

Bridging the Gap: Exploring Explainability in Autonomous and Trusted Computing

As technology continues to advance, autonomous and trusted computing systems are becoming increasingly prevalent in various domains, including healthcare, finance, transportation, and more. However, with this advancement comes the challenge of ensuring that these systems are not only efficient and reliable but also transparent and understandable to users and stakeholders. Explainability in autonomous and trusted computing has emerged as a crucial area of research and development to address this challenge. This workshop aims to explore the significance, methods, and implications of explainability in autonomous and trusted computing systems. Through interactive discussions, presentations, and hands-on activities, participants will share their insights regarding the importance of explainability, current research trends, and practical approaches to enhance transparency and trust in autonomous systems.

Workshop Objectives:

- To understand the concept of explainability in the context of autonomous and trusted computing.
- To explore the importance of explainability for users, developers, and other stakeholders. To discuss current research and development efforts in the field of explainable autonomous systems.
- To identify challenges and opportunities in designing and implementing explainable autonomous and trusted computing systems.
- To facilitate knowledge sharing and collaboration among researchers, practitioners, and industry professionals in the field of autonomous systems.

Workshop Topics:

The workshop will cover, but is not limited to, the following topics:

- Understanding Explainability: Concepts and Definitions
- The Importance of Explainability in Autonomous Systems
- Ethical Considerations in Explainable AI
- Explainable Machine Learning Models for Autonomous Systems
- Human-Centered Design Approaches to Explainability
- Transparency vs. Accuracy: Balancing Trade-offs in Autonomous Computing
- Case Studies: Real-World Applications of Explainable Autonomous Systems
- Explainability Techniques: From Post-hoc Interpretation to Model Design
- Visualizing and Communicating Decisions taken by Autonomous Systems
- Legal and Regulatory Perspectives on Explainability
- User Trust and Acceptance in Explainable Autonomous Systems
- Explainability in Multi-Agent Systems and Collaborative Robotics
- Addressing Bias and Fairness in Explainable AI
- Evaluating the Effectiveness of Explainability Methods
- Security Implications of Explainable Autonomous Systems
- Educational Initiatives for Teaching Explainable AI Concepts
- Future Directions in Explainable Autonomous and Trusted Computing

Workshop Organizers:

- Dr. Hanwei Zhang, a Postdoc Researcher at Saarland University, working on the project of Explainable Intelligent Systems, as well as a project about the adversarial attack against 3D object detection in autonomous driving.
- Prof. Holger Hermanns, Professor of Computer Science at Saarland University in Germany, heading the Dependable Systems and Software group, member of Academia Europaea, and spokesperson of the Center for Perspicuous Computing, SFB TRR 248. The European Research Council has awarded him the ERC Advanced Grant POWVER and the ERC Proof-of-Concept Grant LEOpowver.

Important Dates:

- Workshop Paper Submission Deadline: September 1, 2024
- Notification of Acceptance: October 15, 2024

Papers are to be submitted as PDF via the site: <https://edas.info/N32633>.
Please select the corresponding workshop when submitting your paper.

We look forward to your participation and contributions to the workshop “Bridging the Gap: Exploring Explainability in Autonomous and Trusted Computing”

This workshop is supported by Grant AZ 98514 — [EIS](#) — of VolkswagenStiftung.

