

Special Session Proposal

Title:

Explainable Artificial Intelligence (XAI) for Reliability, Availability, Maintainability and Safety (RAMS) Applications

Organizer(s):

- Piero Baraldi (Politecnico di Milano, Italy)
- Joaquín Figueroa (Politecnico di Milano, Italy)
- Enrico Zio (MINES Paris-PSL, France and Politecnico di Milano, Italy)

Description:

The effective use of Artificial Intelligence (AI) models within frameworks for the assessment of system Reliability, Availability, Maintainability and Safety (RAMS) remains limited. One of the reasons is the "black-box" nature of AI models. This is particularly detrimental in RAMS applications, as the lack of transparency reduces trust and hinders accountability. EXplainable AI (XAI) addresses this issue by making AI models more transparent and interpretable.

This special session is intended to present contributions to the development of XAI for RAMS applications, in an effort to gather the expert community and share knowledge for future advancements and developments in the field.

Main topics are (but not limited to):

Methods:

- Ante-hoc methods (e.g. rule-based systems, decision trees and other inherently explainable models)
- Post-hoc methods (e.g. LIME and SHAP)
- Model-specific methods (e.g. DeepLIFT and integrated gradients)
- Visual explanations (e.g. feature visualization and saliency maps)
- Counterfactual explanations (e.g. WachterCF and DiCE)

Applications:

- Anomaly detection
- Fault diagnostics and prognostics
- Predictive maintenance
- Safety analysis
- Accident anticipation
- Risk assessment and management
- Human-AI collaboration