

# 大模型下的软件质量保障专题前言<sup>\*</sup>

王 赞<sup>1</sup>, 王 莹<sup>2</sup>, 陈碧欢<sup>3</sup>, 姚 远<sup>4</sup>, 张敏灵<sup>5</sup>

<sup>1</sup>(天津大学 智能与计算学部, 天津 300350)

<sup>2</sup>(东北大学 软件学院, 辽宁 沈阳 110169)

<sup>3</sup>(复旦大学 计算机科学技术学院, 上海 200438)

<sup>4</sup>(南京大学 计算机学院, 江苏 南京 210023)

<sup>5</sup>(东南大学 计算机科学与工程学院, 江苏 南京 211189)

通信作者: 王赞, E-mail: [wangzan@tju.edu.cn](mailto:wangzan@tju.edu.cn)



中文引用格式: 王赞, 王莹, 陈碧欢, 姚远, 张敏灵. 大模型下的软件质量保障专题前言. 软件学报, 2025, 36(6): 2401–2403. <http://www.jos.org.cn/1000-9825/7331.htm>

近年来,随着深度学习技术的飞速发展,基于大规模预训练的语言模型 (large language model, LLM) 在自然语言处理领域取得了前所未有的突破,并且在代码生成、软件测试、程序分析与优化等软件工程任务中展现出强大的潜力与广泛的应用前景.作为具备深度语义理解与生成能力的通用人工智能模型,大语言模型的崛起正在重新定义传统的软件质量保障方法,为软件工程的自动化和智能化注入了新的活力.同时,软件质量保障作为确保软件系统在功能性、性能、可靠性和安全性等方面满足设计规范和用户期望的关键环节,其传统方法难以适应现代软件系统日益增长的规模与复杂性.特别是在开源软件、企业级应用和芯片设计程序等高度复杂的应用场景中,传统方法的自动化与智能化程度难以满足实际需求.

在这一背景下,大语言模型赋能的软件质量保障研究逐渐成为学术界与工业界备受瞩目的研究方向.大语言模型不仅具备强大的代码理解与生成能力,且在软件测试、代码优化、代码翻译等关键任务中取得显著效果.其大规模预训练范式和高扩展性,使其成为构建智能化软件质量保障工具的重要技术支撑.然而,如何有效地将大语言模型与现有的软件工程实践深度融合,构建面向实际应用的智能化系统,依然存在诸多挑战.

基于上述背景,本专题聚焦于“大模型赋能的软件质量保障 (LLM4QA)”领域,旨在探讨大语言模型在软件质量保障中的前沿研究与最新进展,关注大模型应用的技术原理、系统实现与实验评估,展示多样化的研究成果与创新方案,探索大模型在促进软件开发和维护过程中的潜力与前景.本专题特别关注以下几个关键研究方向:大语言模型在软件测试技术、代码优化与重构技术、代码翻译技术、代码生成中的许可证合规风险评估,以及多智能体协作等任务中的应用.希望通过本专题的研究成果,促进大语言模型与软件质量保障领域的深度融合,推动理论与实践的协同创新,为软件工程的自动化与智能化发展提供参考与借鉴.

本专题公开征文,共收到投稿 24 篇.所有稿件均通过形式审查,涵盖大语言模型在软件质量保障中的应用与实践等方面.经过特邀编辑的初审与复审、CCF ChinaSoft 2024 会议宣读以及终审环节,历时 6 个月,共有 8 篇高质量论文入选本专题.这些论文从多个视角探讨了大语言模型在软件质量保障中的应用,主要内容分为以下 5 个主题.

## (1) 大模型赋能的软件测试方法

《[基于大语言模型的模糊测试研究综述](#)》对现有的基于大语言模型的模糊测试技术的研究发展现状进行了全面的分析和总结,概述了模糊测试的整体流程和模糊测试研究中常用的大语言模型相关技术,讨论了大模型时代之前的基于深度学习的模糊测试方法的局限性,分析了大语言模型在模糊测试方法中不同环节的应用方式,探

\* 收稿时间: 2024-12-10; jos 在线出版时间: 2024-12-10

讨了大语言模型技术在模糊测试中的主要挑战和今后可能的发展方向.

《Java 程序资源泄露缺陷检测: 传统模型和语言模型的有效性分析》研究了基于深度学习、基于小模型和基于大模型的检测方法在资源泄露检测任务中的有效性, 并探究了提示词增强、微调以及静态分析与大模型结合的多种改进方式, 并从资源泄露根本原因、资源种类、代码复杂度等多个维度分析了不同模型的表现, 通过实验证明了微调技术能够显著提升大模型在资源泄露检测领域的检测效果, 不同模型在识别常规路径下未关闭资源的缺陷的能力, 以及代码复杂度对基于深度学习的检测方法的影响.

《智能化芯片设计程序测试研究综述》从测试输入生成、测试预言构造及测试执行优化这 3 个角度对芯片设计程序智能化测试已有成果进行了系统总结, 重点关注芯片设计程序测试方法从机器学习阶段、深度学习阶段到大语言模型阶段的演化, 探讨了不同阶段方法在提高测试效率和覆盖率、降低测试成本等方面的潜力, 整理了芯片设计程序测试领域的研究数据集和工具, 并展望了大模型下的芯片设计程序质量保障未来的发展方向和挑战.

#### (2) 大模型赋能的代码分析与优化

《大语言模型在代码优化任务中的能力探究及改进方法》探讨了大语言模型和传统小规模预训练模型在自动代码优化任务的表现, 评估了大语言模型在该任务上的综合优势, 提出了基于 Unidiff 的代码优化评估指标, 对优化过程中的变更操作进行量化, 并提出了 LLM-Voter 方法集成不同模型的优势以提升代码优化质量.

《基于大语言模型的长方法分解》提出了一种新的自动分解长方法的模型 Lsplitter, 其通过大语言模型与启发式规则相结合, 在实际 Java 项目中实现了显著的优化效果.

#### (3) 大模型赋能的代码翻译技术

《LLM 赋能的 Datalog 代码翻译技术及增量程序分析框架》提出了一个基于大语言模型的 Datalog 代码翻译框架, 通过分治翻译策略、基于少样本提示和思维链提示的提示工程、基于检查-反馈-修复的迭代纠错机制, 在不同 Datalog 方言之间实现高精度代码翻译, 实现了高精度代码转换和增量程序分析, 加速了跨方言程序分析任务.

#### (4) 大模型下的开源许可证分析

《大模型生成代码的开源许可证违规风险洞察与分析》聚焦于大模型生成代码与开源仓库的许可证违规风险, 研发了一种代码克隆检测框架, 揭示了大模型生成代码的溯源与版权违规问题, 并分析了代码克隆与许可证缺失的实际比例, 探讨了大模型代码生成对开源软件生态的影响.

#### (5) 大模型赋能的多智能体协作系统

《基于大语言模型的多智能体协作代码评审人推荐》提出了一种多智能体协作框架, 基于代码变更和评审人画像, 推荐合适的代码评审人, 显著提高了推荐性能与可解释性.

本专题主要面向系统软件、软件工程、人工智能等领域的研究人员和工程人员, 展示了我国学者在大语言模型赋能的软件质量保障领域的深度探索和前沿进展, 涵盖了软件测试、代码优化、代码翻译、开源合规分析与多智能体协作等关键方向. 感谢《软件学报》编委会、系统软件专委会和软件工程专委会对专题工作的指导和帮助, 感谢专题全体评审专家及时、耐心、细致的评审工作, 感谢踊跃投稿的所有作者. 希望本专题的研究成果能够启发更多学术界与工业界的研究者, 推动大模型在软件质量保障领域的持续发展与技术进步.



王赞(1979—),男,博士,天津大学教授,博士生导师,CCF 专业会员.长期致力于人工智能质量保障、软件工程相关领域的研究,主要研究领域为深度学习系统质量保障,基础软件测试等.近5年在软件工程领域内的国内外高水平会议及期刊(包括 ICSE、FSE、ASE、TSE、TOSEM 及计算机学报、软件学报等)发表学术论文 20 余篇,其中 FSE2020 的文章“Deep Learning Library Testing via Effective Model Generation”获得 ACM SIGSoft Distinguished Paper Award.近5年作为负责人承担包括国家自然科学基金项目在内的多项课题.以第一完成人身份荣获天津市科技进步二等奖一项.主讲包括数据库原理、软件测试等课程,获得天津市教学成果二等奖两项.



王莹(1987—),女,博士,东北大学副教授,博士生导师,CCF 专业会员,CCF 开源发展委员会委员,CCF 女工委委员.荣获微软研究院铸星计划访问学者(2020)、中国计算机学会优秀博士学位论文提名奖(2020)、辽宁省优秀博士学位论文奖(2021)、ACM SIGSOFT 杰出论文奖(ICSE 2021、ESEC/FSE 2023).主要研究领域为智能软件开发技术、开源软件生态治理技术、软件供应链分析等.在多种程序语言软件生态(包括 Java/C#/Python/Go/JavaScript/Android/Rust 等)治理方面发表系列学术成果,形成系列工具平台“英雄联盟” LOL 自动化监控开源软件生态的依赖缺陷.多项技术落地于华为、微软、龙芯等企业平台和 OpenHarmony 开源社区.担任 IEEE Transactions on Software Engineering 期刊编委,ChinaSoft 2023–2024 软件工程女性论坛主席,SANER 2023 Tool Track 联合主席,“计算之美”2021 博士生论坛主席等,热衷参与学术活动鼓励计算机领域女性“研究媛”和“程序媛”坚持科研梦想.



陈碧欢(1986—),男,博士,复旦大学副教授,博士生导师,CCF 高级会员.主要研究领域包括软件供应链、智能网联汽车、AI 系统工程等.主持两项国家自然科学基金项目和多项企业合作项目,参加科技创新 2030—“新一代人工智能”重大项目.研究成果发表在 ICSE、FSE、S&P、SEC、TSE、TIFS 等国际会议和期刊,获 NASAC 青年软件创新奖、3 次 ACM SIGSOFT 杰出论文奖(FSE2016、ASE2018、ASE2022)、2 次 IEEE TCSE 杰出论文奖(ICSME2020、SANER2023).基于相关研究成果,研制了开源风险治理平台伏羲(<http://www.se.fudan.edu.cn/fuxi/>).



姚远(1987—),男,博士,南京大学副教授,博士生导师,CCF 专业会员.近年来研究兴趣主要关注机器学习技术及其在软件智能化等领域的应用.其研究工作获得了一系列国家重大、重点项目支撑,并在华为等企业转化落地.相关研究共发表了论文 60 余篇,包括 ICSE、CCS、S&P、NDSS、ICLR、NeurIPS、KDD 等相关领域的国际旗舰会议,工作受到了计算机业内权威媒体 MIT Technology Review 报道.



张敏灵(1979—),男,博士,东南大学教授,博士生导师,计算机科学与工程学院院长,CCF 杰出会员.主要研究领域为机器学习、数据挖掘.现任中国人工智能学会机器学习专委会副主任、江苏省人工智能学会副理事长等.现任《软件学报》《中国科学:信息科学》《IEEE Transactions on Pattern Analysis and Machine Intelligence》《ACM Transactions on Intelligent Systems and Technology》《Frontiers of Computer Science》《Machine Intelligence Research》等期刊编委.应邀担任 ACML、PAKDD 指导委员会委员,PRICAI/CCF-ICAI/CCFAI 等国内外学术会议程序主席,KDD/IJCAI/AAAI/ICDM 等国际会议领域主席或资深程序委员 60 余次.曾获 CCF-IEEE CS 青年科学家奖(2016)、国家杰出青年科学基金(2022)等.