

Nombre del tutor del proyecto: *Prof. Juan Antonio Rodríguez-Aguilar*

Datos de contacto del tutor del proyecto: jar@iia.csic.es, juanantonio.rodriguez@csic.es

Nombre del Grupo de investigación asociado a AIHUB: Instituto de Investigación en Inteligencia Artificial

Ubicación del centro donde se disfrutará la beca: Campus de la Universitat Autònoma de Barcelona, Bellaterra, Barcelona

Título del proyecto: Multi-objective reinforcement learning for ethical autonomous driving

Descripción del proyecto: The purpose of this project is to investigate how to exploit multi-objective reinforcement learning to help an agent learn ethical behaviours in a driving environment. On the one hand, we will investigate how to extend a state-of-the-art algorithm to create driving environments where the learning of ethical behaviours is guaranteed while considering a user's preferences over different ethical objectives. We will empirically study the learning of ethical policies in a simulated driving environment that will be developed as part of the project. As an important aspect of this empirical evaluation, we will investigate whether it is enough for a learning agent to employ low-cost learning algorithms (e.g. Q-learning), or whether instead it is more convenient to resort to Deep Reinforcement Learning. The project will combine recent work by IIA on multi-objective reinforcement learning for ethical learning with quality guarantees with recent work by ICMAT on decision making models for automated driving systems that consider ethical and operational preferences.