

# 基于大模型微调的安全巡检系统

陈 祥<sup>1</sup> 高吴涵<sup>1</sup>

CHEN Xiang GAO Wuhan

## 摘 要

针对安全巡检中隐患描述文本存在的规范性不足与效率瓶颈问题,文章提出了一种基于大语言模型 LoRA (low-rank adaptation) 微调的安全巡检智能描述系统。通过构建包含 2 万条标准化样本的专业数据集,确立了以“隐患位置、现象描述、标准关联、处置建议、隐患类型、隐患级别”六要素为核心的隐患描述框架,通过对基座模型 DeepSeek-R1-Distill-Qwen-7B 进行 LoRA 微调实现领域适配,并封装 RESTful API 集成至现有巡检系统,实现根据用户输入的关键词、短语或结构化报警信息,实时生成标准化的安全隐患描述文本。实验结果表明,该系统能够解决人工描述主观性强、自动化巡检语义生成质量低的问题,在要素召回率、Safety-ROUGE 指标、分类分级准确性、标准可追溯率及专家评分等方面均显著优于微调前模型。

## 关键词

大模型;指令微调;LoRA;安全巡检

doi: 10.3969/j.issn.1672-9528.2026.03.005

## 0 引言

工业生产的稳定运行以安全管理为根本保障,而定期安全巡检是预防事故的关键手段。当前主流巡检方式主要包括人工巡检与自动化巡检。人工巡检依赖人员经验,存在主观判断偏差风险;自动化巡检依赖智能机器人与传感器,虽能自动识别隐患,却在语义理解与自然语言生成层面存在明显不足。尤其在隐患信息处理、巡检报告生成等环节,如何将检测结果转化为精准、规范、符合行业标准的文本描述,成为制约巡检效能提升的关键瓶颈,由此可能导致隐患记录标准混乱、报告质量参差、后续分析困难,甚至引发管理疏漏<sup>[1]</sup>。因此,开发一种能够自动生成标准化隐患描述的技术,对提升安全管理智能化水平具有迫切现实意义。

规范的隐患描述文本应满足表述精确性、术语规范性和要素完备性三项基本要求,其核心内容应涵盖空间位置信息、风险量化评估及法规条款引用等结构化要素。实践表明,人工撰写的描述文本质量与巡检人员专业素养呈显著正相关,存在明显的个体差异性;而自动化设备输出的检测结果多为坐标边界或类别标签等机器可读数据,无法形成符合工程文档要求的自然语言表述。传统基于固定模板的方法难以适应多变的现场条件,TextCNN、BiLSTM 等早期神经网络虽然在文本分类任务中表现尚可,但无法满足专业领域高质量文

本生成的需求<sup>[2]</sup>。刘婷等<sup>[3]</sup>设计的字符级卷积神经网络虽然能够有效处理口语化表达和文字错误,提升低质量文本的评估效率,但其本质仍是对已有缺陷的被动修正,未能建立主动预防的文本生成机制。

本文研究并实现了一种基于大语言模型微调的安全巡检智能描述系统。通过构建包含 2 万条标准化样本的专业数据集,采用低秩适配 (low-rank adaptation, LoRA) 微调技术对 DeepSeek 开源模型进行轻量化领域适配,成功实现了模型在安全巡检场景的专业化提升。优化后的模型通过 RESTful API 与现有巡检平台深度整合,支持用户通过自然语言输入或结构化查询实时获取符合行业标准的隐患描述内容。该系统显著降低了生成高质量巡检报告的技术门槛与时间成本,有效提升了巡检数据的规范性、一致性与后续分析价值。

## 1 相关技术

### 1.1 传统巡检报告生成技术

随着新一代信息技术与智能装备的深度融合,安全巡检技术体系经历了从传统人工模式向智能感知模式的快速转型。在早期发展阶段,专业技术人员需要亲临现场开展隐患排查工作,凭借经验判断风险等级,并通过图片视频配合文字说明的方式完成信息采集。而在当前智能化阶段,巡检系统已演变为由自主移动机器人、分布式传感节点和空中无人机平台组成的立体监测网络,融合了多源传感数据采集、机器视觉识别和语音交互等智能化技术手段,构建起全天候的异常状态感知体系。

1. 江苏师范大学科文学院 江苏徐州 221132

[基金项目] 江苏省高校实验室研究会课题 (GS2024BZZ20); 江苏师范大学科文学院“人才项目”资助

在技术演进过程中，巡检报告生成技术也经历了基于模板的方法、基于规则引擎的方法以及早期自然语言处理（natural language processing, NLP）技术三个主要发展阶段。基于模板的报告生成技术采用预定义的 Word/Excel 模板框架，通过自动填充检测参数来生成标准化报告。该方法虽然确保了输出文档的格式规范，但在应对不同行业标准或新增检测指标时，往往需要重构整套模板结构，跨领域能力较差。基于规则引擎的技术引入了 Clips、Jess 等规则专家系统工具，运用条件判断规则实现基础决策支持。尽管这类方法在规则配置方面具备一定弹性，但其单维度的 if-then 推理机制难以处理多参数耦合的复杂工况，在跨领域迁移时也面临显著挑战。早期 NLP 技术利用统计语言模型和浅层神经网络改进报告生成，采用 LSTM+Attention 的编码器-解码器架构，初步实现了从结构化数据到自然语言描述的转换。但工程实践表明，这类方法生成的文本在语义连贯性和逻辑严谨性方面仍存在明显不足，制约了其在实际场景中的应用效果。

从迭代成本、跨领域能力、语义理解度三个维度对上述三类报告生成技术进行比较分析，结果如表 1 所示。

表 1 三类巡检报告生成技术对比

| 技术方案   | 迭代成本     | 跨领域能力   | 语义理解度 |
|--------|----------|---------|-------|
| 模板技术   | 高（需重构模板） | 无       | 无     |
| 规则引擎   | 高（规则维护）  | 部分行业可迁移 | 简单推理  |
| 早期 NLP | 中（数据标注）  | 需重新训练   | 浅层关联  |

1.2 大语言模型指令微调

大语言模型是基于 Transformer 架构的自回归深度学习系统，通过海量语料的预训练过程实现对语言统计规律的深度把握。相较于早期基于规则和统计的 NLP 技术，大模型凭借其独特的自注意力机制，能够建立全局上下文关联，在语义理解层面取得质的飞跃，在机器翻译、文本生成等任务上准确率大幅提升<sup>[4]</sup>。然而，当这类通用模型应用于特定专业领域时，由于预训练目标与专业场景需求之间的差异，模型虽然掌握了通用的语言生成模式，却难以针对专业领域的特点作出精准响应，表现出“广度有余而深度不足”的典型特征。这种专业适配性的缺失，严重制约了大模型在垂直领域的实际应用效果。

指令微调作为增强大模型可控性的关键技术，通过在预训练模型基础上，利用人工构建的“指令-输出”数据集进行有监督训练，重构模型的响应逻辑范式，推动大语言模型从通用问答向专业领域深度迁移。当前主流方案采用参数高效微调（parameter-efficient fine-tuning, PEFT）策略，其核心思想是保持基础模型参数不变，仅调整少量新增参数，其轻量化特征大幅降低了计算资源需求。其中，LoRA 技术因近

乎无损的推理效率被广泛采用。该技术的核心创新在于将权重更新矩阵分解为低秩乘积，仅需训练 0.1% 的参数量即可实现 90% 以上的全微调性能，极大提升了技术落地的经济性<sup>[5]</sup>。在医疗领域，Saluja 等<sup>[6]</sup>对 Path-llama3.1-8B 模型采用 LoRA 微调，在癌症类型识别任务中达到 96% 准确率、在 AJCC 分期任务中达到 85% 准确率，验证了轻量化适配的可行性；在软件工程领域，李嘉威等<sup>[7]</sup>通过微调开源大模型 Qwen2，支持用户输入自然语言需求，自动生成 PlantUML 代码并渲染 UML 类图，实现了中文需求到 UML 类图的端到端生成，精确率达 86.9%，展现出较强的工程实用性。

2 本文方法

2.1 数据集构建

高质量领域数据集是微调大型语言模型成为专业化“安全隐患描述专家”的基础。本项目采用多源异构数据融合策略，主体数据来源于某市应急管理局的日常巡检及行政处罚记录、企业 EHS 系统结构化巡检数据以及安全专家构建的标准化隐患模板库。为建立系统化的隐患描述框架，本研究通过三轮德尔菲专家论证，确立了隐患描述六要素体系，包含隐患位置、现象描述、标准关联、处置建议、隐患类型、隐患级别，如表 2 所示。

表 2 隐患描述六要素体系

| 要素分类 | 示例                       | 必要性 | 数据覆盖率 |
|------|--------------------------|-----|-------|
| 隐患位置 | 油罐区 T-103 西侧消防栓          | 必需  | 100%  |
| 现象描述 | 阀门严重锈蚀无法启闭，压力值 0 MPa     | 必需  | 100%  |
| 标准关联 | 《GB 50140—2005》第 3.2.4 条 | 必需  | 97.3% |
| 处置建议 | 除锈润滑并进行 1.6 MPa 水压试验     | 推荐  | 82.6% |
| 隐患类型 | 消防安全类隐患                  | 必需  | 100%  |
| 隐患级别 | 重大 / 较大 / 一般             | 必需  | 100%  |

为适配 LoRA 微调的数据结构要求，需将原始数据集重构为“指令-输出”配对格式<sup>[8]</sup>。该转换需解决两个关键问题，一是消除安全专家个体表述差异导致的语言异构性，生成符合安全隐患描述六要素体系的标准化输出；二是从原始数据中抽炼核心信息构建指令端数据，在保留现场人员非专业表述特征的同时，确保其与人工记录及检测系统输入风格的领域适配性。如图 1 所示数据集构建流程，上述目标通过双模板转换机制实现。

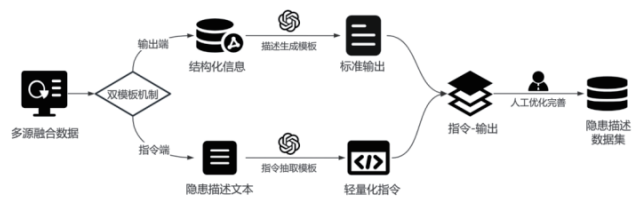


图 1 数据集构建流程

本研究借助大语言模型 ChatGPT-4 API 结合双模板转换机制快速实现原始数据的批量处理，所使用模板如表 3 所示。在输出端，利用描述生成模板，采用规则引导的文本生成框架，结合数据源中结构化数据，将原始描述重构为要素完整、表述精准、格式规范的标准输出；在指令端，利用指令抽取模板，从原始描述中提取位置、现象、处置等核心要素，在保留现场语言分布特征的同时降低信息冗余度，构建轻量化指令。最后，由安全工程师对生成的“指令-输出”进行专业修正，经过人工优化完善，得到规范化的隐患描述数据集。

表 3 提示模板

| 模板名    | 模板作用       | 模板示例                                                                                                                                                                                                                                                                                                        |
|--------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 描述生成模板 | 构建规范化的标准输出 | <p>将以下隐患描述按照模板转换为标准格式：<br/>模板：“[ 隐患位置 ][ 现象描述 ]，违反《[ 标准名称 ]》[ 条款 ]，需 [ 处置建议 ]。\\n 隐患类型：[ 类型 ]\\n 隐患级别：[ 级别 ]”</p> <p>示例输入：“油罐区 T-103 西侧消防栓阀门锈死，压力为 0，需尽快除锈”</p> <p>示例输出：“油罐区 T-103 西侧消防栓阀门严重锈蚀无法启闭，压力表示值 0 MPa，违反《GB 50140—2005》第 3.2.4 条，需在 48 h 内完成除锈润滑并进行 1.6 MPa 水压试验。\\n 隐患类型：消防安全类隐患 \\n 隐患级别：一般”</p> |
| 指令抽取模板 | 构建轻量化指令    | <p>从以下描述中提取隐患位置、现象描述、处置建议等关键信息：</p> <p>示例输入：“油罐区 T-103 西侧消防栓阀门锈死，压力为 0，需尽快除锈”</p> <p>示例输出：“隐患位置：油罐区 T-103 西侧消防栓 \\n 现象描述：阀门锈死，压力为 0 \\n 处置建议：尽快除锈”</p>                                                                                                                                                      |

## 2.2 LoRA 微调与应用部署

本研究选用的基座模型为 DeepSeek 团队研发的 DeepSeek-R1-Distill-Qwen-7B，该模型创新性地融合了强化学习与知识蒸馏的双重优化机制。通过引入 GRPO（grouped relative positional optimization）算法，显著提升了模型在有限训练数据条件下的推理性能。在模型压缩方面，采用渐进式知识迁移策略，将 671 B 参数教师模型的知识蒸馏至 7 B 参数规模，使计算资源消耗降低两个数量级。该架构在 Qwen 原始框架基础上进行了多项针对性改进，包括支持 Ollama 轻量级部署方案和 Unsloth 快速微调工具链，使其特别适合在边缘计算等资源受限环境中处理复杂的工程问题。

LoRA 微调是一种高效的大模型微调技术，其核心在于冻结预训练模型的原始权重，仅通过优化新增的低秩矩阵实

现下游任务适配，技术架构如图 2 所示。

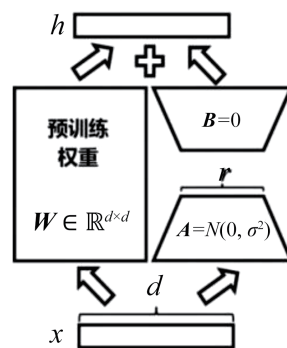


图 2 LoRA 微调技术

该方法的数学表征是将权重更新矩阵  $\Delta W$  分解为两个低秩矩阵的乘积，以参数高效的方式逼近全参数微调效果。

$$h = Wx + \Delta Wx = Wx + BAx \quad (1)$$

具体思路：将预训练大语言模型的权重矩阵  $W \in \mathbb{R}^{d \times d}$  冻结，引入可训练参数的低秩矩阵  $A \in \mathbb{R}^{r \times d}$  与  $B \in \mathbb{R}^{d \times r}$ ，其中  $r \ll d$  表示低秩维度，矩阵  $A$  服从高斯分布  $N(0, \sigma^2)$ ，矩阵  $B$  初始化为零矩阵。

训练完成后，计算  $\Delta W = BA$  构建增量矩阵，并与预训练权重矩阵  $W$  线性叠加，从而实现仅训练少量适配层参数即可完成大模型个性化微调。该算法在保证模型通用性能的同时，显著提升了训练效率，并通过低秩约束防止小数据场景下的过拟合现象，使模型在专业领域任务中仍能保持接近全量微调 95% 的性能表现<sup>[9]</sup>。

在工程化落地阶段，本研究采用 Unsloth 高效训练框架完成模型优化，该方案通过三重技术革新实现快速部署。一是运用 4 bit 量化压缩和显存优化算法，使得基座模型的低秩适配训练可在消费级 GPU 上完成；二是集成参数高效微调整接口与 FlashAttention-3 加速模块，整体训练速度提升 2 倍以上；三是采用模块化设计理念，将训练所得的低秩适配器独立封装，结合高性能 Web 框架构建微服务接口<sup>[10]</sup>。具体实施时，先通过 get\_peft\_model() 方法动态注入可训练参数，待模型收敛后将增量权重提取为独立组件，然后基于异步 IO 架构封装生成式推理服务，通过 RESTful API 提供标准化接口，满足工业场景的高吞吐量需求。

在应用交互层设计中，系统适配人工巡检及自动化巡检。针对人工巡检，通过结构化表单界面引导用户填写隐患位置、现象描述等必填信息及其他选填信息；针对自动化巡检，则从物联网异常报警流中自动提取设备 ID、传感器读数等结构化字段。两类输入经指令格式封装后，调用微调模型的 RESTful API 端点生成规范化描述文本，该文本适当结构化解析后展示给用户审核修订，系统完整流程如图 3 所示。

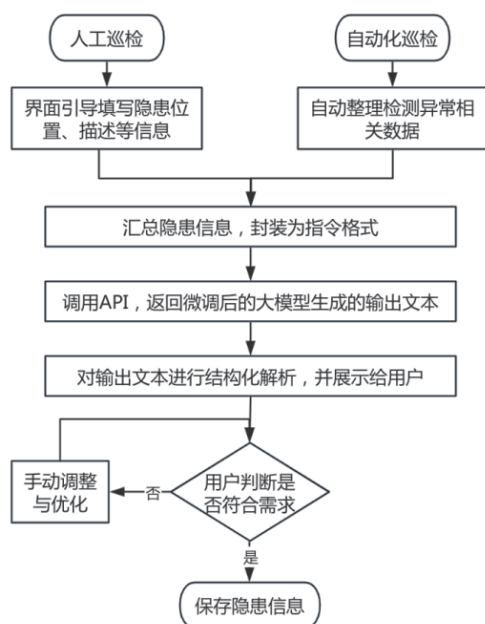


图3 系统流程图

### 3 实验对比及分析

本研究的实验数据采集自多个实际运行的工业巡检系统和EHS管理系统，经过严格筛选和清洗后，构建了含2 000条独立样本的测试集。该数据集采用分层抽样方法确保数据代表性，其中60%为关键词指令、25%为口语化指令、15%为结构化机器报警指令。为确保评估全面性，测试集特别设计包含了5%的罕见工况样本和10%的干扰样本，通过噪声数据的设计检验系统在实际复杂环境中的适应能力。

针对安全巡检文本生成的领域特性，本研究设计了多维评估指标。

(1) 要素完整性：设计要素召回率指标，评估描述文本是否完全覆盖表2中的隐患描述六要素，其计算公式为：

$$\text{Elements-Recall} = \frac{\sum_{i=1}^6 \mathbb{I}(e_i)}{6} \quad (2)$$

式中： $e_i$ 表示六要素的生成状态。

(2) 生成质量：采用改进的Safety-ROUGE指标，在传统ROUGE-L基础上引入要素权重 $\alpha$ 与术语惩罚项 $\beta$ ，加强关键安全要素的高权重匹配，并对专业术语误用施加惩罚。其计算公式为：

$$\text{Safety-ROUGE} = \alpha \cdot \sum_{(s,t) \in \text{LCS}} w(s) + \beta \cdot \left(1 - \frac{\sum_{k=1}^K \mathcal{S}(t_k, r_k)}{K}\right) \quad (3)$$

式中： $w(s)$ 为要素权重函数； $\mathcal{S}(t_k, r_k)$ 为术语相似度函数。实验中取 $\alpha=0.65$ ， $\beta=0.35$ ，实验时对常见安全要素进行了预定义。

(3) 分类分级准确性：通过 $F_1$ -score量化隐患类型、隐患级别的分类一致性，其计算公式为：

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

式中：Precision表示精确率；Recall表示召回率。

(4) 标准可追溯率：验证生成文本中的标准条款的真实性与时效性。

(5) 专家评分：采用人工评估方式，由安全工程师根据逻辑合理性、术语准确性、要素完整性、可操作性、合规性等几个维度综合打分，采用5分制Likert量表评分。

对大语言模型微调前后的性能表现进行对比分析，实验结果如表4所示。

表4 型微调前后对比实验结果

| 指标           | 初始模型 | 微调后模型 | 提升幅度/% |
|--------------|------|-------|--------|
| 要素召回率        | 4.8  | 5.7   | 18.75  |
| Safety-ROUGE | 0.56 | 0.88  | 57.14  |
| 分类分级 $F_1$   | 0.76 | 0.94  | 23.68  |
| 标准可追溯率       | 58%  | 97%   | 67.24  |
| 专家评分         | 3.2  | 4.6   | 43.75  |

实验结果表明，微调后模型在各项指标上均有显著提升。微调后要素召回率达5.7，证明LoRA技术成功注入了六要素生成能力；Safety-ROUGE从0.56提升至0.88，隐瞒描述文本的准确性显著改善，深层次的原因在于大幅降低了术语错误率；标准可追溯率跃升至97%，验证了领域知识注入的有效性；在专家评分环节，模型在常规场景表现优异，但仍受困于多因素耦合的复杂场景中的因果推理。这些数据验证了通用大语言模型在专业领域确实存在大量知识盲区，而LoRA微调能针对性填补领域知识缺口。值得注意的是，模型对极端工况的适应性不足，这一现象为未来研究提供了重要切入点，提示需要开发更强大的多模态推理机制。

### 4 结语

本文研究并实现了一种基于大语言模型LoRA微调的安全巡检智能描述系统，构建了包含2万条标准化样本的领域数据集，采用LoRA微调技术对DeepSeek-R1-Distill-Qwen-7B模型进行高效微调，成功将其转化为安全巡检领域的描述专家。本研究的创新性在于提出“六要素”标准化描述框架、设计结合要素完整性和术语准确性的多维评估体系，并实现轻量化部署方案，大幅降低工业落地门槛。然而，本系统在处理多因素耦合隐患的因果推理方面仍存在局限，对罕见工况的适应性也有待加强。针对这些局限，未来将重点构建知识图谱驱动的风险推演机制、动态标准体系的智能适配机制，并融合视觉传感器实现多模态联合生成，从而在复杂工业场景中提供更精准、鲁棒的安全决策支持，为智能安全管理体系提供核心技术支撑。

# 基于 AI 的地下建筑空间测绘与三维建模研究

程香丽<sup>1</sup>  
CHENG Xiangli

## 摘要

针对地下建筑空间结构复杂、传统测绘效率低的问题,文章提出了一种融合深度学习与多源数据融合的三维建模方法。通过改进点云配准算法与隐式神经表示模型,实现毫米级精度建模,并在某工程项目中验证方法的有效性。实验表明,模型平均误差较传统方法降低 62%,数据处理效率提升 3.8 倍,为地下空间数字化管理提供技术支持。该方法不仅解决传统测绘中数据缺失率高、误差累积严重的问题,还降低人工成本,为地下建筑的全生命周期管理提供了可靠的技术手段。

## 关键词

AI; 地下建筑空间; 工程测绘; 三维建模

doi: 10.3969/j.issn.1672-9528.2026.03.006

## 0 引言

随着城市化进程的加速,地下空间的开发利用已成为缓解城市用地紧张、优化城市功能布局的重要手段。然而地下建筑空间结构复杂、环境封闭,传统测绘技术在实际应用中面临诸多挑战。复杂的地下结构导致数据采集过程中缺失率高达 25% 以上,尤其是在管道密集区域和结构交错部位。设

备信号在地下环境中易受干扰,导致点云配准误差累积,严重影响建模精度。BIM 建模过程中人工成本占比超过 60%,且效率低下,难以满足大规模地下空间开发的需求。针对上述问题,研究基于实际工程需求,提出一种面向地下建筑的 AI 建模框架。该框架通过多传感器数据融合与自适应特征提取算法,建立可支撑运维管理的三维语义模型。本研究不仅具有理论创新性,还在实际工程中得到了验证,为地下空间的数字化管理提供了新的技术路径。

1. 运城职业技术大学建筑工程学院 山西运城 044000

## 参考文献:

- [1] MISHRA A K, AITHAL P S. Job safety analysis during tunnel construction[J]. International journal of applied engineering and management letters(IJAEML), 2021, 5(1): 80-96.
- [2] 程彭圣男, 刘雪梅, 李海瑞. 基于 BERT-BiLSTM 模型的输水工程巡检文本智能分类[J]. 中国农村水利水电, 2023(10): 150-155.
- [3] 刘婷, 张社荣, 李志斌, 等. 基于字符级 CNN 的调水工程巡检文本智能分类方法[J]. 水力发电学报, 2021, 40(6): 89-98.
- [4] 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.
- [5] HU E J, SHEN Y L, WALLIS P, et al. LoRA: low-rank adaptation of large language models[EB/OL]. (2021-10-16) [2026-01-22]. <https://doi.org/10.48550/arXiv.2106.09685>.
- [6] SALUJA R, ROSENTHAL J, ARTZI Y, et al. Cancer type, stage and prognosis assessment from pathology reports using LLMs[J/OL]. Scientific reports, 2025[2026-01-22]. <https://www.nature.com/articles/s41598-025-10709-4>.
- [7] 李嘉威, 邓媛丹, 陈波. 基于微调 Qwen2 自动构建领域 UML 模型[J]. 计算机科学, 2025, 52(S1): 31-34.
- [8] ZHANG S Y, DONG L F, LI X Y, et al. Instruction tuning for large language models: a survey[J]. ACM computing surveys, 2026, 58(7): 1-36.
- [9] 韩霄龙, 曾曦, 刘锐, 等. 基于 LoRA 高效微调通用语言大模型的文本立场检测[J]. 计算机与现代化, 2025(1): 1-6.
- [10] KUMAR S. Overriding safety protections of open-source models[EB/OL]. (2024-09-28)[2026-02-22]. <https://doi.org/10.48550/arXiv.2409.19476>.

## 【作者简介】

陈祥(1986—), 男, 江苏徐州人, 硕士研究生, 副教授, 研究方向: 人工智能、边缘计算。

(收稿日期: 2025-08-07 修回日期: 2026-02-25)