

## PLAN DE INVESTIGACIÓN

DATOS DEL DOCTORANDO	
Datos personales	
Apellidos	Nombre
Lucas Zaragoza	Rubén
DNI, NIE, Pasaporte	Email
06282454G	ruben.lucas.zaragoza@gmail.com
Datos Académicos	
Programa de Doctorado en	Línea de investigación
Tecnologías de la Información y las Comunicaciones	Inteligencia artificial, aprendizaje por refuerzo, Aplicación práctica de Deep reinforcement learning en robots
Título de la tesis	

DATOS DE DIRECCIÓN DE TESIS	
Director de Tesis	
Apellidos	Cañas Plaza
Nombre	José María
DNI	
Email	josemaria.plaza@urjc.es
Departamento/Universidad/Centro	Teoría de la Señal y Comunicaciones y Sistemas Telemáticos/Universidad Rey Juan Carlos
Codirector de Tesis	
Apellidos	Calvo Palomino
Nombre	Roberto
DNI	46850094Z
Email	roberto.calvo@urjc.es
Departamento/Universidad/ Centro	Teoría de la Señal y Comunicaciones y Sistemas Telemáticos/Universidad Rey Juan Carlos
Tutor (si difiere del Director de Tesis)	
Apellidos	
Nombre	
DNI	
Email	
Departamento/Universidad/ Centro	

Los datos personales recogidos serán incorporados y tratados en el fichero *Tesis Doctorales*, cuya finalidad es la gestión de las tesis doctorales, inscrito en el Registro de Ficheros de Datos Personales de la Agencia Española de Protección de Datos (<https://www.agpd.es/portalwebAGPD/index-ides-idphp.php>). El órgano responsable del fichero es el Vicerrectorado de Títulos Propios, Formación Continua y Posgrado, y la dirección donde el interesado podrá ejercer los derechos de acceso, rectificación, cancelación y oposición ante el mismo es, Vicerrectorado de Títulos Propios, Formación Continua y Posgrado, C/Tulipán s/n, 28933-Móstoles, todo lo cual se informa en cumplimiento del artículo 5 de la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.

## 1. Resumen del proyecto de tesis.

Durante la elaboración de esta tesis doctoral se pretende explorar diversas técnicas de inteligencia artificial que permitirán a un vehículo realizar maniobras de diversa índole de manera satisfactoria, centrándose la investigación en la implementación de algoritmos de reinforcement learning, Deep reinforcement learning y Deep learning.

En una primera instancia se resolverán una serie de problemas básicos en un simulador que nos permitirá la realización de una evaluación objetiva en un entorno más controlado para, posteriormente, ser capaces de aplicar estas técnicas en un problema más complejo que se pueda llevar a un entorno real con un robot físico.

Durante la realización de las pruebas se colaborará con el desarrollo de RL-Studio, lo cual permitirá aplicar distintos algoritmos de inteligencia artificial a robots en un entorno simulado. La versión utilizada se compondrá de:

- Gazebo 11 y Carla 0.9.15 para simular el entorno, los robots y las físicas que aplicarán al problema
- Ros noetic para comunicarse con los robots
- Frameworks de deep learning que nos permitirán implementar los algoritmos con los que se va a trabajar en esta tesis, entre los que destacan Tensorflow, Tensorboard y OpenCV.
- Integración con Behavior metrics para evaluar y comparar el rendimiento de los agentes entrenados

A continuación se mencionan algunas alternativas del problema concreto a abordar:

- Negociación de cruces
- Conducción urbana incluyendo semáforos y peatones
- Ejecución de adelantamientos según las condiciones de circulación
- Conducción multiagente utilizando comunicación entre vehículos
- Configuración dinámica de prioridades en base a un agente preentrenado (combustible, comodidad, velocidad, seguridad, etc.)
- Aumento de la seguridad, explicabilidad y predecibilidad de las decisiones del agente ante situaciones críticas potencialmente peligrosas

Mejorar cualquiera de estos aspectos representaría un avance significativo en el estado actual de la conducción autónoma, ya que se abordan algunos de los desafíos más relevantes y urgentes en este campo.

## 2. Justificación del tema de investigación. Antecedentes y estado actual del tema.

La aplicación de inteligencia artificial (IA) para abordar desafíos concretos ha experimentado un crecimiento significativo tanto en el sector privado como en el público, abarcando una amplia gama de industrias, como la medicina, las finanzas, la industria y la automoción. Esta tendencia se debe a la capacidad demostrada de la IA para mejorar la calidad de los productos y servicios, al tiempo que reduce los costos asociados, generando beneficios tanto para las empresas como para la calidad de vida de las personas.

Dentro de este contexto, destaca el aprendizaje por refuerzo, un enfoque que permite a las máquinas aprender de manera similar a los seres humanos, explorando su entorno y obteniendo resultados óptimos. Esto se aplica también a los robots, que pueden aprender mediante ensayo y error, desarrollando políticas óptimas para lograr objetivos específicos. Este avance posibilita la automatización de tareas que anteriormente requerían un esfuerzo significativo, mejorando la eficiencia y precisión en su ejecución.

Una de estas tareas en las que la comunidad lleva tiempo aunando esfuerzos es la conducción autónoma, ámbito en el cual se desarrollará la tesis.

La conducción autónoma es un campo de investigación en constante evolución que promete revolucionar la industria automovilística y mejorar la movilidad en todo el mundo.

Esto es así dado su gran potencial para abordar desafíos significativos y beneficiar a la sociedad en términos de seguridad vial, eficiencia energética y comodidad.

A continuación se mencionan algunos de los avances más significativos en los últimos años

- **Algoritmos de Aprendizaje por Refuerzo Profundo (DRL):**

Los algoritmos DRL, como DQN, DDPG [1] y PPO [2], han logrado avances significativos al utilizar redes neuronales profundas para aprender políticas de conducción autónoma interactuando con un entorno dinámico en tiempo real. Estos algoritmos pueden gestionar la toma de decisiones complejas, como la aceleración, el frenado y la dirección.

- **Negociación de Cruce:**

Las intersecciones se consideran uno de los escenarios más complejos en un marco de conducción autónoma debido a la incertidumbre en los comportamientos de los vehículos circundantes y los diferentes tipos de situaciones que se pueden encontrar.

las estadísticas muestran que el 60% de las lesiones graves relacionadas con el tráfico en Europa están vinculadas a las intersecciones [3]. En los EE. UU., la Administración Nacional de Seguridad del Tráfico en Carreteras (NHTSA) publica un informe anual sobre seguridad vial [4]. Estos datos revelan que el 29% de todos los accidentes automovilísticos y el 18% de las muertes de peatones ocurren en las intersecciones.

- **Resolución de Circuitos Predefinidos con Tráfico y Peditones:**

Los vehículos autónomos deben reconocer y seguir trayectorias específicas en circuitos complejos que incluyen tráfico dinámico y peatones. Esto requiere un control preciso de la velocidad y la dirección. Dada la importancia de proteger al peatón, que resulta el actor más vulnerable en cualquier escenario, podemos encontrar multitud de investigaciones al respecto [5][6].

- **Transferencia de Conocimiento de Agentes Simulados a Vehículos en el Mundo Real:**

Este enfoque implica entrenar modelos de conducción autónoma en simulaciones computacionales realistas y luego transferir este conocimiento a vehículos en el mundo real, lo cual resulta de vital importancia en el momento de sacar provecho de los prometedores resultados obtenidos en simuladores como Gazebo, Carla o Sumo. Esto se logra a través de técnicas de adaptación de dominio y validación cruzada [5][6].

A continuación se enumeran una serie de desafíos que aún podrían ser abordados:

- **Maximizar el nivel de seguridad en la conducción autónoma:**

Uno de los problemas principales que aún no se han resuelto por completo es lograr un alto nivel de seguridad y confiabilidad. Si bien los vehículos autónomos han mostrado resultados prometedores en entornos controlados, navegar por escenarios del mundo real complejos sigue siendo un desafío considerable. El comportamiento humano impredecible, las condiciones climáticas adversas y las carreteras mal señalizadas pueden representar obstáculos significativos para los vehículos autónomos hasta el punto de que empresas como Waymo, Tesla y Cruise aún se encuentran en fase de desarrollo para poder garantizar una solución segura. Lo que ha llevado al lanzamiento de algunos modelos con ciertos niveles de automatización.

Dicho esto, el daño humano, el daño material y la mala reputación que implican los accidentes [7] de estos vehículos autónomos hace de vital importancia centrar esfuerzos en este punto.

- **Cubrir casos de uso adversos mejorando la percepción:**

Otro desafío importante es garantizar una percepción y comprensión robustas del entorno circundante. Los vehículos autónomos se basan en una combinación de sensores, como cámaras, lidar, radar y GPS, para percibir e interpretar de manera precisa su entorno. Sin embargo, estos sensores pueden verse afectados por condiciones climáticas adversas, como lluvia intensa o niebla, lo que provoca un rendimiento degradado. Para abordar esto, los investigadores están explorando técnicas de fusión de sensores que integran datos de múltiples sensores para mejorar las capacidades de percepción en diversas condiciones.

- **Optimizar la conducción:**

Otra línea de investigación en ciernes es la posibilidad de intercomunicar los diferentes vehículos para, prediciendo la secuencia de acciones que el resto de agentes puede tomar, reducir atascos y evitar ineficiencias en la conducción. A esto se podrían añadir distintos inputs como el estado de la vía o el consumo de combustible para también contribuir a una reducción de accidentes y contaminación.

### 3. Hipótesis de trabajo y Objetivos

#### Hipótesis de Trabajo:

- Dando al robot las reglas que rigen el entorno simulado y proporcionándole una serie de estados y acciones, el robot será capaz de aprender y ejecutar acciones óptimas basadas en los estados captados a través de sensores de distinta naturaleza.
- Es posible transferir el conocimiento adquirido por el robot en un simulador realista a un entorno real, lo que implica que las habilidades y comportamientos aprendidos en el simulador serán aplicables y efectivos en situaciones reales.
- La aplicación de aprendizaje por refuerzo resultará en una mejora significativa en la eficiencia, seguridad o calidad de la conducción.
- El agente transferido al vehículo real será capaz de generalizar, adaptándose a diferentes entornos que no son exactamente iguales a los utilizados para entrenar en el simulador.

#### Objetivos:

1. Abordar y resolver problemas específicos de conducción autónoma (sigue carril y autoparking) utilizando en un vehículo simulado técnicas de aprendizaje por refuerzo profundo (PPO, DDPG, DQN, A3C...).
2. Contribuir al desarrollo de una herramienta (RL-Studio) que permita la adaptación sencilla de algoritmos de aprendizaje por refuerzo a diferentes problemas. Esto incluye la configuración de sensores, acciones, objetivos y la naturaleza del problema, de manera que los usuarios futuros puedan resolver problemas de manera rápida y efectiva sin necesidad de implementar algoritmos desde cero.
3. Abordar un problema práctico que aún no ha sido completamente resuelto por la comunidad investigadora o que tiene margen de mejora significativo, lo que implica enfrentarse a desafíos reales y aplicar soluciones innovadoras. A continuación se mencionan algunas posibilidades:
  - Negociación de cruces
  - Conducción urbana incluyendo semáforos y peatones
  - Ejecución de adelantamientos según las condiciones de circulación
  - Conducción multiagente utilizando comunicación entre vehículos
  - Configuración dinámica de prioridades en base a un agente preentrenado (combustible

- Aumento de la seguridad, explicabilidad y predecibilidad de las decisiones del agente ante situaciones críticas potencialmente peligrosas

4. Contribuir al avance teórico en el campo del aprendizaje por refuerzo a través de artículos propios.

5. Contribuir al avance práctico en el campo del aprendizaje por refuerzo demostrando la efectividad de las soluciones desarrolladas transfiriendo el modelo al entorno real.

#### 4. Metodología

La metodología que se emplea a lo largo de esta investigación se basa en la realización de experimentos que arrojen resultados objetivos evaluando las diferentes propuestas utilizadas.

Las métricas utilizadas diferirán en función de la naturaleza del problema a resolver.

En principio, se prevé la utilización en mayor medida de métricas relacionadas con el aprendizaje profundo por refuerzo, pero es posible que exista la necesidad de utilizar otras métricas que permitan la resolución de problemas intermedios e incluso un análisis más minucioso de los resultados obtenidos como velocidad media del vehículo o distancia recorrida.

Entre las posibles métricas de rendimiento de los algoritmos que no utilicen aprendizaje por refuerzo y que serán usados para complementar nuestro sistema podemos encontrar:

- Métricas de clasificación (exactitud, precisión, recuperación, puntuación F1, ROC, AUC,...)
- Métricas de regresión (MSE, MAE)
- Métricas de clasificación (MRR, DCG, NDCG)
- Métricas estadísticas (correlación)
- Métricas de visión por computadora (PSNR, SSIM, IoU)

Entre las posibles métricas de rendimiento específicas de la resolución del problema mediante aprendizaje por refuerzo, serán utilizadas:

##### 1. Métricas intrínsecas:

- Recompensa acumulada media durante un episodio
- Sumatorio de la recompensa descontada media durante un episodio
- Tiempo que tarda el algoritmo en converger a una solución óptima durante el entrenamiento
- Estabilidad y fiabilidad del entrenamiento en función del algoritmo y técnicas utilizadas
- Capacidad del agente a generalizar (comportarse bien ante situaciones o escenarios no observados durante el entrenamiento)
- Porcentaje de episodios completados de manera satisfactoria
- Tiempo medio por episodio en alcanzar el objetivo

##### 2. Métricas extrínsecas específicas (herramienta Behavior metrics)

- Solidez del agente ante perturbaciones externas (lluvia, imprevistos, sensores averiados, etc.)
- Distancia recorrida en un circuito concreto
- Velocidad media del vehículo
- Número y gravedad de infracciones cometidas
- Frecuencia de iteración o frames por segundo



Una vez alcanzado un agente óptimo que resuelva el problema en cuestión en un simulador lo más fiel a la realidad posible, se abordará el reto de transferir el conocimiento adquirido a un entorno real con robots que compartan sensores (inputs) y actuadores (outputs) con el entorno simulado durante el entrenamiento.

5. Revistas científicas en las que tiene previsto de forma tentativa publicar los resultados de la tesis doctoral de acuerdo con lo establecido en el artículo 23.3 de la normativa reguladora de los estudios de doctorado de la URJC.

El tipo de investigación desarrollada, teniendo en cuenta su temática, podría publicarse de forma tentativa en las siguientes revistas:

- Neural Computing and Applications, Springer
- Image and Vision Computing, Elsevier
- Journal of Artificial Intelligence Research
- IEEE Transactions on Robotics
- Journal of Machine Learning Research (JMLR)
- Frontiers in Robotics and AI
- Journal of Intelligent & Robotic Systems

También podría ser expuesta en conferencias como las mencionadas a continuación:

- ICRA 2023, IEEE International Conference on Robotics and Automation
- ROBOT-202X
- European Conference on Connected and Automated Driving

6. Plan de trabajo. Describa brevemente las tareas previstas y los resultados esperados en cada bloqu temporal.

### 1º Año:

1. Estudio de estado de la cuestión para el problema del comportamiento automático de robots.
  2. Estudio en profundidad de técnicas de deep learning y aprendizaje por refuerzo actuales para tener una base clara sobre la que desarrollar la investigación. Dentro de este estudio incluiremos la realización del curso “machine learning” impartido por Andrew NG en coursera.
  3. Implementación y resolución de un problema básico de resolución de un laberinto primero en un simulador propio y luego en Gazebo utilizando técnicas de aprendizaje por refuerzo clásico como Qlearning y Sarsa.
  4. Primeros experimentos en Gazebo y RLStudio con circuitos con restricciones y soluciones basadas en reinforcement learning clásico para un robot terrestre móvil.
- 

### 2º Año:

1. Desarrollo de una herramienta (RL-Studio) que permita la aplicación de algoritmos de reinforcement learning clásico y Deep learning para resolver problemas básicos con los que se pretende recabar técnicas que doten al doctorando de las herramientas necesarias para abordar problemas complejos.
  2. Además, este desarrollo dotará a RL-Studio de la potencia necesaria para facilitar la implementación de problemas más complejos que se abordarán en los próximos años. El plan de iteración es el siguiente:
    - Generalización de componentes para facilitar la integración de distintos problemas, algoritmos y simuladores de manera rápida y homogénea.
    - Implementación de un modo inferencia para ejecutar los algoritmos entrenados sin la sobrecarga relativa a la fase de entrenamiento.
    - Entrenamiento y ejecución de agentes en problemas canónicos.
  3. Realización del curso “Algorithms, Part 1” impartido por la universidad de Princeton en coursera.
- 

### 3º Año:

1. Extensión de RL-Studio para probar con distintos simuladores (Gazebo, openAI y CARLA) y algoritmos (entre ellos DQN, DDPG, PPO y A3C).
2. Resolución de problemas básicos propuestos por openAI (mountain car, cartpole y pendulum)
3. Extensión de problemas básicos propuestos por openAI (cartpole y mountain car) para poder extraer de ellos comparativas y lecciones sólidas acerca de las diferentes técnicas y algoritmos a

aplicar en problemas más complejos. Entre la extensión de estos problemas se incluirá:

- la posibilidad de moldear los entornos ofrecidos por openAI con el objetivo de complicar el problema básico (e.g configuración el mundo simulado a resolver, mejora de las físicas aplicadas, inclusion un ciclo de reloj en el simulador de openAI que obligue al algoritmo a decidir la acción con rapidez, modificación de la naturaleza del problema de acciones discretas acciones continuas)
- Trasladar los problemas a un simulador más realista con físicas más sofisticadas (Gazebo)
- Adición de perturbaciones y condiciones adversas

4. Desarrollo de soluciones a problemas más complejos combinando deep learning y reinforcement learning (entre ellos vehículo con autoaparcamiento, sigue-lineas y sigue-carril)

5. Formalización de los problemas resueltos y los bancos de pruebas realizados con objetivo de servir de base para la resolución de futuros problemas y para ilustrar de manera clara el trabajo realizado de manera que se pueda utilizar con objetivos docentes y en publicaciones de diversa índole. Dichos bancos de pruebas consistirán en:

- Aplicación en RL-Studio de distintos algoritmos (solución programática, QLearning, DQN, DDPG, PPO con acciones discretas y PPO con acciones continuas) a la extensión implementada para el problema cartpole (acciones continuas, físicas mejoradas, aplicación de perturbaciones y condiciones iniciales adversas).
- Aplicación en RL-Studio de distintos algoritmos al problema sigue lineas y sigue carril con el objetivo de comparar su rendimiento haciendo uso de las métricas externas evaluadas en Behavior metrics (velocidad efectiva, desviación con respecto al centro y avance espacial)

6. Realización del curso “Deep reinforcement learning class” impartido por la la compañía Hugging Face.

7. Colaboración en la redacción del artículo [26] en el que se describe RL-Studio

---

#### 4º Año:

1. Refinamiento de los agentes entrenados en años anteriores con objetivo de tener una base lo más sólida posible sobre la que pasar a un problema más complejo en un simulador más realista
2. Aplicación de lo previamente aprendido en un simulador con físicas lo más parecido posible al entorno real (CARLA) con el objetivo de facilitar la transferencia de conocimiento al mundo real
3. Resolución del problema sigue-carril con DDPG y PPO con acciones continuas maximizando la suavidad y velocidad de la conducción en un escenario con líneas discontinuas y curvas de distinta índole en ambas direcciones
4. Extensión de RL-Studio para automatizar el análisis de distintas técnicas de percepción y su impacto en el entrenamiento y el comportamiento del agente
  - Módulos de percepción basados en redes neuronales

- MobileV3Small
- Yolop
- Módulos basados en técnicas clásicas de percepción
  - Thresholding
  - Linear regresion

5. Realización del curso “Deep Learning Specialization” impartido por Andrew NG en Coursera

6. Redacción de un artículo en el ámbito de la conducción autónoma con la comparativa de los distintos sistemas de percepción y de decisión abordados durante este año

---

#### **5º Año:**

1. Definición del problema a abordar en la tesis
  2. Primeros pasos en dirección al problema definitivo a abordar
  3. Presentación de al menos un artículo en el ámbito de la conducción autónoma.
- 

#### **6º Año:**

1. Resolución definitiva del problema definido tras haber experimentado durante el año anterior con distintos entornos, técnicas y algoritmos
  2. Experimentar sobre la viabilidad de trasladar la política aprendida a un robot móvil en el mundo real.
  3. Elaboración de la memoria de la tesis.
-

## 7. Bibliografía relevante relacionada con la investigación

Algunas de las referencias más relevantes relacionadas con las investigación incluyen los siguientes trabajos:

1. Chang, C.-C., Tsai, J., Lin, J.-H., & Ooi, Y.-M. (Título: Autonomous Driving Control Using the DDPG and RDPG Algorithms)
2. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (Título: Proximal Policy Optimization Algorithms)
3. Werneke, J., & Vollrath, M. (Título: How do environmental characteristics at intersections change in their relevance for drivers before entering an intersection: Analysis of drivers' gaze and driving behavior in a driving simulator study. Publicado en Cogn. Technol. en 2014, 16, 157-169.)
4. NHTSA (Título: Traffic Safety Facts 2019; National Highway Traffic Safety Administration: Washington, DC, USA, 2019.)
5. Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (Título: Pedestrian Trajectory Prediction with Structured Memory Hierarchies)
6. Zavodjančík, T., Kasanický, G., & Demčáková, L. (Título: Reduction of Pedestrian Accidents - Automated Road Vehicles)
7. <https://www.nytimes.com/2023/01/31/technology/tesla-autopilot-investigation.html>
8. Chiba, S., & Sasaoka, H. (Título: Effectiveness of Transfer Learning in Autonomous Driving using Model Car. Publicado en Proceedings of the 2021 13th International Conference on Machine Learning and Computing (ICMLC '21). ACM)
9. Akhauri, S., Zheng, L., & Lin, M. (Título: Enhanced Transfer Learning for Autonomous Driving with Systematic Accident Simulation)
10. Russell, S., & Norvig, P. (Título: Artificial Intelligence: A Modern Approach)
11. Goodfellow, I., Bengio, Y., & Courville, A. (Título: Deep Learning)
12. Sutton, R. S., & Barto, A. G. (Título: Reinforcement Learning: An Introduction)
13. Bojarsky, M., et al. (Título: End to end Learning for Self-Driving Cars)
14. LeCun, Y., et al. (Título: Gradient-based learning applied to document recognition)
15. Hochreiter, S., et al. (Título: Long Short-Term Memory)
16. Rojas-Perez, L. O., et al. (Título: DeepPilot: A CNN for Autonomous Drone Racing)
17. Toromanoff, M., Wirbel, E., & Moutarde, F. (Título: End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances)
18. Schulman, J., Moritz, P., Levine, S., Jordan, M. I., & Abbeel, P. (Título: HIGH-DIMENSIONAL CONTINUOUS CONTROL USING GENERALIZED ADVANTAGE ESTIMATION)
19. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & OpenAI (Título: Proximal Policy

## Optimization Algorithms)

20. Schneider, A., Wickert, C., & Marti, E. (Título: Reducing Complexity by Creating Complexity: A Systems Theory Perspective on How Organizations Respond to Their Environments)
21. Rolim, L. A., et al. (Título: Enhanced Reader.pdf)
22. Knox, W. B., et al. (Título: Reward (Mis)design for Autonomous Driving)
23. Gutiérrez-Moreno, R., et al. (Título: Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator)
24. Engstrom, L., et al. (Título: IMPLEMENTATION MATTERS IN DEEP POLICY GRADIENTS: A CASE STUDY ON PPO AND TRPO)
25. Kingma, D. P., & Ba, J. L. (Título: Adam: A Method for Stochastic Optimization)
26. Paniego, S., Lucas, R., Fernández, P., Cañas, J. M., & Arranz, I. (Título: RL-Studio: A tool for reinforcement learning methods in robotic)

Fecha y Firma del  
doctorando:

Rubén