



AFAC2025  
金融智能创新大赛

队名：feiyun

# Contents

## 目录 ▶

01

团队介绍

02

方案综述

03

解决方案

04

技术创新

05

方案成效

# 01 团队介绍

赛题二：金融保险场景下多源文件长上下文一致性校验

队名：feiyun

## 队名： feiyun

### 韩张驰

上海交通大学 硕士

---

上海交大计算机专业研二；  
研究大模型RAG、强化学习方向；  
蓝桥杯大赛国赛二等奖；  
全国大学生数学建模大赛二等奖；  
华为算法精英赛一等奖

### 董勇鹏

上海交通大学 博士

---

上海交大计算机专业博士；  
研究高性能索引方向；  
发表多篇CCF高水平论文；  
对LLM应用有深入理解

### 闻睿

西安交通大学 硕士

---

西安交大动力工程专业研二，研究换热优化算法；  
精通python编程，对金融领域有深入理解；  
数学建模大赛三等奖；

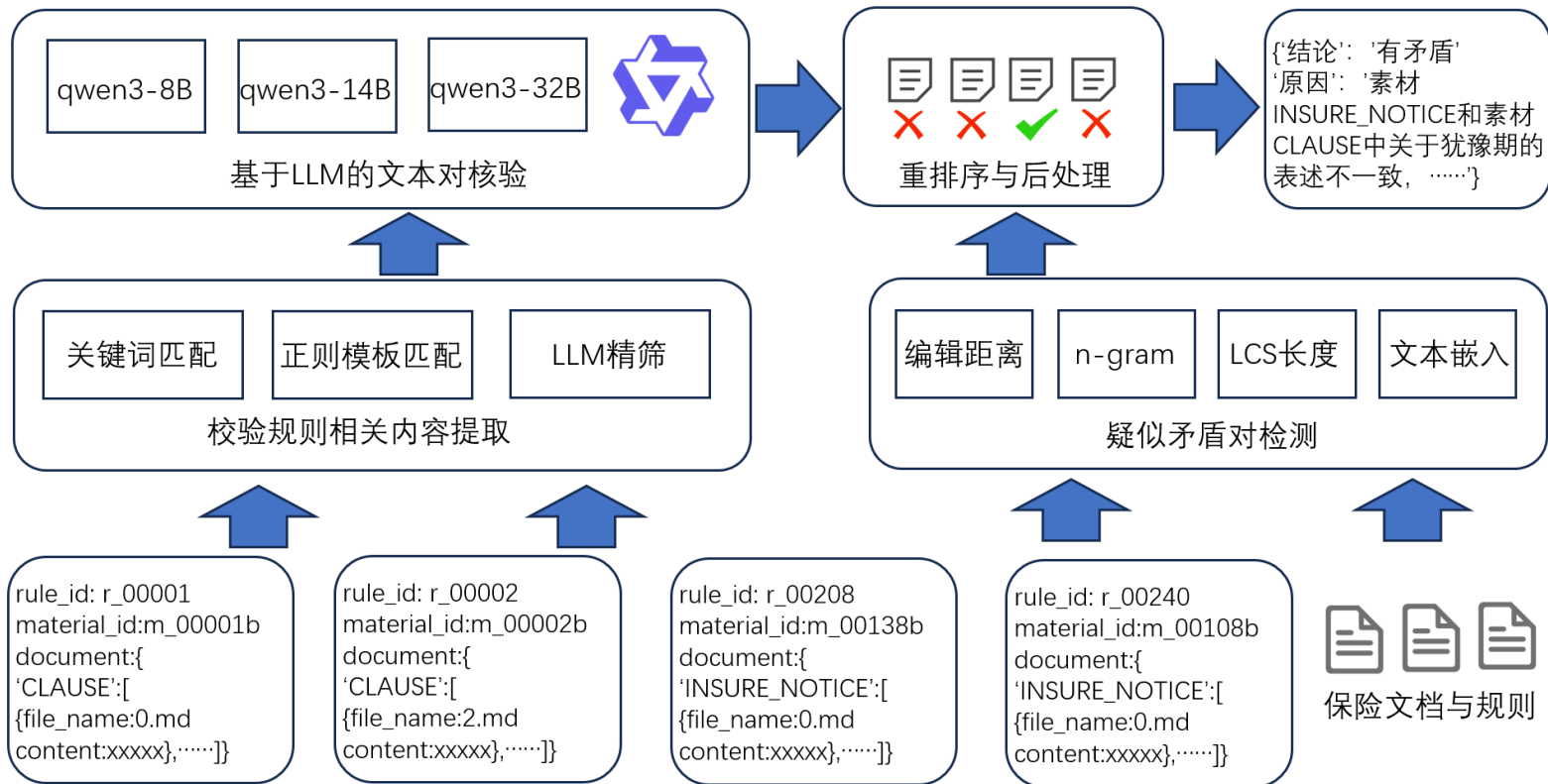


# 02 方案综述

内容提取+文本比对+后处理



# 方案综述





# 03 解决方案

基于prompt的LLM workflow



## 核心目标

从长文档中，找出和**校验规则**强相关的内容片段。

## 实现方案

- 基于**关键词+正则表达式**，初步提取相关内容片段
- 基于**LCS、编辑距离、embedding向量相似度**等方案，筛出疑似矛盾的内容对
- 使用LLM，通过**prompt+few shot**的方式，精确提取出与校验规则相关的内容，并格式化为json格式

## 典型案例

**校验规则:** 基础产品相关信息

**关键词:** (注册编号|注册号)、(保险|条款)

**正则表达式:** "C\d{15,}"

**LLM指令:** 这是一个保险产品的基础信息。其中可能包含一个或多个保险条款的名称及其条款编号、注册号等内容。请你从中提取出所有的保险条款名称及其对应的注册号，并以json格式返回。

请注意，注册号通常以"C"开头，后面跟着若干数字。

你返回的应该是一个列表（即使只有一个元素），每个元素包含两个字段：条款名称和注册号。如果没有任何条款，请返回空列表。如果某个条款缺失名称或注册号，对应项为空字符串，但不能缺少这一项。

保险内容为: {s}

**原始切片:**

中国大地财产保险股份有限公司附加意外伤害救护车费用保险(互联网专属)条款注册号  
C00001031922021122132233总则第一条本保险是意外伤害类保险(互联网专属)(以下简称“主险”)的附加险。在已投保主险的基础上,方可投保本附加险

**提取内容:**

```
{"条款名称": "中国大地财产保险股份有限公司附加意外伤害救护车费用保险(互联网专属)条款",  
"注册号": "C00001031922021122132233" }
```

## 典型案例

校验规则: 责任免除

关键词: (责任免除|免责条款|免责声明)

比对方法: 最长公共子序列 (LCS)



### 可疑矛盾对:

1. (六)任何生物、化学、原子武器、原子能或核能爆炸、辐射或污染;(七)被保险人乘坐民航客机违反承运人关于安全乘坐的规定而导致事故的发生;(八)被保险人乘坐**商业航空飞行器**发生的意外事故
2. (六)任何生物、化学、原子武器、原子能或核能爆炸、辐射或污染;(七)被保险人乘坐民航客机违反承运人关于安全乘坐的规定而导致事故的发生;(八)被保险人乘坐**非民航客机**发生的意外事故"

## 核心目标

使用LLM，对提取出的内容进行不一致性检验，目标是尽可能**检验出所有潜在矛盾**。

## 实现方案

- 针对每个校验规则，精调提示词，让LLM分析不同素材内容是否有矛盾，及矛盾原因
- 异步调用LLM API，最大化吞吐量，提高比对效率；
- json格式化数据流，方便workflow中数据传递；

## system 提示词

你是一个擅长保险领域的智能助手，需要帮助用户完成保险文档的核对工作。请以json格式返回提取结果。

## json schema

```
{  
  "type": "object",  
  "properties": {  
    "结论": {"type": "string", "enum": ["有矛盾", "无矛盾"]},  
    "原因": {"type": "string"}  
  },  
  "required": ["结论", "原因"],  
  "additionalProperties": False  
}
```

## user 提示词

这是一个保险产品中，不同保险素材下，提取出的和赔付给付金额和规则有关的信息，包括保险责任的赔付、给付、领取及免赔细节，如赔付年龄/比例/次数等；

包含CLAUSE, ADDITIONAL\_AGREEMENT, INSURE\_NOTICE, LIABILITY\_EXCLUSION, INTRODUCE\_IMG, PRODUCT\_MANUAL 六个素材。你需要核对不同素材下，是否存在和赔付给付金额和规则有关的内容相互矛盾的情况。

注意，你只需要核对不同素材下是否存在不一致信息，不用核对同一素材中相同或不同文件的内容！只需要核对和各种赔付给付金额，如赔付、给付、领取及免赔细节，如赔付年龄/比例/次数等有关的内容，并且只有当两个条款都提到一个具体的数字，且数字不一致时，才需要处理。不需要管协商约定\保险单载明等情况。只有当确定两个内容有显而易见的矛盾时，才需要处理。不要进行任何推理。

请以json格式返回结果，应该包含“原因”和“结论”两个字段。“结论”只能是“有矛盾”或者“无矛盾”。

提取出的相关信息是：

```
{extracted_chunks}
```

请仔细核对每个细节，认真核对每一条内容！并返回json结果。

## 核心目标

对LLM检验出的潜在矛盾对进行判断，**过滤掉误判的假矛盾对**


## 实现方案

- 在比对过程，让LLM输出**额外信息**，并通过**基于规则**的方式，根据额外信息进行假矛盾对剔除；
- 给LLM提供**工具调用能力**，给LLM提供包括**上下文查询**在内的多种工具，利用LLM的内容理解能力和prompt设计进行过滤。



# 04 技术创新

LLM workflow调试与结构化数据流构建



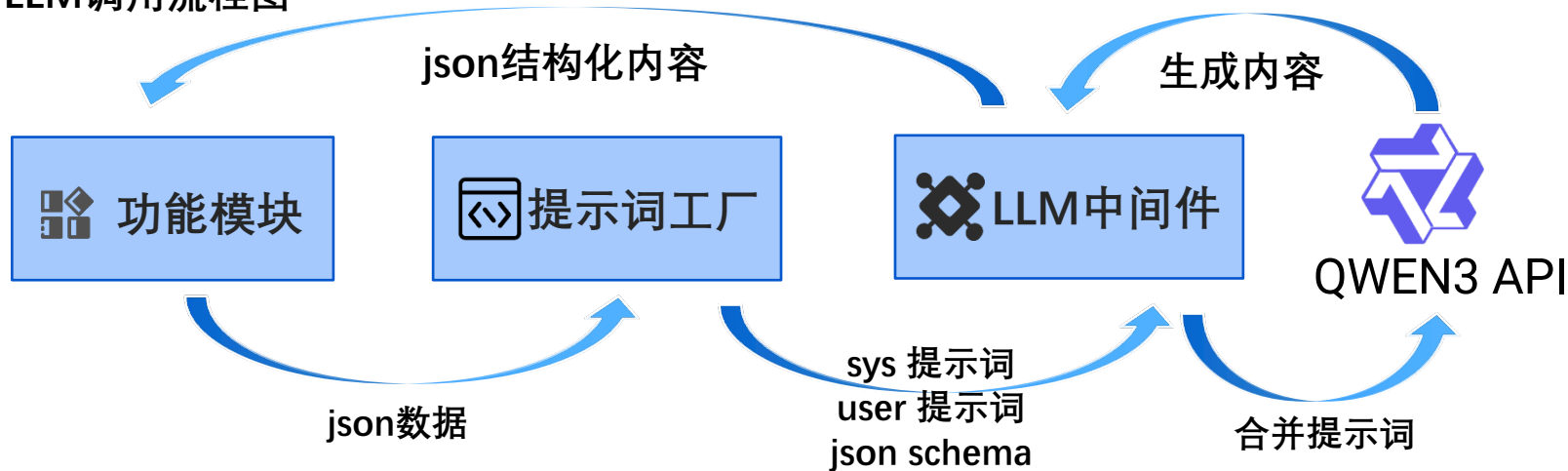
## 1. 为什么选用基于提示词+workflow的方案，而不是fine-tune模型？

- Qwen3模型本身的内容理解和生成能力足够强，可通过精调提示词实现预期目标
- 精调提示词成本小，方便调试和快速迭代
- 灵活，当文档格式或校验规则发生变化时，可精准修改对应提示词

## 2. 为什么使用多阶段+根据校验规则分类的结构，而不是统一使用一套prompt？

- 分阶段、分模块处理，可通过优化每个模块至最优解，从而实现全局最优解，避免强耦合框架下“解决旧的问题，导致新的问题”。
- 实际应用中，方便对校验规则进行新增和修改

## LLM调用流程图



功能模块：workflow中的内容提取、比对、后处理模块

提示词工厂：储存所有提示词模板

LLM中间件：负责API异步调用、异常检测与重试、json规则校验等

## 检验结果可视化 – prompt快速迭代

r\_00010 - 基础产品销售信息

CLAUSE - 0.md

中国大地财产保险股份有限公司  
附加猝死保险(互联网专属)条款  
注册编号 C00001031922021122132253

总则

第一条  
本保险是意外伤害类保险(互联网专属)(以下简称“主险”)的附加险。在已投保主险的基础上,方可投保本附加险。  
凡主险合同内容与本附加险合同相关的部分以及本附加险条款,均为本附加险合同的构成部分。  
本附加险条款未尽事宜,以主险条款为准;若本附加险条款与主险条款内容冲突,则以本附加险条款为准。  
主险合同终止的,本附加险合同同时终止;主险合同无效的,本附加险合同也无效。

第二条  
投保时年龄在出生满二十八日(含)至七十四周岁(含)之间,能正常工作或者正常生活的主险合同的被保险人,可作为本附加险合同的被保险人。

第三条  
1

### 冲突对 1

陈述1: 注册编号 C00001032522021122327803

陈述2: 注册编号 C0000103252202112327803

**原因:** 同一附加险产品注册编号不一致, 数字有明显差异

规则内排名: 1.45

### 冲突对 2

陈述1: 注册编号 C00001032522021122843303

陈述2: 中国大地财产保险股份有限公司附加个人意外伤害住院津贴医疗保险(互联网专属)条款 注册编号 C00001032522022110417831

**原因:** 同一附加险名称在组1和组2中注册编号不一致, 存在明显细节冲突。

规则内排名: 2.36

## 检验结果可视化 – prompt快速迭代

## 组1 切片

中国大地财产保险股份有限公司附加个人意外伤害医疗保险(互联网专属)条款注册编号 C00001032522021122327803

本保险是个人意外伤害类保险(互联网专属)、个人医疗类保险(互联网专属)、个人疾病类保险(互联网专属)(以下简称“主险”)的附加险。在已投保主险的基础上,方可投保本附加险。

凡主险合同内容与本附加险合同相关者及本保险条款,均为本附加险合同的构成部分。

保险期间最长为一年,具体起迄时间由投保人与保险人约定,并于本附加险合同中载明。

## 组2 切片

产品介绍 1、本产品为中国大地财产保险股份有限公司面向全国(港澳台地区除外)销售的室内健身运动意外险,产品条款注册信息:条款名称 注册编号 报备文件编号 中国大地财产保险股份有限公司个人意外伤害保险(A款,互联网专属)条款 C00001032312022090527951 大地财保发〔2022〕610号 中国大地财产保险股份有限公司附加个人意外伤害医疗保险(互联网专属)条款 C0000103252202112327803 大地财保发〔2021〕916号 中国大地财产保险股份有限公司附加个人意外伤害住院津贴医疗保险(互联网专属) C00001032522022110417831 大地财保发〔2022〕723号 中国大地财产保险股份有限公司附加猝死保险(互联网专属)条款 C00001031922021122132253 大地财保发〔2021〕912号 中国大地财产保险股份有限公司附加个人意外骨折或关节脱位津贴保险(互联网专属)条款 C00001032322021122333653 大地财保发〔2021〕913号 中国大地财产保险股份有限公司附加运动期间限制特约保险(互联网专属)条款 C00001031922022012384073 大地财保发〔2022〕62号 中国大地财产保险股份有限公司附加意外伤害救护车费用保险(互联网专属)条款 C00001031922021122132233 大地财保发〔2021〕914号

## 冲突对 1

陈述1: 注册编号 C00001032522021122327803

陈述2: 注册编号 C0000103252202112327803

原因: 同一附加险产品注册编号不一致, 数字有明显差异

规则内排名: 1.45

## 冲突对 2

陈述1: 注册编号 C00001032522021122843303

陈述2: 中国大地财产保险股份有限公司附加个人意外伤害住院津贴医疗保险(互联网专属)条款 注册编号 C00001032522022110417831

原因: 同一附加险名称在组1和组2中注册编号不一致, 存在明显细节冲突。

规则内排名: 2.36

## 1. 使用“格式化输出+规则检查”的方法，代替“提示词硬性要求”。

**场景：**题目要求只检测不同素材中的矛盾对。然而，LLM常会将同一素材中，不同条款的内容当作矛盾对误检。

**方案一：提示词硬性要求。**在提示词中反复强调：“仅检查不同素材中的矛盾内容，不要对同一素材中内容检验”

**效果：**反复强调作用有限。甚至会导致不同素材间的矛盾也被漏检

**方案二：格式化输出+规则检查。**要求LLM输出矛盾内容所属素材，再用if-else判断两个素材是否相同，实现过滤

**效果：**彻底解决了该问题

## 2. 使用异步LLM调用+差异化LLM选型方式，代替单一LLM串行调用

**场景：**baseline方法检验速度慢。一条规则需要2-3分钟。然而，不同文本对、不同规则间的检验并无依赖性。

**方案一：LLM异步调用。**让没有依赖的步骤可以同时调用API并等待返回结果。

**效果：速度提升7-10倍**


**方案二：差异化LLM选型。**对每个LLM调用步骤，首先使用Qwen3-32B得到最优解，再不断尝试参数更小的模型，直到效果相差较大为止。

**效果：**在精度基本不下降的前提下，**降本增效 20% ~ 30%**。



# 05 方案成效

A、B榜第一名



B榜

A榜

排名	参与者	组织	score	最优成绩提交日
1	feiyun	上海交通大学	0.9610	2025-07-19
2	ths	北京理工大学	0.9101	2025-07-22
3	clife	平安财产保险	0.9071	2025-07-11
4	毋庸置疑	四川大学	0.8950	2025-07-22

B榜

A榜

排名	参与者	组织	分数	最优成绩提交日
1	feiyun	上海交通大学	0.8889	2025-07-26
2	ths	北京理工大学	0.8626	2025-07-26
3	数据产业集团	数产	0.8517	2025-07-25
4	clife	平安财产保险	0.8489	2025-07-26

	方案	A榜 score	单条耗时
准确性优化	赛题二baseline	0.6755	约150s
	关键词内容提取+LLM文本比对	0.8270	约90s
	+基于LLM的后处理	0.8701	约115s
	+基于LCS和embedding的矛盾对检测	0.9189	约125s
	+prompt精调+fewshot示例构建	<b>0.9610</b>	约125s
效率优化	差异化LLM选型	<b>0.9610</b>	约95s
	并行化LLM调用	<b>0.9610</b>	<b>15s</b>

# 谢谢

队名: feiyun

答辩人: 韩张驰