IEEE JOURNAL ON
EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS

CALL for PAPERS

Towards Trustworthy AI: Advances in Circuits, Systems, and Applications

Guest editors

Shih-Hsu Huang, Chung Yuan Christian University, Taiwan (shhuang@cycu.edu.tw)

Pin-Yu Chen, IBM T. J. Watson Research Center, USA (Pin-Yu.Chen@ibm.com)

Stjepan Picek, Radboud University, The Netherlands (stjepan.picek@ru.nl)

Chip-Hong Chang, Nanyang Technological University, Singapore (echchang@ntu.edu.sg)

Scope and purpose

Artificial Intelligence (AI) has made remarkable progress in various real-world applications. There is an increasing concern about the security and reliability of circuits and systems that incorporate or deploy AI, particularly when these systems are tasked with executing mission-critical and safety-critical functions for modern society. This special issue is dedicated to showcasing the latest technical advancements in emerging technologies related to trustworthy AI circuits and systems.

Generally, hardware is considered the root of trust. However, in reality, this assumption may not be valid. It is essential to develop safe and secure circuits to reduce the risks associated with the trustworthiness of the hardware that supports AI. Additionally, homomorphic encryption and multi-party computations are essential subjects for hardware-accelerated privacy-preserving model training and edge AI inference, particularly in distributed computing paradigms.

Most hardware designers lack knowledge in hardware security. Moreover, even with expertise in hardware security, hardware designers still need Electronic Design Automation (EDA) tools to perform trade-offs in timing, area, power, and security. Therefore, hardware designers require a security-aware design flow.

AI systems also grapple with new security challenges of heterogeneous computing for AI at the edge. One notable technique is transfer learning, which seeks to identify and use knowledge gained from past tasks or source domains in different but related tasks or target domains. Knowledge distillation is another technique that aids adversarial model extraction, even when the imitation model has different input data or architecture than its target model. This can pose a significant challenge to the intellectual property (IP) protection of large models.

While IP protection techniques such as watermarking, fingerprinting, and active usage control have been well researched in vision-based deep inference networks, less attention has been paid on the authenticity and copyright protection of large natural language processing models and hybrid models. Additionally, generative models offer endless imagination in content creation. Provenance has emerged as a new research direction in ensuring trust, transparency and accountability of generative AI.

The objective of this special issue is to enhance awareness among AI users regarding AI trustworthiness, and to compile the latest research findings from researchers in the field of developing trustworthy circuits and systems for AI.

Topics of interest

To give a comprehensive introduction, we plan to solicit papers presenting the latest developments in circuits, EDA, algorithms, and systems related to the trustworthy AI. With the broad scope, we prioritize different topics as follows:

1. ***Circuits.*** **This special issue is highly interested in the development of secure circuits to mitigate the risks associated with ensuring the trustworthiness of AI. Potential topics include hardware vulnerabilities, attacks, and defenses in AI implementations, as well as topics like homomorphic encryption and multi-party computations.**
2. ***EDA.*** **This special issue looks forward to featuring papers that concentrate on secure-aware EDA algorithms. It is also possible to construct a design flow ensuring the trustworthiness of AI through the integration of EDA tools.**
3. ***Algorithms.*** **At the algorithmic level, the topics include but are not limited to robustness, safety, security, fairness, privacy, and machine interpretability. Additionally, algorithms that explore new vulnerabilities and limitations of AI circuits and systems are highly encouraged.**
4. ***Systems.*** **This special issue anticipates papers presenting design examples of trustworthy AI applications and/or systems. Our aim is to solicit more secure and resilient privacy-preserving solutions.**

More precisely, the topics in any of the following areas or beyond:

- **Hardware vulnerabilities and attacks in AI accelerators**
- **Hardware-based authentication and encryption**
- **Trojan analysis and attacks on AI implementations**
- **Hardware root of trust for secure AI applications**
- **Intellectual property protection for deep learning, generative and large language models**
- **Hardware accelerated privacy-preserving machine learning**
- **Secure-aware EDA algorithms**
- **Secure-aware design flow or design methodology**
- **Trustworthy AI algorithms**
- **Adversarial attacks and defenses**
- **Model extraction attacks and defenses**
- **Fairness, privacy, and machine interpretability**
- **Security of generative models**
- **Security of large language models and hybrid models**
- **Design examples of Trustworthy AI applications and/or systems**
- **Federated learning**
- **Explainable AI (XAI)**

Submission procedure

Prospective authors are invited to submit their papers following the instructions provided on the IEEE JETCAS website: https://ieee-cas.org/publication/JETCAS/manuscript-submission-guide. The submitted manuscripts should not have been previously published, nor should they be currently under consideration for publication elsewhere.
The IEEE JETCAS submission site: https://ieee.atyponrex.com/journal/jetcas

Important dates

- **Manuscript submissions due:** **Jun. 21, 2024 (Extended)**
- **First round of reviews completed:** **Jul. 21, 2024**
- **Revised manuscripts due:** **Aug. 25, 2024**
- **Second round of reviews completed:** **Sep. 15, 2024**
- **Final manuscripts due:** **Oct. 6, 2024**

Request for information

Corresponding Guest Editor: Shih-Hsu Huang (shhuang@cycu.edu.tw)