

大模型与软件工程专题前言^{*}

聂长海¹, 王璐², 王莹³, 姜艳杰⁴, 张敏灵⁵

¹(南京大学 计算机学院, 江苏 南京 210023)

²(西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071)

³(东北大学 软件学院, 辽宁 沈阳 110169)

⁴(天津大学 软件学院, 天津 300072)

⁵(东南大学 计算机科学与工程学院, 江苏 南京 211102)

通信作者: 聂长海, E-mail: changhainie@nju.edu.cn



中文引用格式: 聂长海, 王璐, 王莹, 姜艳杰, 张敏灵. 大模型与软件工程专题前言. 软件学报, 2026, 37(8): 3049–3051. <http://www.jos.org.cn/1000-9825/7599.htm>

我们正在进入智能化时代, 人工智能 (artificial intelligence, AI) 已取得巨大成功, 并被广泛应用于学术和工业领域, 特别是大模型 (large language model, LLM) 取得了突飞猛进的发展, 将人工智能推向了一个新的高度, 使更多不同的应用和工业领域获得智能, 并增强了它们的能力, 例如人与机器的交互更加自然高效, 各行业出现新的服务和业务模式, 科学研究也得到加速. 大模型背后是超大的参数规模、海量的训练数据和强大的计算资源, 软件能力提升的同时其质量保障也越来越困难. 相较于传统软件工程, 大模型时代的软件工程需要面临更多的挑战: 大模型支撑系统与框架的潜在缺陷; 不断变化的基础设施导致系统演化困难; 大模型行为本身缺乏可解释性与可信性; 数据质量问题诱发的模型质量问题; 伦理、道德和法律的约束使得软件应用困难. 在这种局面下, 亟需从软件工程的角度系统性地审视大模型时代下的软件质量保障方法, 从而更好地应对这一新兴领域中的新挑战. 本专题立足于软件工程领域, 探讨大模型时代下软件工程技术的新发展与新趋势, 特别是基于大模型的需求分析、设计和如何利用大模型技术进行自动代码生成、测试和形式化验证等, 同时研究如何利用软件工程理论和方法保证基于大模型的软件系统的质量和安全.

本专题公开征文, 共收到投稿 51 篇 (其中 9 篇直接在软件学报网站上投稿, 42 篇通过 ChinaSoft 2025 会议系统投稿). 论文均通过了形式审查, 内容涉及大模型赋能的软件需求分析、设计、编码、测试与维护等软件全生命周期. 特约编辑先后邀请了 60 多位专家参与审稿工作, 每篇投稿至少邀请 3 位专家进行评审. 稿件经初审、复审、ChinaSoft 2025 会议宣读和终审这 4 个阶段, 历时 6 个月, 最终有 7 篇论文入选本专题.

《SmartGen-AADL: 多智能体系统需求分析与 AADL 模型生成》提出一种面向嵌入式系统的智能建模方法——SmartGen-AADL, 整体方法基于多智能体协同机制构建, 融合语义解析, 结构识别与提示增强生成等关键技术, 实现从自然语言需求到结构化 AADL 模型的高质量转换. 在 15 个嵌入式系统应用场景上的实验结果表明, 相较于仅依赖简单提示工程的方法, 所提多智能体协同建模方法在 4 个主流大语言模型上均展现出显著优势.

《基于大语言模型智能体的代码生成综述》对基于大语言模型的代码生成智能体领域进行了系统性综述, 追溯了该技术自诞生以来的发展脉络, 全面梳理并从方法论的视角对其核心技术 (涵盖单智能体与多智能体系统) 进行归纳和分类. 此外, 还总结了代码生成智能体在软件开发全周期中的各项应用, 整理了主流的评估基准与指标, 并盘点了代表性的工具, 通过分析关键挑战, 展望该领域未来的长期核心研究方向.

《面向 Linux 内核开发知识的大模型问答能力评测》提出一种 LLM 问答能力评测基准数据集构建方法, 构建了面向 Linux 内核的首个高质量问答评测基准 LKQABench, 并提出多裁判协同的代码知识问答评测方法 MJ-CCE.

* 收稿时间: 2025-12-23; jos 在线出版时间: 2025-12-24

LKQABench 基于开发者社区的真实技术问答数据, 经过语义分析和人工审核修订, 构建了 202 个标准问答对, 覆盖 Linux 内核主要模块和不同认知维度. MJ-CCE 方法定义多个裁判大模型的协同评分与投票机制, 从关键知识点覆盖度、事实正确性与表达清晰度等维度对回答进行多维度评估. 在 LKQABench 上对主流大模型的实证研究表明, 当前大模型能较好地回答内核实现的单点知识问题, 但在涉及跨主题知识整合、深度推理及版本演化关联的问题中, 存在知识点遗漏、逻辑链条不完整等不足.

《面向软件测试领域知识问答的大模型评估》对软件测试与 LLM 之间双向评估进行了有趣和深入的探索: 一方面针对软件测试领域中的 100 个概念和方法, 设计了 7 个问题, 利用最有代表性的 5 个大模型应用进行系统的问答实验, 产生了一系列对软件测试教学、教材编写和修订有价值的结果, 即利用大模型服务软件测试. 另一方面利用大模型产生的结果, 根据正确性、完整性、流畅性和创新性这 4 个度量指标, 针对 LLM 的可用性、可信性和鲁棒性等重要特性, 通过评审员进行人工评估, 发现了一系列有价值的问题, 即大语言模型的测评.

《基于并行探索的大模型缺陷定位增强》针对现有方法仅为大模型维护单一决策路径, 导致搜索范围有限, 缺陷定位效果不够理想等问题, 提出一种基于并行探索的大模型缺陷定位增强方法 PRIME. 通过设计缺陷位置的并行探索机制提高大模型的搜索范围, 并结合节点重要性评估方法对大模型预测的多个候选缺陷位置进行排序, 形成优化后的缺陷定位结果. 通过与其他缺陷定位方法的对比分析、全面的消融实验和参数影响分析, 验证了该方法可以有效增强大语言模型的缺陷定位性能.

《大语言模型智能体在软件系统根因分析中的应用综述》聚焦大语言模型智能体在云计算和微服务系统根因分析的研究现状, 进行全面的分析和总结. 概述基于大语言模型的智能体框架构成和根因分析中的数据使用, 从信息获取、根因定位和效果评估这 3 个主要流程, 系统地分析大语言模型智能体在根因分析中的应用方式, 探讨大语言模型智能体技术在根因分析任务中面临的主要挑战和未来发展方向.

《基于文本表达特征分析的大语言模型协议交互抽取》提出一种增强的基于文本表达特征分析的大语言模型协议交互抽取方法. 从真实的协议描述案例出发, 总结协议描述文本中存在的常见语言表达特征; 然后提取能够体现这些特征的典型协议描述案例, 提炼用于协议交互抽取的处理规则; 进一步地, 融合案例与规则, 提出一套规则回溯思维链方法, 使用多路推理和自我验证技术优化任务处理流程. 在多个协议数据集上的实验结果表明, 所提方法在协议交互的抽取精确率和召回率等方面均优于基线方法, 证实了方法的有效性.

本专题主要面向软件工程、系统软件、形式化、人工智能和大模型等多领域的研究人员和工程人员, 反映了我国学者在大模型、人工智能和软件工程等领域相互赋能最新的研究进展. 感谢《软件学报》编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. 希望本专题能够对与大模型相互赋能的软件工程、系统软件与形式化等相关领域的研究工作有所促进.

作者简介

聂长海, 博士, 南京大学计算机系教授, 博士生导师, CCF 杰出会员. 在近 30 年的教学工作中, 多次主讲数学类和计算机类课程 10 多门, 基本上覆盖了数学与计算机科学的所有主干课程. 主要研究领域为软件测试及软件质量保证. 发表及合作发表学术论文 100 多篇, 出版专著《组合测试》, 教材《软件测试的概念与方法》《智能化软件质量保证的概念与方法》《区块链技术基础教程: 原理、方法及实践》, 出版人工智能等新技术应用译著《应用数据分析原理与应用》《创建虚拟原型软件项目: 方法与实践》. 主持国家 863 项目、科技部重点研发项目课题和多项国家自然科学基金项目, 参与的项目多次作为重要成员获得省部级科研奖.

王璐, 博士, 西安电子科技大学副教授, 博士生导师, CCF 高级会员. 西安市智能软件工程重点实验室副主任, 中国计算机学会 (CCF) 软件工程专委会副秘书长, 系统软件专委会执行委员, CCF 首批传播大使, 西安电子科技大学“华山学者”, 西安电子科技大学 OpenHarmony 技术俱乐部主任. 主要研究领域为云操作系统, 微服务与容器技术, 大模型与智能软件工程. 主持国家自然科学基金面上项目等, 入选陕西省青年人才托举工程、陕西省优秀博士学位论文, 在 ICSE、FSE、SANER、ICSME 等国内外期刊与会议上发表论文 37 篇并担任会议 PC 及期刊审稿人, 获陕西省科学技术进步奖二等奖等科技奖励 4 项.

王莹, 博士, 东北大学软件学院教授, 博士生导师, CCF 专业会员. 主要研究领域为开源软件供应链安全, 鸿蒙软件智能迁移适配. 担任 CCF 系统软件专业委员会执行委员、CCF 软件工程专业委员会执行委员、CCF 开源发展技术委员会执行委员. 担任东北大学开源鸿蒙技术俱乐部主任. 中国计算机学会优秀博士论文提名奖、辽宁省优秀博士论文奖、微软亚洲研究院铸星计划、ACM SIGSOFT 杰出论文奖 (ICSE 2021、ESE/FSE2023) 获得者. 软件工程领域期刊《IEEE Transactions on Software Engineering》编委、ChinaSoft 2023–2025 软

件工程女性学者论坛主席. 开源软件生态“依赖地域”问题、跨语言安全漏洞传播影响分析和开源鸿蒙软件供应链智能移植适配解决方案的最早提出者之一. 多项软件缺陷检测与测试方面技术在微软、华为、龙芯中科、微众银行等企业落地应用.

姜艳杰, 博士, 天津大学副教授, 博士生导师, CCF 专业会员. 主要研究领域为人工智能 (大语言模型), 数据挖掘, 智能软件开发, 软件安全, 隐私保护. 在 CCF A 类顶级期刊与会议上发表论文 30 余篇, 受邀担任 ASE、FSE、ICSE、ISSTA、TSE、TOSEM 等 CCF A 类期刊/会议的审稿人. 获得 CCF 科学技术奖、ACM SIGSOFT 杰出论文奖、中国软件大学原型竞赛一等奖、百度学术新星奖、“博士后创新人才”. 主持多项科研项目, 包括华为胡杨林基金、国家自然科学基金 (青年基金) 等.

张敏灵, 博士, 东南大学计算机科学与工程学院教授、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 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, 以及 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.