附件

2026年度湖北省自然科学奖公示信息（候选人二级单位/完成单位/工作单位）

项目名称	对抗环境下智能系统数据安全与鲁棒识别
提名者	武汉大学
提名意见	项目团队依托国家科技创新2030“新一代人工智能”重大项目、国家自然科学基金委员会联合基金重点支持项目等多项国家和省部级项目，历经多年攻关研究，形成“对抗环境下智能系统数据安全与鲁棒识别”系统性成果。该成果聚焦智能系统“训练数据潜藏安全风险”、“模型漏洞导致干扰欺骗”、“复杂环境引发脆弱识别”三个科学问题，揭示了智能系统从模型训练到模型识别再到系统应用面临的各种安全威胁，提出了对抗环境下针对智能系统数据安全与鲁棒识别的攻击与防御新理论、技术和方法，从“对抗”的角度出发，以攻击评测促进防御能力提升，重点解决对抗环境下智能系统数据与模型全周期安全可控问题，理论和实际意义重大。项目成果获最佳论文奖及提名7项。5篇代表性论文发表在Proceedings of the IEEE、IEEE Transactions on Vehicular Technology、IEEE INFOCOM、Engineering、Science China Technological Sciences等国际国内顶级期刊和会议。项目第一完成人被聘为二级教授、弘毅特聘教授，入选“长江学者奖励计划”特聘教授、IEEE Fellow、国家海外高层次人才引进计划（青年项目），获批国家优青项目，获湖北省青年科技创新奖。其他完成人入选ACM/IEEE/AAAS/CCF Fellow、国家海外高层次人才引进计划（创新人才长期项目）、“长江学者奖励计划”特聘教授、科睿唯安“全球高被引科学家”，获腾讯探索奖、阿里青橙奖，获批国家自然科学基金委创新研究群体项目(B类)、国家优青等项目。
项目简介	随着基于人工智能的图像、语音、文本识别在民用和国防领域广泛应用，面向智能系统的数据安全与鲁棒识别问题也日渐凸显，成为影响AI应用安全发展的关键因素。因此，破解AI的“自主性”与“可控性”的对立统一，全面保障智能系统安全是推动AI落地应用健康快速发展的迫切现实要求。 结合国家重大需求和学科国际前沿发展趋势，本项目围绕由智能系统模型训练、模型识别、系统应用三个主要阶段构成的智能系统全生命周期，针对开放、复杂、对抗环境下智能系统面临的安全风险与挑战：“训练数据敏感、模型结构复杂、模型漏洞隐蔽、识别环境复杂”，聚焦“训练数据潜藏安全风险”、“模型漏洞导致干扰欺骗”、“复杂环境引发脆弱识别”三个科学问题，从“对抗”的角度出发，以攻促防，提出了对抗环境下智能系统数据安全与鲁棒识别新理论、技术和方法。主要科学发现如下： 1.阐明了智能系统在模型训练阶段，用户级隐私数据泄露和虚假数据注入攻击的成因与机理，提出了基于梯度信息的隐私训练数据重建、基于状态估计的虚假数据注入攻防、基于零和博弈的攻防对抗等理论和方法，形成了模型训练数据多维安全评估与防护技术体系，提升了模型训练数据威胁检测与防御能力，被国际知名学者评价为“令人印象深刻的”、“更现实和实用的”，“充分研究并证明了隐私风险场景”、“窃取隐私的可用方法”等。 2.发现了智能系统在模型识别阶段，面临的一系列新攻击面和潜在安全漏洞，提出了基于多模态输入数据的白盒/黑盒、非定向/定向攻防理论和方法，展示了此前攻防技术在安全性、有效性、通用性等方面的不足，为模型漏洞深度挖掘提供了新思路，为有效抵御针对模型识别的对抗性攻击、增强模型鲁棒性奠定了理论基

	础，被国际知名学者评价为“人眼不可见的对抗性攻击”、“使得模型识别性能急剧下降”、“代表性防御机制”等。 3.揭示了智能系统在部署应用阶段，多模态输入经物理信道传播导致智能终端识别脆弱性显著增加，证实了从感知、认证、决策等维度智能终端物理域数据安全威胁，提出了结合卷积神经网络与循环神经网络的混合深度学习架构，克服了此前方法在极端情况下的性能局限，为自动驾驶感知算法发展提供了公开基准，被国际知名学者评价为“高精度的”、“常用技术手段”、“具有革新的”等。 本项目获国家科技创新2030“新一代人工智能”重大项目、NSFC重点支持项目等资助。项目第一完成人被聘为二级教授，入选长江学者特聘教授、IEEE Fellow、国家青年千人计划，获批国家优青项目，获湖北省青年科技创新奖。其他完成人入选ACM/IEEE/AAAS/CCF Fellow、国家千人计划、长江学者特聘教授、科睿唯安“全球高被引科学家”，获腾讯探索奖、阿里青橙奖，获批NSFC创新研究群体项目(B类)等项目。
主要完成人 （完成单位）	王骞（武汉大学）、任奎（浙江大学）、沈超（西安交通大学）、王志波（武汉大学）、邹勤（武汉大学）

代表性论文（专著）目录

序号	论文（专著）名称/刊名/作者	年卷页码（xx年xx卷xx页）	发表时间（年月日）	通讯作者（含共同）	第一作者（含共同）	国内作者	他引总次数	检索数据库	论文署名单位是否包含国外单位	是否国内期刊，如是请填写CN号
1	Beyond Inferring Class Representatives: User-Level Privacy Leakage from Federated Learning/IEEE Conference on Computer Communications/Zhibo Wang, Mengkai Song, Zhifei Zhang, Yang Song, Qian Wang, and Hairong Qi	2019年 页:2512-2520	2019年04月01日	Qian Wang	Zhibo Wang	王志波，宋梦凯，王骞	650	Web of Science核心合集	是	否
2	False Data Injection Attacks Against Smart Grid State Estimation: Construction, Detection and Defense/Science China Technological Sciences/Meng Zhang, Chao Shen, Ning He, Sicong Han, Qi Li, Qian Wang, and Xiaohong Guan	2019年卷:62 期:12 页:2077-2087	2019年12月01日	Chao Shen	Meng Zhang	张萌，沈超，贺宁，韩思聪，李琦，王骞，管晓宏	43	Web of Science核心合集	否	CN:11-5845/TH
3	Adversarial Attacks and Defenses in Deep	2020年卷:6 期:3	2020年03月01日	Kui Ren	Kui Ren	任奎，秦湛	410	Web of Science核心合集	是	CN:10–1244/N

	Learning/Engineering/Kui Ren, Tianhang Zheng, Zhan Qin, and Xue Liu	页:346-360								
4	The Security of Autonomous Driving: Threats, Defenses, and Future Direction/Proceedings of the IEEE/Kui Ren, Qian Wang, Cong Wang, Zhan Qin, and Xiaodong Lin	2020 年 卷:108, 期:2 页:357-372	2020 年 02 月 01 日	Kui Ren	Kui Ren	任奎、王骞、王聪、秦湛	145	Web of Science 核心合集	是	否
5	Robust Lane Detection from Continuous Driving Scenes Using Deep Neural Networks/IEEE Transactions on Vehicular Technology/Qin Zou, Hanwen Jiang, Qiyu Dai, Yuanhao Yue, Long Chen, and Qian Wang	2020 年 卷:69 期:1 页: 41-54	2020 年 01 月 01 日	Qian Wang	Qin Zou	邹勤、江瀚文，戴启宇，岳远昊，陈龙、王骞	260	Web of Science 核心合集	否	否