

Preface to the Proceedings of TRUST-AI 2026 – the Second European Workshop on Trustworthy AI

1. Introduction

Trustworthy AI is increasingly important as AI is integrated into the services and practices of an ever-broadening range of application areas.

TRUST-AI 2026 is the second edition of a European workshop to support and strengthen reflection and research on trustworthy AI. The workshop is an arena for researchers and practitioners to meet and discuss applications, approaches, and open challenges in trustworthy AI.

The workshop is conducted August 15-16, 2026 as part of the International Joint Conferences on Artificial Intelligence (IJCAI 2026), in Bremen, Germany – a conference that also encompasses the European Conference on Artificial Intelligence (ECAI 2026).

At TRUST-AI 2026 we approach trustworthy AI from a human-centred perspective, where trustworthiness is considered over the lifetime of the AI system, spanning system technical attributes, business and societal implications, and adherence to ethical and legal requirements.

The interest and engagement in trustworthy AI are reflected in the substantial interest seen in the workshop with a total of 45 papers submitted.

2. Paper Invitation, Review, and Revision

The call for papers for TRUST-AI 2026 was openly distributed through the communication channels of IJCAI-ECAI 2026. Furthermore, it was spread through the networks of the workshop organizers and program committee, relevant email lists and networks, as well as contributors to the previous edition of the workshop.

Contributors could submit their work in one of two formats: short papers or position papers. Short papers were expected to present theoretical contributions, case-based experience, or results from empirical research, with a suggested upper limit of 9 pages (references excluded). Position papers were intended as more brief pieces, presenting a particular stance or raising open questions warranting reflection or debate. These proceedings contain the accepted short papers. Accepted position papers will appear on the workshop's [website](#).

The workshop received 45 submissions overall. After an initial editorial screening for topical fit, 38 of the submissions proceeded to peer review, of which 30 were short papers. Every paper sent for review was assessed by three independent reviewers through the EasyChair system, using a single-blind format. Each paper's review process was overseen by a lead reviewer drawn from the team of organizers and program committee members, with each lead reviewer responsible for 4-5 papers and one organizer coordinating the overall process. Final decisions on each submission were reached collectively at a dedicated lead reviewer meeting following the review stage.

Papers emerging from review could be accepted outright or subject to revision. Among the short papers appearing in this volume, 8 were accepted following minor revisions, 14 following major revisions. Minor revisions were checked by at least one organizer or program committee member to confirm they addressed the reviewers' comments; major revisions were checked by at least two. All conditionally accepted papers were approved after one or two rounds of revisions before final acceptance.

3. Workshop Program

The TRUST-AI 2026 program contained five sessions for plenary presentations of short papers and one session for poster presentations of short and position papers. The program also included a groupwork on the topic of general-purpose AI in knowledge work.

The plenary sessions were structured as five thematically oriented sessions: (1) fairness, bias and explainability, (2) confidence and hallucination detection, (3) risk assessment, governance, and regulation, (4) robustness, safety and security, and (5) trustworthy agents and LLMs. In addition, the short papers presented in the poster session are collated under the thematic heading (6) theoretical and applied perspectives.

We are pleased to note that the topical orientation of the accepted papers covers a range of key areas in trustworthy AI. As such, the program makes a good basis for an engaging and useful workshop. Furthermore, the workshop proceedings help advancing the important field of trustworthy AI.

4. Organizers and Reviewers

The organization of TRUST-AI 2026 is supported by the Horizon Europe research and innovation action THEMIS 5.0 – Human-centered Trustworthiness Optimisation in Hybrid Decision Support. The workshop organizers and program committee members represent several technology research organizations contributing to advancing the field of trustworthy AI. The responsibilities of the organizers and program committee members included workshop chairing, peer review leadership and organization, and the preparation of these proceedings. The organizers and program committee members were:

- Asbjørn Følstad, SINTEF, Norway
- Gregoris Mentzas, National Technical University of Athens, Greece
- Dimitris Apostolou, ICCS & University of Piraeus, Greece
- Steve Taylor, University of Southampton, UK
- Eleni Tsalapati, ATC, Greece
- Giannis Stamatellos, Institute of Philosophy & Technology, Greece
- Andrea Palumbo, KU Leuven, Belgium
- Gabriel González Castañé, ADRA, Belgium
- Henrik Junklewitz, Fraunhofer IAIS, Germany

The peer review process relied on the involvement of 34 peers serving as reviewers. The reviewers were mainly recruited from among the authors of submitted papers, and also from the networks of the organizers and program committee members. Each peer contributed 1-3 reviews. The reviewers were:

- Alexandru Mateescu, Université Paris1 Panthéon-Sorbonne, France
- Astrid Franz, CONTACT-Software GmbH, Germany
- Celine Spannagl, Bavarian Research Institute for Digital Transformation, Germany
- Diogen Babuc, West University of Timișoara, Romania
- Eirik Valseeth, Simula Research Laboratory, Norway
- Evangelia Anagnostopoulou, ICCS, Greece
- Federico Sabbatini, University of Urbino, Italy
- George Bargiannis, University of Huddersfield, United Kingdom
- Hasan Kurban, Hamad Bin Khalifa University, Qatar
- Hayat Ullah, Florida Atlantic University, United States
- Helge Spieker, Simula Research Laboratory, Norway
- Jan K. Argasinski, Sano Centre for Computational Medicine & Jagiellonian University in Kraków, Poland
- Joel Sokol, Georgia Institute of Technology, United States
- Lorenzo Lorusso, ASST Lecco, Reparto di Neurologia, Italy
- Lorenzo Valerio, IIT-CNR, Italy

- Lynn Houthuys, Vrije Universiteit Brussel, Belgium
- Maria Skublewska-Paszkowska, Lublin University of Technology, Poland
- Matteo Gregorini, Università degli Studi di Pisa, Italy
- Md Foysal Ahmed, Bowling Green State University, USA
- Mikhail Tchernikov, Capgemini Engineering Deutschland, Germany
- Mina Gabriel, Temple University, USA
- Nikhil Singhal , Meta, USA
- Nikolaos Antonios Grammatikos, University of Piraeus, Greece
- Nivetha Suntharamoorthy, Norwegian University of Life Sciences (NMBU), Norway
- Pavlo Barkhayev , Capgemini Engineering Deutschland, Germany
- Paweł Powroźnik, Lublin University of Technology, Poland
- Samir Abdaljalil, Texas A&M University, United States
- Sebastian Krügel, University of Hohenheim, Germany
- Shuang Ao, University of Warwick, United Kingdom
- Shukun Tokas, SINTEF Digital, Norway
- Ute Schmid, University of Bamberg, Germany
- Venkat Sai Suman Lamba Karanam, Bowling Green State University, United States
- Zewen Shen, University of Sydney, Australia
- Zexi Wang, Macquarie Business School, Macquarie University, Australia

Acknowledgments

The workshop was supported by EU Horizon Europe, HORIZON-CL4-2022-HUMAN-02-01, under the project THEMIS 5.0 (grant agreement No. 101121042).