ID de Contribution: **5**   Type: **Non spécifié**

Advancing Explainable AI: Testing and Enhancing Techniques Across Multidisciplinary Use-Cases

mardi 28 novembre 2023 10:35 (25 minutes)

Developing and testing methodologies for enhancing the transparency, interpretability, and explainability of AI algorithms is a pressing challenge for the application of artificial intelligence methods in fundamental physics. The Multi-disciplinary Use Cases for Convergent new Approaches to AI explainability (MUCCA) project is an innovative project that aims to address this challenge by bringing together researchers from diverse fields, each contributing complementary skills essential for comprehending AI algorithm behavior. The project centers around the investigation of a wide array of multidisciplinary use-cases, where explainable AI can play a pivotal role. In our presentation we illustrate the MUCCA project in general, to then verticalize with respect to our ongoing research in the field of high energy physics. We showcase its application in both high-energy physics experiment data analysis and its use in detector and real-time systems. Specifically, we delve into our exploration of various explainability methods rooted in different approaches and evaluate their effectiveness across the diverse use-cases. The outcome of our work yields a collection of potentially comprehensible and human-friendly explanations for the predictions made by our models. We conclude by highlighting limitations of existing xAI models for high-energy physics, and brainstorming ideas on how to build novel, explainable-by-design models for accelerating scientific research with AI.

Auteurs principaux: CIARDIELLO, Andrea (Sapienza Università di Roma, Roma, Italy); VOENA, Cecilia (Sapienza Università di Roma, Roma, Italy); BALAN, Corneliu (Polytechnic Institute Bucharest Romania); SEBASTIANI, Cristiano (University of Liverpool); CARMIGNANI, Joseph (University of Liverpool); DIMITROVA, Kalina (Sofia University "St. Kliment Ohridski"); MATTIA, Maurizio (Istituto Superiore di Sanità - Roma, Italy); D'ONOFRIO, Monica (University of Liverpool); MELCHIONNA, Simone (MedLea s.r.l.s.); SCARDAPANE, Simone (Sapienza University of Rome, Roma, Italy); GIAGU, Stefano (Sapienza Università di Roma, INFN Roma, Roma Italy); KOZHUHAROV, Venelin (Sofia University "St. Kl. Ohridski")

Orateur: SCARDAPANE, Simone (Sapienza University of Rome, Roma, Italy)

Classification de Session: Explainable AI

Classification de thématique: Explainable AI, Trustworthy AI