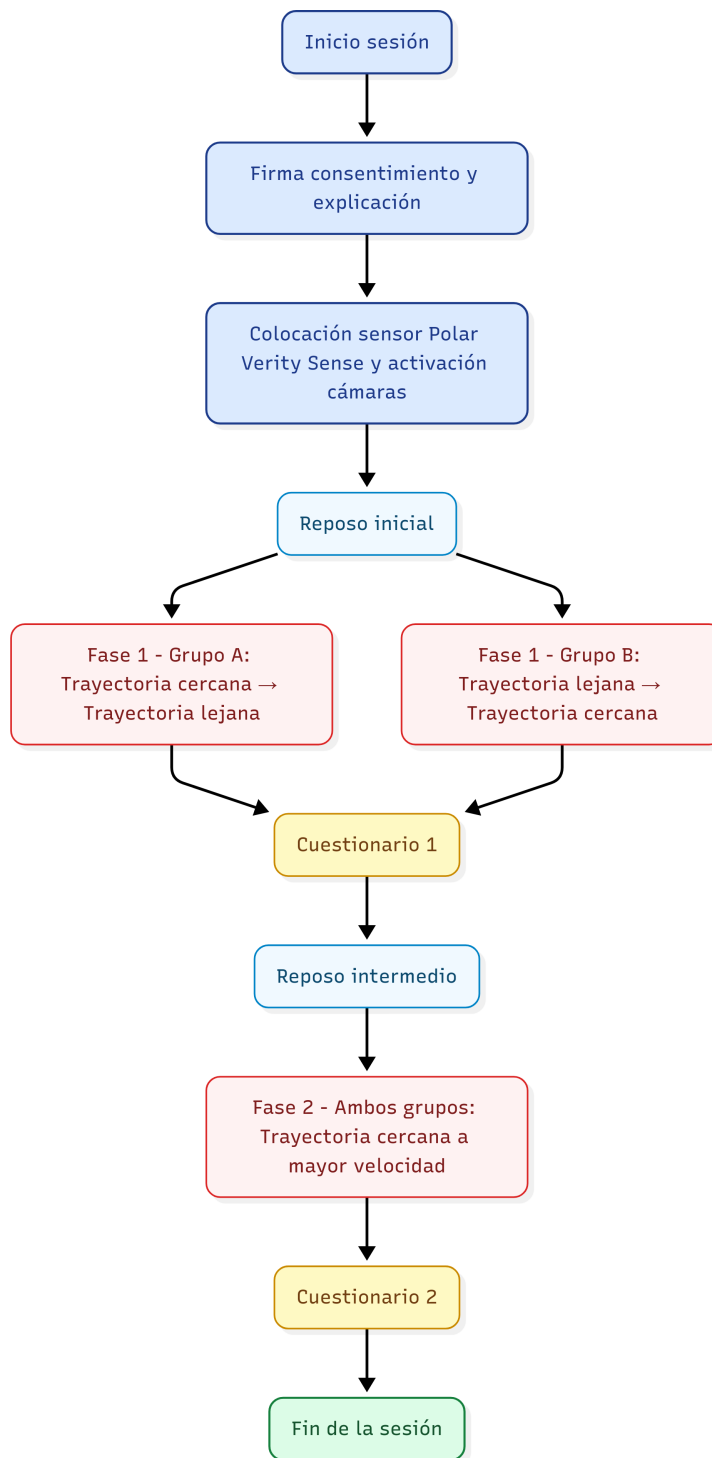


Figura 5.1: Esquema temporal del procedimiento del experimento, mostrando las fases de preparación (azul), exposición a trayectorias (rojo), descansos (azul claro), cuestionarios (amarillo) y finalización de la sesión (verde).

reposo y cumplimentación de cuestionarios. Este tipo de representación gráfica facilita la comprensión del diseño, al proporcionar una visión global y estructurada de las etapas por las que ha pasado cada participante.

En conjunto, el procedimiento se ha configurado de forma que garantiza tanto la homogeneidad entre sujetos como la recogida de datos en condiciones controladas y comparables. Una vez descrita la secuencia temporal de la sesión, el siguiente apartado se centra en la configuración física del entorno del experimento y en los dispositivos empleados para monitorizar las respuestas de los participantes.

5.2 Configuración del entorno del experimento

La configuración del entorno constituye un aspecto esencial para garantizar la validez y la reproducibilidad del experimento. En este apartado se describe la disposición física del laboratorio, incluyendo la posición del robot y de los participantes, así como los sistemas de registro empleados para capturar la interacción desde diferentes perspectivas. Se presentan además los sensores utilizados para monitorizar las respuestas fisiológicas de los usuarios y el marco tecnológico que permite la sincronización de todos los dispositivos. De este modo, se ofrece una visión global de la infraestructura sobre la que se asienta el estudio, antes de detallar en secciones posteriores las variables específicas y los procedimientos seguidos.

El laboratorio se ha organizado de forma que el robot UR3e permanezca fijo en su espacio de trabajo, ejecutando las trayectorias previamente entrenadas bajo condiciones controladas. Frente a él, se ha delimitado una posición marcada en el suelo que indica al participante dónde situarse, garantizando así la repetibilidad de las pruebas y la comparabilidad de los resultados entre sujetos. Para documentar la interacción, se ha instalado un sistema de tres cámaras: dos ubicadas lateralmente, destinadas a registrar expresiones faciales y movimientos de la parte superior del cuerpo, y una adicional con visión de cuerpo entero para captar la postura y la dinámica global del participante durante la tarea. En la Figura 5.2 se muestra una vista general del montaje del experimento, donde pueden observarse el robot UR3e, la posición señalada con las huellas para el participante y la disposición de las cámaras empleadas en el registro.



Figura 5.2: Vista general del laboratorio con el robot UR3e, la posición asignada al participante y las cámaras.

El sistema de cámaras se ha configurado con el objetivo de proporcionar una visión completa de la interacción. En total se han empleado tres dispositivos de grabación en alta definición: dos orientados lateralmente hacia el rostro del participante y una adicional que ofrece una vista completa de su postura corporal. Estas grabaciones no solo sirven como registro visual, sino que constituyen una fuente fundamental para análisis posteriores, ya que podrán procesarse mediante modelos de reconocimiento facial, estimación de pose y seguimiento ocular. De este modo, se facilita la identificación de expresiones emocionales, gestos y patrones de movimiento, complementando las señales fisiológicas con indicadores de conducta y postura. A modo ilustrativo, en la Figura 5.3 se presenta un collage con ejemplos de las imágenes obtenidas simultáneamente por las tres cámaras durante los ensayos.



Figura 5.3: Capturas de las tres cámaras utilizadas en el experimento, mostrando de manera simultánea el rostro y la postura completa del participante.

Para complementar la información obtenida mediante las cámaras, se ha empleado un sensor fisiológico portátil Polar Verity Sense, colocado en la parte superior del brazo de cada participante. Este dispositivo ha permitido registrar de forma continua la frecuencia cardíaca y la variabilidad de la frecuencia cardíaca, métricas utilizadas como indicadores de activación fisiológica y estrés en entornos de interacción humano-robot. En la Figura 5.4 se muestra el sensor colocado en el brazo de un participante, tal y como se utilizó durante el desarrollo del experimento.



Figura 5.4: Sensor fisiológico Polar Verity Sense colocado en el brazo de un participante durante el experimento.

La coordinación de todos los dispositivos se ha gestionado mediante el framework *Robot Operating System* (ROS), que ha permitido sincronizar en tiempo real los comandos enviados al robot, las señales fisiológicas captadas por el Polar Verity Sense y las grabaciones de vídeo procedentes de las cámaras. Para facilitar el control del experimento se ha desarrollado una aplicación de escritorio que posibilita iniciar, pausar y detener cada sesión desde un único punto de gestión, asegurando que los tiempos de registro se mantengan homogéneos en todas las pruebas. El código de control, junto con los *launch files* de ROS y los scripts de procesamiento de datos, se encuentra disponible en el repositorio público del proyecto, de modo que otros investigadores puedan replicar o adaptar el procedimiento en sus propios entornos.

Con todo lo anterior se ha descrito la disposición física del experimento, incluyendo el laboratorio, el robot, los sistemas de captura visual y el sensor fisiológico, así como la infraestructura de control y sincronización mediante ROS. Este apartado permite comprender cómo se ha configurado el espacio de inter-

acción y qué dispositivos se han empleado para registrar de manera fiable las respuestas de los participantes. No obstante, el diseño del experimento no se limita a la infraestructura técnica; resulta necesario caracterizar la muestra de voluntarios y definir las variables de interés que guiarán el análisis. De este modo, una vez establecida la idea general del experimento y el entorno en el que se lleva a cabo, se puede avanzar hacia los próximos apartados de la descripción de los participantes implicados y de los indicadores objetivos y subjetivos que permiten evaluar el confort en la interacción.

5.3 Selección y características de los participantes del experimento

La caracterización y selección de los participantes constituye un elemento clave para asegurar la validez del estudio. Dado que el objetivo principal es evaluar la influencia del confort en la interacción humano-robot, resulta imprescindible contar con una muestra diversa que permita obtener resultados representativos y minimizar sesgos asociados a variables individuales. En este apartado se detallan los criterios seguidos para la selección de los voluntarios, la justificación del tamaño de la muestra y la organización de los grupos de estudio.

El experimento cuenta con la participación de cincuenta voluntarios adultos, reclutados de manera voluntaria y sin compensación económica. El tamaño de la muestra no se elige de forma arbitraria, sino que está respaldado por un análisis de potencia diseñado para un ANOVA mixto (Analysis of Variance, análisis de varianza) con medidas repetidas de un factor que incluye tres condiciones intra-sujeto y dos grupos entre sujetos. Bajo parámetros habituales en este tipo de estudios ($\alpha = 0,05$, potencia $1 - \beta = 0,95$, tres mediciones por sujeto y una correlación esperada de $r \approx 0,5$), el cálculo indica que entre 40 y 50 participantes resultarían suficientes para detectar efectos de tamaño medio. La decisión de ampliar la cifra hasta 50 participantes válidos, incorporando un pequeño margen frente a posibles pérdidas de datos o abandonos, garantiza la robustez y estabilidad de los análisis posteriores.

Con el fin de controlar posibles variables de confusión, la muestra se estratifica atendiendo a tres criterios: edad, género y conocimiento previo de robótica. Para la variable edad se establecen cinco intervalos homogéneos (18–24, 25–29, 30–40, 41–49 y 50–65 años), con diez participantes en cada uno de ellos. Dentro de cada grupo de edad se mantiene un equilibrio exacto entre hombres y mujeres (cinco por categoría), lo que asegura una representación balanceada a lo largo de todo el ciclo adulto. En cuanto a la experiencia previa en robótica, esta es intencionadamente heterogénea, incluyendo tanto participantes sin

formación específica como sujetos con conocimientos técnicos, de modo que se recoge una diversidad de perfiles más cercana a escenarios reales de interacción.

Una vez completada la estratificación, los participantes son asignados aleatoriamente a dos grupos contrabalanceados Jones y Kenward (2003), con el objetivo de evitar efectos derivados del orden de presentación de las condiciones. El Grupo A realiza en primer lugar la condición *cercana–lenta* seguida de la condición *lejana–lenta*, mientras que el Grupo B completa las mismas condiciones en el orden inverso. Esta manipulación se centra en la variable de distancia, dado que la velocidad se evalúa en una fase posterior de carácter exploratorio. Como complemento, se incluye una condición adicional *cercana–rápida*, destinada a explorar de manera preliminar los efectos de la velocidad en la percepción de confort. En este sentido, se opta por excluir la condición *lejana–rápida*, puesto que el interés principal reside en validar el efecto de la proximidad robot–humano, previamente modelada en simulación.

Tras definir la muestra, el siguiente apartado se centra en las variables y condiciones del experimento que guiarán el análisis. Este apartado aborda tanto las métricas objetivas obtenidas a partir de las señales fisiológicas y las grabaciones en vídeo, como las valoraciones subjetivas recogidas mediante cuestionarios. De este modo, se establece el marco de indicadores sobre el que se apoyará el análisis posterior del capítulo de resultados.

5.4 Variables y señales del experimento

El análisis del experimento se sustenta en un conjunto de variables que permiten evaluar de manera objetiva y subjetiva la influencia del confort en la interacción humano–robot. En este apartado se describen los indicadores extraídos durante las sesiones, incluyendo tanto medidas fisiológicas como registros audiovisuales y valoraciones declaradas por los participantes. Aunque no todas estas variables puedan aportar información concluyente en el análisis final, se documentan de forma completa para constituir la base metodológica sobre la que se apoya la interpretación de los resultados.

En primer lugar, se han registrado indicadores fisiológicos relacionados con la frecuencia cardíaca. Para cada participante se ha medido un valor de referencia durante la fase inicial de relajación con música, así como una respuesta de control frente a un estímulo sonoro inesperado introducido en ese mismo periodo, con el objetivo de detectar posibles anomalías o reacciones bruscas. Durante las trayectorias del robot se ha recogido la frecuencia cardíaca inmediatamente antes de cada ejecución y el valor máximo alcanzado durante su desarrollo, lo

que permite evaluar los cambios asociados a cada condición.

Como complemento, se ha codificado el grado de movimiento corporal observado en las grabaciones de vídeo. Esta variable se ha categorizado en una escala de 0 a 3, donde el valor más bajo corresponde a la ausencia de movimiento, niveles intermedios reflejan gestos de cabeza o desplazamientos parciales del tronco, y el valor más alto representa una reacción global, acompañada de retroceso o conductas de marcada inquietud. De este modo, es posible relacionar la respuesta fisiológica con la cantidad de movimiento de los participantes.

Más allá de las medidas puntuales, también se ha recogido información continua de toda la señal de frecuencia cardíaca a lo largo del experimento, alineada temporalmente con la posición del robot, el estado de los motores y el resto de flujos de información. Esta sincronización, gestionada mediante ROS, permite realizar análisis más minuciosos en los que se pueden vincular de manera precisa los cambios fisiológicos de los participantes con eventos concretos del movimiento del robot. De este modo, se dispone de un conjunto de datos detallado que enriquece la interpretación de los resultados y abre la posibilidad de explorar correlaciones finas entre las trayectorias ejecutadas y las reacciones humanas.

En el caso de las grabaciones frontales, se ha empleado un análisis automatizado mediante técnicas de visión por computador con el fin de extraer información relevante sobre la respuesta emocional y expresiva de los participantes. Concretamente, se ha utilizado la librería *DeepFace*, que integra diversos modelos preentrenados para la detección de emociones básicas a partir de expresiones faciales, así como el detector *RetinaFace* para la localización precisa del rostro en cada fotograma. Esta combinación ha permitido procesar de manera eficiente las secuencias de vídeo y obtener indicadores adicionales de la interacción humano-robot.

Junto con las medidas fisiológicas, se han recogido valoraciones subjetivas mediante cuestionarios administrados al finalizar cada fase del experimento. Estas encuestas han incluido escalas de 1 a 5 para evaluar tanto el nivel de confort como la percepción de seguridad experimentada durante cada trayectoria. Asimismo, se ha solicitado a los participantes que comparasen las trayectorias observadas entre sí, indicando con cuál se habían sentido más cómodos y seguros en términos relativos. De esta manera, los cuestionarios aportan una medida directa de la experiencia subjetiva, complementando los indicadores objetivos y permitiendo contrastar las reacciones fisiológicas con la percepción declarada por cada usuario.

Tabla 5.1: Resumen simplificado de variables.

Categoría	Ejemplos de medidas
Fisiológicas	Frecuencia cardíaca, variabilidad de la frecuencia cardíaca
Conductuales	Movimientos de cabeza, tronco o retroceso completo
Visuales	Detección facial (RetinaFace), emociones (DeepFace)
Subjetivas	Confort (1–5), seguridad (1–5), preferencia de trayectoria

Tal y como se resume en la Tabla 5.1, en el diseño del experimento, se ha contemplado un conjunto amplio y diverso de variables que abarcan dimensiones fisiológicas, conductuales, visuales y subjetivas. Esta estructura proporciona un marco de análisis en el que las señales internas de activación se contrastan con la conducta observada, las expresiones faciales y las valoraciones declaradas por los propios participantes, permitiendo una interpretación de la influencia de las trayectorias del robot en la percepción de confort y seguridad.

Con el fin de clarificar también de forma explícita el flujo seguido desde la adquisición de señales hasta la obtención de las métricas analíticas empleadas en el capítulo de resultados, la Tabla 5.2 presenta una síntesis estructurada del proceso completo de captura, preprocesado y variables derivadas, reforzando la trazabilidad entre los datos en crudo y las métricas finales utilizadas en el análisis.

Tabla 5.2: Síntesis estructurada de señales, captura, preprocesado y métricas derivadas del experimento.

Señal / Variable	Fuente / Sensor	Momento de captura	Preprocesado principal	Métrica derivada
Frecuencia cardíaca (HR)	Sensor fisiológico en el brazo	Fase de relajación y durante cada trayectoria	Cálculo de línea base y detección de máximos	Δ HR por condición
Movimiento corporal	Cámara lateral	Durante la ejecución de cada trayectoria	Codificación manual en escala (0-3)	Índice de movimiento ponderado (WMI)
Expresión facial	Dos cámaras frontales	Durante la ejecución de cada trayectoria	Uso de modelos como <i>DeepFace</i> y <i>RetinaFace</i>	Distribución temporal de emociones
Estado del robot	Controlador (ROS)	Durante toda la ejecución	Sincronización temporal y filtrado básico	Identificación temporal y contextual de la trayectoria
Valoración subjetiva	Cuestionarios	Tras cada fase del experimento	Normalización de escalas Likert (1-5)	Confort y seguridad medios

Además de eso, todas las señales adquiridas durante el experimento se registran en formato estructurado y sincronizado mediante marcas temporales comunes, lo que permite reconstruir con precisión la evolución temporal de la interacción y garantizar la trazabilidad completa del proceso. Los detalles técnicos relativos al formato de los datos en crudo, su sincronización y las etapas de preprocesado se documentan de manera detallada en el Anexo C.

5.5 Consideraciones éticas de la experimentación

Las consideraciones éticas asociadas al diseño del experimento descrito en los distintos apartados de este capítulo abarcan el tratamiento de aquellas características sensibles derivadas de la participación de personas y de la recogida de datos fisiológicos en contextos de interacción humano-robot. Este apartado tiene como objetivo detallar de forma explícita los principios, criterios y medidas adoptadas para garantizar la seguridad, la privacidad y el bienestar de los participantes, así como el uso responsable de la información recopilada durante las diferentes fases del experimento.

El marco ético definido para este estudio se ajusta a los principios de la investigación responsable y a la normativa vigente en materia de protección de datos personales. En su definición se presta especial atención a minimizar los riesgos, al carácter voluntario de la participación y al tratamiento adecuado de la información recopilada. Los procedimientos y medidas descritos en este apartado han sido evaluados y aceptados por el Comité de Ética de la Investigación de la Universidad de Deusto, asegurando su idoneidad para la realización de experimentos con participantes humanos.

A partir de este marco ético, se contemplan decisiones orientadas a minimizar el impacto de la experimentación sobre los participantes y a limitar la recogida de información sensible a lo estrictamente necesario para los objetivos del estudio. En este sentido, se prioriza el uso de herramientas no invasivas para la adquisición de datos fisiológicos, se establecen sesiones de duración reducida y se garantiza en todo momento la posibilidad de abandonar el experimento de forma voluntaria, sin penalización ni justificación adicional.

En relación con la privacidad y la protección de los datos, se establecen medidas específicas para garantizar un tratamiento responsable de la información recopilada durante las sesiones. La recogida de datos se limita a variables fisiológicas y respuestas subjetivas directamente relacionadas con los objetivos del estudio, evitando la adquisición de información innecesaria. Asimismo, los datos se gestionan mediante procedimientos de anonimización y se almacenan en sistemas seguros con control de acceso restringido, reduciendo el riesgo de

identificación de los participantes y de un uso indebido de la información.

La seguridad física y emocional de los participantes se considera un aspecto central del diseño del experimento. Para ello, las pruebas se plantean exclusivamente con robots colaborativos diseñados para operar de forma segura en entornos compartidos con personas, incorporando mecanismos de parada y protocolos de actuación ante posibles incidencias. Durante el desarrollo de las sesiones, se establece una supervisión constante por parte del investigador, con el objetivo de garantizar que la interacción se mantiene dentro de condiciones controladas y seguras en todo momento.

En cuanto a la selección de participantes y al análisis de los resultados, se garantizan criterios de equidad, objetividad y transparencia. La muestra se define de manera homogénea en aquellas características relevantes para el estudio, evitando cualquier forma de coerción o exclusión injustificada. Del mismo modo, el tratamiento y análisis de los datos se plantean con criterios estadísticos neutros, asegurando que los resultados, tanto positivos como negativos, se reporten con el mismo rigor científico y se utilicen exclusivamente con fines de investigación, contribuyendo a la validez y transparencia de los resultados obtenidos.

5.6 Configuración del entorno complementario en VR

Este apartado describe la configuración de un entorno complementario en realidad virtual, diseñado como alternativa cuando no se dispone de un robot físico ni de un espacio de experimentación real. El objetivo de este entorno es permitir la ejecución controlada de los mismos experimentos planteados en este capítulo, manteniendo las condiciones básicas de interacción y facilitando tanto la evaluación de los comportamientos aprendidos como la recogida de valoraciones por parte de los usuarios. De este modo, se ofrece una solución práctica y accesible para reproducir el estudio en escenarios donde la disponibilidad de recursos materiales resulta limitada.

El flujo de experimentación en VR reproduce los mismos pasos y ajustes que en el experimento real propuesto, incluyendo la modificación de los parámetros del nivel de confort. De este modo, la dinámica de prueba y la lógica de interacción se mantienen equivalentes, lo que permite comparar los resultados de forma coherente. No obstante, existen algunas limitaciones propias de esta modalidad, siendo la más relevante la imposibilidad de emplear la cámara frontal para registrar la cara del participante, ya que las gafas de realidad virtual la ocultan. En consecuencia, el análisis de los modelos fisiológicos y de confort

debe realizarse sin esta fuente de información, lo que reduce parcialmente las posibilidades de evaluación respecto al experimento físico.

Un aspecto adicional de la utilización del entorno en VR es que las colisiones no suponen ningún riesgo para los participantes, lo que convierte al sistema en un espacio seguro para la experimentación. Además, este formato facilita la introducción de interacciones más complejas, como la posibilidad de coger la pelota o de acercarse físicamente al robot virtual, aumentando así el realismo de la colaboración sin comprometer la seguridad. También permite escalar la complejidad del experimento de manera más sencilla, incorporando nuevos gestos o patrones de movimiento del humano digital sin necesidad de modificar el montaje físico. En la Figura 5.5 se muestra un entorno simplificado de VR que, en términos de comportamiento, velocidad y nivel de confort, reproduce las mismas condiciones que el propuesto para el laboratorio.

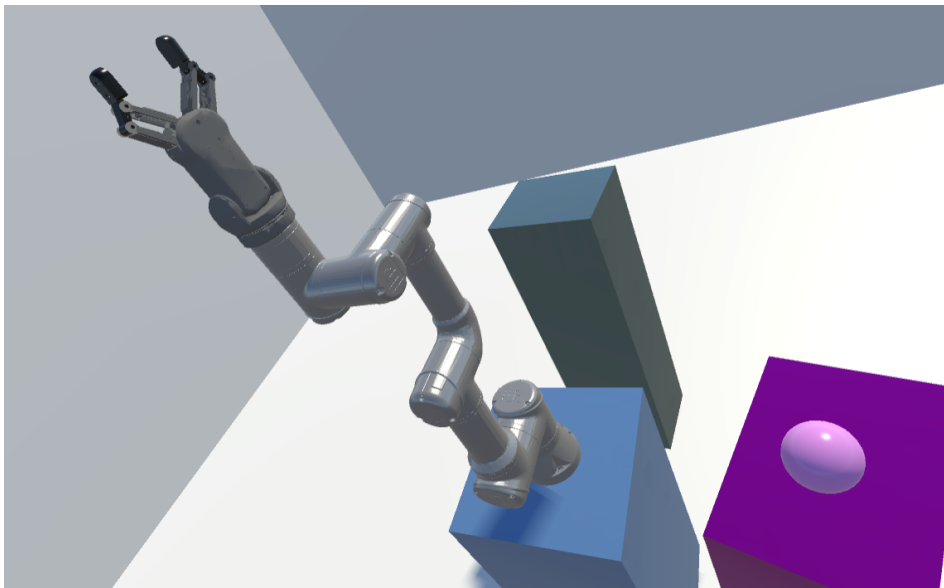


Figura 5.5: Entorno complementario en VR empleado para la validación del experimento en condiciones controladas sin necesidad de robot físico.

Se han realizado varias pruebas preliminares con un total de cuatro participantes con el fin de validar la viabilidad de este tipo de entornos de realidad virtual. Estos ensayos han permitido confirmar que el flujo del experimento puede reproducirse en condiciones controladas, aunque se han identificado diferencias relevantes respecto a la experiencia con el robot físico. Entre ellas, los mareos ocasionales asociados al uso prolongado de las gafas de VR y la percepción ficticia de seguridad, derivada de saber que se está en un entorno simulado, en

lugar de estar físicamente frente al robot.

En relación con las valoraciones subjetivas, los participantes que también habían experimentado el procedimiento en el laboratorio real coincidieron en señalar que las sensaciones eran más intensas en el escenario físico, especialmente en lo relativo a la proximidad del robot y la percepción de riesgo durante la interacción. Estos comentarios refuerzan la idea de que la VR constituye una alternativa útil y complementaria, pero que no sustituye completamente a la experiencia con un robot real en términos de realismo y carga perceptiva. Este contraste ha sido señalado en estudios previos sobre el uso de entornos inmersivos en interacción humano-robot, donde se destaca que la presencia física del robot genera un impacto diferencial en la experiencia subjetiva de los usuarios (Li et al., 2019; Plomin et al., 2023).

Un ejemplo de la experimentación en VR puede observarse en la Figura 5.6. En esta configuración, el procedimiento resulta especialmente sencillo de ejecutar: basta con un sensor de pulsaciones colocado en el brazo y una cámara externa para registrar los movimientos corporales y laterales del participante. La sincronización de estas señales con el entorno virtual se gestiona de manera integrada a través de Unity, lo que permite reproducir el flujo de experimentación ágil y con un coste de implementación reducido.

En definitiva, la realidad virtual se presenta como una alternativa útil cuando no se dispone del equipamiento necesario para ejecutar el experimento en un laboratorio con robot físico. Este entorno permite replicar los procedimientos de manera bastante fiel, manteniendo el flujo de la experimentación y posibilitando la evaluación de los comportamientos aprendidos. No obstante, persisten diferencias sutiles respecto a la experiencia real, especialmente en lo relativo a la percepción de riesgo y la intensidad subjetiva de la interacción. Asimismo, este entorno puede emplearse también para tareas de clasificación y análisis, como se detallará en apartados posteriores.

Una línea de investigación abierta y de gran interés consiste en comprobar de manera sistemática las diferencias percibidas entre la interacción en un entorno físico frente a un entorno virtual. Estudios previos han señalado que la presencia real del robot introduce matices que afectan tanto a la percepción como a la respuesta fisiológica de los usuarios, lo que refuerza la necesidad de explorar comparaciones directas y cuantitativas entre ambas modalidades. Este tipo de análisis permitiría delimitar con mayor precisión el alcance y las limitaciones de la VR como herramienta de validación.



Figura 5.6: Participante humano interactuando con el entorno en VR: a la izquierda en la configuración estándar y a la derecha en una versión mejorada con interacción directa con las pelotas.

Cabe destacar, además, que tanto el entorno complementario en VR como el entorno industrial de pruebas han sido desarrollados con un diseño modular que permite replicar los experimentos descritos en esta tesis. Ambos se encuentran disponibles en repositorios asociados al proyecto, lo que facilita su reutilización por parte de otros investigadores y favorece la reproducibilidad de los resultados. Estos recursos permiten a la comunidad contar con plataformas flexibles para la validación de modelos y la exploración de nuevas configuraciones experimentales.

Finalmente, aunque en este trabajo se ha optado por priorizar el despliegue en el entorno físico real, se considera que esta elección aporta una mayor riqueza en términos de información, al integrar múltiples fuentes como las cámaras adicionales y el registro directo de comportamientos. Asimismo, la interacción con un robot físico introduce una complejidad inherente que resulta esencial para evaluar la viabilidad de los comportamientos aprendidos en condiciones reales. En conjunto, la realidad virtual se consolida como un recurso de apoyo complementario que amplía las posibilidades, aunque por ahora no reemplaza la riqueza de la validación en un entorno físico real.

5.7 Resumen del diseño y procedimiento del experimento

Dado que el diseño y procedimiento del experimento constituye un pilar central en esta investigación, resulta necesario ofrecer una síntesis global que integre los principales elementos abordados en este capítulo. El experimento se ha diseñado con el propósito de validar, en un escenario real con usuarios humanos, las políticas de aprendizaje por refuerzo entrenadas previamente en simulación bajo distintos niveles de confort. Para ello, se ha diseñado un procedimiento estructurado en fases, que garantiza la comparabilidad entre participantes y el control de posibles factores de confusión, complementado con un sistema de contrabalanceo en la presentación de trayectorias.

La infraestructura del experimento se ha configurado en torno a un robot UR3e y un entorno de laboratorio especialmente adaptado, equipado con cámaras de alta definición y sensores fisiológicos sincronizados mediante ROS. Esta disposición ha permitido registrar de manera multimodal la interacción, combinando datos fisiológicos, movimientos corporales, expresiones faciales y valoraciones subjetivas. La recogida continua de pulsaciones cardíacas junto con la posición del robot y el estado de sus motores ha proporcionado además un marco preciso para relacionar eventos específicos con reacciones fisiológicas.

En cuanto a la muestra, el experimento ha contado con la participación de cincuenta voluntarios adultos, estratificados por edad, género y experiencia previa en robótica, lo que ha asegurado diversidad y equilibrio en la composición del grupo. La participación ha sido voluntaria y sin compensación económica, y todos los procedimientos se han realizado bajo aprobación ética y consentimiento informado.

El conjunto de variables recogidas abarca dimensiones objetivas y subjetivas. Entre las primeras destacan las medidas de frecuencia cardíaca y variabilidad, los índices de movimiento y las métricas derivadas del análisis automático de las imágenes frontales mediante técnicas de visión por computador. En el plano subjetivo, los cuestionarios administrados tras cada fase han permitido recoger la percepción de confort, seguridad y preferencia de trayectoria expresada por los participantes. Esta combinación de indicadores constituye la base metodológica sobre la que se apoyará el análisis de resultados.

En síntesis, el diseño aquí presentado proporciona un marco integral que permite evaluar de manera rigurosa cómo influyen los distintos niveles de confort en la interacción humano-robot. Al integrar participantes diversos, un procedimiento controlado y un conjunto completo de medidas fisiológicas, de con-

ducta, visuales y subjetivas, el estudio establece las condiciones necesarias para comprobar si los patrones identificados en simulación generan efectos observables y diferenciados en escenarios reales. El capítulo siguiente se centrará en la realización práctica del experimento y en los resultados obtenidos, analizando cómo cada una de estas variables contribuye a validar la hipótesis de que los niveles de confort definidos en simulación tienen un reflejo tangible en la interacción real con los usuarios.

Extraordinary claims require extraordinary evidence.

CARL SAGAN

Capítulo

6

Resultados del experimento y análisis de perfiles de confort

Una vez definido el diseño y la preparación del experimento, este capítulo presenta los resultados obtenidos en su implementación con participantes humanos. El estudio se ha llevado a cabo con un total de cincuenta y tres voluntarios, cada uno de los cuales ha completado una sesión de entre 30 y 40 minutos de duración. Los resultados presentados en este capítulo se basan en una muestra final de $N = 50$, tras el proceso de depuración de los datos experimentales. Durante este proceso se ha recogido un volumen considerable de información, incluyendo señales fisiológicas, corporales, visuales y valoraciones subjetivas, cuya depuración y análisis han permitido generar un conjunto de datos sólido y representativo.

El propósito de este capítulo no se limita a mostrar de manera descriptiva los resultados obtenidos, sino que busca ofrecer una interpretación crítica que permita valorar hasta qué punto las políticas de aprendizaje por refuerzo entrenadas en simulación se traducen en efectos observables en la interacción real. Para ello, se analizan de forma diferenciada los patrones de frecuencia cardíaca, la intensidad del movimiento corporal, las expresiones faciales procesadas mediante visión por computador y las respuestas recogidas a través de cuestionarios. Este análisis, además de proporcionar resultados cuantitativos, permite explorar de qué manera los distintos niveles de proximidad y velocidad del robot influyen en la percepción de confort y seguridad de los usuarios.

Más allá de estas métricas individuales, un elemento central de este capítulo es el análisis de agrupamiento de las respuestas, cuyo objetivo es identificar

perfiles de reactividad diferenciados entre los participantes. La aplicación de técnicas de *clustering* permite revelar tendencias ocultas en los datos, mostrando cómo ciertas personas responden con mayor estabilidad y confianza, mientras que otras presentan patrones de activación más acusados frente a trayectorias rápidas o cercanas. Este enfoque aporta un marco novedoso para personalizar el comportamiento del robot en función de las características de cada usuario y constituye una validación empírica de la hipótesis de partida.

El capítulo actual se estructura en varias secciones. En primer lugar, se presentan los resultados descriptivos iniciales que permiten caracterizar de manera general la respuesta de los participantes. A continuación, se aborda el análisis de perfiles mediante técnicas de *clustering*, seguido del estudio de la relación entre estos perfiles y las medidas objetivas y subjetivas recogidas. Posteriormente, se describe un procedimiento de asignación de nuevos usuarios a los perfiles identificados y, finalmente, discuten las implicaciones y los resultados del experimento. Esta organización permite ofrecer una visión progresiva, desde la descripción básica de los datos hasta la interpretación avanzada de patrones de comportamiento, proporcionando una visión integral del experimento para comprender la influencia del confort en la interacción humano-robot.

6.1 Resultados iniciales y observaciones exploratorias

Antes de profundizar en los análisis cuantitativos y en la identificación de perfiles de reactividad, resulta interesante presentar una serie de resultados iniciales que permiten ilustrar de manera exploratoria cómo se ha desarrollado el experimento y qué tendencias generales pueden observarse en los datos. En este apartado se incluyen representaciones visuales de las señales recogidas durante las sesiones, así como algunas observaciones cualitativas derivadas de las grabaciones y de la evolución fisiológica de los participantes. El objetivo no es extraer conclusiones definitivas, sino ofrecer una primera aproximación a los patrones emergentes, destacar aspectos llamativos del comportamiento de los voluntarios y contextualizar el análisis detallado que se desarrolla en las secciones posteriores.

La Figura 6.1 muestra la evolución media de la frecuencia cardíaca de los cincuenta participantes en las condiciones de trayectoria lejana y cercana, referentes a la fase 1 del experimento. Cabe recordar que, en la trayectoria cercana, la distancia mínima fue de 22 cm y la distancia media de 49 cm, mientras que en la lejana, la distancia mínima fue de 45 cm y la media de 71 cm. Tal y como se observaba durante la propia ejecución del experimento, la trayectoria cercana tiende a provocar un incremento más acusado en las pulsaciones, mientras

que la condición lejana mantiene valores más estables. En torno al segundo 5, coincidiendo con el inicio del movimiento del robot, ambas curvas presentan un comportamiento muy similar, pero a medida que avanza la interacción la respuesta promedio en la condición cercana se intensifica y alcanza un pico ligeramente superior al de la condición lejana alrededor del segundo 10. Este patrón preliminar sugiere que la proximidad del robot influye de forma perceptible en la activación fisiológica de los participantes.

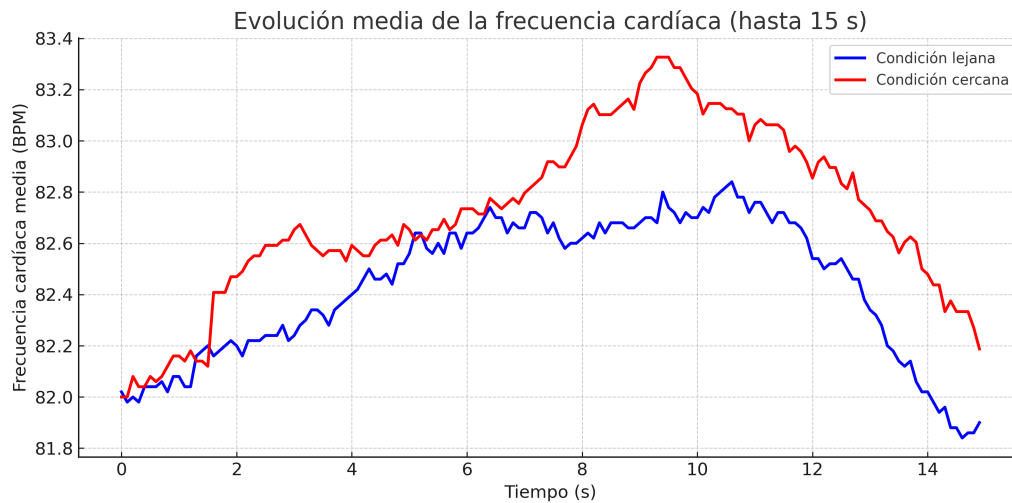


Figura 6.1: Media de la frecuencia cardíaca de los 50 participantes bajo las condiciones lejana (azul, $\bar{d} = 71$ cm) y cercana (rojo, $\bar{d} = 49$ cm).

Con el fin de analizar si este efecto se intensifica ante estímulos más dinámicos, en la segunda fase del experimento se incorporó una tercera condición cercana-rápida, caracterizada por un incremento del 67 % en la velocidad del movimiento del robot. Esta variación permite explorar si, además de la distancia, la velocidad constituye un factor determinante en la respuesta fisiológica de los participantes.

La Figura 6.2 representa el cambio medio de la frecuencia cardíaca respecto al valor de referencia inicial para las tres condiciones analizadas. El valor de referencia o *baseline* se ha calculado como la media de los primeros cinco segundos de cada registro, correspondientes a la fase de reposo anterior al inicio del movimiento del robot (indicado mediante una línea discontinua naranja). Este periodo representa el estado de calma inicial de los participantes, durante el cual no se produce ningún estímulo visual ni motor. El valor de *baseline* no puede ser cero, ya que la frecuencia cardíaca en reposo difiere entre individuos y el análisis debe realizarse en términos de variación relativa respecto al reposo

específico de cada participante, de modo que se capture correctamente el incremento fisiológico individual inducido por cada interacción robótica. A partir de este valor medio se han estimado los cambios relativos de frecuencia cardíaca en las tres condiciones consideradas. Se observa nuevamente que la trayectoria cercana genera una variación más pronunciada que el resto, confirmando que la proximidad constituye un factor relevante en la activación fisiológica. La condición cercana-rápida presenta un incremento algo más rápido en el inicio, pero su magnitud global no supera de forma clara a la condición cercana, lo que podría deberse al aplanamiento propio del promedio entre participantes o al hecho de tratarse de la última trayectoria presentada en la sesión. En cualquier caso, ambas condiciones se diferencian claramente de la trayectoria lejana, que es la que menor variación produce en términos de pulsaciones.

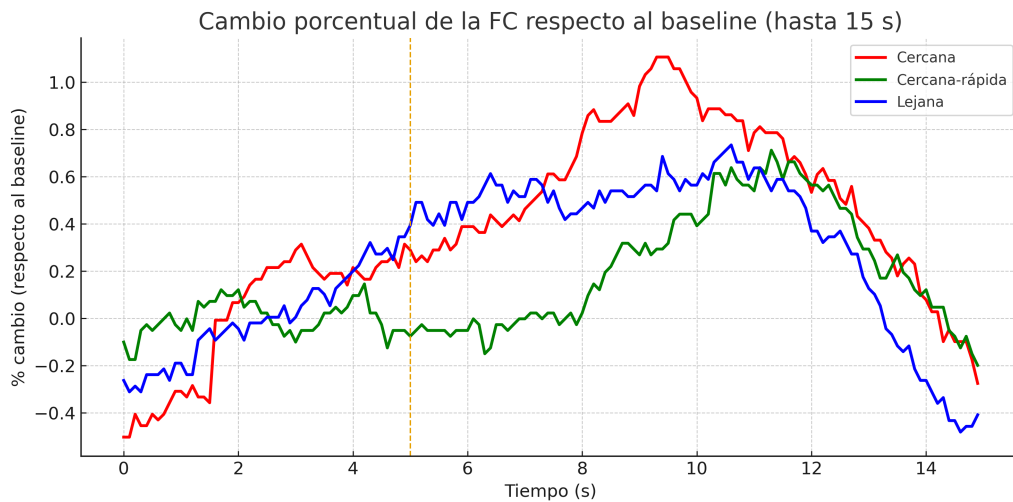


Figura 6.2: Cambio medio de la frecuencia cardíaca respecto al baseline en las condiciones lejana (azul, $\bar{d} = 71$ cm), cercana (rojo, $\bar{d} = 49$ cm) y cercana-rápida (verde, +67% de velocidad).

Es importante subrayar que estos resultados deben considerarse preliminares y que no permiten aún establecer conclusiones definitivas. La media agregada ofrece una visión global de la tendencia, pero oculta las diferencias individuales que pueden ser decisivas a la hora de comprender los perfiles de reactividad. Por este motivo, será necesario un análisis más exhaustivo que permita desagregar las respuestas y explorar hasta qué punto la combinación de proximidad y velocidad modula la experiencia fisiológica de los participantes. Este análisis se aborda en los apartados siguientes, donde se estudiará la variabilidad entre individuos y la consistencia de estos patrones en relación con otras medidas ob-

jetivas y subjetivas.

Aunque se recogieron también las imágenes frontales de los participantes con el objetivo de analizar las emociones mediante visión por computador, en la práctica esta métrica se ha mostrado de utilidad limitada en el contexto de este experimento. Dado que cada sesión estaba compuesta únicamente por trayectorias breves y muy controladas, la variabilidad en las expresiones faciales ha resultado escasa. Por este motivo, tras un primer análisis exploratorio se optó por descartar la emoción facial como indicador principal de confort, manteniendo estos valores únicamente como referencia cualitativa. A pesar de eso, estas medidas pueden resultar útiles para aportar interpretabilidad adicional en casos concretos, como la identificación de *outliers* o situaciones poco convencionales que no queden reflejadas de manera clara en las métricas fisiológicas.

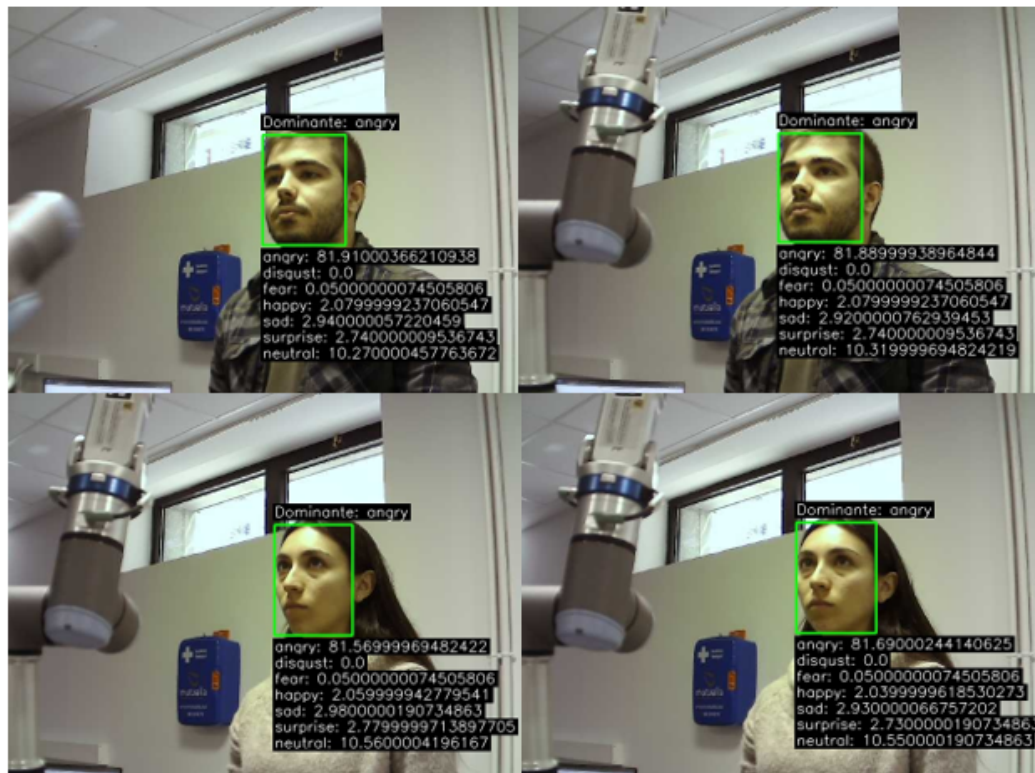


Figura 6.3: Ejemplos de capturas frontales de dos participantes en diferentes momentos de las trayectorias, con anotaciones automáticas de emociones estimadas.

En lo que respecta a los cuestionarios, los resultados apuntan a una clara preferencia de los participantes por aquellas trayectorias percibidas como menos invasivas y, en general, por aquellas que se ejecutan a menor velocidad. Este

patrón coincide con lo observado de manera cualitativa durante el desarrollo de las sesiones y refuerza la hipótesis de que la proximidad y la rapidez del movimiento influyen en la percepción de confort. No obstante, estos resultados deben ser interpretados con cautela, ya que su validez depende de la coherencia con el resto de métricas recogidas. En las secciones posteriores se compararán estas valoraciones subjetivas con los indicadores fisiológicos, con el fin de analizar el grado de correlación existente y determinar hasta qué punto las respuestas declaradas se alinean con las reacciones objetivas registradas.

Durante la propia ejecución del experimento ya se apreciaban diferencias notables entre los participantes. Por un lado, se observaban personas que apenas mostraban cambios ante las trayectorias del robot, manteniendo un comportamiento estable incluso en las condiciones más cercanas. Por otro, aparecían participantes más sensibles, que tendían a mostrar reacciones de sobresalto o incremento fisiológico tanto con la proximidad como con la velocidad. Estas observaciones sugieren que no todas las personas reaccionan de la misma manera ante un mismo estímulo y que, por tanto, resulta necesario aplicar técnicas de agrupamiento que permitan identificar perfiles diferenciados de reactividad. Este enfoque no solo facilita un análisis más preciso de los datos, sino que también abre la puerta a la personalización del comportamiento del robot en función de las características de cada usuario.

En síntesis, los resultados iniciales muestran que la proximidad del robot tiende a generar una activación fisiológica más marcada y que, en general, las trayectorias percibidas como menos invasivas y más lentas son mejor valoradas por los participantes. No obstante, las diferencias observadas entre individuos evidencian que no todos reaccionan del mismo modo, lo que limita la interpretación de las medias agregadas. Por este motivo, la siguiente sección se centra en llevar a cabo un análisis más profundo basado en técnicas de agrupamiento, con el objetivo de identificar perfiles diferenciados de reactividad y explorar cómo estas diferencias individuales condicionan la percepción de confort en la interacción humano-robot.

6.2 Análisis de perfiles de reactividad mediante *clustering*

En el desarrollo de la experimentación se observó, a partir de la interacción directa con los participantes, la existencia de patrones de respuesta fisiológica heterogéneos que no quedaban reflejados de forma clara en un primer análisis agregado. Tras realizar un análisis inicial de carácter general, basado

en descripciones estadísticas básicas y comparaciones entre las condiciones del experimento, y constatar que este enfoque no permitía destacar comportamientos diferenciados de manera concluyente, se optó por incorporar un análisis mediante *clustering*. Este análisis se plantea con un carácter exploratorio y no predictivo, sin objetivo de clasificación u optimización, y orientado a identificar posibles perfiles de reactividad fisiológica entre participantes. Para ello, se emplea como variable de entrada la variación de la frecuencia cardíaca respecto al *baseline*, permitiendo agrupar las respuestas en perfiles diferenciados y obtener una caracterización más estructurada de la variabilidad observada, cuya interpretación e implicaciones se desarrollan en la sección 6.4

El *clustering* es una técnica de análisis no supervisado que permite agrupar observaciones en función de su similitud, de manera que los elementos pertenecientes a un mismo clúster presentan patrones más próximos entre sí que respecto a los de otros grupos. Entre los distintos métodos disponibles, el algoritmo de *k-means* es uno de los más empleados por su sencillez y eficacia, ya que busca particionar el conjunto de datos en k grupos optimizando la proximidad a un centroide representativo de cada agrupamiento. Una cuestión clave en este procedimiento es la determinación del número óptimo de clústeres, ya que un valor demasiado bajo puede enmascarar diferencias relevantes mientras que un valor excesivo puede fragmentar en exceso los patrones. Para abordar esta decisión se recurre habitualmente a la denominada *regla del codo*, que consiste en representar la inercia intra-clúster en función de k y localizar el punto a partir del cual el descenso de la inercia se atenúa, configurando una forma de codo que indica el equilibrio entre simplicidad del modelo y capacidad explicativa.

A partir de los datos recogidos en el experimento se calcula la gráfica correspondiente a la regla del codo, representando la inercia intra-clúster en función del número de agrupamientos considerados. Tal y como se observa en la Figura 6.4, la disminución de la inercia es pronunciada en los primeros valores de k y comienza a estabilizarse a partir de $k = 3$, configurando un punto de inflexión claro. Este resultado sugiere que tres agrupamientos constituyen la solución más adecuada.

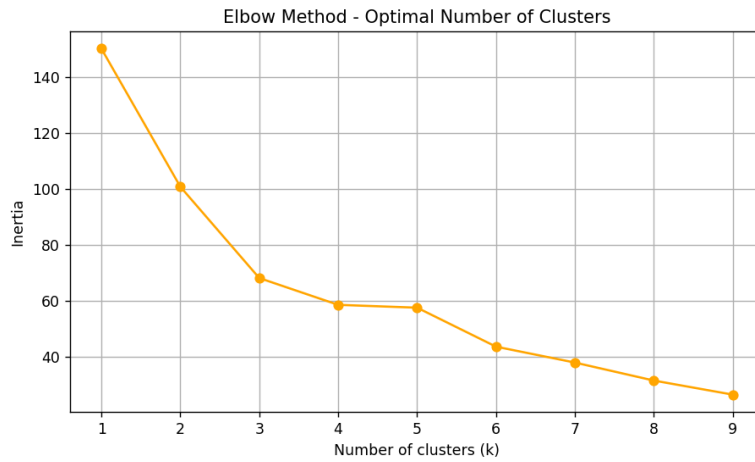


Figura 6.4: Método del codo. Evolución de la inercia en función del número de clústeres, con un punto de inflexión en $k=3$.

Una vez fijado el valor de $k = 3$, se aplica el algoritmo de *k-means* sobre el espacio estandarizado de características. Este método busca particionar el conjunto de observaciones minimizando la suma de distancias cuadráticas entre cada participante y el centroide de su clúster, generando así agrupamientos internos compactos y claramente diferenciados entre sí. Si bien el procedimiento depende en cierta medida de las condiciones iniciales, en este análisis se ha comprobado que la configuración de tres clústeres emerge de manera estable en distintas ejecuciones, lo que refuerza la consistencia y validez del resultado. Con el fin de facilitar su interpretación y ofrecer una representación gráfica intuitiva, los datos se proyectan en un espacio bidimensional mediante Análisis de Componentes Principales (PCA). Esta reducción de dimensionalidad no interviene en el ajuste del modelo, sino que se utiliza exclusivamente para fines ilustrativos, tal y como se muestra en la Figura 6.5, donde se aprecia la separación de los tres perfiles identificados.

A partir del análisis mediante *k-means* y de los resultados posteriores, emergen tres perfiles diferenciados de confort. El Clúster 1 (*Calm & Confident* - Calmados y confiados) agrupa a los participantes con una reactividad autonómica mínima a lo largo de todas las trayectorias (bajo ΔHR) y valoraciones consistentemente altas de confort y seguridad. El Clúster 2 (*Reactive but Tolerant* - Reactivos pero tolerantes) incluye a individuos que presentan respuestas moderadas de ΔHR , especialmente en la condición cercana-rápida, pero que aun así informan de elevados niveles subjetivos de confort y seguridad. Por último, el Clúster 3 (*Highly Reactive & Saturable* - Altamente reactivos) está compuesto por participantes con variaciones marcadas de ΔHR en todas las condiciones y

con efectos de habituación en el segundo ensayo; este grupo altamente reactivo es además el que reporta las valoraciones más bajas de confort y seguridad entre todos los clústeres.

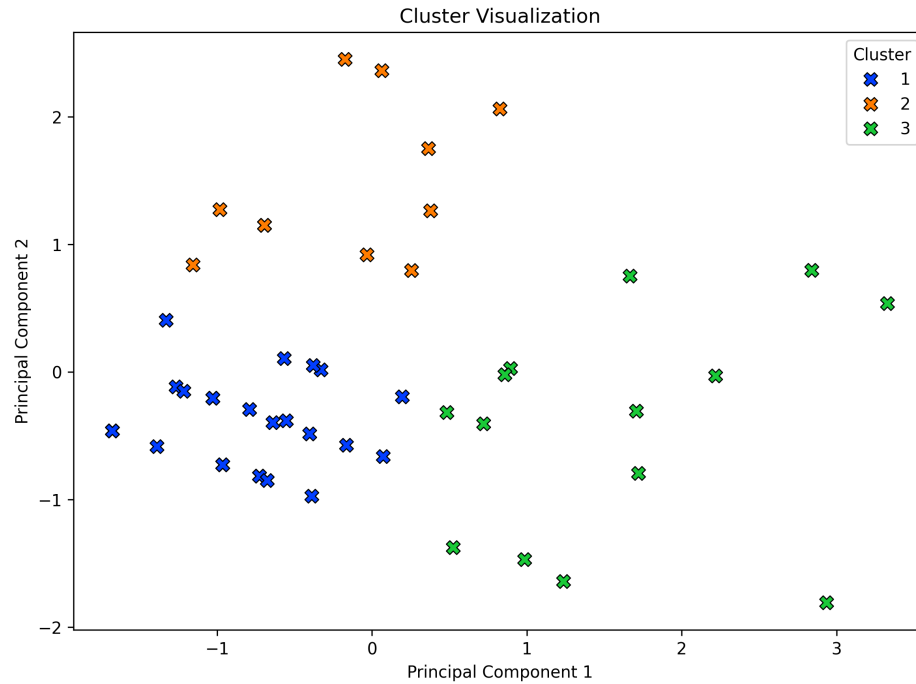


Figura 6.5: Clústeres identificados. Proyección bidimensional mediante PCA de los participantes, coloreados según su pertenencia a cada clúster ($k = 3$). La PCA se emplea únicamente con fines ilustrativos.

Tabla 6.1: Centroides de los clústeres para ΔHR (bpm)

Clúster	ΔHR_{near}	ΔHR_{far}	ΔHR_{fast}
1 (Calm & Confident)	2.04	0.84	0.80
2 (Reactive but Tolerant)	1.82	0.73	5.00
3 (Highly Reactive)	6.07	4.71	2.64

La Tabla 6.1 presenta los centroides exactos de reactividad autonómica. El Clúster 1 (*Calm & Confident*) incluye a 25 participantes, mientras que los Clústeres 2 (*Reactive but Tolerant*) y 3 (*Highly Reactive & Saturable*) engloban a los 25 restantes. Esta distribución 25/25 se mantiene constante en los dos grupos de contrabalance (A y B), confirmando que la pertenencia a un clúster es independiente del orden de los ensayos y que la muestra se encuentra equilibrada

para comparaciones posteriores.

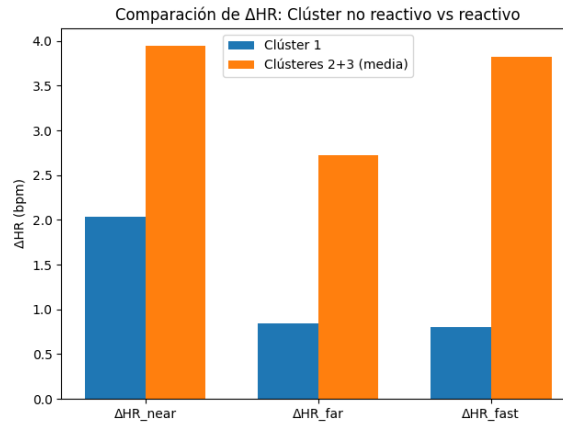


Figura 6.6: Comparación de la variación de frecuencia cardíaca (ΔHR) entre el Clúster 1 (no reactivo) y el grupo reactivo (Clúster 2 y Clúster 3).

Si además se combinan los Clústeres 2 y 3 en un grupo reactivo y se considera el Clúster 1 como grupo no reactivo, se obtiene una dicotomía clara: un 50 % de los participantes muestra baja activación autonómica, mientras que el otro 50 % presenta alta reactividad. Esta simplificación binaria resulta útil para los análisis posteriores y respalda el diseño de controladores adaptativos capaces de alternar entre trayectorias más conservadoras o más dinámicas en función de la monitorización fisiológica en tiempo real (ver Figuras 6.6 y 6.7).

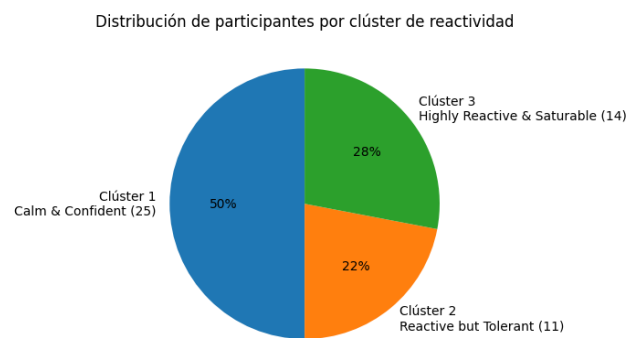


Figura 6.7: Distribución de participantes por clúster de reactividad fisiológica. El Clúster 1 (Calm & Confident) agrupa al 50 % de la muestra, mientras que los Clústeres 2 (Reactive but Tolerant) y 3 (Highly Reactive & Saturable) representan el 22 % y 28 % respectivamente.

Es importante señalar que, durante la exploración de los datos, se observaron respuestas de hiper-reactividad especialmente intensas en algunos casos. Estas observaciones no forman parte del análisis principal presentado, pero permiten apuntar la posible existencia de un perfil minoritario que no pudo caracterizarse en detalle debido a su baja frecuencia. Este aspecto constituye una limitación y abre la puerta a que futuros estudios con muestras más amplias exploren la validez de un cuarto agrupamiento y perfil de hiper-reactividad.

De manera complementaria, se calculó un índice ponderado de movimiento (Weighted Movement Index, WMI) para cada ensayo, definido en la Ecuación 6.1:

$$WMI = \frac{\sum(\Delta HR \times (M + 1))}{\sum(M + 1)} \quad (6.1)$$

donde ΔHR representa el cambio en la frecuencia cardíaca (bpm) y M es una puntuación categórica de movimiento (0 = sin movimiento, 1 = cabeza, 2 = torso, 3 = cuerpo completo). El término $(M + 1)$ asegura que incluso los intervalos sin movimiento visible contribuyan al denominador, al tiempo que otorga un mayor peso a los desplazamientos amplios para reflejar su mayor esfuerzo biomecánico. En promedio, el WMI fue más bajo en el Clúster 1 (*Calm & Confident*; $M = 1.38$), intermedio en el Clúster 2 (*Reactive but Tolerant*; $M = 3.12$), y más elevado en el Clúster 3 (*Highly Reactive & Saturable*; $M = 4.26$). No obstante, durante las entrevistas posteriores al experimento, varios participantes informaron haber minimizado conscientemente sus movimientos, lo que sugiere que el WMI puede ser menos fiable como métrica para modelos de clasificación futuros. Pese a esta limitación, la clara separación de valores medios de WMI entre clústeres respalda su utilidad para diferenciar perfiles reactivos y no reactivos (ver Figura 6.8).

Un análisis de correlaciones de Pearson (Schober et al., 2018) reveló relaciones negativas moderadas entre ΔHR y las valoraciones subjetivas de seguridad y confort: $r = -0,25$ (condición cercana) y $r = -0,28$ (condición lejana) para seguridad, y $r = -0,35$ (condición cercana) y $r = -0,14$ (condición lejana) para confort. Estos resultados se corresponden con los clústeres identificados: en el Clúster 1 (*Calm & Confident*) las autoevaluaciones se concentraron en la parte alta de la escala, atenuando los coeficientes de correlación, mientras que en los clústeres reactivos las variaciones más intensas de frecuencia cardíaca coincidieron con puntuaciones subjetivas más comprimidas. Las entrevistas posteriores también indicaron que los esfuerzos de algunos participantes por proyectar confort y seguridad podrían haber sesgado sus respuestas. En consecuencia, resulta recomendable priorizar índices objetivos como ΔHR para sustentar una clasifi-

cación fiable en tiempo real.

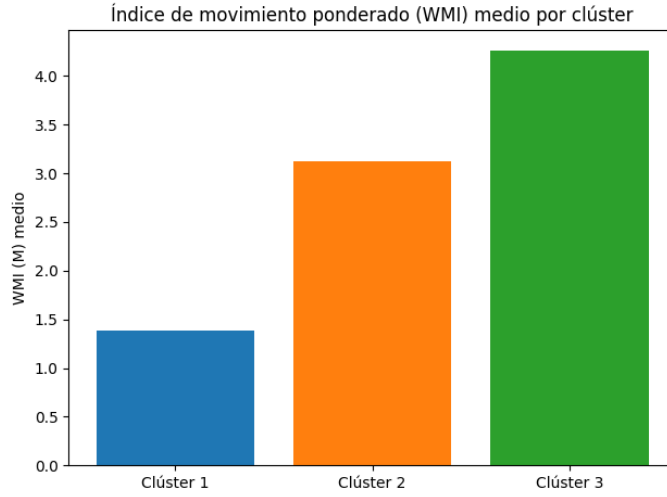


Figura 6.8: Índice de movimiento ponderado medio (WMI, M) por clúster de reactividad fisiológica. El Clúster 1 presenta valores bajos, el Clúster 2 valores intermedios y el Clúster 3 los valores más elevados, reflejando una mayor implicación corporal en los perfiles reactivos.

La asignación de un nuevo participante a uno de los tres perfiles de confort puede realizarse mediante el cálculo de la distancia euclídea a cada centroide (Hastie et al., 2009):

$$d_j = \sqrt{\sum_{k=1}^3 (x_k - c_{jk})^2} \quad (6.2)$$

donde $x = [\Delta HR_{\text{near}}, \Delta HR_{\text{far}}, \Delta HR_{\text{fast}}]$ es el vector de características del nuevo usuario y $c_j = [c_{j1}, c_{j2}, c_{j3}]$ es el centroide previamente obtenido mediante *k-means*. El participante se asigna al clúster j con el menor valor de d_j , y en consecuencia al meta-grupo correspondiente (reactivo o no reactivo). La implementación de este clasificador resulta directa utilizando el marco disponible en GitHub (Gonzalez-Santocildes, 2025a), lo que permite asignar niveles de confort o perfiles de comportamiento de forma automática a partir de datos autónomos, sin requerir sesiones adicionales. Este procedimiento posibilita que otros investigadores entrenen sus propios modelos y desplieguen optimización de confort en tiempo real con base en los tres niveles definidos o en una clasificación binaria reactivo/no reactivo (Gonzalez-Santocildes et al., 2024a).

Conviene subrayar que el procedimiento de *clustering* es de naturaleza exploratoria y no constituye un modelo de clasificación supervisado. El método de asignación por distancia a centroides debe entenderse, por tanto, como una heurística que aproxima la similitud de los usuarios con los perfiles de reactividad observados, más que como un predictor estadísticamente validado. No se han calculado métricas de validación cruzada ni de precisión supervisada, puesto que el objetivo es posibilitar la personalización en tiempo real basándose en patrones previamente identificados. En este sentido, la validez se apoya en la consistencia de los clústeres con las tendencias fisiológicas y las valoraciones subjetivas recogidas en los participantes.

En resumen, el análisis mediante *k-means* ha permitido identificar tres perfiles de reactividad fisiológica diferenciados, cuya consistencia a través de las distintas condiciones experimentales refuerza su validez como tipologías interpretables. Estos clústeres no deben entenderse como categorías rígidas, sino como agrupaciones empíricas que capturan tendencias predominantes en la respuesta de los participantes, lo que constituye un primer paso hacia la personalización de la interacción en función de la reactividad individual. En la sección 6.4 se profundiza en cómo aprovechar estos perfiles en entornos de aplicación real, abordando la clasificación de nuevos usuarios y explorando las implicaciones prácticas de adaptar el comportamiento del robot en función del perfil detectado.

6.3 Caracterización de los perfiles de reactividad

Los resultados del análisis de *clustering* permiten identificar tres perfiles diferenciados de reactividad fisiológica frente a la interacción con el robot. Estos perfiles reflejan patrones consistentes de respuesta tanto en las métricas fisiológicas como en las valoraciones subjetivas, evidenciando una variabilidad interindividual relevante en la percepción de confort.

En términos cuantitativos, el Clúster 1 (*Calm & Confident*) se caracteriza por variaciones mínimas de frecuencia cardíaca ($\Delta HR_{near} = 2,04$, $\Delta HR_{far} = 0,84$, $\Delta HR_{fast} = 0,80$) y un índice de movimiento ponderado bajo ($M = 1,38$). Este grupo presenta valoraciones subjetivas elevadas tanto en comodidad como en seguridad, lo que indica una percepción estable de control y confort a lo largo de las distintas condiciones de la experimentación.

El Clúster 2 (*Reactive but Tolerant*) muestra una activación fisiológica moderada, especialmente ante la condición rápida ($\Delta HR_{fast} = 5,00$), mientras que las variaciones asociadas a la proximidad resultan más contenidas ($\Delta HR_{near} =$

1,82; $\Delta HR_{far} = 0,73$). El índice de movimiento alcanza valores intermedios ($M = 3,12$), reflejando una mayor implicación corporal durante la interacción. A pesar de esta activación, las valoraciones subjetivas se mantienen en niveles elevados, lo que sugiere una buena tolerancia a estímulos más exigentes.

Por último, el Clúster 3 (*Highly Reactive & Saturable*) presenta las mayores variaciones fisiológicas en todas las condiciones ($\Delta HR_{near} = 6,07$, $\Delta HR_{far} = 4,71$, $\Delta HR_{fast} = 2,64$), junto con un índice de movimiento elevado ($M = 4,26$). Aunque sus valoraciones subjetivas de comodidad y seguridad son inferiores a las de los otros perfiles, se mantienen en rangos intermedios, lo que apunta a una mayor sensibilidad inicial acompañada de posibles efectos de habituación tras la exposición repetida.

En conjunto, estos resultados muestran que los perfiles identificados no solo difieren en la magnitud de la activación fisiológica, sino también en la relación entre dicha activación y la percepción subjetiva de confort. Esta caracterización proporciona una base sólida para la interpretación integrada de los resultados que se desarrollará en la siguiente sección.

6.4 Discusión e implicaciones de los resultados del experimento

Los resultados presentados en los apartados anteriores muestran que, si bien la proximidad y la velocidad del robot influyen de manera consistente en la activación fisiológica y en la percepción subjetiva de confort, la respuesta de los participantes presenta una variabilidad considerable. Esta variabilidad se traduce en perfiles de reactividad claramente diferenciados, que permiten interpretar los efectos observados de forma más precisa que un análisis basado únicamente en valores medios.

Desde esta perspectiva, los perfiles de reactividad identificados no deben entenderse únicamente como una herramienta de análisis, sino como un elemento operativo que permite explorar distintas estrategias de adaptación del comportamiento robótico. A partir de los resultados presentados, es posible analizar de qué manera estos perfiles pueden emplearse para guiar decisiones de diseño y personalización de la interacción.

Un primer uso práctico de los perfiles consiste en replicar el procedimiento de detección y asignar a un nuevo usuario a uno de los tres clústeres utilizando únicamente las diferencias de pulsaciones. Dado que en los párrafos anteriores se ha validado la coherencia entre las métricas fisiológicas y subjetivas, y se han

presentado los centroides (Tabla 6.1) y la regla de decisión (Ecuación 6.2), la clasificación puede implementarse de forma directa en tiempo real sin necesidad de recurrir a cámaras ni cuestionarios adicionales.

En la práctica, el flujo es sencillo: primero se estima la línea base de frecuencia cardíaca, después se calculan las variaciones en las condiciones cercana, lejana y rápida, y finalmente se construye el vector de características $x = [\Delta HR_{near}, \Delta HR_{far}, \Delta HR_{fast}]$. Este vector se compara con los centroides previamente estimados, y el usuario se asigna al clúster con menor distancia euclídea. De este modo se determina el perfil de reactividad, lo que permite personalizar parámetros de interacción (distancia, velocidad o estilo de movimiento) desde el inicio de la tarea.

Cuando el objetivo operativo no requiere distinguir entre tres perfiles, puede recurrirse a una clasificación binaria en dos meta-grupos: no reactivo (Clúster 1) frente a reactivo (Clústeres 2 y 3). Para ello basta con comparar la distancia al centroide no reactivo con la mínima distancia entre los centroides reactivos. Esta implicación resulta especialmente útil en escenarios donde es necesario decidir de manera rápida entre una trayectoria conservadora o una más dinámica, y también permite establecer de forma simple la condición de partida del usuario en términos de confort.

En consecuencia, el uso exclusivo de las pulsaciones es suficiente para detectar el perfil de un nuevo participante. Las correlaciones y coherencias mostradas en los apartados anteriores respaldan que ΔHR capta de forma fiable la reactividad relevante para la personalización de la interacción, evitando depender de sensores adicionales. Este procedimiento constituye, por tanto, la base para aplicaciones en las que se ajusta el comportamiento del robot en función del estado del usuario, tanto en la versión con tres clústeres como en la versión binaria, según el nivel de precisión que se requiera.

Además de la detección inicial, una implicación directa de los perfiles identificados es la posibilidad de entrenar controladores con distintos niveles de confort predefinidos. Si se realiza un entrenamiento con tres niveles diferenciados (correspondientes a los clústeres obtenidos) o con dos niveles simplificados (reactivo/no reactivo), en un entorno modificado y adaptado a una actividad específica similar a la presentada, el modelo resultante puede transferirse a otros escenarios.

Esta estrategia permite que, en contextos tanto de investigación como industriales, la interacción del robot se personalice en función del perfil detectado en cada usuario. De este modo, un mismo sistema puede operar con diferentes

configuraciones de comportamiento, asegurando que la experiencia se ajuste al nivel de confort individual. En consecuencia, los perfiles de reactividad fisiológica no solo validan la correcta implementación de la función de recompensa y la generación de comportamientos adaptativos en entornos simulados y reales, sino que abren la puerta a sistemas adaptativos generalizables a múltiples aplicaciones colaborativas.

Una aplicación especialmente interesante, que no requiere ajustes complejos ni experimentación directa con personas, consiste en el uso de una interfaz sencilla, como una tablet o un dispositivo móvil, con tres símbolos por ejemplo, un semáforo con caritas verde, naranja y roja (Ver Figura 6.9) o únicamente dos (reactivo/no reactivo). A través de esta interfaz, el usuario puede indicar de manera rápida y no invasiva su nivel percibido de confort, lo que permite al sistema ajustar en tiempo real parámetros como la distancia de operación, la velocidad del robot o incluso la carga de trabajo asignada. De esta forma, es posible compensar estados de fatiga o cansancio, manteniendo la interacción segura y eficiente sin necesidad de recurrir a mediciones fisiológicas continuas.

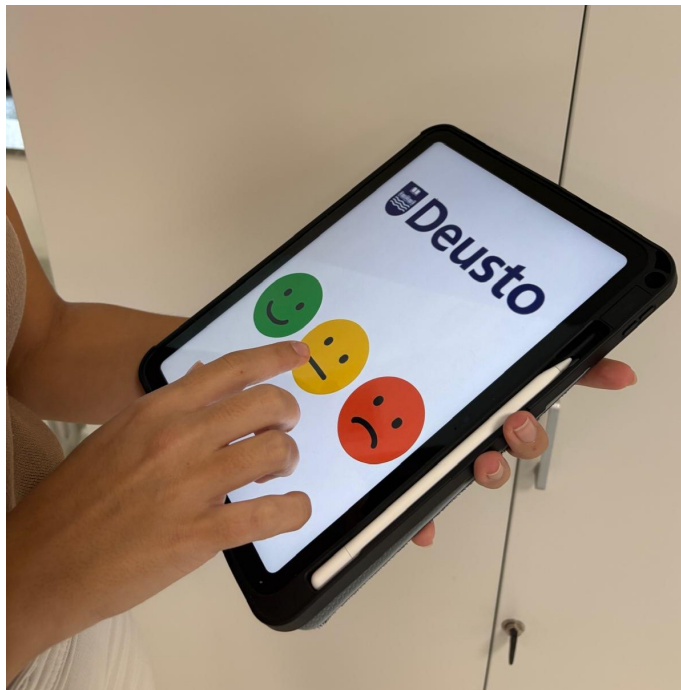


Figura 6.9: Ejemplo de interfaz simplificada con tres niveles de confort (verde = alto, amarillo = medio, rojo = bajo).

Esta propuesta constituye un punto intermedio entre un despliegue plenamente automático basado en muestras fisiológicas y un control manual previo.

La principal ventaja radica en la simplicidad, es suficiente con un entrenamiento previo de niveles de confort (o con los modelos derivados del análisis realizado en este estudio) para que el sistema pueda responder dinámicamente a la retroalimentación del trabajador mediante una interfaz accesible. Al no requerir sensores adicionales ni procedimientos invasivos, esta solución resulta fácilmente transferible a entornos industriales o colaborativos muy diversos, contribuyendo a mejorar la productividad y el bienestar de los usuarios con un coste de implementación mínimo.

Finalmente, puede plantearse un despliegue puramente automatizado basado en la monitorización fisiológica continua y en el cálculo de centroides para reasignar perfiles y ajustar parámetros de interacción de manera automática. Esta opción representa el nivel más avanzado de personalización, ya que permitiría adaptar en tiempo real la distancia, la velocidad o la carga de trabajo sin requerir intervención explícita del usuario.

No obstante, este enfoque conlleva importantes desafíos. En primer lugar, exige que el trabajador porte de manera constante un sensor fisiológico, lo que puede resultar invasivo desde el punto de vista práctico y generar resistencias en contextos laborales. En segundo lugar, plantea cuestiones éticas relacionadas con la privacidad, el control y el posible uso indebido de datos biométricos. Por ello, si bien este despliegue se perfila como una vía prometedora para el futuro, en el presente debe considerarse con cuidado.

Estas consideraciones establecen el marco para las conclusiones generales del trabajo y para la definición de las líneas futuras orientadas a una integración del confort humano en sistemas robóticos colaborativos.

So after all is said and done, after all the complainin' and the cryin'... is this all there is?

TONY SOPRANO (James Gandolfini), *The Sopranos* (David Chase)

Capítulo

7

Conclusiones y trabajo futuro

Este capítulo recoge las principales conclusiones derivadas del trabajo desarrollado en esta tesis doctoral y ofrece una visión de conjunto de los avances alcanzados. En primer lugar, se presenta un resumen global de la investigación realizada (incluyendo también una reflexión sobre sus limitaciones). A continuación, se exponen las contribuciones más relevantes, tanto científicas como técnicas, destacando aquellas que se han materializado en publicaciones en revistas de impacto y presentaciones en congresos internacionales. Esta revisión permite situar los resultados en el contexto de la investigación actual y subrayar la aportación original de la tesis en el ámbito de la robótica colaborativa centrada en el confort humano.

Del mismo modo, se lleva a cabo un balance crítico de la hipótesis inicial y de los objetivos planteados en la introducción, analizando el grado de cumplimiento alcanzado a lo largo del proceso. Finalmente, se proponen diversas líneas de trabajo futuro que permitirán consolidar y ampliar los resultados obtenidos, abriendo el camino a nuevas investigaciones y aplicaciones que profundicen en la integración del confort humano dentro de sistemas robóticos colaborativos adaptativos.

7.1 Resumen del trabajo y conclusiones

El trabajo desarrollado en esta tesis se ha centrado en dar respuesta a un vacío detectado en la evolución de la robótica colaborativa, la falta de consideración del confort humano como variable explícita en el diseño de comportamientos robóticos. Si bien las investigaciones previas han explorado ampliamente la

seguridad, la precisión o la eficiencia, el bienestar del usuario en cuanto al comportamiento robótico, ha quedado relegado a un segundo plano. La tesis parte de la premisa de que, en escenarios de cooperación y colaboración, la calidad de la interacción adquiere un papel crítico y exige que los sistemas robóticos sean capaces de ajustar dinámicamente su comportamiento en función del estado del usuario. Bajo esta perspectiva, se planteó como objetivo general el desarrollo de un marco basado en aprendizaje por refuerzo que integrara métricas de confort en la función de recompensa, posibilitando así la generación de políticas más adaptativas y centradas en la persona.

Para perseguir este objetivo, se ha seguido una estrategia metodológica progresiva que parte de la simulación y culmina en la experimentación con usuarios reales. En primer lugar, se diseñaron y validaron entornos simulados que permitieron entrenar y evaluar agentes bajo diferentes grados de complejidad colaborativa. A continuación, las políticas aprendidas se trasladaron a un robot físico y se implementaron en un experimento controlado con participantes humanos. Esta progresión permitió no solo comprobar la viabilidad técnica del enfoque, sino también extraer conclusiones relevantes. En particular, se observó que los robots que adaptan su comportamiento en función del nivel de confort generan percepciones diferenciadas en los usuarios, influyendo en su aceptación, confianza y sensación de seguridad durante la interacción.

Los resultados obtenidos a lo largo del trabajo confirman la viabilidad de integrar el confort humano como criterio explícito dentro de marcos de aprendizaje por refuerzo aplicados a la robótica colaborativa. En los entornos de simulación, la inclusión de métricas de confort en la función de recompensa permitió generar políticas más adaptativas y alineadas con el bienestar del usuario. Por su parte, los experimentos en un robot físico demostraron que estos modelos pueden trasladarse a escenarios reales, mostrando un impacto positivo tanto en la aceptación subjetiva de los participantes como en la fluidez y seguridad de la interacción. En conjunto, estos hallazgos evidencian que el confort no es un factor accesorio, sino una dimensión esencial para avanzar hacia sistemas colaborativos más sostenibles y centrados en las personas.

Si bien los resultados obtenidos en esta tesis han sido ampliamente positivos, todo proyecto de investigación presenta limitaciones que conviene reconocer de manera explícita. En primer lugar, la captura y análisis de señales fisiológicas conlleva una complejidad inherente, marcada tanto por la variabilidad entre individuos como por el ruido presente en las mediciones, lo que introduce incertidumbre en la interpretación de los resultados. A ello se suma el hecho de que los experimentos con participantes humanos se han llevado a cabo en un contexto controlado y con un número limitado de sujetos, lo cual restringe la

generalización de las conclusiones a entornos industriales de mayor escala.

Por otro lado, la transferencia desde la simulación hacia el robot real, aunque viable, aún enfrenta retos importantes para su despliegue en escenarios dinámicos o con múltiples usuarios. El estudio se ha centrado en interacciones individuales humano-robot y no contempla situaciones grupales ni la integración en sistemas productivos continuos, aspectos que pueden ser abordados en futuras investigaciones. Estas limitaciones, lejos de restar valor a los resultados alcanzados, delimitan el alcance del trabajo y abren líneas claras de investigación que se van a mencionar en el apartado 7.4 para consolidar y ampliar el marco propuesto. Conviene señalar que la mayoría de estas limitaciones no surgieron de manera imprevista durante el desarrollo de la investigación, sino que fueron ya identificadas en el planteamiento inicial de la tesis (ver el apartado 1.2). De este modo, más que debilidades inesperadas, constituyen restricciones acotadas que han permitido definir el alcance realista del trabajo y orientar de forma clara las líneas de investigación futura.

En conjunto, este apartado ha sintetizado los principales logros alcanzados y ha reconocido las limitaciones que acompañan a cualquier proyecto de investigación. A partir de esta visión general, resulta importante revisar de manera crítica el grado en que la hipótesis formulada al inicio del desarrollo de la tesis y los objetivos derivados de ella han sido cumplidos. Este análisis permitirá valorar con mayor detalle la coherencia entre las preguntas planteadas y las respuestas obtenidas, así como delimitar el alcance real de las contribuciones presentadas.

7.2 Balance de la hipótesis y los objetivos

Al inicio de esta tesis, concretamente en la sección 1.2 se formuló la siguiente hipótesis de investigación:

Hipótesis. *La incorporación de funciones de recompensa centradas en el confort humano dentro de entornos de aprendizaje por refuerzo permite desarrollar comportamientos robóticos adaptativos capaces de mejorar la calidad de la interacción colaborativa. Estos comportamientos no solo reducen el estrés percibido por el usuario, sino que también incrementan su confianza y sensación de seguridad, incluso en contextos dinámicos y no previstos. Esta adaptación personalizada puede lograrse mediante el uso de señales fisiológicas y modelos previamente entrenados, sin necesidad de calibraciones extensas ni sensores invasivos durante la fase de despliegue.*

Además de plantear la hipótesis, con el fin de poder validarla, se planteó el siguiente objetivo principal de investigación:

Objetivo principal. *Diseñar, implementar y validar un framework de entrenamiento robótico basado en aprendizaje por refuerzo que integre parámetros humanos, en particular métricas de confort fisiológico y subjetivo, con el fin de generar comportamientos personalizados y adaptativos en entornos colaborativos. Este sistema será entrenado y evaluado en simulación y experimentado posteriormente con usuarios reales, explorando la viabilidad de modelos de confort transferibles, escalables y fácilmente desplegables en escenarios reales.*

El desarrollo de esta tesis ha permitido confirmar la hipótesis planteada. La integración de métricas fisiológicas y subjetivas en la función de recompensa mediante la variable de nivel de confort ha demostrado ser viable, dando lugar a políticas robóticas capaces de ajustar su comportamiento en función del estado del usuario. Tanto en los entornos de simulación como en el robot físico se ha evidenciado que dependiendo del nivel de confort, este enfoque tiende a producir interacciones percibidas como más seguras, fluidas y confiables por parte de los participantes. En consecuencia, la hipótesis inicial se valida en sus fundamentos esenciales, al mostrar que el confort puede constituir un criterio explícito y efectivo para guiar el aprendizaje por refuerzo en contextos colaborativos. Con todo, los resultados obtenidos aportan evidencias sólidas de que el confort humano puede convertirse en un eje fundamental en el diseño de la robótica colaborativa del futuro.

Respecto al objetivo principal de la tesis, el trabajo desarrollado ha permitido diseñar e implementar un framework de entrenamiento basado en aprendizaje por refuerzo que integra de manera efectiva el confort humano como parámetro en la función de recompensa. Para ello, se construyeron entornos de simulación progresivos que facilitaron el entrenamiento de agentes bajo diferentes grados de complejidad colaborativa, modelando la recompensa a partir del nivel de confort como variable de entrada. Estas políticas fueron posteriormente transferidas y validadas en un experimento controlado con 50 participantes humanos, lo que permitió constatar su aplicabilidad en condiciones reales. Los resultados obtenidos confirman que el framework propuesto es capaz de generar comportamientos adaptativos y personalizados, cumpliendo de manera satisfactoria el objetivo planteado y consolidando la validez del enfoque en escenarios colaborativos reales. Esto supone un avance respecto a los enfoques tradicionales centrados únicamente en métricas de rendimiento técnico. Además, el marco metodológico desarrollado presenta un alto grado de reproducibilidad, lo que refuerza su potencial de transferencia hacia otros escenarios colaborativos.

A continuación, se presenta el grado de cumplimiento de cada uno de los objetivos específicos planteados en la sección 1.2:

1. Se han diseñado e implementado entornos de simulación progresivos que permitieron entrenar agentes en escenarios con distintos grados de colaboración, desde tareas simples de manipulación autónoma hasta configuraciones con interacción humano-robot en el mismo espacio de trabajo (Capítulo 3).
2. Se han desarrollado funciones de recompensa que integran el confort humano como criterio explícito en el proceso de aprendizaje (Capítulo 4).
3. Los agentes robóticos fueron entrenados en los entornos simulados con distintos niveles de complejidad, mostrando sensibilidad a los cambios en el nivel de confort durante el aprendizaje (Capítulos 3 y 4).
4. Se evaluó el impacto de la introducción de métricas de confort frente a controladores no adaptativos (Capítulo 4).
5. Se implementó un experimento con 50 participantes humanos, lo que permitió validar la respuesta del sistema en condiciones reales de interacción (Capítulo 5).
6. Se recopilaron y analizaron datos fisiológicos (frecuencia cardíaca) y valoraciones subjetivas, confirmando su utilidad para medir la experiencia de confort durante la colaboración (Capítulo 6).
7. El análisis de los datos recogidos permitió identificar distintos perfiles de usuario según su reactividad fisiológica y subjetiva ante los comportamientos del robot (Capítulo 6).
8. La aplicabilidad real del enfoque quedó demostrada, ya que el sistema fue capaz de generar comportamientos personalizados y adaptativos que mejoraron la percepción de seguridad y la experiencia global de los usuarios (Capítulo 6 y 7).

Con el fin de facilitar la consulta y ofrecer una visión sintética del grado de cumplimiento de los objetivos específicos, se ha elaborado un esquema visual que muestra la correspondencia entre cada objetivo planteado y el capítulo de la tesis en el que se aborda su desarrollo. Este recurso permite identificar de manera rápida dónde se encuentra la evidencia asociada a cada objetivo, sirviendo como guía de lectura y referencia cruzada dentro del documento (ver Figura 7.1).

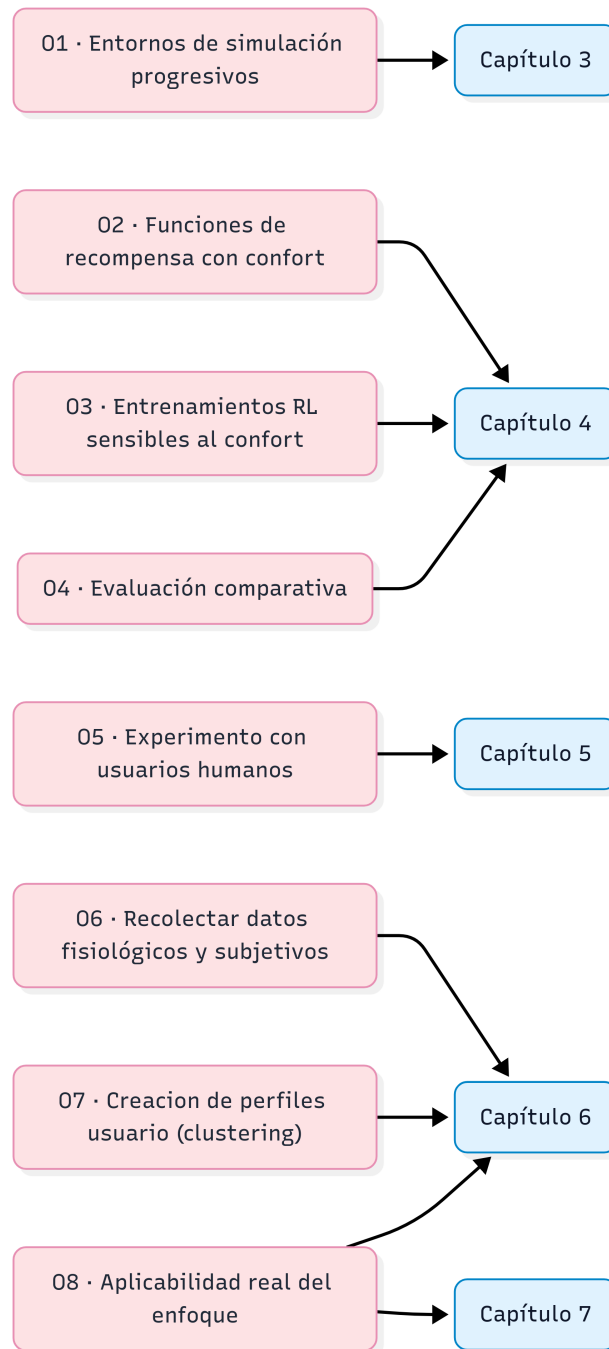


Figura 7.1: Esquema de correspondencia entre los **objetivos específicos de la tesis** y los **capítulos** en los que se desarrolla su cumplimiento.

En conjunto, el análisis realizado confirma que tanto la hipótesis como los objetivos, generales y específicos, han sido cumplidos en los términos plantea-

dos. Este balance refuerza la solidez del marco propuesto y prepara el terreno para destacar, en el siguiente apartado, las contribuciones científicas derivadas de este trabajo.

7.3 Publicaciones asociadas a la investigación

El trabajo desarrollado en esta tesis ha dado lugar a diversas contribuciones científicas y recursos de acceso abierto. Estos resultados reflejan tanto la relevancia académica de la investigación como el compromiso con la transparencia y la transferencia de conocimiento. A continuación, se presentan las publicaciones y repositorios asociados.

Repositorios abiertos

1. **Comfort-RL.** Repositorio de código y scripts de entrenamiento para la integración de funciones de recompensa centradas en el confort humano en entornos de aprendizaje por refuerzo. Disponible en:
<https://github.com/AsierGonz/Comfort-RL>
2. **Human-Robot-Comfort-Experiment.** Repositorio con datos y scripts asociados a los estudios con participantes humanos y robot colaborativo UR3e. Disponible en:
<https://github.com/AsierGonz/Human-Robot-Comfort-Experiment>

Artículos publicados en revistas internacionales

1. A. Gonzalez-Santocildes, J.-I. Vazquez, A. Eguiluz. *Adaptive Robot Behavior Based on Human Comfort Using Reinforcement Learning*. **IEEE Access**, 2024. DOI: 10.1109/ACCESS.2024.3451663.
2. A. Gonzalez-Santocildes, J.-I. Vazquez, A. Eguiluz. *Enhancing Robot Behavior with EEG, Reinforcement Learning and Beyond: A Review of Techniques in Collaborative Robotics*. **Applied Sciences**, vol. 14, 46345, 2024. DOI: 10.3390/app14146345.

Artículo en preparación

1. A. Gonzalez-Santocildes, J.-I. Vazquez, A. Eguiluz, I. Fidalgo Astorquia. *Can Robots Adapt to Human Comfort? Experimental Validation of a Reinforcement Learning Approach*. En preparación para su envío a revista indexada en JCR, 2026.

Conferencias internacionales

1. A. Gonzalez-Santocildes, J.-I. Vazquez, A. Eguiluz, P. García Bringas. *Reinforcement Learning for Dynamic Trajectory Adjustment in Human Robot Interaction Within Virtual Simulation*. En: *Hybrid Artificial Intelligent Systems*, Lecture Notes in Computer Science, Springer, Cham, 2024. DOI: 10.1007/978-3-031-74186-9_20.

Con el fin de ofrecer una visión integrada del desarrollo de esta tesis, se ha elaborado un esquema que relaciona los repositorios abiertos, las publicaciones científicas y los capítulos del documento. De este modo, se visualiza cómo los recursos generados (código y datos) han servido de base para las distintas contribuciones académicas y, a su vez, qué apartados de la memoria sustentan cada uno de estos resultados. Este mapa de trazabilidad facilita la comprensión global del flujo de la investigación y del proceso de transferencia desde los desarrollos técnicos hasta las aportaciones científicas.

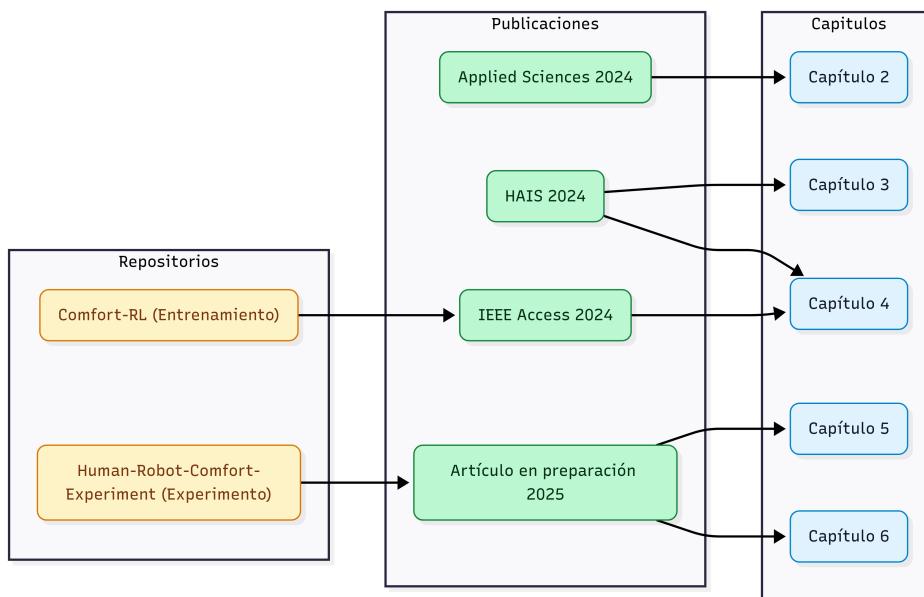


Figura 7.2: Mapa de trazabilidad entre los repositorios abiertos, las publicaciones científicas y los capítulos de la tesis.

En conjunto, los repositorios y publicaciones generados refuerzan la coherencia del trabajo desarrollado y facilitan su continuidad. Estos resultados ofrecen un punto de partida útil para las líneas futuras (apartado 7.4), orientadas a ampliar y consolidar el papel del confort humano en la robótica colaborativa.

7.4 Líneas futuras

Las contribuciones de esta tesis abren un abanico de posibilidades para avanzar en la integración del aprendizaje por refuerzo en la robótica colaborativa centrada en el confort humano. Si bien los resultados obtenidos han permitido validar la viabilidad de este enfoque en escenarios controlados, los desafíos asociados a la escalabilidad, la multimodalidad y la transferencia a entornos industriales reales sugieren varias direcciones de investigación. En esta sección se destacan algunas de las líneas futuras más relevantes, que pretenden consolidar y extender los hallazgos presentados.

En esta tesis se ha dado un paso inicial hacia el despliegue de trayectorias de modelos de aprendizaje por refuerzo en robots reales, demostrando que es posible trasladar los resultados de las políticas aprendidas en entornos simulados a escenarios físicos. No obstante, llevar este planteamiento a un nivel de producción plenamente funcional supone un reto considerable. Los sistemas robóticos que operan en entornos industriales deben afrontar condiciones de incertidumbre, interrupciones y dinámicas cambiantes que exigen una mayor robustez en la toma de decisiones. Además, el hecho de que las políticas aprendidas deban ejecutarse en tiempo real introduce limitaciones computacionales y riesgos asociados a la seguridad, lo que plantea la necesidad de investigar estrategias específicas para garantizar la fiabilidad de estos sistemas en contextos de alta exigencia (Zhao et al., 2020).

Una línea de investigación reciente consiste, por tanto, en explorar cómo desplegar políticas de aprendizaje por refuerzo de manera segura y eficiente en robots colaborativos que operen en entornos reales. Esto implica analizar en detalle los mecanismos de toma de decisiones en tiempo real, desarrollar marcos de supervisión que permitan detectar y corregir comportamientos no deseados, y diseñar arquitecturas capaces de adaptarse dinámicamente a las condiciones de trabajo (Salimpour et al., 2025). Profundizar en estos aspectos permitirá avanzar hacia el objetivo de contar con modelos que no solo funcionen en entornos experimentales controlados, sino que sean capaces de ofrecer un rendimiento estable y confiable en situaciones de producción (Tang et al., 2025; Da et al., 2025). En este sentido, la robustez técnica constituye un prerequisite esencial que conecta directamente con otras líneas de investigación orientadas a la transferencia y la validación en contextos reales.

De manera complementaria, otra dirección prometedora consiste en profundizar en el papel de los entornos de realidad virtual como alternativa o complemento a los experimentos físicos. A lo largo de esta tesis se ha mostrado que la VR puede ser un recurso útil cuando no se dispone del equipamiento robótico

necesario, ya sea para validar hipótesis preliminares o para acelerar las fases iniciales de entrenamiento. Sin embargo, todavía es necesario analizar en detalle hasta qué punto los resultados obtenidos en VR son transferibles a escenarios reales, y qué limitaciones surgen debido a las diferencias entre ambos contextos (Zhu et al., 2025).

En este sentido, futuras investigaciones deberían evaluar de manera sistemática la viabilidad de utilizar VR no solo como banco de pruebas, sino también como plataforma de entrenamiento de agentes en tiempo real. Este enfoque abriría la posibilidad de ajustar políticas de aprendizaje por refuerzo con retroalimentación directa del usuario en entornos virtuales inmersivos, reduciendo costes y riesgos asociados al uso del robot físico. Asimismo, comparar rigurosamente las discrepancias de comportamiento entre VR y la ejecución en entornos reales permitirá establecer métricas de fidelidad y diseñar mecanismos de compensación que faciliten una transferencia más fiable de los modelos entrenados (Albeaino et al., 2025). La exploración de la VR, por tanto, se cobra sentido con la búsqueda de mecanismos de validación y adaptación que refuercen la robustez de los sistemas en escenarios reales.

Más allá de los aspectos relacionados con los entornos de entrenamiento y despliegue, resulta fundamental profundizar en el análisis de los parámetros que el robot puede controlar de manera directa, como la velocidad de sus movimientos o la distancia respecto a la persona. Estas variables constituyen dimensiones críticas para garantizar interacciones seguras y aceptables, pero su influencia en la experiencia del usuario todavía puede explorarse con mayor detalle. Evaluar cómo variaciones en estos parámetros impactan en la percepción de confort, así como identificar posibles interacciones con otros factores contextuales (por ejemplo, el tipo de tarea, el nivel de carga del operador o la complejidad del entorno), representa una vía de investigación valiosa para el diseño de comportamientos colaborativos más adaptativos (Apraiz et al., 2025; Spurny et al., 2025). Este análisis se relaciona de manera directa con la necesidad de caracterizar con mayor precisión la experiencia subjetiva del usuario, permitiendo cerrar la brecha entre variables técnicas y percepciones humanas (Wong y Crowe, 2024).

En paralelo, resulta necesario avanzar en la definición y caracterización del confort humano. Mientras que en este trabajo se han establecido niveles de confort, la transición desde estados internos de la persona hacia un nivel cuantificable constituye un reto aún abierto (Mongile, 2025). Desarrollar metodologías que permitan inferir estos niveles de manera sistemática, ya sea a través de métricas fisiológicas, cuestionarios subjetivos o enfoques multimodales, se presenta como una línea futura especialmente prometedora. En este contexto, cobran re-

levancia metodologías emergentes como el *experience sampling* (Larson y Csikszentmihalyi, 2014; Kucuk y Sisman, 2020), el seguimiento ocular (*eye-tracking*) o la integración de señales de EEG, que ofrecen la posibilidad de capturar de forma más precisa y dinámica el estado del usuario (Ren et al., 2024; Gonzalez-Santocildes et al., 2024b). Este avance facilitaría la integración de medidas de confort más precisas en las funciones de recompensa, favoreciendo la personalización y la adaptación dinámica de los sistemas robóticos colaborativos.

Finalmente, una línea de trabajo a más largo plazo apunta hacia la integración de estas capacidades en escenarios colaborativos complejos que incluyan múltiples robots y varios usuarios humanos interactuando de forma simultánea. En dichos contextos, emergen desafíos adicionales relacionados con la coordinación, la equidad en la adaptación y la gestión de conflictos entre distintos estilos de interacción. A su vez, será necesario abordar cuestiones éticas y de aceptación social, incorporando principios de transparencia, explicabilidad y confianza en las decisiones tomadas por los agentes autónomos. Estos aspectos, junto con la exploración de la transferencia cultural y contextual de las políticas de aprendizaje, representan un campo de investigación que permitirá avanzar hacia robots colaborativos capaces de desenvolverse en entornos reales de manera segura, eficiente y socialmente aceptada.

Además, todas las líneas futuras aquí planteadas se ven reforzadas por la reciente proliferación de robots humanoides en entornos domésticos, asistenciales e industriales. En estos casos, el confort del usuario adquiere una dimensión adicional asociada a la apariencia y el comportamiento cuasi-humano del robot. La presencia de rasgos antropomórficos, expresiones faciales o gestos naturales puede modificar de forma significativa la percepción de cercanía, confianza o incomodidad del usuario, lo que introduce un nuevo factor a considerar en la modelización del confort. Analizar cómo la forma, el movimiento y la expresividad del robot influyen en la experiencia subjetiva del humano constituye, por tanto, una nueva línea de investigación emergente (Tong et al., 2024; Dai et al., 2025).

En conjunto, las líneas futuras aquí planteadas no deben entenderse como caminos aislados, sino como ámbitos complementarios que conviene abordar de manera integrada. Se recomienda reforzar la robustez técnica de las políticas de aprendizaje para garantizar su despliegue en escenarios reales; aprovechar los entornos de realidad virtual como recurso para acelerar la validación y transferencia de modelos; analizar de forma sistemática los parámetros de interacción controlados por el robot y su impacto en la experiencia del usuario; y avanzar en la caracterización del confort humano mediante metodologías emergentes y multimodales, integrándolas en funciones de recompensa basadas en aprendi-

zaje por refuerzo. Esta aproximación permitirá avanzar hacia robots más adaptativos, fiables y centrados en la persona.

7.5 Consideraciones finales

Las investigaciones presentadas en esta tesis han demostrado que la incorporación del confort humano como variable explícita en el diseño de sistemas robóticos colaborativos constituye una vía sólida para avanzar hacia interacciones más adaptativas y centradas en la persona. A lo largo del trabajo se ha demostrado la viabilidad de integrar métricas relacionadas con el nivel de confort del usuario dentro de marcos de aprendizaje por refuerzo, validando su impacto en entornos simulados y, especialmente, en escenarios reales con un robot físico. Esta aproximación no solo amplía el alcance de la robótica colaborativa tradicional, orientada principalmente a la productividad y la seguridad, sino que también añade una dimensión de bienestar que resulta clave para la aceptación y sostenibilidad de estas tecnologías en contextos cotidianos e industriales.

En este sentido, los resultados obtenidos aportan valor en distintos niveles. Desde el punto de vista científico, se han propuesto metodologías originales para cuantificar y modelar el confort, así como su integración en funciones de recompensa, contribuyendo a la literatura en interacción humano-robot y aprendizaje por refuerzo. Desde la perspectiva aplicada, el despliegue en un entorno físico real demuestra que estas aproximaciones no se limitan a un ejercicio teórico, sino que pueden implementarse en robots colaborativos comerciales, con implicaciones directas en la mejora de la aceptación por parte de los usuarios, en la mejora del bienestar y en el aumento de la eficiencia de las tareas compartidas. De este modo, la tesis se sitúa en la intersección entre la investigación básica y la transferencia tecnológica, reforzando la relevancia de su contribución.

Un aspecto adicional que merece ser destacado se encuentra en los entornos de entrenamiento y validación desarrollados a lo largo de esta tesis. Estos escenarios, diseñados tanto en simulación como en condiciones reales, constituyen una aportación metodológica de gran relevancia, ya que permiten reproducir y comparar diferentes enfoques de aprendizaje por refuerzo en tareas colaborativas. La posibilidad de transferir comportamientos desde la simulación al robot físico, junto con la disponibilidad de métricas específicas, aporta un marco sólido y reutilizable que refuerza la consistencia de los resultados. De este modo, los entornos creados trascienden el ámbito de esta investigación concreta y se consolidan como una base sobre la cual otros investigadores podrán continuar explorando nuevas aplicaciones y configuraciones.

En conjunto, esta tesis doctoral ha mostrado que la integración del confort humano en la robótica colaborativa no solo es posible, sino que aporta un valor diferenciador para avanzar hacia sistemas más responsables, adaptativos y sostenibles. La combinación de metodologías innovadoras, validación experimental y la construcción de entornos de entrenamiento sienta las bases para un nuevo paradigma en el que la productividad y la seguridad se equilibran con el bienestar y la aceptación de los usuarios en entornos robóticos.

Así, el trabajo aquí presentado contribuye a estrechar la distancia entre el desarrollo técnico y las necesidades reales de las personas abriendo el camino hacia una robótica más humana, donde eficiencia y confort conviven como pilares inseparables del progreso tecnológico.



Repositorio Comfort-RL

Este anexo presenta el repositorio *Comfort-RL*, desarrollado como parte de esta tesis doctoral y creado para facilitar la experimentación con funciones de recompensa centradas en el confort humano en entornos de aprendizaje por refuerzo. El repositorio incluye la implementación de los entornos utilizados en las simulaciones y los scripts de entrenamiento, así como herramientas de visualización de resultados. Además de describir las funcionalidades actualmente disponibles, se señalan actualizaciones futuras orientadas a ampliar su usabilidad y adaptabilidad a diferentes escenarios experimentales. Finalmente, se ofrece una revisión del código más relevante, con el fin de guiar al lector en la comprensión de su estructura interna y de los módulos principales que lo componen.

La estructura principal del repositorio es la siguiente:

- **Comfort-RL_Source_code/**: contiene el proyecto completo en Unity, que permite la personalización de entornos, la definición de recompensas y la configuración de los agentes de entrenamiento mediante ML-Agents.
- **pre_release_test_1_0/**: incluye un ejecutable preliminar (.exe) que facilita la exploración del entorno sin necesidad de compilar el proyecto.
- **VR-Robotic-SetUp/**: proporciona una escena específica en realidad virtual, diseñada para evaluar de forma inmersiva las políticas aprendidas y analizar la interacción humano-robot en entornos simulados.
- **utils_readme/**: carpeta de apoyo que contiene imágenes, animaciones y otros recursos gráficos empleados en la documentación y en el README del repositorio.

Cada uno de estos elementos se ha desarrollado con el objetivo de garantizar la reproducibilidad, facilitar la exploración rápida de resultados y permitir la extensión del entorno a nuevas tareas o configuraciones de confort. Como ejemplo ilustrativo, en la Figura A.1 se muestra una captura del *README* principal del repositorio, donde se resumen sus funcionalidades y características más destacadas.

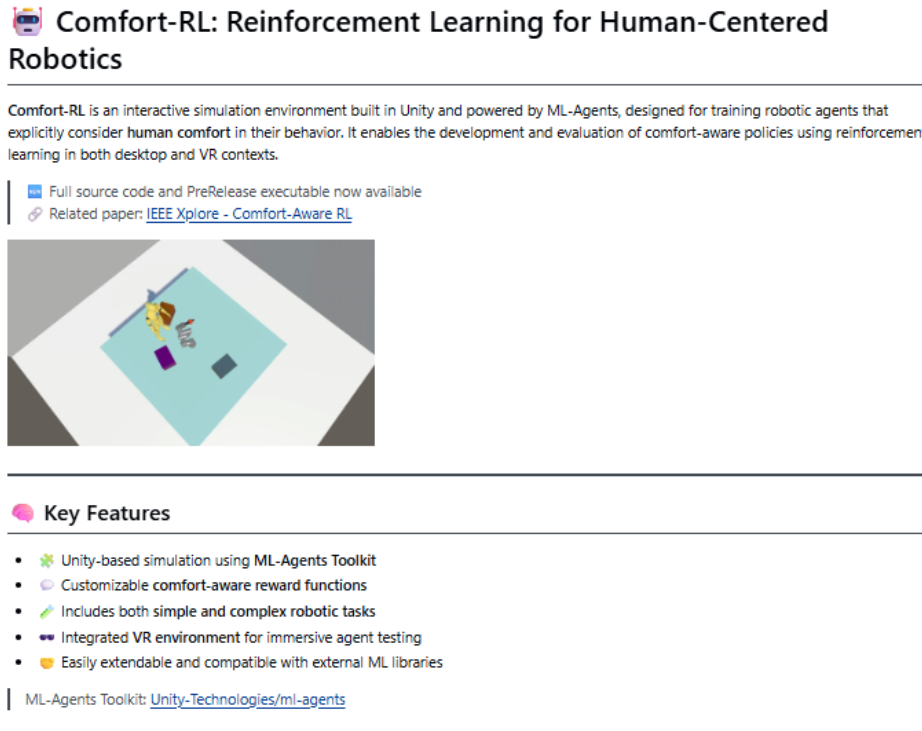


Figura A.1: Captura de pantalla del readme principal del repositorio.

De cara a futuras versiones, el repositorio *Comfort-RL* continuará ampliándose con nuevas funcionalidades. Entre las mejoras previstas se incluyen la documentación detallada de la configuración de recompensas y procedimientos de entrenamiento (extraído de la presente tesis), la incorporación de mecanismos de personalización adaptativa del confort y el desarrollo de pipelines sencillos de utilizar de exportación e importación de modelos de redes neuronales (formato .nn y .onnx). Estas actualizaciones buscan reforzar tanto la usabilidad como la capacidad de adaptación del entorno a escenarios cada vez más complejos.



Figura A.2: Captura de pantalla del entorno es su forma simple.



Figura A.3: Captura de pantalla del entorno industrial para test en VR.

```
{
  "goals": {
    "precisionGoal": 3.0
  },
  "penalties": {
    "activatePenalty": false
  },
  "phases": {
    "startInFirstPhaseOnly": false,
    "startInSecondPhaseOnly": false,
    "startInThirdPhaseOnly": false,
    "activateThirdPhase": true
  },
  "humanInteraction": {
    "humanMove": false,
    "personMovement": false,
    "teleportationDistance": 0.015,
    "humanMovementMax": 0.5,
    "spawnHuman": [6]
  },
  "environment": {
    "pinktablePosition": {
      "x": 3.9498484134674074,
      "y": -0.7431505918502808,
      "z": -4.255302429199219
    },
    "greytablePosition": {
      "x": 3.9498484134674074,
      "y": -0.7431505918502808,
      "z": -4.255302429199219
    },
    "orangetablePosition": {
      "x": 3.9498484134674074,
      "y": -0.7431505918502808,
      "z": -4.255302429199219
    },
    "colorBall": 1,
    "routeSaveFile": "save_file.txt",
    "numberOfEnvs": 4
  },
  "comfortParameters": {
    "temperature": 0.0,
    "numberOfComfortLevels": 10.0,
    "currentComfortLevel": 8.0
  }
}
```

Figura A.4: Captura de pantalla del formato del archivo de configuración.

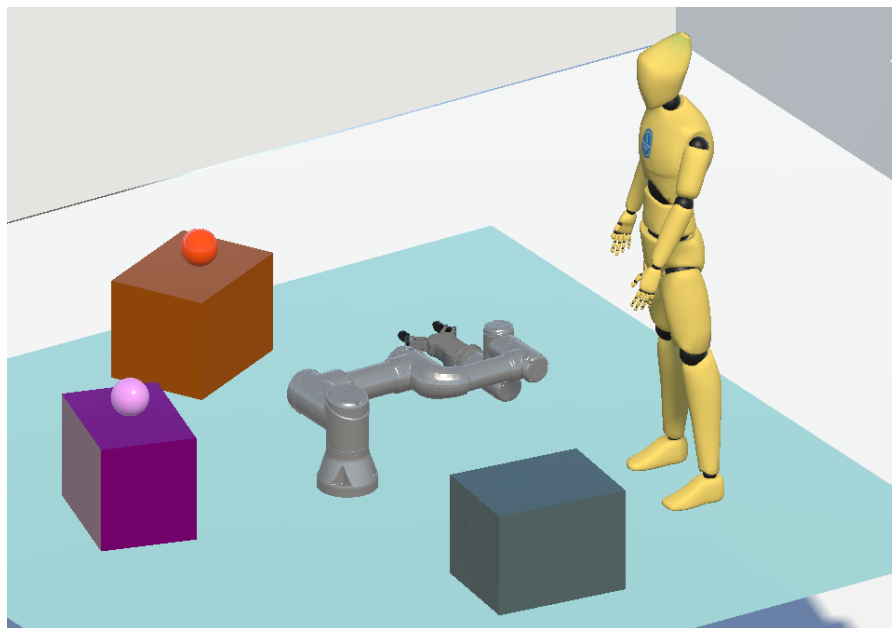


Figura A.5: Captura de pantalla del entorno de entrenamiento final con dummy humano.

Las Figuras A.2–A.5 ilustran distintos componentes y configuraciones del entorno desarrollado en el repositorio. En ellas se muestran ejemplos de la versión simplificada del escenario, el entorno industrial adaptado para pruebas en VR, el formato del archivo de configuración utilizado para personalizar recompensas y parámetros de entrenamiento, así como la versión final con la incorporación de un *dummy* humano para evaluar la interacción colaborativa. Estas capturas permiten visualizar de manera sintética la evolución del entorno y la diversidad de recursos disponibles para su uso experimental.

A continuación, se presentan los bloques de código más relevantes del repositorio, con el objetivo de resaltar su funcionamiento interno y facilitar la comprensión de la lógica implementada. Esta revisión no pretende ser exhaustiva, sino ofrecer una guía clara de los módulos principales sobre los que se apoya el desarrollo.

```
public override void OnEpisodeBegin()
{
    // inicializa contadores y limpia buffers de distancia
    steps = 0;
    ClearDistances();
}
```

```

// umbrales dependientes de confort
thresholdTemp = calculateThresholdBasedOnComfortLevelAndDistance();
thresholdAngle = calculateIncrementAnglePerStepBasedOnComfortLevel();

// reset del cobot y objetivos (fase 1 por defecto)
cobotControl.ResetCobot(1);
primerGoal = false;
spawnManager.MoveRandom();
colorPelota = spawnManager.ChangeMaterial(checkVR_socket);
SetMaterialBasedOnColor(colorPelota);
meshRendererGoal.enabled = true;

// activa movimiento humano aleatorio
if (humanMove) {
    humanMover.MoverHaciaEsferaAleatoria();
}

// activar tercera fase si procede
if (activarTerceraFase) {
    enTerceraFase = inicioEnTerceraFaseSolo;
}

// conmutacion de modelo (p.ej. para fases distintas)
if (modelSwitch) {
    switcher.SwitchModel(1);
}
}

```

Lista A.1: Inicio de episodio y gestion de fases

Comentario al Listado A.1. Se inicializa el episodio, se recalibran umbrales dependientes del confort y se configura el estado inicial del entorno. Se habilita el movimiento humano y, si procede, se selecciona el modelo de política adecuado.

```

public override void OnActionReceived(ActionBuffers actionBuffers)
{
    steps++;

    // aplica acciones continuas a 6 DoF con saturacion segura
    MoveAgentContinious(actionBuffers.ContinuousActions);

    // guarda articulaciones para ejecucion en robot real
    cobotControl.SaveJointPositions_deg(false, pathRealRobot);
}

```



```

// movimiento de la persona (exploracion espacial acotada)
if (movimientoPersona) {
    humanMover.MoveRandom();
    humanMover.MovePersonInRadius();
}

// penalizacion temporal por paso para promover eficiencia
if (activatePenalty) {
    AddReward(-0.1f);
}
}

public void MoveAgentContinious(ActionSegment<float> a)
{
    float j1 = Mathf.Clamp(a[0], -2f, 2f);
    float j2 = Mathf.Clamp(a[1], -2f, 2f);
    float j3 = Mathf.Clamp(a[2], -2f, 2f);
    float j4 = Mathf.Clamp(a[3], -2f, 2f);
    float j5 = Mathf.Clamp(a[4], -2f, 2f);
    float j6 = Mathf.Clamp(a[5], -2f, 2f);

    cobotControl.UpdateJointPosition(cobotControl.joint1, angleStep: j1);
    cobotControl.UpdateJointPosition(cobotControl.joint2, angleStep: j2);
    cobotControl.UpdateJointPosition(cobotControl.joint3, angleStep: j3);
    cobotControl.UpdateJointPosition(cobotControl.joint4, angleStep: j4);
    cobotControl.UpdateJointPosition(cobotControl.joint5, angleStep: j5);
    cobotControl.UpdateJointPosition(cobotControl.joint6, angleStep: j6);

    // gestor de recompensas por fase
    if (!enTerceraFase) {
        RewardManagerLeaveBall();
    } else {
        RewardTerceraFase();
    }
}
}

```

Lista A.2: Accion del agente y control articular

Comentario al Listado A.2. Se procesan acciones continuas, se actualiza el brazo y se invoca el gestor de recompensas según la fase (deposito o pose articular). Se incluye penalización por paso y registro de articulaciones para su uso en UR3e.

```

public override void CollectObservations(VectorSensor sensor)

```

```

{
    // juntas (estado propio del cobot)
    sensor.AddObservation(cobotControl.joint1.jointPosition[0]);
    sensor.AddObservation(cobotControl.joint2.jointPosition[0]);
    sensor.AddObservation(cobotControl.joint3.jointPosition[0]);
    sensor.AddObservation(cobotControl.joint4.jointPosition[0]);
    sensor.AddObservation(cobotControl.joint5.jointPosition[0]);
    sensor.AddObservation(cobotControl.joint6.jointPosition[0]);

    // persona: posicion relativa o punto clippeado
    if (!clipPositionPerson) {
        sensor.AddObservation(person.transform.position.x -
            transform.position.x);
        sensor.AddObservation(person.transform.position.z -
            transform.position.z);
    } else {
        sensor.AddObservation(
            calculateClippingPoint(person.transform.position)[0] -
                transform.position.x);
        sensor.AddObservation(
            calculateClippingPoint(person.transform.position)[1] -
                transform.position.z);
    }

    // objetivos (vectores de distancia) y etiqueta de color
    sensor.AddObservation(CalculateDistanceVectorTake());
    sensor.AddObservation(CalculateDistanceVectorGeneric(leaveOranje));
    sensor.AddObservation(CalculateDistanceVectorGeneric(leavePink));
    sensor.AddObservation(colorPelota);
}

```

Lista A.3: Vector de observaciones para la politica

Comentario al Listado A.3. El estado de observaciones del agente. incluye joints del robot, representación relativa de la persona (o punto clippeado) y vectores a objetivos, ademas de una etiqueta de color para desambiguar destinos de deposito.

```

public void RewardManagerLeaveBall()
{
    // fase 1: acercamiento al objeto
    if (!primerGoal) {
        float d = CalculateDistanceGeneric(goalReachSphere2);
        if (d * 10 < presitionGoal) {

```

```

        GoalReachedNewTask(); // transicion a fase 2
    }
}
// fase 2: deposito segun color
else {
    GameObject target = (colorPelota == 1) ? leavePink : leaveOranje;
    float d = CalculateDistanceGeneric(target);
    if (d * 10 < presitionGoal) {
        GoalReachedNewTask(); // transicion a fase 3 o fin
    }
}
}

public void RewardTerceraFase()
{
    // fase 3: validacion de pose articular objetivo
    List<float> jointDiff = cobotControl.CalculateAllJointAngleDifferences();
    if (cobotControl.JointsInPosition(jointDiff)) {
        GoalReachedNewTask(); // recompensa final y fin de episodio
    }
}

public void GoalReachedNewTask()
{
    if (!primerGoal) {
        meshRendererGoal.enabled = false;
        meshRendererPinza.enabled = true;
        GiveReward(50.0f); // recompensa intermedia
        primerGoal = true; // pasa a fase 2
    }
    else {
        if (!activarTerceraFase) {
            GiveReward(100.0f); // exito en fase 2
            EndEpisode();
        } else if (!enTerceraFase) {
            GiveReward(100.0f); // transicion a fase 3
            enTerceraFase = true;
            if (modelSwitch) switcher.SwitchModel(3);
            meshRendererPinza.enabled = false;
            if (colorPelota == 1) pinkBallLeaved.enabled = true; else
                orangeBallLeaved.enabled = true;
            meshRendererGoal.enabled = false;
        } else {
            GiveReward(150.0f); // exito en fase 3
        }
    }
}

```

```

        EndEpisode();
    }
}
}

```

Lista A.4: Gestion de objetivos por fases

Comentario al Listado A.4. Varias de las funciones de recompensa utilizadas durante las experimentaciones. Se estructura la tarea en tres fases: tomar, dejar y volver a pose articular. Las recompensas escalonadas (50/100/150) facilitan el aprendizaje por hitos y el cambio de modelo permite especializar la política en la fase final.

```

public float GeometricDistanceReward(float value)
{
    // umbrales dependientes del confort
    float thresholdSign = thresholdTemp; // cambio de signo
    float thresholdMax = distanceMaxThreshold; // distancia de confort plena

    if (value == 0) return -1.0f;

    // forma tanh suave y acotada
    float factor1 = (value - thresholdSign) / Mathf.Max(value, 1e-10f);
    float factor2 = Mathf.Max(thresholdMax - value, 1e-10f);
    float exponent = 2* (factor1 / factor2);
    float eToTheX = Mathf.Exp(exponent);
    float tanhValue = (eToTheX - 1) / (eToTheX + 1);

    // penaliza cercania, recompensa distancia aceptable segun confort
    return tanhValue;
}

float CalculateFloatRewardAngle(float j1, float j2, float j3, float j4, float
j5, float j6)
{
    // suavidad: penaliza pasos angulares grandes
    if (j1>thresholdAngle || j2>thresholdAngle || j3>thresholdAngle ||
j4>thresholdAngle || j5>thresholdAngle || j6>thresholdAngle)
        return -0.1f;
    return 0.0f;
}

float calculateThresholdBasedOnComfortLevelAndDistance()
{

```

```
// umbral de distancia escalado por nivel de confort
float increment = distanceMaxThreshold / numberOfComfortLevels;
return distanceMaxThreshold - increment * CurrentComfortLevel;
}

float calculateIncrementAnglePerStepBasedOnComfortLevel()
{
    // umbral de paso angular escalado por nivel de confort
    return 2.0f / numberOfComfortLevels * CurrentComfortLevel;
}
```

Lista A.5: Recompensa basada en confort (tangente hiperbolica) y umbrales

Comentario al Listado A.5. La función `GeometricDistanceReward` implementa una no linealidad tipo *tanh* que mapea distancia a recompensa según umbrales de confort (cambio de signo y distancia plena aceptable). Los umbrales se escalan con el nivel de confort y se añade un término de suavidad angular para desalentar movimientos bruscos. Es parte de la recompensa que resulto clave en el aprendizaje del robot en distintos niveles de confort.

En resumen, el repositorio *Comfort-RL* contiene en torno a 40.000 líneas de código, que abarcan tanto la lógica de los entornos en Unity como los scripts de entrenamiento, control y visualización. En este anexo se han mostrado únicamente los fragmentos más relevantes del agente y de las funciones de recompensa, como complemento a las descripciones ofrecidas en la tesis. De este modo, se proporciona una visión práctica y representativa de la implementación, sin necesidad de reproducir en detalle la totalidad del código disponible en el repositorio.

Repositorio Human-Robot-Comfort-Experiment

Este anexo presenta el repositorio *Human-Robot Comfort-Based Interaction Experiment (UR3e)*, desarrollado como parte de esta tesis doctoral para la validación experimental de los modelos y políticas entrenadas en el entorno *Comfort-RL*. Su objetivo es ofrecer un protocolo completo de replicación que permita reproducir el estudio empírico con participantes humanos y analizar el impacto del comportamiento del robot sobre el confort percibido.

El repositorio recopila el material, código y documentación necesarios para ejecutar los experimentos realizados con el robot colaborativo UR3e, incluyendo la sincronización de sensores, la configuración de trayectorias, los formularios de recogida de datos y las herramientas de análisis fisiológico. Además, complementa el entorno de simulación previo, estableciendo la conexión entre los resultados obtenidos mediante aprendizaje por refuerzo y su validación con humanos reales.

La estructura principal del repositorio es la siguiente:

- `ros_launch/`: archivos de lanzamiento de ROS para la ejecución de trayectorias del UR3e.
- `sensors/`: scripts adicionales para la sincronización de sensores fisiológicos y cámaras HD.
- `camera_calibration/`: diagramas de calibración, disposición de los sensores en el entorno experimental y scripts para el test de cámaras.

- `experiment_program/`: implementación en consola del programa experimental, responsable de la selección de trayectorias, la temporización y la velocidad de ejecución de cada sesión. Utilizado durante la ejecución de los experimentos para un control manual.
- `data_samples/`: ejemplos anonimizados de los datos recogidos (vectores AHR, WMI y cuestionarios subjetivos).
- `analysis/`: scripts de Python para el procesamiento de señales, extracción de características y análisis estadístico de los resultados.
- `trajectories/`: coordenadas espaciales de las trayectorias del robot empleadas en el experimento.
- `questionnaires/`: versiones editables en PDF de los formularios de evaluación subjetiva. Disponibles los formularios en el siguiente anexo.

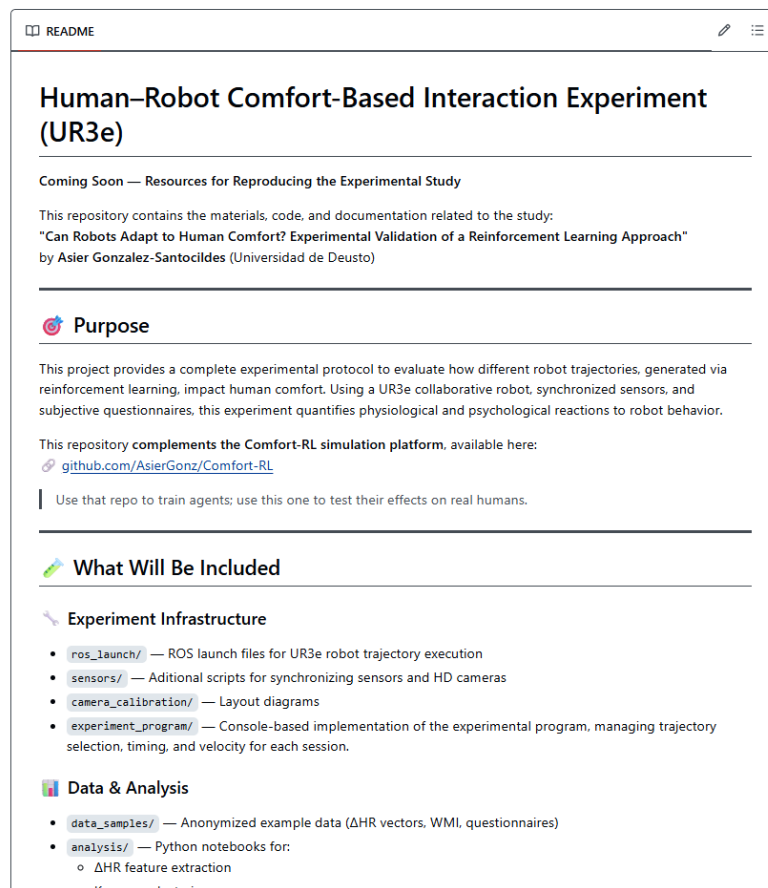


Figura B.1: Captura de pantalla del *README* principal del repositorio *Human-Robot Comfort-Based Interaction Experiment (UR3e)*.

Cada una de estas carpetas ha sido diseñada para garantizar la reproducibilidad del estudio y facilitar el uso del sistema tanto en nuevos experimentos como en entornos colaborativos similares. En la Figura B.1 se muestra una captura del *README* principal del repositorio, donde se resumen sus objetivos, estructura y contenidos.

Durante 2026 está previsto completar la documentación y los recursos disponibles en este repositorio, cuya publicación en su versión completa (*full release*) se encuentra planificada en paralelo a la difusión del tercer artículo asociado a esta línea de investigación. No obstante, en el presente anexo se describen y analizan los bloques de código y las aportaciones más relevantes del repositorio, con el fin de ofrecer una visión detallada de su estructura y de los elementos técnicos que sustentan la fase experimental del estudio.

A continuación, se presentan los bloques de código más relevantes del repositorio, con el objetivo de resaltar su funcionamiento interno y facilitar la comprensión de la lógica implementada. Esta revisión no pretende ser exhaustiva, sino ofrecer una guía clara de los módulos principales sobre los que se apoya el desarrollo.

```
class SincronizadorSuscriptor:
    def __init__(self):
        rospy.init_node("SincronizadorDatos")
        self.delay_entre_muestras = delay
        self.info_subscripcion = {
            "camara1": {"topic": "/usb_cam1/image_raw", "tipo":
                Image},
            "robot_joints": {"topic": "/joint_states", "tipo":
                JointState},
            "polar": {"topic": "/polar/heart_rate", "tipo":
                HeartBeat}
        }
        self.bridge = CvBridge()
        self.__data_list = []
        self.__registrar_subscribers()

    def __callback(self, *args):
        self.__data_list.append(args)
```

Lista B.1: Nodo ROS de sincronización de datos multimodales.

Comentario al Listado B.1. El nodo ROS SincronizadorSuscriptor coordina la adquisición simultánea de datos procedentes de tres cámaras USB, el estado articular del robot UR3e y el sensor de frecuencia cardíaca del participante. Los

mensajes se agrupan por marcas temporales y se almacenan en estructuras unificadas junto con las imágenes comprimidas en formato JPEG. Esta estrategia permite garantizar la coherencia temporal entre los distintos canales y reproducir con precisión la evolución de cada sesión experimental.

```
class PolarROSDriver:
    def __init__(self):
        rospy.init_node('polar_ros_driver')
        self.publisher = rospy.Publisher("/polar/heart_rate",
            HeartBeat, queue_size=10)
        self.heart_rate = 0
        self.stop_event = threading.Event()

    def heart_rate_handler(self, sender, data):
        """Callback ejecutado al recibir una notificacion BLE."""
        flags = data[0]
        hr_format = flags & 0x01
        hr = data[1] if hr_format == 0 else int.from_bytes(data
            [1:3], 'little')
        self.heart_rate = hr
        rospy.loginfo(f"Frecuencia cardiaca: {hr} bpm")

    async def ble_loop(self):
        """Conexion BLE y recepcion continua de datos."""
        async with BleakClient(ADDRESS) as client:
            await client.start_notify(HEART_RATE_MEASUREMENT_UUID
                , self.heart_rate_handler)
            while not self.stop_event.is_set():
                await asyncio.sleep(0.5)
            await client.stop_notify(HEART_RATE_MEASUREMENT_UUID)
```

Lista B.2: Nodo ROS para la lectura BLE de la frecuencia cardíaca (Polar Verity Sense).

Comentario al Listado B.2. El nodo `polar_ros_driver` establece una conexión BLE con el sensor *Polar Verity Sense* mediante la librería *Bleak*, extrayendo la frecuencia cardíaca y publicándola en ROS en forma de mensajes del tipo `HeartBeat`. La función `heart_rate_handler` interpreta los paquetes binarios recibidos y actualiza el valor de frecuencia cardíaca en tiempo real, que posteriormente se difunde a 10 Hz a través del tópico `/polar/heart_rate`. Este nodo constituye el vínculo entre la señal fisiológica y el sistema robótico, permitiendo analizar la respuesta del participante durante las distintas fases de interacción.

```
def pintar_menu(self, persona: str, delay: float, velocidad:
float) -> str:
    os.system('clear')
    print(f"Para {persona} == Delay: {delay} s == Velocidad: {
        velocidad}")
    print("      N: Trayectoria close.")
    print("      F: Trayectoria far.")
    print("      M: Trayectoria close + grabar.")
    print("      G: Trayectoria far + grabar.")
    print("      D: Anyadir delay al principio y final.")
    print("      V: Velocidad trayectorias. (0->1]")
    print("      X: Acabar persona.")
    print("      S: Screenshot.")
    print("      Q: Salir.\n")
    return input("Introduce la opcion: ")
```

Lista B.3: Interfaz de control en consola del nodo principal del experimento.

Comentario al Listado B.3. El nodo de control ofrece una interfaz de consola que permite seleccionar de forma manual las trayectorias a ejecutar, activar la grabación de datos sincronizados, modificar la velocidad de movimiento o introducir delays antes y después de la ejecución. Durante las sesiones experimentales, este menú fue empleado por el operador para gestionar la secuencia de tareas de cada participante, manteniendo una comunicación directa con el nodo de sincronización y el robot UR3e.

Lista B.4: Archivo de lanzamiento principal del sistema ROS (sistema.launch).

```
<launch>
  <arg name="sim" default="false" />
  <param name="sim" value="$(eval sim)" />
  <include file="$(find sistema_asier)/launch/cams.launch"/>
  <node name="sistema_asier_sincronizador" pkg="sistema_asier"
    type="suscriptor_sincronizador.py" output="log"/>
  <group if="$(eval sim)">
    <include file="$(find launcher_robots_lab_robotica)/
      launch/sim_205.launch"/>
  </group>
  <group if="$(eval not sim)">
    <include file="$(find launcher_robots_lab_robotica)/
      launch/robot_205.launch" />
  </group>
</launch>
```

Comentario al Listado B.4. El archivo sistema.launch constituye el punto de entrada del sistema experimental. Define el parámetro sim que permite alternar

entre el modo de simulación y el robot físico, e incluye de forma jerárquica los distintos nodos del sistema: las cámaras, el sincronizador de datos y el modelo del UR3e (real o simulado). De este modo, la ejecución del comando `roslaunch sistema sistema.launch` inicializa automáticamente todos los componentes necesarios para el desarrollo de cada sesión experimental.

```
# 1) Escalado
scaler = StandardScaler()
X = df_raw[["X1", "X2", "X3"]]
X_scaled = scaler.fit_transform(X)

# 2) Regla del codo
inertia, K_range = [], range(1, 10)
for k in K_range:
    km = KMeans(n_clusters=k, random_state=42, n_init='auto')
    km.fit(X_scaled)
    inertia.append(km.inertia_)
plt.figure(figsize=(6,4))
plt.plot(K_range, inertia, marker='o'); plt.tight_layout()
plt.savefig("elbow_plot.png", dpi=300); plt.close()

# 3) PCA (2 componentes)
pca = PCA(n_components=2, random_state=42)
scores = pca.fit_transform(X_scaled)
df_raw["PCA1"], df_raw["PCA2"] = scores[:,0], scores[:,1]
```

Lista B.5: Estandarización, estimación de k por regla del codo y reducción PCA.

Comentario al Listado B.5. Este fragmento de código constituye una implementación básica y una aproximación analítica inicial para el análisis de los resultados obtenidos en el experimento. A través de técnicas de escalado, reducción de dimensionalidad (*PCA*) y agrupamiento no supervisado (*K-Means*), se exploran patrones generales en las respuestas de los participantes, sirviendo como punto de partida para los estudios e interpretaciones más detalladas durante la tesis.

Finalmente, se aprovecha este anexo para incorporar una selección de imágenes ilustrativas del entorno experimental, el equipamiento empleado y la disposición general del escenario durante las sesiones con participantes. Estas fotografías permiten obtener una visión más global del proceso de experimentación, mostrando la configuración real del robot colaborativo UR3e, la ubicación de las cámaras y sensores fisiológicos, así como la interacción del voluntario en el espacio de trabajo. Con ello, se busca complementar la descripción técnica presentada a lo largo del anexo con una representación visual del entorno en el

que se desarrollaron las pruebas, facilitando la comprensión integral del montaje y de las condiciones bajo las cuales se recogieron los datos.



Figura B.2: Vista general del entorno experimental, mostrando el robot UR3e y dos cámaras empleadas para la captura de expresiones faciales.

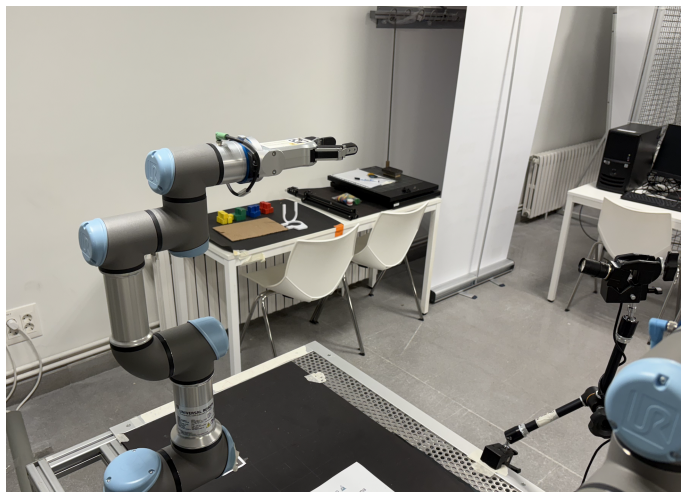


Figura B.3: Configuración del entorno experimental con el robot UR3e y una cámara dedicada al registro de expresiones faciales.

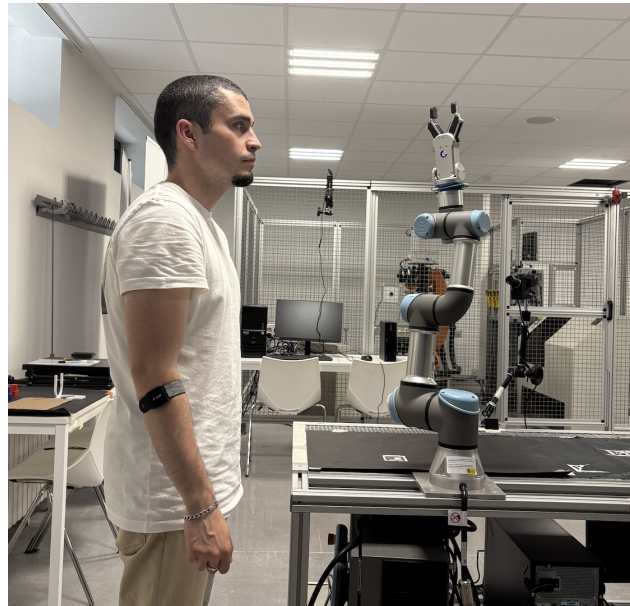


Figura B.4: Participante y robot en posición inicial del experimento. Se aprecia el sensor colocado en el brazo derecho del participante y el robot en su posición de origen.



Figura B.5: Entorno experimental con el robot UR3e, dos cámaras, una facial y una de movimiento al fondo.



Figura B.6: Vista panorámica del laboratorio con las tres cámaras instaladas y la disposición de las huellas que marcan las posiciones del participante.

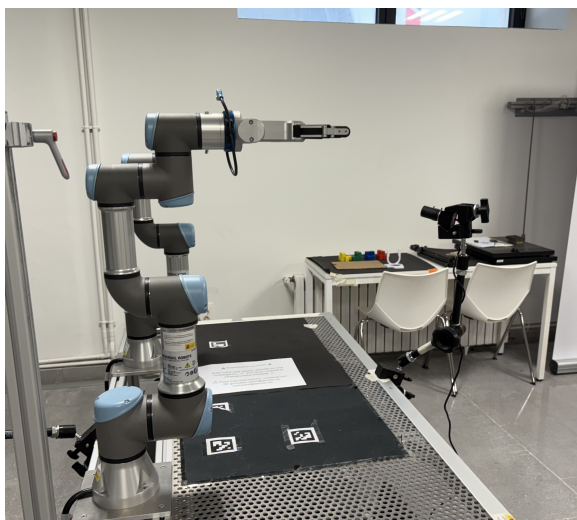


Figura B.7: Configuración del entorno experimental con una cámara para la captura de expresiones y seguimiento visual.

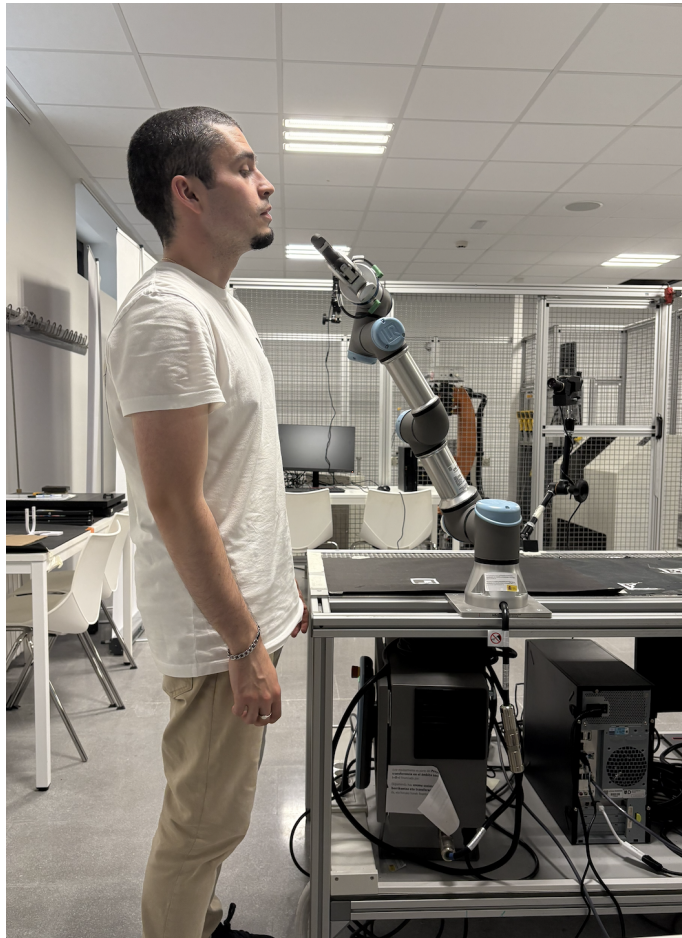


Figura B.8: Imagen capturada durante la ejecución del experimento, mostrando al participante y al robot UR3e en la distancia mínima aproximada de interacción.

Estructura, adquisición y preprocesado de los datos del experimento

Este anexo documenta de manera detallada la estructura de los datos en crudo adquiridos durante los experimentos con participantes humanos, así como el proceso de sincronización, preprocesado y generación de métricas derivadas.

El objetivo de esta documentación es reforzar la transparencia metodológica del estudio, garantizar su reproducibilidad y facilitar la reutilización del conjunto de datos en futuros trabajos. Para evitar sobrecargar el cuerpo principal del manuscrito, los detalles técnicos relativos al formato de los registros y a su gestión temporal se presentan de forma complementaria en este anexo.

Durante cada sesión de experimentación se registraron de forma simultánea múltiples flujos de información procedentes de sensores fisiológicos, cámaras de vídeo y del propio robot colaborativo. Todos los datos fueron gestionados mediante una arquitectura basada en ROS, que permitió la publicación y suscripción sincronizada de los distintos topics, así como el almacenamiento estructurado de la información en formato JSON.

Cada muestra registrada corresponde a una instantánea temporal del sistema, en la que se integran:

- Capturas de imagen procedentes de tres cámaras sincronizadas.
- Estado completo del robot, incluyendo posiciones, velocidades y esfuerzos articulares.

- Medidas fisiológicas instantáneas de frecuencia cardíaca.

La sincronización temporal entre estos flujos se realiza mediante marcas de tiempo expresadas en nanosegundos, lo que permite una alineación precisa entre eventos y respuestas humanas.

Cada registro se almacena como un objeto JSON que agrupa la información procedente de todos los sensores activos en un mismo instante temporal. A continuación se muestra un ejemplo representativo de un bloque de datos:

```
{
  "camara1": {
    "time_stamp": 1741088128738172684,
    "path_imagen": ".../camara1/imagen_2025_03_04_11_35_28.jpg"
  },
  "camara2": {
    "time_stamp": 1741088128742447339,
    "path_imagen": ".../camara2/imagen_2025_03_04_11_35_28.jpg"
  },
  "camara3": {
    "time_stamp": 1741088128740903187,
    "path_imagen": ".../camara3/imagen_2025_03_04_11_35_28.jpg"
  },
  "robot": {
    "time_stamp": 1741088128768124103,
    "joint_positions": [...],
    "joint_velocities": [...],
    "joint_efforts": [...]
  },
  "heart_beat": {
    "time_stamp": 1741088128770031690,
    "heart_beat": 92
  }
}
```

Cada uno de los campos del registro cumple una función específica dentro del análisis posterior:

- **Cámaras (camara1, camara2, camara3):** contienen la marca temporal de captura y la ruta al archivo de imagen almacenado. Las imágenes se procesan posteriormente para la detección facial y el análisis de expresiones.

- **Robot:** incluye el estado articular completo del robot en ese instante, con posiciones, velocidades y esfuerzos articulares, permitiendo relacionar de forma precisa el movimiento del robot con las reacciones humanas.
- **Frecuencia cardíaca (heart_beat):** registra el valor instantáneo de pulsaciones por minuto junto con su marca temporal, constituyendo la base para el cálculo de métricas fisiológicas derivadas.

La utilización de marcas temporales comunes derivadas del uso de ROS, permite alinear de forma precisa los distintos flujos de información. Esta sincronización resulta importante para asociar cambios fisiológicos a eventos concretos del movimiento del robot y para realizar análisis temporales continuos más allá de medidas puntuales.

A partir de los datos en crudo se aplicaron distintas etapas de preprocesado, adaptadas a cada tipo de señal, con el objetivo de reducir ruido, normalizar valores y extraer métricas relevantes para el análisis experimental.

- Se estimó una línea base de frecuencia cardíaca durante la fase inicial de relajación. Para cada trayectoria se calculó el valor máximo alcanzado, obteniendo la variación de frecuencia cardíaca según:

$$\Delta HR = HR_{max} - HR_{baseline}$$

- El movimiento corporal observado en las grabaciones de vídeo se codificó manualmente en una escala ordinal de 0 a 3. A partir de esta codificación se calculó un índice de movimiento ponderado (WMI) por condición experimental.
- Las imágenes frontales se procesaron mediante detección facial con *RetinaFace* y clasificación de emociones básicas utilizando *DeepFace*. Las emociones detectadas se agregaron temporalmente para cada trayectoria.
- Las valoraciones subjetivas de confort y seguridad se normalizaron en escalas Likert de 1 a 5, calculando medias por condición y participante, así como comparaciones relativas entre trayectorias.

Las etapas de preprocesado descritas dan lugar a un conjunto de métricas derivadas que constituyen la base del análisis presentado en el Capítulo 6. Estas métricas se sintetizan en la Tabla 5.2, donde se recogen de forma estructurada las señales originales, los momentos de captura, las etapas de preprocesado y

las variables finales empleadas en el análisis.

La disponibilidad de los datos en crudo, junto con la documentación detallada de su estructura y preprocesado, permite reconstruir de manera íntegra el flujo de análisis seguido en esta tesis. De este modo, las métricas presentadas en el capítulo de resultados pueden trazarse directamente hasta las señales originales adquiridas durante el experimento, reforzando la validez y la reproducibilidad del estudio.

Cuestionarios de evaluación del experimento

Con el fin de valorar la experiencia subjetiva de los participantes durante las pruebas, se diseñaron dos formularios breves centrados en la percepción de confort y seguridad en la interacción con el robot. Ambos instrumentos fueron empleados tras las distintas fases de la sesión experimental para recoger impresiones inmediatas sobre distintos aspectos de la trayectoria ejecutada y las diferencias percibidas entre condiciones.

A continuación se incluyen, en su formato original, los dos cuestionarios administrados:

3/11/25, 11:20

Formulario Experimentación Robótica Colaborativa & Confort - Formulario1

Formulario Experimentación Robótica Colaborativa & Confort - Formulario1

Definición de Confort:

El confort en un ámbito de trabajo con robots se refiere a tu sensación de comodidad y seguridad durante la interacción con el robot.

- **Comodidad:** Se refiere a cómo de agradable o libre de molestias percibes la interacción. En la escala del cuestionario, 1 significa "Nada cómodo - Nivel de confort muy bajo" y 5 significa "Muy cómodo - Nivel de confort muy alto".
- **Seguridad:** Se refiere a cómo de protegido o libre de riesgos percibes la interacción. En la escala del cuestionario, 1 significa "Nada seguro - Nivel de seguridad muy bajo" y 5 significa "Muy seguro - Nivel de seguridad muy alto".

gonzalez.asier@deusto.es [Cambiar de cuenta](#)



 No compartido

Código

Tu respuesta

Edad

Tu respuesta

Evaluación de Comodidad: En una escala del 1 al 5, donde 1 es "Nada cómodo" y 5 es "Muy cómodo", ¿Cómo evaluarías tu nivel de confort durante la **primera** trayectoria?

	1	2	3	4	5	
Nada cómodo	☐	☐	☐	☐	☐	Muy cómodo



https://docs.google.com/forms/d/e/1FAIpQLSeiH7NFTSs5X5E_WAL-Xq8MwH6CaTS5-_IDPjo1uSh6PWtQ/viewform

1/4

Figura D.1: Formulario 1 - parte 1: Evaluación de confort y seguridad durante la interacción con el robot.

3/11/25, 11:20

Formulario Experimentación Robótica Colaborativa & Confort - Formulario1

Evaluación de Comodidad: En una escala del 1 al 5, donde 1 es "Nada cómodo" y 5 es "Muy cómodo", ¿Cómo evaluarías tu nivel de confort durante la **segunda** trayectoria?

1 2 3 4 5

Nada cómodo ☐ ☐ ☐ ☐ ☐ Muy cómodo

Evaluación de Seguridad: En una escala del 1 al 5, donde 1 es "Nada seguro" y 5 es "Muy seguro", ¿Cómo evaluarías tu nivel de seguridad durante la **primera** trayectoria?

1 2 3 4 5

Nada seguro ☐ ☐ ☐ ☐ ☐ Muy seguro

Evaluación de Seguridad: En una escala del 1 al 5, donde 1 es "Nada seguro" y 5 es "Muy seguro", ¿Cómo evaluarías tu nivel de seguridad durante la **segunda** trayectoria?

1 2 3 4 5

Nada seguro ☐ ☐ ☐ ☐ ☐ Muy seguro

Preferencia de Trayectoria: ¿Con cuál de las dos trayectorias (primera y/o segunda) te encontraste mejor?

- ☐ Primera trayectoria
- ☐ Segunda trayectoria
- ☐ Ambas por igual
- ☐ Ninguna

https://docs.google.com/forms/d/e/1FAIpQLSei-17NFTSs5X5E_WAL-Xq8MwH6CaTS5-_IDPjoi1uSh6PWtQ/viewform

2/4

Figura D.2: Formulario 1 - parte 2: Evaluación de confort y seguridad durante la interacción con el robot.

6/10/25, 13:28

Formulario Experimentación Robótica Colaborativa & Confort - Formulario 2

Formulario Experimentación Robótica Colaborativa & Confort - Formulario 2

Definición de Confort:

El confort en un ámbito de trabajo con robots se refiere a tu sensación de comodidad y seguridad durante la interacción con el robot.

- **Comodidad:** Se refiere a cómo de agradable o libre de molestias percibes la interacción. En la escala del cuestionario, 1 significa "Nada cómodo - Nivel de confort muy bajo" y 5 significa "Muy cómodo - Nivel de confort muy alto".
- **Seguridad:** Se refiere a cómo de protegido o libre de riesgos percibes la interacción. En la escala del cuestionario, 1 significa "Nada seguro - Nivel de seguridad muy bajo" y 5 significa "Muy seguro - Nivel de seguridad muy alto".

gonzalez.asier@deusto.es [Cambiar de cuenta](#)



No compartido

Código

Tu respuesta

Evaluación de Comodidad: En una escala del 1 al 5, donde 1 es "Nada cómodo" y 5 es "Muy cómodo", ¿Cómo evaluarías tu nivel de confort durante la trayectoria observada con **velocidad ajustada**?

	1	2	3	4	5	
Nada cómodo	☐	☐	☐	☐	☐	Muy cómodo



https://docs.google.com/forms/d/e/1FAIpQLSf2d1vixG8_NQFaRGNTfIVvM7gKKdSJTwxn1Xz9ahCqhlEFBw/viewform

1/3

Figura D.3: Formulario 2 - parte 1: Evaluación comparativa tras la trayectoria con velocidad ajustada.

6/10/25, 13:28

Formulario Experimentación Robótica Colaborativa & Confort - Formulario 2

Evaluación de Seguridad: En una escala del 1 al 5, donde 1 es "Nada seguro" y 5 es "Muy seguro", ¿Cómo evaluarías tu nivel de confort durante la trayectoria observada con **velocidad ajustada**?

1 2 3 4 5

Nada seguro ☐ ☐ ☐ ☐ ☐ Muy seguro

Comparación con la Trayectoria Inicial: Comparando esta trayectoria con la que observaste anteriormente, ¿Con cuál de las dos trayectorias te encontraste mejor?

- ☐ Primera trayectoria
- ☐ Segunda trayectoria (Trayectoria velocidad ajustada)
- ☐ Ambas por igual
- ☐ Ninguna

[Enviar](#)[Borrar formulario](#)

Nunca envíes contraseñas a través de Formularios de Google.

Este formulario se creó en Universidad de Deusto. - [Contactar con el propietario del formulario](#)

Google Formularios



https://docs.google.com/forms/d/e/1FAIpQLSf2d1vixG8_NQFaRGNTfVVm7gKKdSJTwxn1Xz9ahCqhlEFBw/viewform

2/3

Figura D.4: Formulario 2 - parte 2: Evaluación comparativa tras la trayectoria con velocidad ajustada.

Bibliografía

- ALBEAINO, G., JEELANI, I., GHEISARI, M. y ISSA, R. R. Assessing proxemics impact on human-robot collaboration safety in construction: A virtual reality study with four-legged robots. *Safety Science*, vol. 181, página 106682, 2025. (ref. pág. 134)
- AMERSHI, S., ÇAKMAK, M., KNOX, W. B. y KULESZA, T. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, vol. 35(4), página 105–120, 2014. ISSN 2371-9621. (ref. pág. 29)
- ANDRYCHOWICZ, M., RAICHUK, A., STAŃCZYK, P., ORSINI, M., GIRGIN, S., MARI-NIER, R., HUSSENOT, L., GEIST, M., PIETQUIN, O., MICHALSKI, M., GELLY, S. y BACHEM, O. What matters in on-policy reinforcement learning? a large-scale empirical study. 2020. (ref. pág. 23)
- APRAIZ, A., LASA, G., MAZMELA, M., ARANA-AREXOLALEIBA, N., SERRANO MU-ÑOZ, A., ELGUEA, Í. y ETXABE, A. Evaluating the effect of speed and accelera-tion on human factors during an assembly task in human–robot interaction (hri). *International Journal of Social Robotics*, vol. 17(2), página 211–256, 2025. ISSN 1875-4805. (ref. pág. 134)
- AYOUBI, Y., LARIBI, M. A., ARSICAULT, M. y ZEGHLOUL, S. Safe phri via the varia-ble stiffness safety-oriented mechanism (v2som): Simulation and experimen-tal validations. *Applied Sciences*, vol. 10(11), página 3810, 2020. (ref. pág. 12)
- BARTNECK, C., KULIĆ, D., CROFT, E. y ZOGHBI, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, vol. 1(1), página 71–81, 2008. ISSN 1875-4805. (ref. pág. 35)
- BAUER, W., BENDER, M., BRAUN, M., RALLY, P. y SCHOLTZ, O. Lightweight robots in manual assembly–best to start simply. *Frauenhofer-Institut für Arbeitswir-tschaft und Organisation IAO, Stuttgart*, vol. 1, 2016. (ref. pág. 3)

- BECK, A. T., STEER, R. A. y BROWN, G. Beck depression Inventory–II. 2011. Title of the publication associated with this dataset: PsycTESTS Dataset. (ref. pág. 19)
- BEDDIAR, D. R., NINI, B., SABOKROU, M. y HADID, A. Vision-based human activity recognition: a survey. *Multimed. Tools Appl.*, vol. 79(41-42), páginas 30509–30555, 2020. (ref. pág. 19)
- BETHEL, C. L., SALOMON, K., MURPHY, R. R. y BURKE, J. L. Survey of psychophysiology measurements applied to human-robot interaction. En *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, páginas 732–737. IEEE, 2007. (ref. pág. 28)
- BIBBO, D., CORVINI, G., SCHMID, M., RANALDI, S. y CONFORTO, S. The impact of human-robot collaboration levels on postural stability during working tasks performed while standing: Experimental study. *JMIR Human Factors*, 2025. (ref. pág. 13, 67)
- BLOSS, R. Collaborative robots are rapidly providing major improvements in productivity, safety, programing ease, portability and cost while addressing many new applications. *Industrial Robot the International Journal of Robotics Research and Application*, vol. 43(5), páginas 463–468, 2016. (ref. pág. 12)
- BORBONI, A., ELAMVAZUTHI, I. y CUSANO, N. EEG-based empathic safe cobot. *Machines*, vol. 10(8), página 603, 2022. (ref. pág. 31)
- BOSTON DYNAMICS. Boston dynamics - industrial inspection solutions. 2025. Acceso el 18 de julio de 2025. (ref. pág. xiii, 26)
- VAN DEN BRULE, R., DOTSCH, R., BIJLSTRA, G., WIGBOLDUS, D. H. y HASELAGER, P. Do robot performance and behavioral style affect human trust? a multi-method approach. *International journal of social robotics*, vol. 6(4), páginas 519–531, 2014. (ref. pág. 33, 67)
- BRUNKE, L., GREEFF, M., HALL, A. W., YUAN, Z., ZHOU, S., PANERATI, J. y SCHOELLIG, A. P. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5(1), páginas 411–444, 2022. (ref. pág. 4)
- CHARALAMPOUS, K., KOSTAVELIS, I. y GASTERATOS, A. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, vol. 93, páginas 85–104, 2017. (ref. pág. 34)
- COHEN, A., TENG, E., BERGES, V.-P., DONG, R.-P., HENRY, H., MATTAR, M., ZOOK, A. y GANGULY, S. On the use and misuse of absorbing states in multi-agent

- reinforcement learning. *RL in Games Workshop AAAI 2022*, 2022. (ref. pág. 41, 42)
- COUMANS, E. y BAI, Y. Pybullet, a python module for physics simulation for games, robotics and machine learning. <http://pybullet.org>, 2016–2021. (ref. pág. 42)
- DA, L., TURNAU, J., KUTRALINGAM, T. P., VELASQUEZ, A., SHAKARIAN, P. y WEI, H. A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models. *arXiv preprint arXiv:2502.13187*, 2025. (ref. pág. 133)
- DAI, N., ZHANG, K., ZHANG, F., LI, J., ZHONG, J., HUANG, Y. y DING, H. Ai-assisted flexible electronics in humanoid robot heads for natural and authentic facial expressions. *The Innovation*, vol. 6(2), 2025. (ref. pág. 135)
- DE GRAAF, M. M., ALLOUCH, S. B. y KLAMER, T. Sharing a life with harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior*, vol. 43, páginas 1–14, 2015. ISSN 0747-5632. (ref. pág. 35)
- DE LA FUENTE, N. y GUERRA, D. A. V. A comparative study of deep reinforcement learning models: Dqn vs ppo vs a2c. *arXiv preprint arXiv:2407.14151*, 2024. (ref. pág. 23)
- DE SANTIS, A., SICILIANO, B., DE LUCA, A. y BICCHI, A. An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, vol. 43(3), páginas 253–270, 2008. ISSN 0094-114X. (ref. pág. 12)
- DRAGAN, A. D., LEE, K. C. y SRINIVASA, S. S. Legibility and predictability of robot motion. En *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, páginas 301–308. 2013. (ref. pág. 16, 33, 69)
- DUNCAN, J. A., ALAMBEIGI, F. y PRYOR, M. A survey of multimodal perception methods for human–robot interaction in social environments. *ACM Transactions on Human-Robot Interaction*, vol. 13, páginas 1 – 50, 2024. (ref. pág. 17)
- EIMER, T., LINDAUER, M. y RAILEANU, R. Hyperparameters in reinforcement learning and how to tune them. En *International conference on machine learning*, páginas 9104–9149. PMLR, 2023. (ref. pág. 23)
- ESTERWOOD, C., ESSENMACHER, K., YANG, H., ZENG, F. y ROBERT, L. A meta-analysis of human personality and robot acceptance in human-robot interaction. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. (ref. pág. 35)

- FAN, J., WANG, Z., XIE, Y. y YANG, Z. A theoretical analysis of deep q-learning. En *Learning for dynamics and control*, páginas 486–489. PMLR, 2020. (ref. pág. 23)
- FLORENSA, C., HELD, D., WULFMEIER, M., ZHANG, M. y ABBEEL, P. Reverse curriculum generation for reinforcement learning. En *Conference on robot learning*, páginas 482–495. PMLR, 2017. (ref. pág. 27)
- GERVASI, R., CAPPONI, M., MASTROGIACOMO, L. y FRANCESCHINI, F. Does size matter? exploring the effect of cobot size on user experience in human–robot collaboration. *The International Journal of Advanced Manufacturing Technology*, vol. 133(11–12), página 5777–5791, 2024. ISSN 1433-3015. (ref. pág. 14)
- GIALLANZA, A., LA SCALIA, G., MICALE, R. y LA FATA, C. M. Occupational health and safety issues in human-robot collaboration: State of the art and open challenges. *Safety science*, vol. 169, página 106313, 2024. (ref. pág. 28)
- GONZALEZ-SANTOCILDES, A. Comfort-rl. <https://github.com/AsierGonz/Comfort-RL>, 2025a. Accessed: November 10, 2025. (ref. pág. 9, 118)
- GONZALEZ-SANTOCILDES, A. Human-robot-comfort-experiment. <https://github.com/AsierGonz/Human-Robot-Comfort-Experiment>, 2025b. Accessed: October 28, 2025. (ref. pág. 9)
- GONZALEZ-SANTOCILDES, A., VAZQUEZ, J.-I. y EGUILUZ, A. Adaptive robot behavior based on human comfort using reinforcement learning. *IEEE Access*, vol. 12, páginas 122289–122299, 2024a. (ref. pág. 3, 118)
- GONZALEZ-SANTOCILDES, A., VAZQUEZ, J.-I. y EGUILUZ, A. Enhancing robot behavior with eeg, reinforcement learning and beyond: A review of techniques in collaborative robotics. *Applied Sciences*, vol. 14(14), página 6345, 2024b. ISSN 2076-3417. (ref. pág. 17, 31, 69, 135)
- GOODRICH, M. y SCHULTZ, A. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, vol. 1, páginas 203–275, 2007. (ref. pág. 15, 16, 69)
- GRIFFITH, S., SUBRAMANIAN, K., SCHOLZ, J., ISBELL, C. L. y THOMAZ, A. L. Policy shaping: Integrating human feedback with reinforcement learning. *Advances in neural information processing systems*, vol. 26, 2013. (ref. pág. 30)
- GULLETTA, G., SILVA, E. C. E., ERLHAGEN, W., MEULENBROEK, R., COSTA, M. F. P. y BICHO, E. A human-like upper-limb motion planner: Generating naturalistic movements for humanoid robots. *International Journal of Advanced Robotic Systems*, vol. 18(2), página 1729881421998585, 2021. (ref. pág. 33, 70)

- HAARNOJA, T., ZHOU, A., HARTIKAINEN, K., TUCKER, G., HA, S., TAN, J., KUMAR, V., ZHU, H., GUPTA, A., ABBEEL, P. ET AL. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018. (ref. pág. 23)
- HANCOCK, P. A., BILLINGS, D. R., SCHAEFER, K. E., CHEN, J. Y. C., DE VISSER, E. J. y PARASURAMAN, R. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 53(5), página 517–527, 2011. ISSN 1547-8181. (ref. pág. 17, 67, 69)
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. y FRIEDMAN, J. H. *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009. (ref. pág. 118)
- HENDERSON, P., ISLAM, R., BACHMAN, P., PINEAU, J., PRECUȚ, D. y MEGER, D. Deep reinforcement learning that matters. 2017. (ref. pág. 23)
- HENTOUT, A., MAOUDJ, A. y AOUACHE, M. A review of the literature on fuzzy-logic approaches for collision-free path planning of manipulator robots. *Artificial Intelligence Review*, vol. 56(4), páginas 3369–3444, 2023. (ref. pág. 17)
- HOFFMAN, G. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, vol. 49(3), páginas 209–218, 2019. (ref. pág. 34)
- HOFFMAN, G. y BREAZEAL, C. Collaboration in human-robot teams. *AIAA 1st Intelligent Systems Technical Conference*, 2004. (ref. pág. 16)
- HOWELL, P., KOLB, J., LIU, Y. y RAVICHANDAR, H. The effects of robot motion on comfort dynamics of novice users in close-proximity human-robot interaction. 2023. (ref. pág. 68, 69)
- HUSSEIN, A., GABER, M. M., ELYAN, E. y JAYNE, C. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, vol. 50(2), páginas 1–35, 2017. (ref. pág. 30)
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Robots and robotic devices – collaborative robots. 2016. Technical Specification. (ref. pág. 14, 69)
- ITURRATE, I., MONTESANO, L. y MINGUEZ, J. Robot reinforcement learning using EEG-based reward signals. En *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010. (ref. pág. 2, 30)

- JACKSON-KOKU, G. Beck depression inventory. *Occup. Med. (Lond.)*, vol. 66(2), páginas 174–175, 2016. (ref. pág. 19)
- JENSEN, L. C., FISCHER, K., SUVEI, S.-D. y BODENHAGEN, L. Timing of multimodal robot behaviors during human-robot collaboration. En *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, páginas 1061–1066. IEEE, 2017. (ref. pág. 34)
- JEON, H. J., LOSEY, D. P y SADIGH, D. Shared autonomy with learned latent actions. *arXiv preprint arXiv:2005.03210*, 2020. (ref. pág. 28)
- JONES, B. y KENWARD, M. G. *Design and Analysis of Cross-Over Trials*. Chapman and Hall/CRC, 2003. ISBN 9780429126000. (ref. pág. 96)
- JUMANI, A. K., SIDDIQUE, W. A., LAGHARI, A. A., ABRO, A. y KHAN, A. A. Virtual reality and augmented reality for education. En *Multimedia Computing Systems and Virtual Reality*, páginas 189–210. CRC Press, Boca Raton, 2022. (ref. pág. 19)
- KAR, R., GHOSH, L., KONAR, A., CHAKRABORTY, A. y NAGAR, A. K. EEG-induced autonomous game-teaching to a robot arm by human trainers using reinforcement learning. *IEEE Trans. Games*, vol. 14(4), páginas 610–622, 2022. (ref. pág. 2, 31)
- KAUFMANN, E., BAUERSFELD, L., LOQUERCIO, A., MÜLLER, M., KOLTUN, V. y SCARAMUZZA, D. Champion-level drone racing using deep reinforcement learning. *Nature*, vol. 620(7976), página 982–987, 2023. ISSN 1476-4687. (ref. pág. 26)
- KHAVAS, Z. R. A review on trust in human-robot interaction. 2021. (ref. pág. 17, 67, 69)
- KIM, S., ANTHIS, J. R. y SEBO, S. A taxonomy of robot autonomy for human-robot interaction. En *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, páginas 381–393. 2024. (ref. pág. 16)
- KNUDSEN, M. S. y KAIVO-OJA, J. Collaborative robots: Frontiers of current literature. *Journal of Intelligent Systems Theory and Applications*, 2020. (ref. pág. 12)
- KOBER, J., BAGNELL, J. A. y PETERS, J. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, vol. 32(11), páginas 1238–1274, 2013. (ref. pág. 4, 27)

- KOENIG, N. y HOWARD, A. Design and use paradigms for gazebo, an open-source multi-robot simulator. En *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)*(IEEE Cat. No. 04CH37566), vol. 3, páginas 2149–2154. Ieee, 2004. (ref. pág. 43)
- KORMUSHEV, P., CALINON, S. y CALDWELL, D. Reinforcement learning in robotics: Applications and real-world challenges. *Robotics*, vol. 2(3), páginas 122–148, 2013. (ref. pág. 4, 15, 26, 28)
- KUCUK, S. y SISMAN, B. Students' attitudes towards robotics and stem: Differences based on gender and robotics experience. *International Journal of Child-Computer Interaction*, vol. 23, página 100167, 2020. (ref. pág. 135)
- KUMAR, A. Reinforcement learning for robotic-assisted surgeries: Optimizing procedural outcomes and minimizing post-operative complication. *Int. J. Res. Publ. Rev*, vol. 6, páginas 5669–5684, 2025. (ref. pág. 25)
- LARSON, R. y CSIKSZENTMIHALYI, M. The experience sampling method. En *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*, páginas 21–34. Springer, 2014. (ref. pág. 135)
- LASI, H., FETTKE, P., KEMPER, H.-G., FELD, T. y HOFFMANN, M. Industry 4.0. *Business & information systems engineering*, vol. 6(4), páginas 239–242, 2014. (ref. pág. 12)
- LASOTA, P. A., FONG, T., SHAH, J. A. ET AL. A survey of methods for safe human-robot interaction. *Foundations and Trends® in Robotics*, vol. 5(4), páginas 261–349, 2017. (ref. pág. 33)
- LEUSMANN, J., FELDER, L., WANG, C. y MAYER, S. Understanding preferred robot reaction times for human-robot handovers supported by a deep learning system. *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, páginas 4–13, 2025. (ref. pág. 34)
- LEVINE, S., KUMAR, A., TUCKER, G. y FU, J. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020. (ref. pág. 27)
- LI, R., VAN ALMKERK, M., VAN WAVEREN, S., CARTER, E. y LEITE, I. Comparing human-robot proxemics between virtual reality and the real world. En *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)*, páginas 431–439. IEEE, 2019. (ref. pág. 103)
- LI, S. E. Deep reinforcement learning. En *Reinforcement learning for sequential decision and optimal control*, páginas 365–402. Springer, 2023. (ref. pág. 22)

- LI, Y. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017. (ref. pág. 22)
- LIM, V., ROOKSBY, M. y CROSS, E. S. Social robots on a global stage: Establishing a role for culture during human–robot interaction. *International Journal of Social Robotics*, vol. 13, páginas 1307 – 1333, 2020. (ref. pág. 35)
- LIN, Y., CHEN, L., ALI, A., NUGENT, C., CLELAND, I., LI, R., DING, J. y NING, H. Human digital twin: A survey. *Journal of Cloud Computing*, vol. 13(1), página 131, 2024. (ref. pág. 30)
- LOU, G., WANG, C., XU, Z., LIANG, J. y ZHOU, Y. Controlling soft robotic arms using hybrid modelling and reinforcement learning. *IEEE Robotics and Automation Letters*, vol. 9, páginas 7070–7077, 2024. (ref. pág. 27)
- MARISCAL, M. A., ORTIZ BARCINA, S., GARCÍA HERRERO, S. y LÓPEZ PEREA, E. M. Working with collaborative robots and its influence on levels of working stress. *Int. J. Comput. Integr. Manuf.*, páginas 1–20, 2023. (ref. pág. 18)
- MATHESON, E., MINTO, R., ZAMPIERI, E. G., FACCIO, M. y ROSATI, G. Human–robot collaboration in manufacturing applications: A review. *Robotics*, vol. 8(4), página 100, 2019. (ref. pág. 12)
- MEHRI SHERVEDANI, A., LI, S., MONAIKUL, N., ABBASI, B., ŽEFRAN, M. y DI EUGENIO, B. Multimodal reinforcement learning for robots collaborating with humans. *International Journal of Social Robotics*, 2025. ISSN 1875-4805. (ref. pág. 2)
- MNIH, V., KAVUKCUOGLU, K., SILVER, D., RUSU, A. A., VENESS, J., BELLEMARE, M. G., GRAVES, A., RIEDMILLER, M., FIDJELAND, A. K., OSTROVSKI, G. ET AL. Human-level control through deep reinforcement learning. *nature*, vol. 518(7540), páginas 529–533, 2015. (ref. pág. 23)
- MONGILE, S. Towards adaptive human-robot interaction: A comfort-driven architecture for social robots. 2025. (ref. pág. 134)
- MUMM, J. y MUTLU, B. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. En *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, páginas 331–338. 2011. (ref. pág. 34, 67, 69)
- MUSIĆ, S. y HIRCHE, S. Control sharing in human-robot team interaction. *Annu. Rev. Control.*, vol. 44, páginas 342–354, 2017. (ref. pág. 16)
- NGUYEN, H. y LA, H. Review of deep reinforcement learning for robot manipulation. En *2019 Third IEEE International Conference on Robotic Computing (IRC)*, páginas 590–595. 2019. (ref. pág. 26)

- NIKOLAIDIS, S., HSU, D. y SRINIVASA, S. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, vol. 36(5-7), páginas 618–634, 2017. (ref. pág. 28)
- NVIDIA. Isaac Sim. 2025. (ref. pág. 42)
- ONNASCH, L., SCHWEIDLER, P. y SCHMIDT, H. The potential of robot eyes as predictive cues in HRI-an eye-tracking study. *Front. Robot. AI*, vol. 10, página 1178433, 2023. (ref. pág. 18)
- ONOSE, G., GROZEA, C., ANGHELESCU, A., DAIA, C., SINESCU, C. J., CIUREA, A. V., SPIRCU, T., MIREA, A., ANDONE, I., SPÂNU, A., POPESCU, C., MIHĂESCU, A.-S., FAZLI, S., DANÓCZY, M. y POPESCU, F. On the feasibility of using motor imagery EEG-based brain-computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. *Spinal Cord*, vol. 50(8), páginas 599–608, 2012. (ref. pág. 18)
- PACCHIEROTTI, E., CHRISTENSEN, H. I. y JENSELT, P. Evaluation of passing distance for social robots. En *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, páginas 315–320. 2006. (ref. pág. 34, 67, 69)
- PADURARU, C., MANKOWITZ, D., DULAC-ARNOLD, G., LI, J., LEVINE, N., GOWAL, S. y HESTER, T. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, vol. 110, páginas 2419 – 2468, 2021. (ref. pág. 26)
- PAL, S., MUKHOPADHYAY, S. y SURYADEVARA, N. Development and progress in sensors and technologies for human emotion recognition. *Sensors*, vol. 21(16), página 5554, 2021. (ref. pág. 17)
- PATALAS-MALISZEWSKA, J., DUDEK, A., PAJAŁ, G. y PAJAŁ, I. Working toward solving safety issues in human–robot collaboration: A case study for recognising collisions using machine learning algorithms. *Electronics*, vol. 13(4), página 731, 2024. (ref. pág. 14)
- PAULÍKOVÁ, A., BABEJOVÁ, Z. G. y UBÁROVÁ, M. Analysis of the impact of human–cobot collaborative manufacturing implementation on the occupational health and safety and the quality requirements. *International Journal of Environmental Research and Public Health*, vol. 18(4), página 1927, 2021. (ref. pág. 14)
- PEREIRA, E., SIGCHA, L., SILVA, E., SAMPAIO, A., COSTA, N. y COSTA, N. Capturing mental workload through physiological sensors in human–robot collaboration: A systematic literature review. *Applied Sciences*, vol. 15(6), página 3317, 2025. (ref. pág. 28)

- PÉREZ, L., RODRÍGUEZ, Í., RODRÍGUEZ, N., USAMENTIAGA, R. y GARCÍA, D. F. Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors (Basel)*, vol. 16(3), página 335, 2016. (ref. pág. 19)
- PLOMIN, J., SCHWEIDLER, P y OEHME, A. Virtual reality check: a comparison of virtual reality, screen-based, and real world settings as research methods for hri. *Frontiers in Robotics and AI*, vol. 10, página 1156715, 2023. (ref. pág. 103)
- RAFFIN, A., HILL, A., GLEAVE, A., KANERVISTO, A., ERNESTUS, M. y DORMANN, N. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, vol. 22(268), páginas 1–8, 2021. (ref. pág. 40, 41)
- REDONDO, M. E. L., NIEWIADOMSKI, R., REA, F., INCAO, S., SANDINI, G. y SCIUTTI, A. Comfortability analysis under a human–robot interaction perspective. *International Journal of Social Robotics*, vol. 16(1), página 77–103, 2023. ISSN 1875-4805. (ref. pág. 69)
- REN, H., ZHENG, Z., ZHANG, J., WANG, Q. y WANG, Y. Electroencephalography (eeg)-based comfort evaluation of free-form and regular-form landscapes in virtual reality. *Applied Sciences*, vol. 14(2), página 933, 2024. (ref. pág. 135)
- RETZLAFF, C. O., DAS, S., WAYLLACE, C., MOUSAVI, P., AFSHARI, M., YANG, T., SARANTI, A., ANGERSCHMID, A., TAYLOR, M. E. y HOLZINGER, A. Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities. *Journal of Artificial Intelligence Research*, vol. 79, páginas 359–415, 2024. (ref. pág. 30, 68)
- ROY, R., DROUGARD, N., GATEAU, T., DEHAIS, F. y CHANEL, C. P. C. How can physiological computing benefit human-robot interaction? *Robotics*, vol. 9, página 100, 2020. (ref. pág. 19)
- RUDENKO, A., PALMIERI, L., HERMAN, M., KITANI, K. M., GAVRILA, D. M. y ARRAS, K. O. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, vol. 39(8), páginas 895–935, 2020. (ref. pág. 33)
- SALIMPOUR, S., PEÑA-QUERALTA, J., PAEZ-GRANADOS, D., HEIKKONEN, J. y WESTERLUND, T. Sim-to-real transfer for mobile robots with reinforcement learning: from nvidia isaac sim to gazebo and real ros 2 robots. *arXiv preprint arXiv:2501.02902*, 2025. (ref. pág. 133)
- SALVATO, E., FENU, G., MEDVET, E. y PELLEGRINO, F. A. Crossing the reality gap: a survey on sim-to-real transferability of robot controllers in reinforcement learning. *IEEE Access*, vol. PP, páginas 1–1, 2021. (ref. pág. 27)

- SCHOBBER, P., BOER, C. y SCHWARTE, L. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, vol. 126, página 1763–1768, 2018. (ref. pág. 117)
- SCHULMAN, J., WOLSKI, F., DHARIWAL, P., RADFORD, A. y KLIMOV, O. Proximal policy optimization algorithms. 2017. (ref. pág. 22)
- SHERIDAN, T. B. Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41(4), páginas 662–667, 2011. (ref. pág. 16)
- SHERIDAN, T. B. Human–robot interaction: status and challenges. *Human factors*, vol. 58(4), páginas 525–532, 2016. (ref. pág. 1, 4, 15, 28, 67, 69)
- SHI, T., WU, Y., SONG, L., ZHOU, T. y ZHAO, J. Efficient reinforcement fine-tuning via adaptive curriculum learning. *arXiv preprint arXiv:2504.05520*, 2025. (ref. pág. 27)
- SHU, L., XIE, J., YANG, M., LI, Z., LI, Z., LIAO, D., XU, X. y YANG, X. A review of emotion recognition using physiological signals. *Sensors*, vol. 18(7), página 2074, 2018. ISSN 1424-8220. (ref. pág. xiii, 18)
- SICILIANO, B., SCIAVICCO, L., VILLANI, L. y ORIOLO, G. Robotics: Modelling, planning and control. En *Springer Handbook of Robotics* (editado por B. Siciliano y O. Khatib), páginas 37–80. Springer, 2016. Mención histórica al robot Unimate. (ref. pág. 12)
- SINGAMANENI, P. T., BACHILLER-BURGOS, P., MANSO, L. J., GARRELL, A., SANFELIU, A., SPALANZANI, A. y ALAMI, R. A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*, vol. 43(10), páginas 1533–1572, 2024. (ref. pág. 34)
- SLADE, P., ATKESON, C., DONELAN, J. M., HOUDIJK, H., INGRAHAM, K. A., KIM, M., KONG, K., POGGENSEE, K. L., RIENER, R., STEINERT, M. ET AL. On human-in-the-loop optimization of human–robot interaction. *Nature*, vol. 633(8031), páginas 779–788, 2024. (ref. pág. 29, 68)
- SPURNY, T., BABJAK, J., BOBOVSKY, Z. y VYSOCKY, A. Dynamic proxemic model for human–robot interactions using the golden ratio. *Applied Sciences*, vol. 15(15), página 8130, 2025. ISSN 2076-3417. (ref. pág. 134)
- STAMLER, J. S., GOLDMAN, M. E., GOMES, J., MATZA, D. y HOROWITZ, S. F. The effect of stress and fatigue on cardiac rhythm in medical interns. *J. Electrocardiol.*, vol. 25(4), páginas 333–338, 1992. (ref. pág. 19)

- SUTTON, R. S. y BARTO, A. G. *Reinforcement learning: An introduction*. MIT press, 2018. (ref. pág. xiii, 1, 4, 20, 21)
- TANG, C., ABBATEMATTEO, B., HU, J., CHANDRA, R., MARTÍN-MARTÍN, R. y STONE, P. Deep reinforcement learning for robotics: A survey of real-world successes. En *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, páginas 28694–28698. 2025. (ref. pág. 25, 26, 133)
- TODOROV, E., EREZ, T. y TASSA, Y. Mujoco: A physics engine for model-based control. En *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, páginas 5026–5033. IEEE, 2012. (ref. pág. 41)
- TOICHOA EYAM, A., MOHAMMED, W. M. y MARTINEZ LASTRA, J. L. Emotion-driven analysis and control of human-robot interactions in collaborative applications. *Sensors (Basel)*, vol. 21(14), página 4626, 2021. (ref. pág. 17, 30)
- TONG, Y., LIU, H. y ZHANG, Z. Advancements in humanoid robots: A comprehensive review and future prospects. *IEEE/CAA Journal of Automatica Sinica*, vol. 11(2), página 301–328, 2024. ISSN 2329-9274. (ref. pág. 135)
- TOWERS, M., KWIATKOWSKI, A., TERRY, J., BALIS, J. U., DE COLA, G., DELEU, T., GOULÃO, M., KALLINTERIS, A., KRIMMEL, M., KG, A. ET AL. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024. (ref. pág. 41)
- UNITY TECHNOLOGIES. Documentación oficial de Unity ML-Agents (versión 2019.4). 2019. Accedido el 26 de julio de 2025. (ref. pág. 44)
- UNITY TECHNOLOGIES. Unity urdf importer. <https://github.com/Unity-Technologies/URDF-Importer>, 2023. Accedido el 26 de julio de 2025. (ref. pág. 45)
- UNITY TECHNOLOGIES. Unity ML-Agents Documentation. 2025. Accedido el 26 de julio de 2025. (ref. pág. 44)
- VILLAGRÁN, N. V., PESADO, P. M. y ESTÉVEZ, E. Cloud robotics for industry 4.0 - a literature review. *IFIP Advances in Information and Communication Technology*, vol. 566, páginas 3–14, 2020. (ref. pág. 13)
- VILLANI, V., PINI, F., LEALI, F. y SECCHI, C. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, vol. 55, páginas 248–266, 2018. (ref. pág. 1, 67, 69)
- WANG, T., ZHENG, P., LI, S. y WANG, L. Multimodal human–robot interaction for human-centric smart manufacturing: a survey. *Advanced Intelligent Systems*, vol. 6(3), página 2300359, 2024. (ref. pág. 35)

- WANG, Y., WU, J., HE, H., WEI, Z. y SUN, F. Data-driven energy management for electric vehicles using offline reinforcement learning. *Nature Communications*, vol. 16(1), página 2835, 2025. (ref. pág. 27)
- WIESE, K., BOCK, E. y MÜLLER, G. Human-robot collaboration in industrial assembly: Trends and developments. *Procedia CIRP*, vol. 81, páginas 122–127, 2019. (ref. pág. 13)
- WONG, S. W. y CROWE, P. Cognitive ergonomics and robotic surgery. *Journal of Robotic Surgery*, vol. 18(1), página 110, 2024. (ref. pág. 134)
- XINTARAKOU, A., SOUSONIS, V., ASVESTAS, D., VARDAS, P. E. y TZEIS, S. Remote cardiac rhythm monitoring in the era of smart wearables: Present assets and future perspectives. *Front. Cardiovasc. Med.*, vol. 9, página 853614, 2022. (ref. pág. 19)
- YAN, Y. y JIA, Y. A review on human comfort factors, measurements, and improvements in human-robot collaboration. *Sensors*, vol. 22(19), página 7431, 2022. (ref. pág. 69)
- YAN, Y., SU, H. y JIA, Y. Modeling and analysis of human comfort in human-robot collaboration. *Biomimetics*, vol. 8(6), página 464, 2023. ISSN 2313-7673. (ref. pág. 69)
- YANCO, H. y DRURY, J. A taxonomy for human-robot interaction. páginas 111–119, 2002. (ref. pág. 16)
- ZHAO, K., HE, H., WANG, Y. y XU, W. Eeg-based human state prediction for adaptive human-robot collaboration. *IEEE Transactions on Human-Machine Systems*, vol. 52(5), páginas 450–459, 2022. (ref. pág. 15)
- ZHAO, W., QUERALTA, J. P. y WESTERLUND, T. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. En *2020 IEEE symposium series on computational intelligence (SSCI)*, páginas 737–744. IEEE, 2020. (ref. pág. 25, 133)
- ZHU, S., MOU, L., LI, D., YE, B., HUANG, R. y ZHAO, H. Vr-robo: A real-to-sim-to-real framework for visual robot navigation and locomotion. *IEEE Robotics and Automation Letters*, 2025. (ref. pág. 134)
- ZHU, X., LIANG, Y., SUN, H., WANG, X. y REN, B. Robot obstacle avoidance system using deep reinforcement learning. *Ind. Rob.*, vol. 49(2), páginas 301–310, 2022. (ref. pág. 19)

Esta tesis fue terminada de escribir en Bilbao el 20 de enero de 2026

