University of Kent | Institute of Cyber Security for Society (iCSS)

Co-Chair (TPC Chair)'s Report

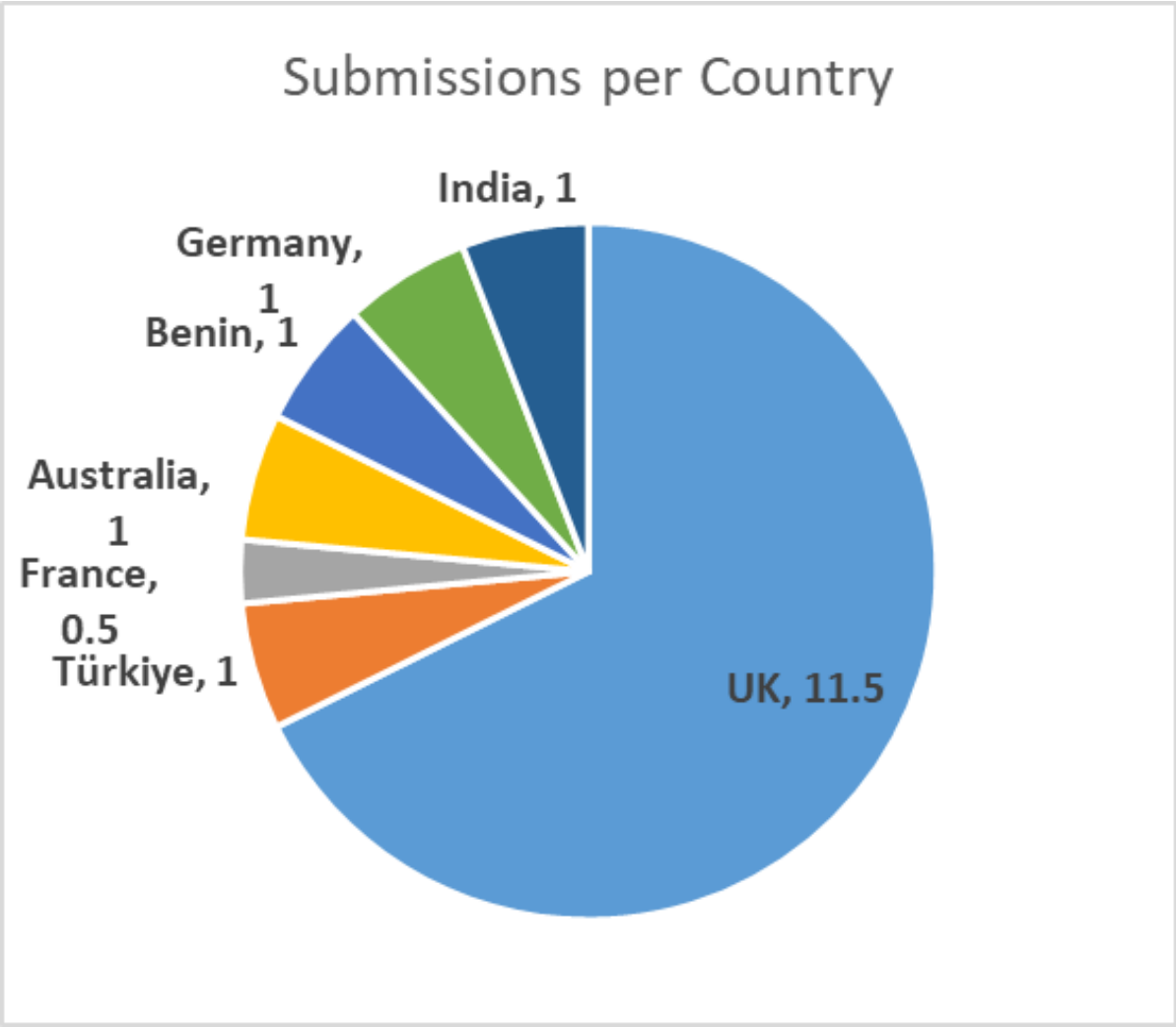
AI-SS 2026 (1st International Workshop on AI Safety and Security)

Shujun Li
Institute of Cyber Security for Society (iCSS) & School of Computing
University of Kent
Canterbury, Kent, UK

Paper submissions

University of Kent | Institute of Cyber Security for Society (iCSS)

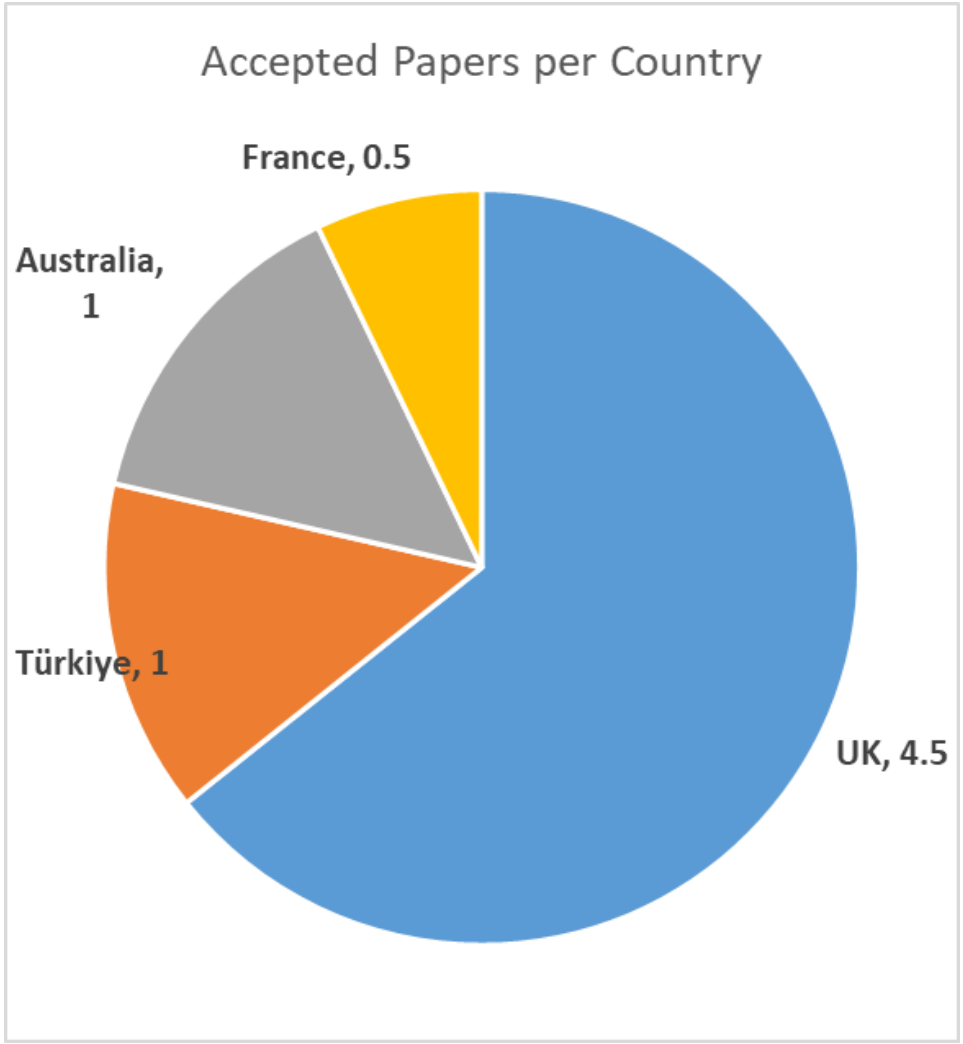
- 18 submissions from authors in 7 countries
- The most submissions (11) from researchers in the UK
- One paper with authors from the UK and France, so counted 0.5 for each country



Papers accepted, rejected and withdrawn

University of Kent | Institute of Cyber Security for Society (iCSS)

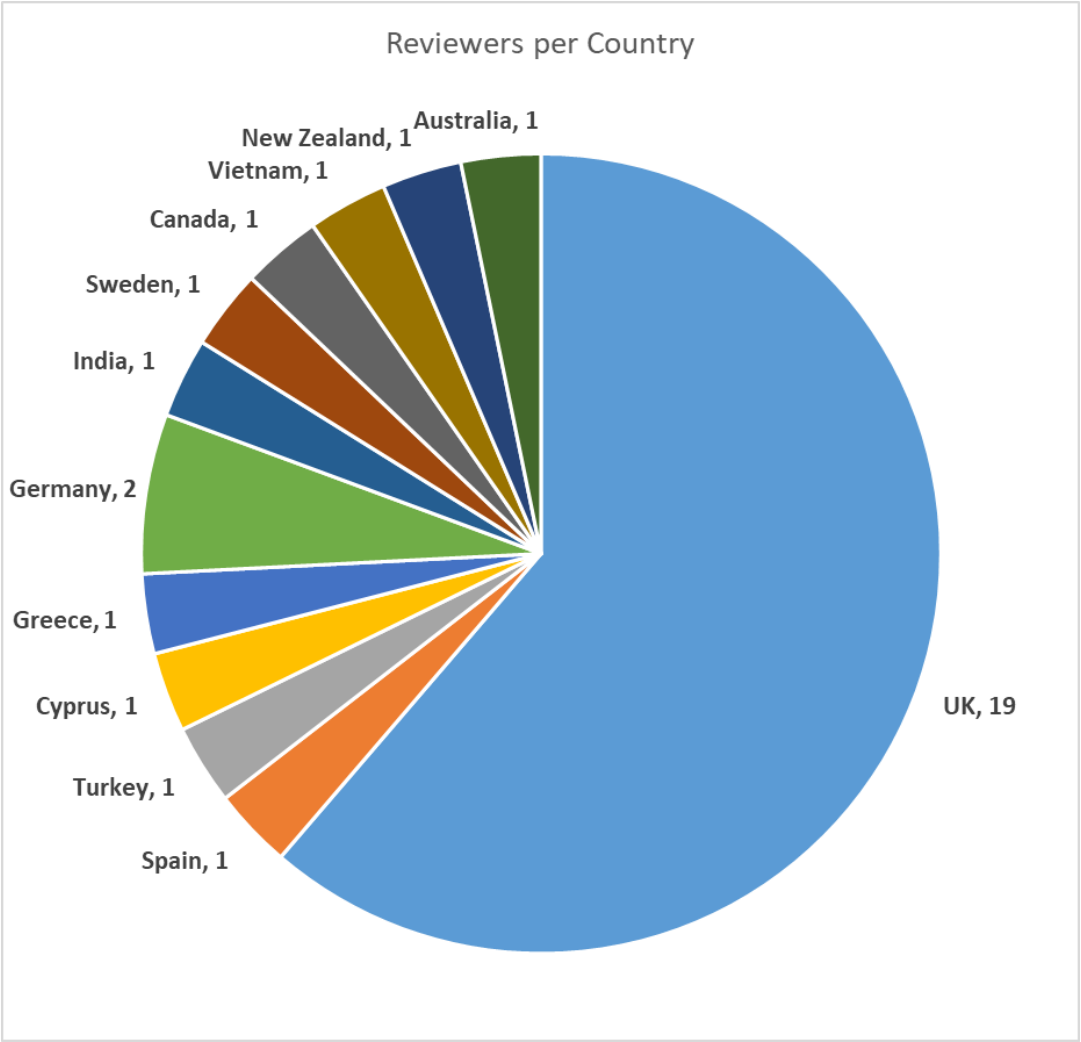
- Submissions: 18
- Papers accepted: 7
- Papers rejected: 9
- Papers withdrawn (without peer reviews): 2
- Acceptance rates: including withdrawn papers (7/18≈38.9%), excluding withdrawn papers (7/16=43.75%)
- Statistics of accepted papers per country (see the pie chart)



TPC members, reviewers and reviews

University of Kent | Institute of Cyber Security for Society (iCSS)

- 29 TPC members from 12 countries/regions
- 30 peer reviewers
  - 28 TPC member
  - Workshop Chair (Professor He)
  - An additional reviewer
- Number of reviews
  - 3 reviews: 10 submissions
  - 4 reviews: 5 submissions
  - 5 reviews: 1 submission
  - Total reviews: 55 (≈1.83 per reviewer)



AI-SS 2026 proceedings

University of Kent | Institute of Cyber Security for Society (iCSS)

- Published by IEEE Computer Society CPS (Conference Publishing Services) as a companion proceedings of the EDCC 2026 Conference
- Will be submitted to IEEE Xplore for inclusion

Keynote (9:30-10:30)

University of Kent | Institute of Cyber Security for Society (iCSS)

- **Speaker**: Professor Aad van Moorsel, Chair in Decentralised Systems & Head of School, School of Computer Science, University of Birmingham, UK
- **Title**: The Uneasy Marriage of AI and Dependability
- **Abstract**:
  - In this presentation, the speaker will discuss the relation between designing traditional dependable, fault-tolerant computer systems and enhancing the robustness of modern-day AI-based services. On the one hand, AI mechanisms such as ensembles and reject option have direct counterparts in hardware and software dependability, even though their motivation, justification and implementation are quite different. On the other hand, AI shifts the emphasis to a new type of faults that are hard to identify and, in the case of Generative AI, underspecified. The speaker will reflect on how to bridge between the world of dependability and AI, and provide a number of ways forward to design highly dependable AI-based systems.

Technical sessions (11-12:30 + 16-17:10)

University of Kent | Institute of Cyber Security for Society (iCSS)

- **Technical Session 1 – AI and Cyber Threats**
  1. *HaiR – A Human-Centric AI Remediation Framework* (University of Salford, UK)
  2. *An Empirical Evaluation of Prompt Injection Vulnerabilities in Large Language Models Across Multilingual and Obfuscated Attack Scenarios* (Sabanci Univeristy, Türkiye
  3. *Resilience to Disinformation Among Cyber Security and Information Warfare Professionals: Why Experts May Get it Wrong* (University of Kent, UK; Université Bourgogne Europe, Burgundy, France; ESSCA, France; Université Bourgogne Europe, Burgundy, France)
  4. *Leveraging Sparse Intelligence: Large Language Models Can Provide Actionable Priorities From Limited Cyber Threat Data* (Macquarie University, Australia; Data61, CSIRO, Australia)
- **Technical Session 2 – Agentic AI Safety and Security**
  1. *Agentic AI vs Non-Agentic AI: Motivation, Security Implications, and Research Foundations* (University of Kent, UK)
  2. *Exploring Agentic AI in Anti-Forensics: Simulation of Evasion Tactics in Digital Investigations* (University of Salford, UK)
  3. *Agentic Knowledge Distillation: Autonomous Training of Small Language Models for SMS Threat Detection* (University of Kent, UK)

Panel discussion (14-15:30)

University of Kent | Institute of Cyber Security for Society (iCSS)

- **Facilitator**:
  - Hongmei (Mary) He, University of Salford, UK
- **Panellists**:
  - Ali Hessami, Director of R&D and AI, Vega Systems, UK; Chair of IEEE UK and Ireland Society on Social Implications of Technology (SSIT) Chapter
  - Gareth Howells, Professor of Cyber Security, University of Essex, UK; Founder, Director and CTO, Metrarc Ltd, UK
  - Carl Shaw, Codasip, UK
  - Harish Vundavalli, Senior Technical Architect, Strategic Education, Inc., USA
- **Title**:
  - Towards Dependable and Secure AI for Critical Systems

Best Paper & Best Student Paper Awards

University of Kent | Institute of Cyber Security for Society (iCSS)

- One paper will be selected to receive the **Best Paper Award**.
- One paper whose first author is a student will be been selected to receive the **Best Student Paper Award**.
- The Workshop Chair and Co-Chair will be the judges.
- **Criteria**: paper reviews and scores, the quality of the presentation, and feedback from the audience
- Award winners will be announced at the closing session (17:40-18).
- In addition to **a printed certificate**, there is also **a cash prize** for each award, sponsored by University of Kent | Institute of Cyber Security for Society (iCSS)

Enjoy your stay at AI-SS 2026!