前　言

孟国柱[1], 陈　恺[1], 卜　磊[2], 蒲戈光[3]

[1]中国科学院信息工程研究所 北京 中国 100093
[2]南京大学 软件学院 南京 中国 210023
[3]华东师范大学 软件工程学院 上海 中国 200062

随着人工智能技术在社会、经济和生活领域的不断渗透和应用, 人工智能的安全问题也得到研究人员的广泛关注。以深度学习为代表的人工智能技术存在鲁棒性、模型后门、公平性和隐私等问题, 并且由于神经网络模型的高复杂性和难解释性, 导致这些安全隐患无法得到有效的检测和防御。特别在航空航天、智慧医疗、无人驾驶等安全攸关领域对人工智能的可信、可靠和可解释上提出更高的要求, 因此如何保障人工智能的安全成为国内外研究的趋势及热点。

本期“人工智能安全”专刊是 2022 年中国软件大会(ChinaSoft)和《信息安全学报》共同举办, 主题涵盖人工智能系统的攻防技术、自动化验证和测试、计算框架安全分析与检测、安全标准与评估方法、隐私保护技术、可解释性理论和方法、公平性和社会伦理研究以及人工智能在软件工程和信息安全的应用。本次专刊相继通过中国软件大会和《信息安全学报》的评审, 通过的稿件在 2022 年 11 月 25 日举办的中国软件大会进行论文报告, 最后根据评审意见完成论文的最终修订。经过本专刊审稿专家的专业和及时的评阅, 最终从众多稿件中遴选出 9 篇优秀稿件, 其中包括 1 篇综述性文章和 8 篇技术性文章。

在人工智能安全方向, 本专刊共收录 5 篇文章, 内容涵盖针对神经网络的对抗样本防御、鲁棒性验证、缺陷修复和隐私保护。其中: 《基于特征分布差异的对抗样本检测》探讨了神经网络模型面对对抗样本攻击时的应对措施, 提出了一种利用特征分布进行对抗样本检测的框架, 其中包括广义对抗样本检测和条件对抗样本检测方法。《基于健壮半径求解的循环神经网络形式化验证方法》介绍了一种基于健壮半径求解的循环神经网络形式化验证方法(VR-RRS), 通过逐层回溯迭代的方式得到循环神经网络各层神经元近似区间上下界关于输入的线性表达式, 利用赫尔德不等式推导出各层神经元的近似上下界关于扰动半径的解析解, 并采用改进的二分法对健壮半径进行求解, 最终得到了更为精确的近似区间和验证成功率。《基于分治法的神经网络修复方法》提出一种基于分治法的神经网络修复方法, 通过不断划分目标样本集合并逐个修复得到局部补丁, 最后整合得到对整个特征空间的补丁。《基于 BFV 同态加密神经网络参数设置实证研究》研究了基于 BFV 同态加密方案的隐私安全神经网络在不同参数设置下的影响, 测定了同态加密中多项式模数与神经网络模型的预测准确度、时间复杂度和空间复杂度的关联关系, 以及同态加密过程中不同神经网络层的耗时分析。《保护用户数量信息的安全虹膜识别方案》针对虹膜识别系统中合法用户数量这一隐私, 通过每个用户的虹膜特征随机选择和系统参数决定用户的注册模板数量, 有效抵御攻击者对用户数量的推断攻击, 提升现有安全虹膜识别方案的隐私性。

在人工智能辅助安全的方向, 共收录 4 篇文章, 其中包括跨语言自动生成、同源性分析和漏洞检测。其中: 《基于深度学习的跨自然语言与程序语言生成任务综述》系统地梳理了跨自然语言和程序语言生成技术的研究进展, 设计了一个基于深度学习的跨自然语言和程序语言的通用实现模型, 并从程序代码表示方法、网络模型结构、安全问题、常用数据集和模型效果等方面对已有的研究成果进行了分析和总结, 探讨了该领域未来的发展方向。《基于模型驱动的分治并行函数式程序生成及自动验证》提出一种基于模型驱动的分治并行函数式程序生成及自动验证方法, 融合形式化方法, 从问题描述出发得到功能规约, 利用分划递推法和循环不变式开发新策略推导出用 Radl 语言描述的串行算法, 然后通过验证框架验证算法连接函数满足同态定理, 最后转换为 Haskell 并行函数式可执行程序, 解决并行程序生成过程中缺乏解释性、易错和低可信等问题。《IoT 固件同源性智能检测研究》为解决 IoT 程序开发过程中容易引入存在缺陷的第三方库问题, 利用同源性分析技术挖掘程序间的关联关系, 并实现漏洞的智

能溯源。文章介绍了两种数据来源和涉及的特征选择、表示和检测方法, 并对方案的特点、局限性以及在不同类型 IoT 设备程序的适配性进行了比较和总结。《基于联邦学习的第三方库流量识别》针对移动应用开发过程中来自第三方库的安全风险, 提出了一种用于第三方库流量识别的框架——LibCapture。通过利用动态插桩与第三方库检测技术自动生成加密流量数据集, 采用基于卷积神经网络的联邦学习模型识别 TPL 流量, 在对 2327 个真实应用的流量测试中, 取得较高的识别准确率。

作为本次专刊的特邀编委, 我们首先感谢《信息安全学报》编委会和 2022 年中国软件大会对本期专刊工作的大力支持和指导, 感谢编辑部的各位同事不辞辛苦, 从论文征稿、审稿、定稿和出版过程中付出的辛勤工作, 非常感谢本次专刊的审稿专家, 占用了宝贵的时间为本次投稿的论文提供了专业和及时的评审意见, 保证每一篇收录文章的学术水准。我们还要感谢向本次专刊投稿的各位作者, 感谢你们在各自研究领域内深入钻研、默默付出, 为国家的科技事业贡献自己的力量。最后, 衷心地希望本次专刊的文章能够给读者朋友们带来一些思想的启迪和研究的帮助, 祝大家身体健康、工作顺利。