



肖卫

求职意向

算法工程师

深圳市

¥

1周内到岗

工作经验

2023.8-2025.2

联丰工程咨询有限公司佛山市分公司

算法工程师

- 负责公司内部业务相关数据的挖掘与分析，包括建筑合同、部门文档、员工咨询文本等多种类型的数据采集、清洗和标注，确保数据质量满足模型训练需求；
- 使用深度学习框架（PyTorch、TensorFlow等），完成文本分类、意图识别、实体抽取等NLP模型的设计与实现；
- 开发面向公司内部智能化应用系统，包括智能客服、合同风险识别系统和文档分类系统；
- 结合建筑行业的实际业务需求，与法务、行政、商务等部门紧密协作，收集需求并制定技术解决方案；

2022.7-2023.8

中国建筑第八工程局有限公司

算法工程师

- 数据收集与处理：收集公司过往的建筑合同数据，包括施工合同、项目会议、总分包合同、保密协议等，标注合同中的关键法律条款和潜在风险点。对合同文本进行预处理，去除无关信息，进行分句和分词处理，确保文本结构清晰。
- 使用编程语言（Python、C++、Java等）和框架（TensorFlow、PyTorch、Scikit-learn等）实现算法逻辑，设计、开发和优化算法模型

2020.3-2022.7

中国建筑第八工程局有限公司

土建工程师

2016.7-2019.12

深圳市润和天泽环境科技发展有限公司

土建工程师

32岁

2年经验

18676224424

技能特长

Pytorch

OpenCV

Transformer

Bert

教育背景

2012.9-2016.6

土木工程

井冈山大学

2024.9-至今

人工智能

伊利诺伊理工大学

在职研究生

项目经验

项目一：基于深度学习的公司内部文档分类系统

项目描述：为提升公司内部文档管理效率，项目开发了一套基于深度学习的文本分类系统，专注于自动识别并分类公司施工部门、设计部门、商务部门、法务部门、技术部门等相关的文件。该系统旨在解决公司内部文档繁多、人工归类耗时、分类准确性不足的问题，助力各部门快速获取所需文档并规范管理流程。

主要技术栈：pandas、BERT、DataLoader、AdamW

技术实现：

1、数据准备

1) 数据收集：从公司内部管理系统中抽取历史归档文档，包括 PDF、Word 和纯文本格式的文件。

2) 数据清洗：使用正则表达式对文档内容进行无关信息过滤（如页眉页脚、模板表格）。针对不同部门特定术语（如施工部门的“施工进度表”、法务部门的“合同条款”）。

3) 标签处理：结合标注的文件类别，完成训练数据的标注和验证。

2、模型开发

1) 预训练模型：采用 BERT 生成上下文嵌入，捕获语义信息。

2) 分类器设计：在 BERT 的基础上添加全连接层，通过 Softmax 实现文档类别的多分类任务。

3、长文档处理：针对长文本截断问题，设计基于动态滑动窗口的文本分块方法，提高模型对全局信息的理解能力。

4、训练与优化

1) 使用 AdamW 优化器，并结合余弦退火学习率调度器提升训练效果。

2) 利用类别加权损失函数平衡各类文档数量分布不均的问题。

3) 实现数据增强策略，如随机删除无关句子、同义词替换等，增加模型鲁棒性。

项目成果：分类模型部署在公司内部知识管理系统，帮助快速分类公司施工、设计、商务、法务、技术等部门的文件，大幅提升了文档检索效率。减少了 60% 的人工审核和归档工作量，有效降低人力成本。在测试集上，模型达到了 95.6% 的准确率，极大提高了业务效率。

项目二：建筑合同自动审查与风险识别系统

项目描述：本项目通过RAG技术，构建了一个自动化合同审查系统，能够在建筑合同中识别潜在的法律风、约定条款不合理之处以及与公司政策不一致的内容。系统通过智能问答形式提供风险提示，并生成相关的审查结果，帮助商法部门和项目快速完成合同分析，减少人工检查工作，提升效率和准确度。

主要技术栈：BM25、NER、RAG、BERT、jieba 技术实现：

- 1、数据收集与处理：收集公司过往的建筑合同数据，包括施工合同、项目协议、总分包合同、保密协议等，标注合同中的关键法律条款和潜在风险点。对合同文本进行预处理，去除无关信息，进行分句和分词处理，确保文本结构清晰。
 - 2、RAG模型架构：检索阶段使用 BM25技术建立合同文本的索引系统，能够快速检索到与输入 问题相关的合同条款和内容。生成阶段使用GPT-3生成模型，基于检索到的相关合同片段，结合法律知识库生成审查报告或风险提示。
 - 3、检索生成：系统首先根据用户提供的问题或合同中的关键信息进行检索，获取相关的条款，然后通过生成模型结合上下文，生成相关的审查建议、风险分析或合规性检查结果。
 - 4、法律风险识别：使用命名实体识别（NER）技术从合同中提取关键实体（如公司名称、时间、金额、违约条款等）。基于规则引擎和模型，识别合同中的潜在风险，如合同违约条款不明确、付款方式不合理、处罚条款过于严苛等。对识别的风险点生成提示，提示用户可能存在的法律问题，并根据相关法规或公司政策提供优化建议。
 - 5、问答系统：用户可以通过输入自然语言问题（如 "合同中的付款条款是否合规？"）与系统进行交互。系统首先检索与问题相关的合同条款，然后生成回答，提供关于合同条款合法性、合规性、合理性等的分析。
 - 6、报告生成：基于RAG系统的输出，自动生成合同审查报告，包括发现的风险点、需要修改的条款、与公司政策的对比分析等。报告支持定制化格式输出，如 PDF、Word 等，便于法务部门进一步处理和存档。
- 项目成果：通过对合同条款的智能审查，系统帮助公司发现了多项潜在的法律风险，如不明确的付款条款、过于宽泛的免责条款等。通过智能生成的审查报告，商法部门及项目人员能够迅速了解合同中的问题，并提出修改意见，提高了合同合规性。合同审查时间减少了70%，商法部门及项目人员能够专注于风险更高的关键问题，而非逐条检查合同内容。自动化风险识别准确率达到 90%，大大减少了漏审和误审的概率。

我的优势

- 1)掌握Transformer、BERT、ChatGLM等大语言模型技术，具有一定的应用积累。
- 2)熟悉PyTorch、TensorFlow等常用深度学习框架，能够高效实现和优化模型。
- 3)掌握传统图像处理算法，熟练使用OpenCV，skimage，Matplotlib等进行相关应用开发。
- 4)了解常用机器学习算法与NLP技术，具备一定的跨领域应用能力。
- 5)熟练使用Python编程语言，具备较强的编程与开发能力。
- 6)具备良好的沟通能力、学习能力，和质量意识，能够高效推动项目进展。