==
Call for Contributions
6th International Workshop on Artificial Intelligence Safety Engineering (WAISE2023), co-located with SAFECOMP
Toulouse, France, September 19, 2023
Submission Deadline: 5 May 2023
WAISE2023 will be run as an in-person event
https://www.waise.org
==

The Workshop on Artificial Intelligence Safety Engineering (WAISE) explores new ideas on safety engineering for AI-based systems, ethically aligned design, regulation and standards for AI-based systems. In particular, WAISE will provide a forum for thematic presentations and in-depth discussions about safe AI architectures, ML safety, safe human-machine interaction, bounded morality and safety considerations in automated decision-making systems, in a way that makes AI-based systems more trustworthy, accountable and ethically aligned.

WAISE aims at bringing together experts, researchers, and practitioners, from diverse communities, such as AI, safety engineering, ethics, standardization and certification, robotics, cyber-physical systems, safety-critical systems, and application domain communities such as automotive, healthcare, manufacturing, agriculture, aerospace, critical infrastructures, and retail.

The workshop proceedings will be provided as a complimentary book to the SAFECOMP Proceedings in Springer Lecture Notes in Computer Science (LNCS) series, same as in previous years.

TOPICS

Contributions are sought in (but are not limited to) the following topics:

* Safety in AI-based system architectures: safety by design
* Regulating AI-based systems: safety standards and certification
* Runtime AI safety monitoring and adaptation
* Safe machine learning
* Formal methods for trustworthy AI
* Safety constraints and rules in decision-making systems
* AI-based system predictability
* Testing, verification and validation of safety properties in AI-based systems
* Algorithmic bias and AI discrimination
* Model-based engineering approaches to AI safety
* Ethically aligned design of AI-based systems
* Machine-readable representations of ethical principles and rules
* Uncertainty in AI
* Accountability, responsibility and liability of AI-based systems
* AI safety risk assessment and reduction
* Confidence, self-esteem and the distributional shift problem
* Reinforcement learning reward hacking and training corruption

* Self-explanation, self-criticism and the transparency problem
* Safety in the exploration vs exploitation dilemma
* Simulation for safe exploration and training
* Human-in-the-loop and the scalable oversight problem
* AI applied to safety engineering
* AI safety education and awareness
* Shared autonomy and human-autonomy teaming
* Experiences in AI-based safety-critical systems, including industrial processes, health, automotive systems, robotics, critical infrastructures, among others

SUBMISSION GUIDELINES

You are invited to submit full scientific contributions (max. 12 pages), short position papers (6 pages), proposals of technical talk/sessions (short abstracts, max. 2 pages) or doctoral papers (6 - 8 pages max, authored only by the PhD student describing their PhD dissertation research). Please keep your paper format according to Springer LNCS formatting guidelines (single-column format). The Springer LNCS author kit can be downloaded from: https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines . For publication in Springer LNCS, 6 pages are the minimum for a paper.

Manuscripts must be submitted as PDF files via EasyChair online submission system: https://easychair.org/my/conference?conf=waise2023

All papers will be peer-reviewed by the Program Committee (minimum of 3 reviewers per paper). The workshop follows a single-blind reviewing process.

ORGANIZING COMMITTEE
----
* Simos Gerasimou, University of York, UK
* Orlando Avila-García, Arquimea Research Center, Spain
* Mauricio Castillo-Effen, Lockheed Martin, USA
* Chih-Hong Cheng, Fraunhofer IKS, Germany
* Zakaria Chihani, CEA LIST, France

IMPORTANT DATES
----
* Full paper submission: 5 May 2023
* Notification of acceptance: 31 May 2023
* Camera-ready submission: 7 June 2023

All questions about submissions should be emailed to "waise2023 (at) easychair (dot) org"

For further details see https://www.waise.org .