

Call for Papers

ETRI Journal Special Issue on **Trustworthy and Safe AI: Interpretability, Robustness, and Security**

Recent advances in artificial intelligence, particularly in large language models (LLMs) and physical robotics, are rapidly reshaping various aspects of our daily lives and industrial applications. While these technologies hold immense potential, they also introduce critical challenges in ensuring that AI systems are safe, trustworthy, and aligned with human values. Addressing these challenges requires not only mitigating harmful or unintended behaviors but also promoting robustness and transparency in increasingly autonomous and complex AI systems.

This special issue aims to bring together high-quality, original research that advances both the theoretical foundations and practical approaches to AI Safety and Trustworthy AI. We invite submissions on, but not limited to, the following topics:

- Robustness and Resilience of AI Algorithms and Systems
- Explainable AI (XAI) and Interpretable AI for AI Safety
- Alignment Algorithms including Reinforcement Learning from Human Feedback (RLHF), Cooperative Inverse Reinforcement Learning
- Human-AI interaction, Collaboration, Oversight
- Bias Detection, Measurement, and Mitigation Techniques
- Deepfake Techniques for Image, Audio, and Video Generation and Detection
- Adversarial Attacks and Corresponding Defenses for AI Models
- Jailbreaking and Defense Strategies for LLMs and VLMs
- Simulation Environments and Scenario-based Testing for Agentic AI
- Efficient and Scalable Techniques for Real-World Safety Evaluation
- Comparative and Empirical Perspectives on AI Governance and Regulation in Practice

Important Dates (tentative)

Paper submission due: March 10, 2026

First decision: May 20, 2026

Final decision: July 20, 2026

Publication: October 20, 2026

Paper Submission

Papers should be submitted at <https://mc.manuscriptcentral.com/etrij>, and should adhere to the journal's [Author Guidelines](#).

Section Editor

Prof. DaeHun Nyang, Ewha Womans University, Rep. of Korea, nyang@ewha.ac.kr

Guest Editors

Dr. Sungwon Yi, ETRI, Rep. of Korea, sungyi@etri.re.kr

Prof. Jaehyung Kim, Yonsei University, Rep. of Korea, jaehyungk@yonsei.ac.kr

Prof. Jaegul Choo, KAIST, Rep. of Korea, jchoo@kaist.ac.kr

The Editorial Office can be contacted at etrij@etri.re.kr.