统计与人工智能安全研讨会

会议时间：2025.11.24（周一） 14:30-17:30

会议方式：数学楼 133 会议室（腾讯会议：851-287-084）

11/24

14:30-17:30 **统计与人工智能安全研讨会**

14:30-15:00 【1】**报告题目：**人机物融合智能系统安全

报告人：蔺琛皓，西安交通大学

个人简介：蔺琛皓，西安交通大学网络空间安全学院教授、博士生导师，青年长江学者，教育部工程研究中心副主任。长期从事人工智能安全、AI4Science 等相关研究，在该方向发表论文 70 余篇，包括 IEEE TPAMI、TDSC、TIFS、TIP、USENIX Security、S&P、ICML、NeurIPS、AAAI、CVPR、ICCV 等，获得了 IJCAI-DCM 等国际学术会议最佳论文奖 3 次、达摩院青橙奖最具潜力奖、吴文俊人工智能优秀青年奖、中国自动化学会自然科学一等奖、人社部高层次留学人才回国资助等；入选了陕西省高层次人才（青年）、陕西省青年科技新星、小米青年学者、思源学者等；主持了重点研发计划项目课题，科技创新 2030--“新一代人工智能”重大项目课题，国家自然科学基金重点类项目课题、面上、青年项目等，基金委创新研究群体 B 类骨干成员；担任中国自动化学会人工智能与安全专委会副主任委员，以及 ACM SIGSAC China、中国人工智能学会人工智能与安全等多个专委会委员。

15:00-15:30 【2】**报告题目：**多模态大模型安全与可信

报告人：杨乐，西安交通大学

个人简介：杨乐 2021 年博士毕业于清华大学自动化系。现任西安交通大学电信学部特聘研究员，博士生导师，入选“博新计划”（合作导师：管晓宏院士），获 CSIG 自然科学一等奖，长期从事人工智能安全、计算机视觉等方向研究，主持国自然青基等项目 10 余项，发表高水平学术论文 30 余篇，谷歌学术引用 2500 余次。其中以第一/通讯作者发表 CVPR 等高水平会议/期刊论文 20 余篇，获最佳学术论文奖 3 项，授权/受理国家发明专利4项，授权美国发明专利 1 项，参与国际 ITU 标准立项 1 项，担任《电子学报》青年编委。

14:30-17:30 **统计与人工智能安全研讨会**

15:30-16:00 【3】**报告题目：人工智能安全：小模型→大模型→智能供应链**

报告人：赵正宇，西安交通大学

个人简介：赵正宇，西安交通大学网络空间安全学院教授/博士生导师，国家自然科学基金-优秀青年科学基金项目（海外）获得者。博士毕业于荷兰 Radboud 大学，获 CCF-A 类会议 CVPR 2021 博士论坛奖，随后于德国 CISPA 亥姆霍兹信息安全中心担任博士后。主要从事人工智能安全研究，发表相关学术论文 40 余篇，包括 NeurIPS、ICML、ICLR、CVPR、ICCV、USENIX Security、CCS、NDSS、TPAMI等，获最佳论文奖 1 项。担任 ICML、NeurIPS、AAAI、MM 等多个 CCF-A 类会议领域主席/组织成员等，以及 10 余个 CCF-A 类会议/期刊审稿人，获杰出审稿人奖 4 次。多次带队获得 CCF-A 类会议人工智能安全国际挑战赛前 3 名。作为负责人或研究骨干参与多项国家级科研项目。

16:00-16:30 【4】**报告题目：模型组合中的双下降现象与泛化理论研究**

报告人：姜丹丹，西安交通大学

个人简介：姜丹丹教授任西安交通大学数学与统计学院教授、博士生导师，国家级青年人才计划入选者、陕西省高层次人才引进青年计划入选者、陕西基础科学研究院副院长、西安数学与数学技术研究院统计学与大数据技术中心副主任。主要从事随机矩阵、高维统计推断等理论研究及其应用研究，研究成果发表在 Annals of Statistics、Biometrika、Bernoulli、Statistica Sinica、Science China Mathematics 等期刊。主持千万级国家重点研发计划课题1项；主持国家自然科学基金面上项目 3 项；主持省部级科研项目 6 项；先后主持华为横向项目 3 项，研究成果以第一发明人申请专利 2 项，1 项国际专利已授权并获华为“突出贡献奖”。作为第一获奖人曾获吉林省自然科学学术成果奖特别奖、全国百篇优博论文提名奖等。担任中国现场统计研究会随机矩阵理论及其应用分会秘书长、中国现场统计研究会大数据统计分会副理事长、CSIAM 青年工作委员会委员、CSIAM 大数据与人工智能专委会委员等。

14:30-17:30 **统计与人工智能安全研讨会**

16:30-17:00 【5】报告题目：**Robust Safety Guarantee for Large Language Models via Preference-Augmented Distributional Alignment**

报告人：严晓东，西安交通大学

个人简介：严晓东，西安交通大学数学与统计学院教授，博士生导师，入选国家级青年人才项目和校内青拔 A 类支持计划，荣获“华为火花奖”，“滴滴盖亚学者”, 研究方向为统计决策、统计推断和统计计算等。学术成果发表在著名期刊 JRSSB, AOS, JASA, JOE 以及人工智能顶级会议 NeurIPS, ICMI, AAAI 等 50 余篇。 在“高等教育出版社出版”以独立主编出版了《机器学习》、《数据科学实践基础-基于R》两部教材。

17:00-17:30 【6】专家交流.