CALL FOR PAPERS
SPECIAL SESSION ON

Explainable and Trustworthy AI for Safety-Critical Industrial Systems: Verification, Interpretability, and Adversarial Robustness

for ICCAD 2026
July 7-9, 2026, Lisbon, Portugal

Session Co-Chairs:
- Dr. Ali Teta, Applied Automation and Industrial Diagnostics Laboratory LAADI, Djelfa University, Algeria. ali.teta@univ-djelfa.dz
- Mr. Ahmed Chennana, Dept. Electrical Engineering, Laboratory of LI3C, University of Mohamed Khider, Biskra, Algeria. ahmed.chennana@univ-biskra.dz
- Prof. Lotfi Saidi, SIME Laboratory (LR 13ES03), University of Sousse, Higher School of Sciences and Technology of Sousse, Sousse, Tunisia; Lotfi.Saidi@esstt.rnu.tn

Session description:
This special session deals with the problem of deploying artificial intelligence and machine learning models in safety-critical control, automation, and diagnostic systems where transparency, interpretability, and reliability are paramount.

As AI-driven systems increasingly make autonomous decisions in high-stakes domains such as healthcare, transportation, industrial processes, and energy infrastructure, there is a critical need to understand, verify, and trust these systems' decision-making processes. The lack of explainability and formal guarantees poses significant barriers to regulatory approval, user acceptance, and safe deployment in environments where failures can have catastrophic consequences.

The goal of this session is to advance the development of explainable, verifiable, and trustworthy AI methodologies that can provide transparent reasoning, quantifiable confidence measures, and formal safety guarantees for control and diagnostic systems.
This session brings together researchers and practitioners to share innovative approaches for creating interpretable models, validating AI behavior under uncertainty, defending against adversarial threats, and establishing certification frameworks that bridge the gap between cutting-edge AI capabilities and the stringent requirements of safety-critical applications.

The topics of interest include, but are not limited to:

- Interpretable and explainable AI models for fault diagnosis, anomaly detection, predictive maintenance, and process monitoring in industrial systems
- Formal verification, validation, and certification of learning-enabled control and automation systems for safety-critical applications

- Uncertainty quantification, confidence estimation, and probabilistic guarantees for AI-driven industrial decision-making and control
- Adversarial robustness and defense mechanisms: detection, prevention, and mitigation of adversarial attacks on industrial AI systems
- Causal inference and counterfactual reasoning for root cause analysis, transparent decision support, and explainable diagnostics
- Security testing and cyber-physical protection for AI-enabled industrial IoT, edge computing, and critical infrastructure systems
- Runtime verification, safety monitoring, and fail-safe mechanisms for deployed AI systems in industrial environments
- Neuro-symbolic AI and transparent neural architectures for human-interpretable reasoning and knowledge integration
- Human-in-the-loop frameworks and visualization techniques for validation, oversight, and collaborative AI-assisted decision-making
- Trustworthy AI lifecycle management: benchmarking, standardization, bias detection, fairness assessment, and regulatory compliance for industrial AI deployment

---

SUBMISSION

Papers must be submitted electronically for peer review by**: January 31, 2026**
https://www.iccad-conf.com/submission/

All papers must be written in English and should describe original work. The length of the paper is limited to a maximum of 6 pages (in the standard IEEE conference double column format).