2026 第九届 IEEE 国际无人系统大会
特邀专题简介表

特邀专题名称

可解释人工智能

组织者

1. 周鋆，副教授，国防科技大学
2. 老明瑞，副教授，国防科技大学
3. 左浩，博士，国防科技大学

个人简介

周鋆，伦敦大学玛丽皇后学院计算机科学博士，国防科技大学信息系统工程重点实验室主任助理，副教授，硕士生导师。主要研究方向为机器学习与贝叶斯网络理论，主持包括国家自然科学基金面上项目在内的国家级、省部级科研项目 7 项。2019 年入选国防科技大学首批创新人才（卓越青年）培养对象，2020 年入选长沙市杰出创新青年培养计划，2021 年入选湖湘青年科技创新人才，以第一作者、通信作者身份发表论文 30 余篇，其中包括领域知名会议 UAI、KDD、ICML、IJCAI、ICDM 等，第一发明人授权国家发明专利 20 项，学术成果目前被引 1200 余次，并担任 TPAMI、TKDE、TNNLS 等十余个 SCI 期刊的审稿人，以及中国指挥与控制学会 C4ISR 理论与技术专委会秘书长，获 CICC 科技进步奖一等奖 1 项，二等奖 1 项。

老明瑞，荷兰莱顿大学计算机科学博士，国防科技大学信息系统工程重点实验室副教授，硕士生导师。主要研究方向为多模态学习与大模型智能体技术，入选湖南省优青、中国指挥与控制学会青年人才托举计划，主持国家自然科学基金青年项目、湖南省重点研发计划等科研项目 5 项，在 NeurIPS、CVPR、ICCV、ACM MM、KDD、SIGIR、AAAI、TMM、计算机研发等中国计算机学会 A 类期刊/会议上发表长文 20 篇，出版专著 2 部，第一发明人授权国家发明专利 6 项，担任中国计算机学会分布式计算与系统执行委员，获中国指挥控制学

会科技进步二等奖 1 项。

左浩，国防科技大学系统工程学院博士研究生，主要研究方向为深度学习、因果发现与因果推断。共计发表学术论文 5 余篇，其中 SCI 检索论文 3 篇、国家软件著作权 1 项、国家发明专利 1 项,并担任 Human-Centric Intelligent Systems 期刊的审稿人。

特邀专题简介

随着人工智能技术的快速发展，机器学习、深度学习和大模型等方法已广泛应用于医疗、金融、制造等领域，但其“黑盒”性质导致了决策过程不透明、机理难以理解等问题。可解释人工智能（XAI）旨在研究能够揭示模型决策依据、推理过程和作用机制的理论、方法与技术，提升人工智能系统的安全性、可信性与公平性，使其能够对预测、分类和决策结果提供合理的解释，已成为人工智能领域的重要研究方向。XAI 研究逐步拓展到可自解释建模、人机协同解释、因果发现与因果推断、知识增强解释、公平性与鲁棒性分析等多个方面。开展相关研究，对于推动人工智能从高性能向高可信发展，促进其在复杂场景中的安全应用，具有重要的理论意义与实际价值。

本特邀专题面向可解释人工智能相关前沿问题，征集具有创新性和应用价值的原创论文。征稿范围包括但不限于：

- 可解释人工智能基础理论与方法
- 机器学习与深度学习的可解释建模
- 大模型的可解释方法
- 基于智能体的可解释系统
- 面向可解释性的贝叶斯模型
- 因果发现与因果推断方法
- 知识驱动与神经符号推理方法
- 人工智能中的公平性、鲁棒性与可信性
- 可解释人工智的应用

IEEE ICUS 2026
Invited Session Summary

Title of Session

Explainable Artificial Intelligence

Name

1. Prof. Yun Zhou
National University of Defense Technology, China
2. Prof. Mingrui Lao
National University of Defense Technology, China
3. Dr. Hao Zuo
National University of Defense Technology, China

Biosketches of Organizers

Yun Zhou received his Ph.D. in Computer Science from Queen Mary University of London. He is currently an Associate Professor, Master's Supervisor, and Assistant to the Director of the Key Laboratory of Information Systems Engineering at the National University of Defense Technology. His research mainly focuses on machine learning and Bayesian network theory.He has led seven national- and provincial-level research projects, including a General Program grant from the National Natural Science Foundation of China. He was selected for the first cohort of the Innovation Talent (Outstanding Young Scholars) Program at the National University of Defense Technology in 2019, the Changsha Distinguished Young Innovators Program in 2020, and the Huxiang Young Scientific and Technological Innovation Talent Program in 2021. As first author or corresponding author, he has published more than 30 papers, including those at leading conferences such as UAI, KDD, ICML, IJCAI, and ICDM. As first inventor, he has been granted 20 Chinese invention patents. His work has received more than 1,200 citations. He also serves as a reviewer for more than ten SCI-indexed journals, including TPAMI, TKDE, and TNNLS, and as Secretary-General of the C4ISR Theory and Technology Committee of the Chinese Institute of Command and Control. He has received one First Prize and one Second Prize of the CICC Science and Technology Progress Award.

Mingrui Lao, a Ph.D. in Computer Science from Leiden University, Netherlands, is an associate professor and master's supervisor at the Key Laboratory of Information System Engineering, National University of Defense Technology. His main research focuses include multimodal learning and large model intelligent agent technologies. He has been selected as a Young Outstanding Talent in Hunan Province and included in the Youth Talent Support Program of the China Society for Command and Control. He has led five scientific research projects, including the National Natural Science Foundation Youth Project and the Hunan Provincial Key Research and Development Plan. He has published 20 long papers in A-class journals/conferences of the China Computer Federation, such as NeurIPS, CVPR, ICCV, ACM MM, KDD, SIGIR, AAAI, TMM, and Computer Research and Development. He has authored two monographs, obtained six authorized national invention patents as the first inventor, and serves as a member of the Distributed Computing and System Execution Committee of the China Computer Federation. He has also received the Second Prize for Scientific and Technological Progress from the China Society for Command and Control.

Hao Zuo is a Ph.D. candidate at the College of Systems Engineering, National University of Defense Technology. His research interests include deep learning, causal discovery, and causal inference. He has published more than five academic papers, including three SCI-indexed papers. He has also obtained one software copyright and one national invention patent. In addition, he serves as a reviewer for the journal Human-Centric Intelligent Systems.

Details of Session

With the rapid development of artificial intelligence technologies, methods such as machine learning, deep learning, and large language models have been widely applied in fields including healthcare, finance, and manufacturing. However, their “black-box” nature has led to problems such as opaque decision-making processes and difficulty in understanding underlying mechanisms. Explainable Artificial Intelligence (XAI) aims to investigate theories, methods, and technologies that can reveal the basis of model decisions, reasoning processes, and operating mechanisms, thereby improving the safety, trustworthiness, and fairness of AI systems. By enabling models to provide reasonable explanations for prediction, classification, and decision-making results, XAI has become an important research direction in artificial intelligence. Research on XAI has gradually expanded to multiple areas, including self-explainable modeling, human-centered collaborative explanation, causal

discovery and causal inference, knowledge-enhanced explanation, and fairness and robustness analysis. Advancing research in these areas is of great theoretical significance and practical value for promoting the development of artificial intelligence from high performance to high trustworthiness, and for facilitating its safe deployment in complex scenarios.

This special issue focuses on frontier problems related to Explainable Artificial Intelligence and solicits original papers with innovation and practical value. Topics of interest include, but are not limited to, the following:

- Fundamental theories and methods of Explainable Artificial Intelligence
- Explainable modeling for machine learning and deep learning
- Explainability methods for large models
- Agent-based explainable systems
- Bayesian models for explainability
- Methods for causal discovery and causal inference
- Knowledge-driven and neuro-symbolic reasoning methods
- Fairness, robustness, and trustworthiness in artificial intelligence
- Applications of Explainable Artificial Intelligence