

Hands-on in XAI

4th Iberian Gravitational Waves Meeting (IGWM2025)

26 June 2025

Miguel Cárdenas-Montes

CIEMAT

Deep Learning and XAI

- **Deep neural networks** (DNNs) rely on multiple processing layers to learn intricate patterns.
 - While this capability allows models to make highly **accurate decisions**, it often **lacks transparency**, making it difficult to trace their reasoning.
 - The **large number of parameters and nonlinear interactions** between neurons contribute to the so-called '**black box**' problem, fostering **skepticism in critical applications**.
 - The success of these models has led to significant challenges related to **explainability**, prompting the scientific community to develop tools that address their inherent opacity.
- Explainable artificial intelligence (**XAI**) has emerged as a key area of research to tackle the interpretability of neural networks.
 - XAI techniques can be broadly categorized into **post-hoc** and ante-hoc methods.
 - Ante-hoc requires of special architectures/layers impacting in its performance. **Avoid them!**

Contents of XAI session at IGWM 2025

1. **Grad-CAM Algorithm — Rationale, Strengths, and Weaknesses**

This module demonstrates the Grad-CAM algorithm applied to a classification task, using a reduced version of the *Gravity Spy* dataset.

2. **LIME Algorithm — Rationale, Strengths, and Weaknesses**

This module illustrates the use of the LIME algorithm in a classification task, also based on a reduced version of the *Gravity Spy* dataset.

3. **SHAP Algorithm — Rationale, Strengths, and Weaknesses**

This module presents the *SHAP* algorithm applied to a classification task, using a subset of galaxies from the COSMOS2015 catalogue.

- 1 hour per block, with short breaks in between
- All codes are fully executable – no placeholders or empty sections.
- Practice-oriented approach.
- Focus on XAI, not performance in NN.

Datasets

The selected datasets were chosen based on several key criteria: their small size, facilitating hands-on demonstrations; the interpretability of their outputs; and their generality, which enhances the transferability of insights to other domains.

<https://agenda.ciemat.es/event/5116/>

Routine

1. Go to the provided link
2. For each block
 - a. Download the presentation and the code
 - b. Upload the code to your Google Drive
 - c. Open and run it in Google Colaboratory (Colab)
3. All code is fully operational
4. Try to run it using a T4 GPU in Colab.



Hands-on course in Explainable Artificial Intelligence

We are pleased to announce that a *Hands-on Course in Explainable Artificial Intelligence (XAI)* will take place in the context of the [14th Iberian Gravitational Waves Meeting \(IGWM2025\)](#), hosted at [CIEMAT](#) (Madrid). Participation in IGWM2025 is not required to sign up, and it is open exclusively to members of [REFORMA](#) and [IAGUGG](#).

