

同行专家业内评价意见书编号：20250854480

## 附件1

# 浙江工程师学院（浙江大学工程师学院） 同行专家业内评价意见书

姓名：                                王冠淇

学号：                                22260292

申报工程师职称专业类别（领域）：                电子信息

浙江工程师学院（浙江大学工程师学院）制

2025年05月30日

## 填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

## 一、个人申报

(一) 基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

### 1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

在长期参与大语言模型（LLM）异常行为分析与系统性调试过程中，我系统掌握了人工智能与软件工程交叉领域的核心理论与关键技术，具备扎实的专业基础。具体包括自然语言处理、深度学习模型结构（如Transformer架构）、大规模预训练模型的性能优化策略、分布式系统调度，以及软件缺陷检测与根因定位等方面的知识。在专业实践中，我聚焦于LLM模型在复杂输入条件下出现Bug的根因追溯与表现分析，结合因果推断、语义信息流追踪、多模态干预实验等方法，建立了一整套面向大模型的Bug实证分析技术框架。

我对模型异常的语义演化路径、Prompt设计对生成逻辑的影响、输入扰动对输出稳定性的干扰机制等问题具有深入理解，并能够结合可观测现象进行量化建模和机制解释。与此同时，我深入学习并应用了业界先进的调试工具链，如TraceLang、OpenLLM

Debugger、以及GitHub

Copilot调试日志分析系统，并探索其在工业场景中的实用性和通用性。此外，我关注开源社区中各类LLM系统的工程实现细节，包括模型推理流程优化、低资源场景下的鲁棒性对比分析等，持续拓展了我在模型工程和复杂系统调试方面的知识广度与深度。

综上，我已能够将基础理论与工程知识贯通应用于复杂AI系统中的实际问题，并在行业主导的真实项目中不断打磨和提升跨学科工程实践能力。

### 2. 工程实践的经历(不少于200字)

在本次专业实践中，我以核心成员身份参与了某开源大语言模型平台在生产环境下出现多次响应异常的根因排查与修复工作。面对模型出现无关回答、生成偏离预期、甚至潜在安全风险的情况，我首先构建了一套系统化的Bug归因分析流程，结合异常日志溯源、Prompt语义结构解析、以及输入输出分布漂移检测等技术手段，对多轮交互中的异常生成行为进行逐一定位与分析。

在此基础上，我与团队合作开发了一款轻量级的Bug注入与敏感性测试框架，可对模型在特定任务上的行为稳定性进行自动化评估，有效发现Prompt注入攻击、上下文干扰引发的错误响应等隐患，显著提升了模型的可解释性与工程可靠性。此外，我还主导了多个典型Bug案例的技术复盘，并撰写了调试文档与测试规范，推动所在实验室与企业联合团队构建起一套覆盖开发—测试—上线的LLM系统调试流程。

通过对实际问题的深入分析与跨团队协作，我不仅将理论知识有效应用于复杂工程问题的解决，也对LLM系统工程调试与鲁棒性评估建立了更加系统的认知，并具备了在企业级系统中推动技术落地的实战经验。

### 3. 在实际工作中综合运用所学知识解决复杂工程问题的案例(不少于1000字)

首先，我基于模型生成过程的特性，从输入端构建“语义保持扰动测试框架”，在保证语义不变的前提下，对提示词（Prompt）进行结构化扰动，形成多个逻辑等价但表达方式不同的输入集合。通过对这些扰动输入在模型输出端的响应差异性进行度量，我提出了“生成稳定性评分”指标，用于评估模型对输入变化的鲁棒性。这一机制在多个任务中成功发现模型存在对条件变化过度敏感、信息聚焦能力差等问题，尤其在多轮交互式对话和程序生成任务中更为明显。

其次，为进一步定位Bug出现的内部机制，我设计了结合注意力机制可视化、位置编码分布分析和中间层激活状态对比的多模态行为编码方法。通过将每一次模型生成过程中的注意力

矩阵、激活向量进行标准化表示，并使用聚类算法（如KMeans、HDBSCAN）进行行为聚合，我构建了“异常行为簇检测系统”。该系统能够高效地识别出模型在不同输入下的“异常反应群体”，例如在一些带有否定条件或嵌套逻辑的指令中，模型经常误解核心语义，导致回答方向严重偏差。

针对上述分析结果，我进一步借鉴程序静态分析与动态污点跟踪思想，在Transformer架构基础上开发了“语言模型信息流追踪模块”。该模块通过hook模型的多层注意力机制，追踪输入token对最终输出token的影响路径，重建输入到输出的语义传递链条，揭示模型生成过程中的“因果路径”。例如在代码生成任务中，模型在函数定义与返回语义不一致时，我通过信息流路径发现模型在处理“return语句”时对上文变量作用域的关注明显减弱，进而定位出模型未能有效捕捉前文逻辑结构是造成Bug的根本原因。

为了将上述分析流程工程化、系统化落地，我基于Huggingface Transformers框架和PyTorch环境，开发了“LLM异常行为分析工具链”。该工具支持对主流大模型（如LLaMA、BLOOM、ChatGLM等）进行Prompt自动生成、异常评分、行为路径可视化、日志提取与聚类分析等功能，并通过配置文件方式实现模型无缝切换，方便企业与实验室实际部署与调试使用。在一个具体项目中，该工具帮助企业成功定位并修复了一个长期存在的代码生成异常问题。该问题表现为模型生成函数时常在无返回值的情形下仍保留return None语句。通过分析生成日志和行为路径，我们发现模型对“无返回值需求”的指令识别存在盲区，在优化了Prompt结构并微调小样本后，Bug出现率下降超过70%。

此外，我还撰写了详细的Bug分析手册与测试规范，协助构建了从模型开发到上线的闭环调试流程，推动实验室与企业在大模型系统工程方向形成长期合作机制。整个项目过程中，我不仅要在大模型的底层实现有深入理解，还需熟练掌握自然语言处理、可解释性建模、异常检测、深度学习框架调试等技术，真正实现了理论知识与工程实践的融合。

**(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】**

**1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】**

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注
一种针对代码仓库的 README 文档自动生成方法	发明专利申请	2024年09 月14日	申请号: CN20241129 3030.6	3/4	

**2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】**

<b>(三) 在校期间课程、专业实践训练及学位论文相关情况</b>	
课程成绩情况	按课程学分核算的平均成绩： 83 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1 年（要求1年及以上） 考核成绩： 79 分
<b>本人承诺</b>	
<p>个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！</p> <p style="text-align: right;">申报人签名： </p>	



浙江大学研究生院  
攻读硕士学位研究生成绩表

学号: 22260292	姓名: 王冠淇	性别: 男	学院: 工程师学院	专业: 计算机技术	学制: 2.5年						
毕业时最低应获: 24.0学分		已获得: 28.0学分		入学年月: 2022-09	毕业年月:						
学位证书号:			毕业证书号:			授予学位:					
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2022-2023学年秋季学期	研究生英语		2.0	免修	公共学位课	2022-2023学年冬季学期	工程中的有限元方法		2.0	98	专业选修课
2022-2023学年秋季学期	工程技术创新前沿		1.5	92	专业学位课	2022-2023学年秋冬季学期	高阶工程认知实践		3.0	81	专业学位课
2022-2023学年秋季学期	数据科学技术与软件实现		2.0	88	专业学位课	2022-2023学年冬季学期	新时代中国特色社会主义思想理论与实践		2.0	89	公共学位课
2022-2023学年秋季学期	数值计算方法		2.0	89	专业选修课	2022-2023学年冬季学期	产业技术发展前沿		1.5	92	专业学位课
2022-2023学年秋季学期	研究生英语能力提升		1.0	免修	跨专业课	2022-2023学年春季学期	自然辩证法概论		1.0	77	公共学位课
2022-2023学年秋季学期	研究生英语基础技能		1.0	免修	公共学位课	2023-2024学年秋季学期	人工智能算法与系统		2.0	80	跨专业课
2022-2023学年秋冬季学期	工程伦理		2.0	63	公共学位课	2023-2024学年秋季学期	电子与信息工程技术管理		2.0	70	跨专业课
2022-2023学年秋冬季学期	研究生论文写作指导		1.0	84	专业学位课		硕士生读书报告		2.0	通过	

说明: 1. 研究生课程按三种方法计分: 百分制, 两级制 (通过、不通过), 五级制 (优、良、中、及格、不及格)。

2. 备注中 "\*" 表示重修课程。

学院成绩校核章:

成绩校核人: 张梦依

打印日期: 2025-06-03





# 国家知识产权局

310013

浙江省杭州市西湖区竞舟路1号筑品金座501室 杭州天勤知识产权代理有限公司  
王琛(0571-87755911)

发文日:

2024年09月14日



申请号: 202411293030.6

发文序号: 2024091401313640

## 专利申请受理通知书

根据专利法第28条及其实施细则第43条、第44条的规定,申请人提出的专利申请已由国家知识产权局受理。现将确定的申请号、申请日等信息通知如下:

申请号: 2024112930306

申请日: 2024年09月14日

申请人: 浙江大学

发明人: 邓水光,王冠淇,李国昌,韩俊晓

发明创造名称: 一种针对代码仓库的README文档自动生成方法  
经核实,国家知识产权局确认收到文件如下:

权利要求书 1份3页,权利要求项数: 9项

说明书 1份9页

说明书附图 1份1页

说明书摘要 1份1页

专利代理委托书 1份2页

发明专利请求书 1份5页

实质审查请求书 文件份数: 1份

申请方案卷号: 24124F1687

提示:

1.申请人收到专利申请受理通知书之后,认为其记载的内容与申请人所提交的相应内容不一致时,可以向国家知识产权局请求更正。

2.申请人收到专利申请受理通知书之后,再向国家知识产权局办理各种手续时,均应当准确、清晰地写明申请号。

审查员: 自动受理

联系电话: 010-62356655

审查部门: 初审及流程管理部



200101  
2023.03

纸件申请,回函请寄: 100088 北京市海淀区蓟门桥西土城路6号 国家知识产权局专利局受理处收  
电子申请,应当通过专利业务办理系统以电子文件形式提交相关文件。除另有规定外,以纸件等其他形式提交的文件视为未提交。