Contribution ID: **3** Type: **Oral (Regular)**

When Explainable AI is not enough: Informed Machine Learning to Combine Fidelity and Interpretability

Tuesday 9 September 2025 10:00 (25 minutes)

Recently the huge amounts of data collected in modern large experiments have motivated the deployment of machine learning (ML) in physics. Unfortunately, ML models are typically black boxes almost impossible to interpret. Consequently, in the last years eXplainable Artificial Intelligence, whose objective consists of improving the transparency of ML tools, has received a lot of attention. However in science, explainability assumes a flavour different from that in commercial applications and cannot be reduced to pure user convenience. Indeed the priority is also fidelity, which means developing models that reflect the actual mechanisms at play in the investigated phenomena. Genetic Programming supported Symbolic Regression (GPSR) has some obvious competitive advantages in trying to find an optimal trade-off between interpretability and realism. Unfortunately, the search spaces are typically too large; therefore the algorithms have to be steered to converge on meaningful solutions. The present work describes techniques to constrain GPSR and to combine it with deep leaning tools, so that the final models are expressed in terms of interpretable mathematical equations but at the same time reflect the physics at play in the phenomena under study. The strategies to combine fidelity and explainability include dimensional analysis, integration of prior information about symmetries and conservation laws and improvements of the fitness function. Great attention has been paid to devising practical solutions and they cover all the essential aspects of the data analysis process, from the treatment of the uncertainties to the quantification of the model complexity. All the main applications of ML, from regression to density estimation and classification, benefit from the proposed improvements. Theoretical considerations, systematic numerical tests, simulations with multiphysics codes and the results of actual experiments prove the potential of the devised upgrades.

Speaker's email address

andrea.murari@istp.cnr.it

Speaker's Affiliation

ENEA Consorzio RFX

Member State or International Organizations

Italy

Author: MURARI, Andrea (Consorzio RFX)

Presenter: MURARI, Andrea (Consorzio RFX)

Session Classification: Physics-Based Machine Learning

Track Classification: Physics-Based Machine Learning