

目 次

人工智能应用中的数据安全专刊

编者按·····	2643
黄欣沂, 何德彪	
SDFL: 一种隐私保护和拜占庭鲁棒的去中心化联邦学习方案·····	2645
全韩或, 钱彦屹, 田晖, 刘雪峰, 李越, 王健宗, 钟迦	
自适应拜占庭鲁棒的差分隐私联邦学习·····	2663
王玉画, 张沁楠, 邱望洁, 柴子川, 高胜, 朱建明, 童咏昕, 郑志明	
一种适用于无辅助计算车载网络的联邦学习隐私保护方案·····	2683
罗向阳, 陈星星, 马玉千, 程庆丰, 陈晓峰	
基于更新残差的差分隐私联邦遗忘学习机制·····	2704
王腾, 翟林东, 禹勇, 杨腾飞, 张雪锋, 任雪斌	
基于模型分解与加权聚合的联邦元遗忘·····	2722
王亚杰, 范青, 潘梓杰, 应作斌, 张子剑, 祝烈煌	
博弈论驱动的多层次扰动集成对抗攻击·····	2741
陈淑红, 唐猛猛, 李汉俊, 周志立, 王国军, 王田	
DMS-MIA: 面向 TEE 保护机器学习模型的磁盘重放与多指标序列成员推理攻击·····	2759
董一凡, 冯伟, 李昊, 张敏, 秦宇, 冯登国	
基于改进知识蒸馏的黑盒攻击方法·····	2780
石家乐, 宋亚飞, 吴晓佰, 李天鹏	
无需触发器与辅助数据集的模型后门攻击·····	2798
王嘉豪, 张翔龙, 章宜乐, 马小博, 成秀珍, 胡鹏飞, 张国明	
支持双外包的轻量级函数隐藏内积加密方案·····	2817
钱心缘, 袁帅, 徐国文, 李洪伟	
Bandage: 针对图对抗攻击的双向嵌套的防御算法·····	2834
刘生昊, 徐科, 王婧, 邓贤君, 何媛媛, 杨天若	
训练—推理协同的图神经网络成员推理防御框架·····	2852
戚富琪, 全星屹, 高海昌, 王萍	
基于并行深度神经网络的物联网边缘侧隐私数据检测·····	2870
周少鹏, 王滨, 李超豪, 周政演, 张峰, 吴春明	
从编译到反编译: 基于源码级转换的高效水印去除方法·····	2886
黄昊, 周瑞桦, 罗家棠, 马晓欢, 李运鹏, 刘玉岭	
基于神经网络决策边界与一致性分析的对抗样本攻击检测方法·····	2902
谢盈, 张小松, 丁旭阳, 郎蕊霞	
M ³ -SafetyBench: 多领域多场景多维度的大语言模型安全评估体系·····	2923
杨伟平, 程豪杰, 周百顺, 文煜鑫, 刘雨帆, 刘宇阳, 李兵, 郎丛妍, 陈乃月, 张伟, 胡卫明	

Contents

Special Topic: Data Security in Artificial Intelligence Applications

Editorial.....	2643
Xinyi HUANG & Debiao HE	
SDFL: a privacy-preserving Byzantine-robust decentralized federated learning scheme.....	2645
Hanyu QUAN, Yanyi QIAN, Hui TIAN, Xuefeng LIU, Yue LI, Jianzong WANG & Jia ZHONG	
Adaptive Byzantine-robust differentially private federated learning	2663
Yuhua WANG, Qinnan ZHANG, Wangjie QIU, Zichuan CHAI, Sheng GAO, Jianming ZHU, Yongxin TONG & Zhiming ZHENG	
A federated learning privacy preserving scheme for infrastructure-less vehicular networks.....	2683
Xiangyang LUO, Xingxing CHEN, Yuqian MA, Qingfeng CHENG & Xiaofeng CHEN	
Differentially private federated unlearning mechanism based on update residuals	2704
Teng WANG, Lindong ZHAI, Yong YU, Tengfei YANG, Xuefeng ZHANG & Xuebin REN	
Federated meta unlearning based on model decomposition and weighted aggregation.....	2722
Yajie WANG, Qing FAN, Zijie PAN, Zuobin YING, Zijian ZHANG & Liehuang ZHU	
Game theory-driven multi-level perturbation ensemble adversarial attacks	2741
Shuhong CHEN, Mengmeng TANG, Hanjun LI, Zhili ZHOU, Guojun WANG & Tian WANG	
DMS-MIA: disk replay-based and multi-metric sequence membership inference attack on TEE-protected machine learning models.....	2759
Yifan DONG, Wei FENG, Hao LI, Min ZHANG, Yu QIN & Dengguo FENG	
Black-box attack method based on improved knowledge distillation	2780
Jiale SHI, Yafei SONG, Xiaobai WU & Tianpeng LI	
Trigger-free and data-free backdoor attacks on deep neural networks	2798
Jiahao WANG, Xianglong ZHANG, Huanle ZHANG, Xiaobo MA, Xiuzhen CHENG, Pengfei HU & Guoming ZHANG	
A lightweight function-hiding IPE scheme with support for dual outsourcing.....	2817
Xinyuan QIAN, Shuai YUAN, Guowen XU & Hongwei LI	
Bandage: bidirectional and nested defense algorithm for graph adversarial attack.....	2834
Shenghao LIU, Ke XU, Jing WANG, Xianjun DENG, Yuanyuan HE & Tianruo YANG	
A training-inference collaborative defense framework against membership inference attacks for graph neural networks.....	2852
Fuqi QI, Xingyi QUAN, Haichang GAO & Ping WANG	
PDNN-PDE: privacy data detection for the edge IoT based on P-DNN	2870
Shaopeng ZHOU, Bin WANG, Chaohao LI, Zhengyan ZHOU, Feng ZHANG & Chunming WU	
From compile to decompile: efficient watermark removal via source-to-source transformations	2886
Hao HUANG, Ruihua ZHOU, Jiatang LUO, Xiaohuan MA, Yunpeng LI & Yuling LIU	
A detection method against adversarial example attack based on neural network decision boundary and consistency analysis	2902
Ying XIE, Xiaosong ZHANG, Xuyang DING & Ruixia LANG	
M ³ -SafetyBench: a comprehensive benchmark for evaluating the safety of large language models across multiple domains, scenarios, and dimensions.....	2923
Weiping YANG, Haojie CHENG, Baishun ZHOU, Yuxin WEN, Yufan LIU, Yuyang LIU, Bing LI, Congyan LANG, Naiyue CHEN, Wei ZHANG & Weiming HU	