Università Ca'Foscari Venezia

PROJECT ACRONYM AND TITLE : **REXlearn - Reliable and Explainable Adversarial Machine Learning**

FUNDING PROGRAMME: **PRIN 2017**

SCIENTIFIC FIELD: **PE6**

HOST DEPARTMENT: **European Centre for Living Technologies**

SCIENTIFIC RESPONSIBLE: **Marcello Pelillo**

FINANCIAL DATA:

Project total costs	Overall funding assigned to UNIVE
764.640 €	179.483 €

ABSTRACT:
Machine-learning technologies have become pervasive, and even able to outperform humans on specific tasks. However, it has been shown that they suffer from hallucinations known as adversarial examples, i.e., imperceptible, adversarial perturbations to images, text and audio that fool these systems into perceiving things that are not there. This has severely questioned their suitability for mission-critical applications, including self-driving cars and autonomous vehicles. The defense strategies proposed to overcome this issue have been shown to be ineffective against more sophisticated attacks carefully crafted to bypass them, highlighting the challenging nature of this problem. In this project, we formulate three main challenges that demand for novel learning paradigms, able to take reliable and explainable decisions, to assess and mitigate the security risks associated to such potential misuses of machine learning. This project will pave the way towards the design of reliable and explainable machines that are also useful beyond adversarial settings. We will indeed develop tools and prototypes that can face the challenges posed not only by cybersecurity applications with a clear adversarial nature, but also by recent computer-vision and deep-learning technologies.

Start date	End date
29/08/2019	29/08/2022

PARTNERSHIP:

1. **Università degli Studi di Cagliari**	**Coordinator**
2. **Università Ca' Foscari Venezia**	
3. **Università degli Studi di Firenze**	
4. **Università degli Studi di Siena**	