

AFAAC 2024

金融智能创新大赛

Advanced FinTech AI Competition





Contents 目录

01

团队介绍

02

方案综述

03

解决方案

04

方案成效

05

技术创新



01 团队介绍

赛题四：金融规则长文本中的矛盾识别与漏洞发现
队名：不知道叫啥

队名：不知道叫啥



队长 顾国勤

厦门大学在读、神思电子实习生

- ◆ 厦门大学金融工程专业大三；
- ◆ 神思电子OR组算法实习生；
- ◆ 全国大学生数学建模大赛一等奖、全国大学生统计建模大赛一等奖、第七届工业互联网大赛前2%。
- ◆ 厦门银行杯金融科技创新大赛一等奖；厦门数字金融创新创业大赛全国十强。



指导老师 宋玉平

上海师范大学商学院副教授

- ◆ 浙江大学概率论与数理统计博士，目前主持国家自科青年项目、教育部青年项目、全国统计科学重点项目及上海哲社一般项目等，上海重点市属高校副教授。
- ◆ 研究方向：金融大数据分析；人工智能前沿与金融交叉研究等。累计论文总篇数40多篇，其中人工智能在金融中应用方向以第一作者（通讯作者）在Expert System with Application、Applied Soft Computing、Journal of Forecasting及Computational Economics等SCI与SSCI期刊发表论文二十多篇；担任多个SCI与SSCI期刊匿名审稿人。



队员1 柏路路

华南师范大学在读、神思电子实习生

- ◆ 华南师范大学人工智能专业研二、熟悉NLP和时序预测任务；
- ◆ 神思电子OR组算法实习生；
- ◆ 中国研究生数学建模(华为杯)三等奖、亚太杯数学建模一等奖、全国大学生数学竞赛二等奖、蓝桥杯大赛一等奖、CCPC程序设计赛铜奖、天梯赛全国总决赛铜奖。



队员2 陈真琦

齐鲁工业大学在读、神思电子实习生

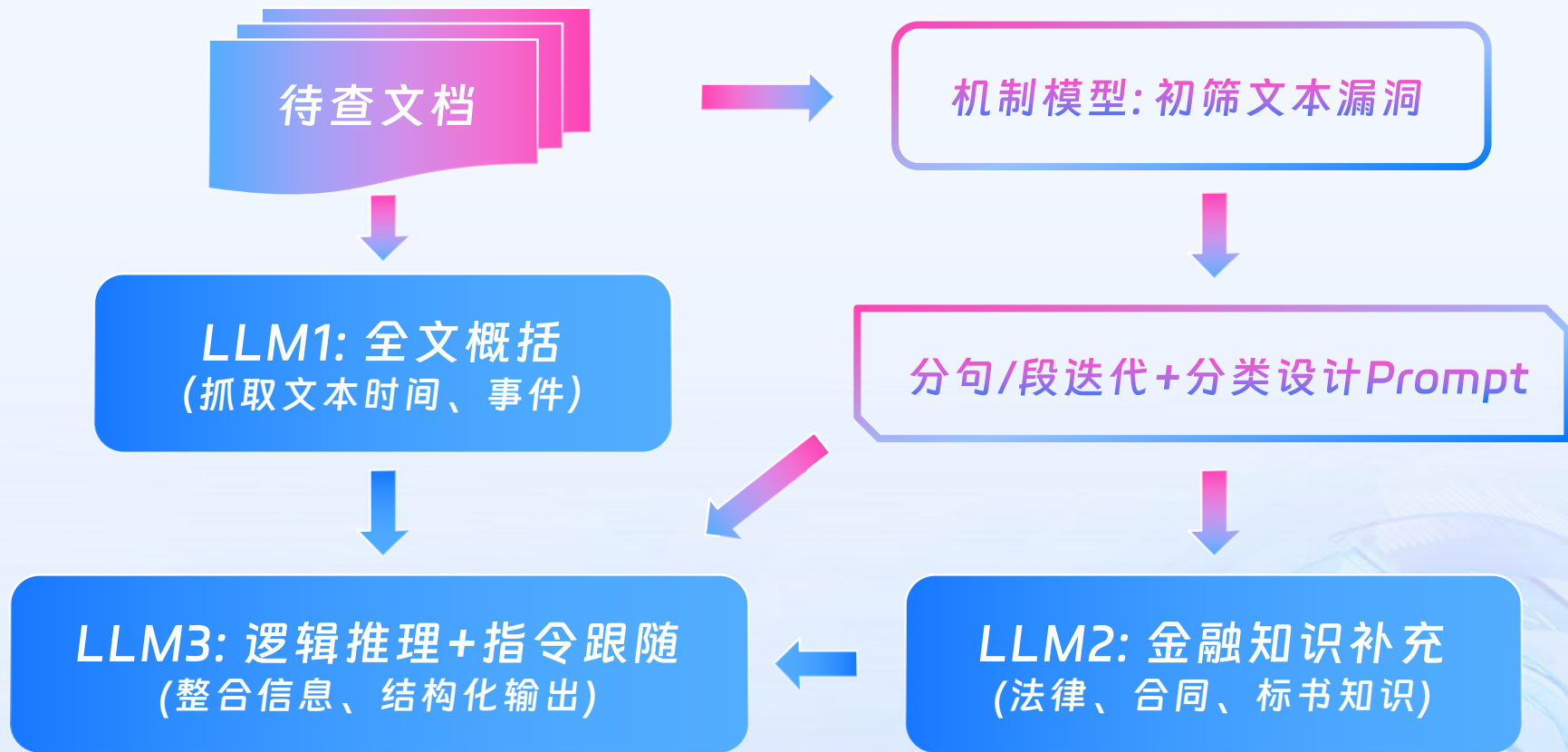
- ◆ 齐鲁工业大学控制工程专业研二，熟悉数字图像处理、机器视觉以及深度学习领域；
- ◆ 神思电子技术研究院算法工程师实习生；
- ◆ 第二十五届中国机器人及人工智能大赛一等奖，“兆易创新杯”第十九届中国研究生电子设计大赛华北赛区三等奖。



02 方案综述

逻辑筛选+

文本概括+金融知识补充+推理（指令跟随）





03 解决方案

逻辑筛选+

文本概括+金融知识补充+推理（指令跟随）

应用场景

在金融领域中，规章制度，合规指南，合同，研报等长文本文档可能存在常识错误、矛盾、模糊，甚至漏洞，需要设计一套通用的检测模型自动识别问题，保障金融机构的稳健运行。

任务目标

设计算法，理解长文本中的矛盾和漏洞，识别出文档中的问题句子，或者存在矛盾的问题句子对，以标点符号或换行符为切割点。

任务分析

对文本数据进行预处理，再用模型理解并识别在金融领域中各种类型的长文本中出所有的问题句子以及矛盾句子对，最终筛选输出文本中存在问题的句子以及存在矛盾的句子对。

问题分析

方案设想

数据预处理

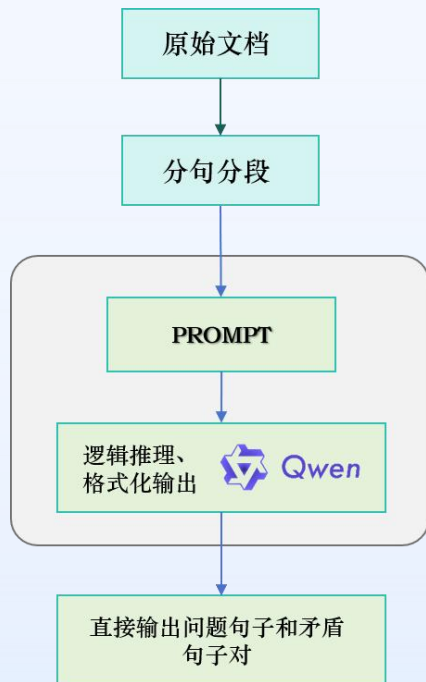
对给出的长文本数据进行分句或分段操作，设计机制模型，以问题类型对句子模式匹配并标记可能出现的问题类型。

以大模型为主的算法模型

由于大语言模型的涌现现象使我们对文本的理解能力强，我们的目标也是理解并识别文本中的错误句子，所以采用大语言模型以及prompt技术作为此次解题方案。

集成多LLM

由于我们没有找到高质量，覆盖法规、标书、保险条例的数据集进行微调任务，所以采用多个LLM分工完成任务，以提高识别精度。



方案1.0

通义金融-14B是专为金融领域训练的大语言模型，我们首选此模型对文本进行问题句子的识别。

在这个版本的方案中，我们设置了一套通用的提示词模板，在prompt中对每种错误给出定义，遍历文档获取识别结果。

存在问题：

- 1：通义金融的指令跟随能力略差，无法同时输出两种格式的答案（问题句子&矛盾句子对），并且无法保持按照给定的Json模板进行输出。
- 2：采用通用的prompt模板并对八种错误类型给出定义的方式让大语言模型出现了幻觉现象，混淆不同错误类型并且极易误判。

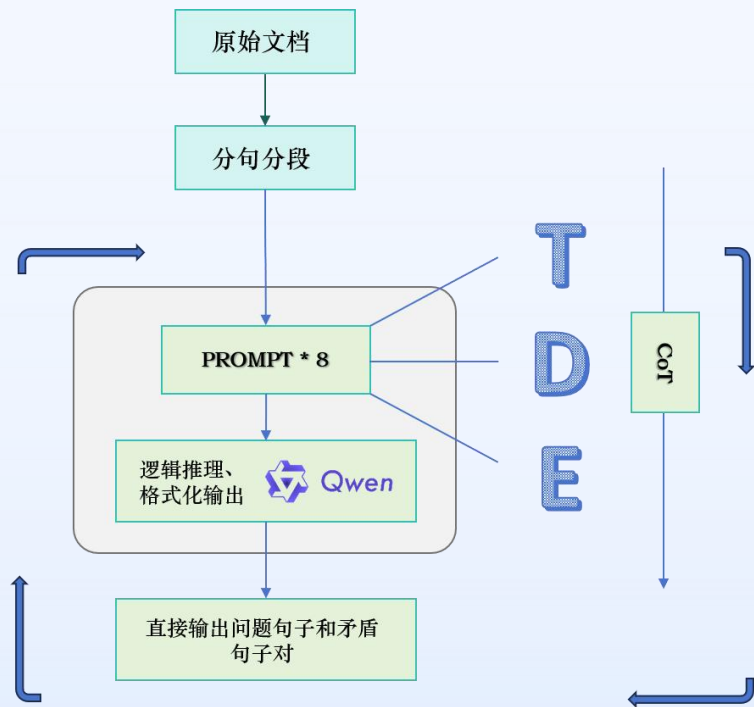
Prompt

- 作为一位识别金融文本中的漏洞和矛盾的专家，您的任务是从给定的上下文中找出存在漏洞或矛盾的句子。
 - ❑ 给定的上下文: "{para}\n"
- 请综合上述信息，你给出的回复需要包含以下两个字段：
 - ❑ 1. error_type: 基于上下文内容，指出这段文本中存在的漏洞或者矛盾类型，文本漏洞包含五种：常识错误、数值单位错误、数据不完整、计算错误、语句重复；文本矛盾包含三种：逻辑矛盾、时间矛盾、数值前后矛盾。
 - ❑ 2. error_sentence: 根据你找到的漏洞或者矛盾类型，写出原文中存在漏洞或矛盾的句子，请用 markdown 格式。
 - ❑ 如果你认为有一个存在漏洞的句子，请按照以下JSON格式来回答：...
 - ❑ 如果你认为有两个存在矛盾的句子，请按照以下JSON格式来回答：...
- 注意：如果碰到上下文内容中没有存在漏洞或矛盾，无法回答问题的情况，type和error_sentence可以返回空。
- 最后强调一下：你的回复将直接用于javascript的JSON.parse解析...

Answer

【误判-混淆矛盾类型】此处可能存在矛盾错误：['预算金额：2822.07元', '最高限价：2660.333311万元']

【误判-不遵循格式输出】矛盾错误：['预算金额：1551.555000元;第1包:675.000000万元;第2包:576.555000万元'与'第1包:675万元;第2包:576.555万元']之间存在单位矛盾



方案2.0

依然使用通义金融-14B对文本进行问题句子的识别。
在原来的Prompt基础上，针对八种错误类型（Type）设计八种不同定义（Definition）并分别提供多组示例（Example），按照思维链（CoT）的方式组织提示词，引导模型的思考方向。

存在问题：

- 1: 循环八次文本耗费时间较大，即使在Vllm框架下加速推理，每次完成输出也需要耗费近四个小时，这并不令人满意。
- 2: Few-Shot下Qwen的指令跟随能力较差，无论是把示例安排在提示词开始或是结束，当输入的段落本身没有错误的时候，模型会把示例当做错误输出。

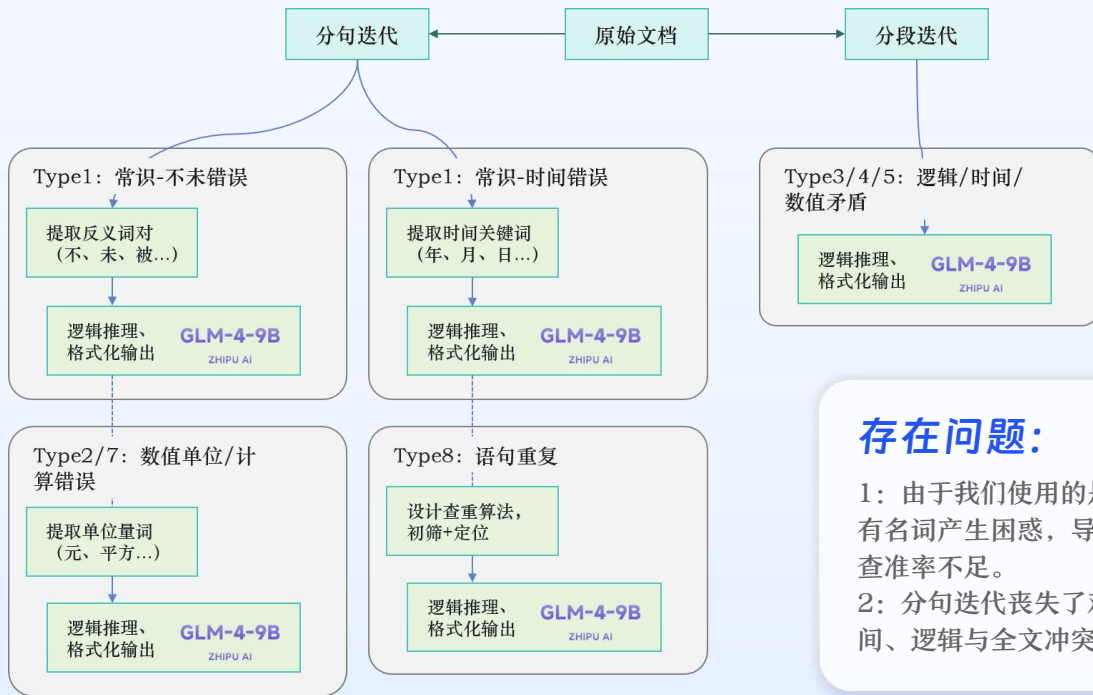
Prompt

- 作为一位识别金融文本中的漏洞和矛盾的专家，您的任务是从给定的上下文中找出存在漏洞或矛盾的句子。
 - ❑ 给定的上下文： "{para} \n"
- 请综合上下文信息，以上文章可能包含以下错误，请根据示例来回答错误句子：
 - ❑ "T": "常识错误"
 - ❑ "D": 常识错误是指文本中存在普遍认识上的错误，比如一年有12个月份、每月最多不超过31天、时间前后顺序应合理，又比如标书、法规中'不会'-'会'、'未被'-'被'、'可以'-'不得'等词的错用。存在这个错误则输出常识出错的句子。
 - ❑ "E": "错误句子示例输出：['合同签订日期：2022-07-32 合同公告日期：2022-07-06 00:00:00']"，"错误句子示例输出：['2024年0-4月，TOP100房企销售总额为12464.4亿元']"
- 注意：如果碰到上下文内容中没有存在漏洞或矛盾，无法回答问题的情况，type和error_sentence可以返回空。
- 最后强调一下：你的回复将直接用于javascript的JSON.parse解析…

Answer

【误判-原句正确时输出示例的错误】"错误句子：['2024年0-4月，TOP100房企销售总额为12464.4亿元']"

【误判-不遵循格式输出】["错误句子：“在恒生12个一级子行业中，医疗保健行业周涨跌幅排名为第14位。”"]



方案3.0

更换指令跟随、推理逻辑能力更强的GLM-4-9B模型。

将八种错误和矛盾类型整合为五种识别模式，设计机制模型进行初筛。

存在问题:

- 1: 由于我们使用的是无微调方案，GLM-4-9B会对一些金融专有名词产生困惑，导致判别错误。模型结果表现为查全率较高而查准率不足。
- 2: 分句迭代丧失了对全文语义的理解能力。金融文本的局部时间、逻辑与全文冲突的情况无法兼顾。

Prompt

- 作为一位识别金融文本中的漏洞和矛盾的专家，您的任务是从给定的上下文中找出存在漏洞或矛盾的句子。
 - ❑ "给定的上下文：" + "上文：{context_upper}\n" + "下文：{context_lower}\n"
- 这段文本来自于研报、招标书或者法律条文，你需要判断以下这组逻辑词相互矛盾的句子中哪个不符合上下文语义："
 - ❑ "句子序号1：{possible_error_sent}\n" + "句子序号2：{possible_error_sent_antonymy}\n"
 - ❑ "根据上下文语义，判断逻辑词`{err}`与其反义词`{errs_antonymy[k]}`哪个更符合语境。"
- 请综合上述信息，你给出的回复需要包含以下两个字段：
 - ❑ 1. num: 不符合段落语义的句子的序号。
 - ❑ 2. error_sentence: 指出逻辑词不符合段落语义的那个句子，输出包含错误逻辑词的最小粒度分句。
- 请按照以下JSON格式来回答：...
- 最后强调一下：你的回复将直接用于javascript的JSON.parse解析...

Answer

【误判-不理解金融语境下的复杂逻辑】“num”：“1”，“error_sentence”:[投标人应承诺近三年内未发生以下情况或失信行为：…投标人或其法定代表人、主要负责人、实际控制人被法院采取限制高消费或相关消费措施]

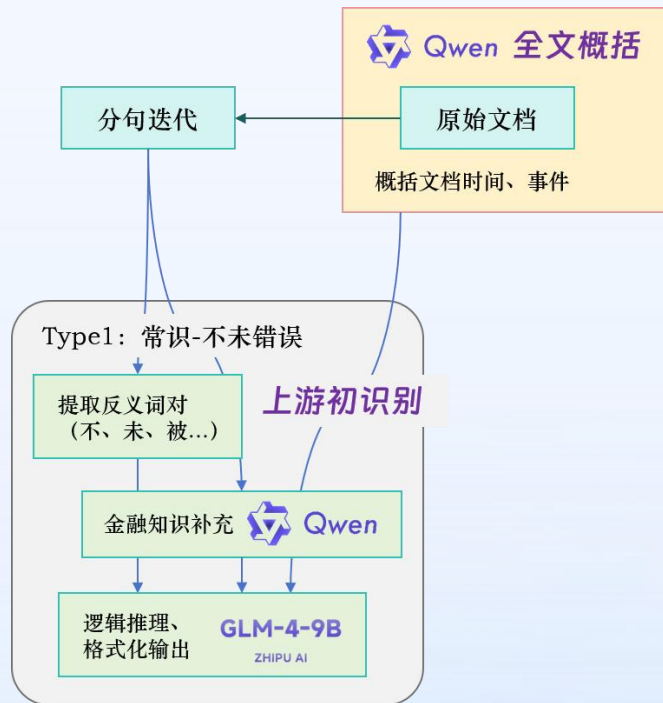
【正确判断】“num”：“2”，“error_sentence”:[["第八条 因下列原因造成被保险人身故、伤残或医疗费用支出的，保险人不承担给付保险金责任：”，“（六）被保险人（未）遵医嘱服用、涂用、注射药物；"]]



使用指令跟随、推理逻辑能力更强的 GLM4-9B模型进行答案输出。

- ❑ 在迭代之前对每篇原始文档进行全文概括，抽取文档基本时间、事件，作为后续步骤大模型的输入。
- ❑ 将八种错误和矛盾类型整合为五种识别模式，设计机制模型进行初筛，针对不同种类的错误提供更为精确的 Prompt。
- ❑ 对标书文档进行后处理，剔除标书开头和结尾容易造成大模型混淆的时间、人名、联系方式等，将标书按点分段进行重新筛查。
- ❑ 所有模型通过VLLM框架部署。

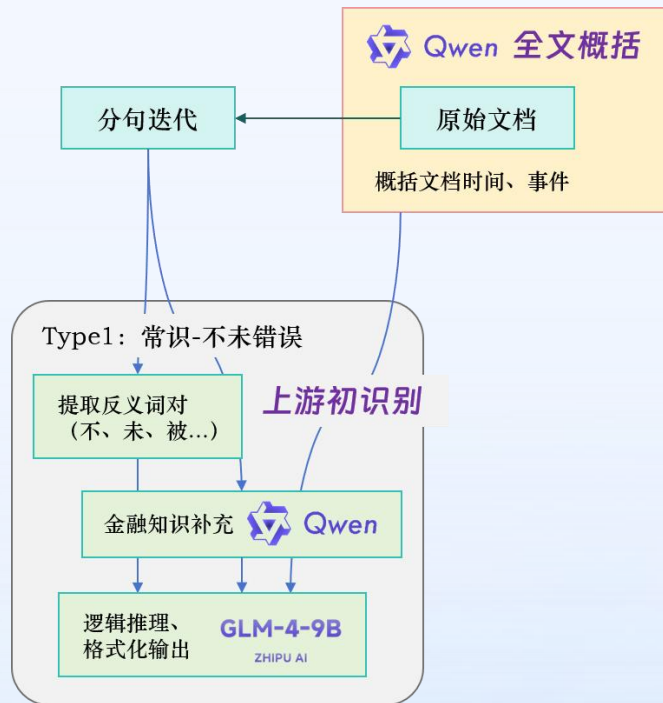
Type1: 常识-不未错误



定义

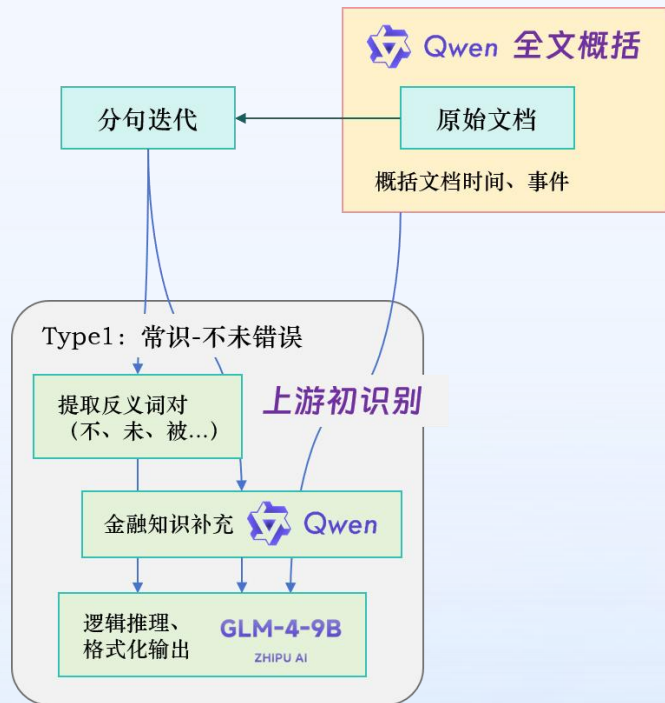
➤ 我们把这种错误定义为：错误地使用了“不、未、被、没有、须”等逻辑词的错误，比如以下示例（来自A榜数据集），这种错误也包含了大部分逻辑错误，因此我们就不单独识别逻辑错误：

- ❑ 中标人**不**为石化公司各装置提供密封相关服务，对于现场密封的稳定运行及防泄漏管理起到重要支撑作用。
- ❑ 参与投标的供应商必须具备以下资格，任何一条不满足将**不**会被否决参与投标。
- ❑ 会计帐簿登记，必须以**未**经过审核的会计凭证为依据。



解决方案

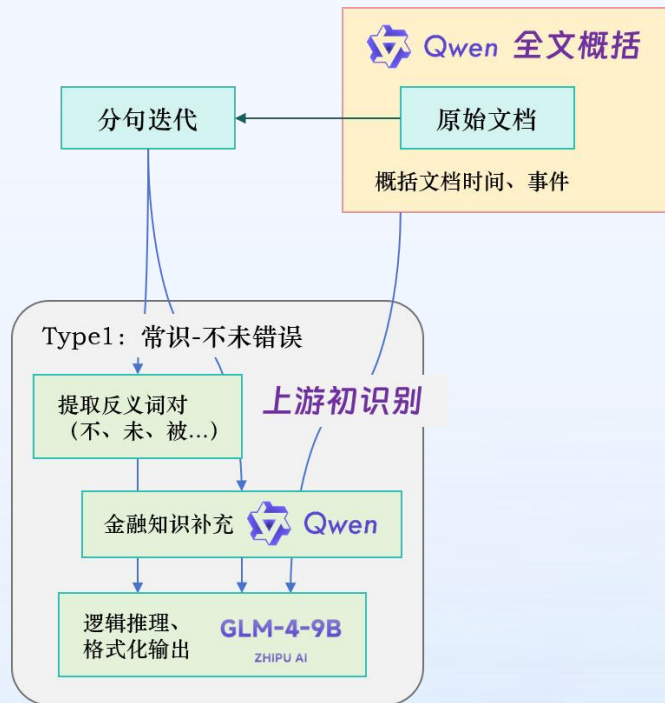
- 针对这种错误，我们的方案是对分句后的文档进行错误关键词识别，我们手动构建了关键词列表，如下：
 - ❑ `errs = ['不', '不为', '不会', '不能', '不得', '不具有', '会', '未', '无', '没有', '不可以', '可以', '免除', '被', '未被', '不被', '具备', '不具备', '免除' ...]`
 - ❑ `errs_antonymy = ['', '为', '会', '可以', '可以', '具有', '不会', '有', '有', '有', '可以', '不得', '不免除', '未被', '被', '被', '不不具备', '具备', '不免除' ...]`
- 我们使用构建好的反义词列表对句子进行反义词替换，从而为大模型提供了一组句子，减轻其幻觉，使其能更加准确地针对这种类别错误进行判断。



Prompt

- 给定的上下文:
 - 上文: {context_upper}; 下文: {context_lower}
- 这段文本来自于研报、招标书或者法律条文, 你需要判断以下这组逻辑词相互矛盾的句子中哪个不符合上下文语义:
 - "句子序号1: {possible_error_sent}"
 - "句子序号2: {possible_error_sent_antonymy}"
- 你的金融助手对于句子所涉及的金融知识给出了以下补充或修正:
 - "{qwen_answer}"
- 示例:
 - 若上文提到'承诺要求投标人应承诺近三年内未发生…'
- 请综合上述信息, 你给出的回复需要包含以下这两个字段:
 - 1. num: 不符合段落语义的句子的序号
 - 2. error_sentence: 指出逻辑词不符合段落语义的那个句子, 输出原句中包含错误逻辑词的最小粒度分句。
- 请按照以下JSON格式来回答: ...
- 警告: 你的回复将直接用于javascript的JSON.parse解析...

Type1: 常识-不未错误



Answer

【正确判断】“num”：“1”，“error_sentence”：[审计机关通过政务信息系统和数据共享平台取得的电子数据等资料不能够满足需要的，不得要求被审计单位重复提供。]

【正确判断】“num”：“1”，“error_sentence”：[不可以依法申请行政复议或者提起行政诉讼]

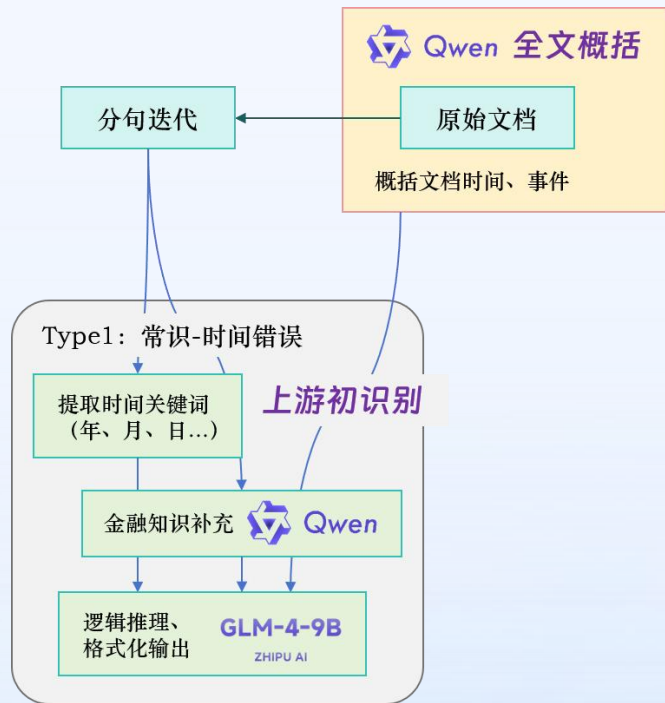
【正确判断】“num”：“1”，“error_sentence”：[不能责令改正]

【正确判断】“num”：“1”，“error_sentence”：[不可以组织潜在投标人踏勘项目现场]

【正确判断】“num”：“1”，“error_sentence”：[不满足★号技术指标将不被否决投标]

【正确判断】“num”：“1”，“error_sentence”：[若无，不用提交质量管理体系文件。]

【误判】“num”：“1”，“error_sentence”：[本次招标不接受联合体申请。]

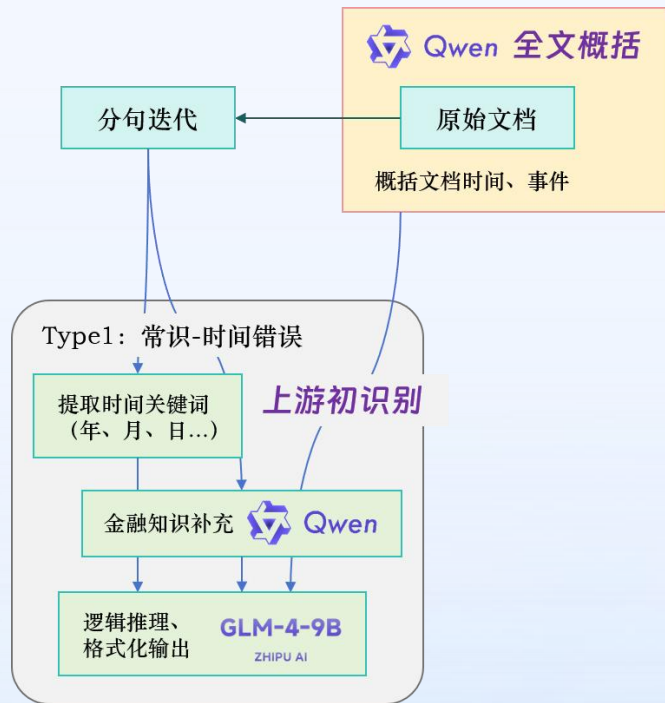


定义、解决方案

- 我们把这种错误定义为，句子中的时间信息违背常识、上下文逻辑或存在缺失，也即融合了时间矛盾和数据不完整，不再单列。
- 我们对分句后的文档进行进一步的模式识别，匹配时间信息，如下所示：

□ `pattern_list = [r'\d+年', r'\d+月', r'\d+个月', r'\d{4}-\d{2}-\d{2}']`

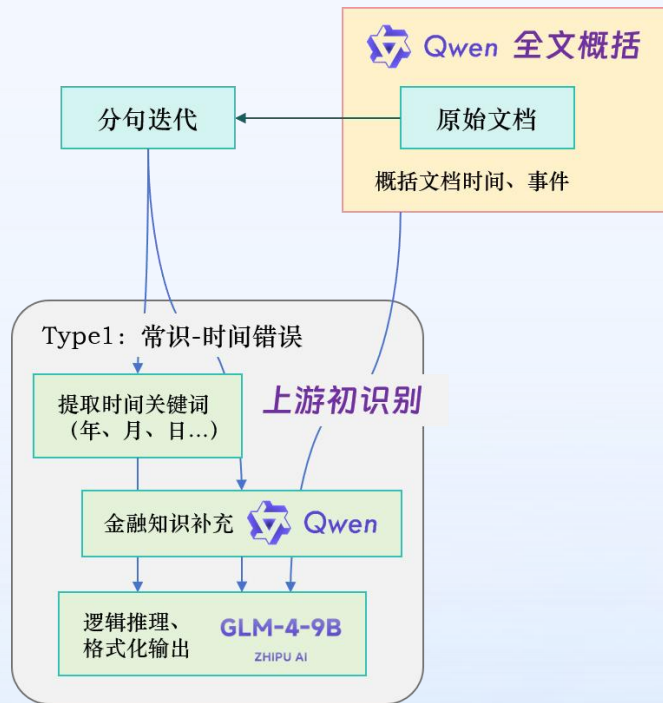
- 最终获得每篇文档的只含时间信息的句子列表，我们对文档进行迭代，直接向大模型传入**该文档中全部包含时间信息的句子**进行推断。



Prompt

- 以下句子来自同一篇文档，每个句子都包含至少一处时间信息，请检查是否句子存在时间信息上的错误，例如时间与常识不符、时间前后颠倒等。
 - ❑ "句子列表: \n"
 - ❑ for sentence in doc:
 - ❑ prompt += f"- {sentence}\n "
- 你的金融助手对于文章所涉及的金融知识给出了以下补充或修正:
 - ❑ "{qwen_answer}"
- 请综合上述信息，你给出的回复需要包含以下这个字段：
 - ❑ 1. error_sentence: 如果所有句子都没有时间错误，字段留空；如果某句有时间错误，输出包含错误时间的最小粒度分句；如果多个句子存在错误，则将错误之处都写入列表中。
- 请按照以下JSON格式来回答: ...
- 警告：你的回复将直接用于javascript的JSON.parse解析...

Type1: 常识-时间错误



Answer

【正确判断】“error_sentence”:[本条例自190年10月1日起施行。]

【正确判断】“error_sentence”:[合同履行期限: 2024-05-20 00:00:00至1925-05-21 00:00:00]

【正确判断】“error_sentence”:[4024年5月31日09时00分]

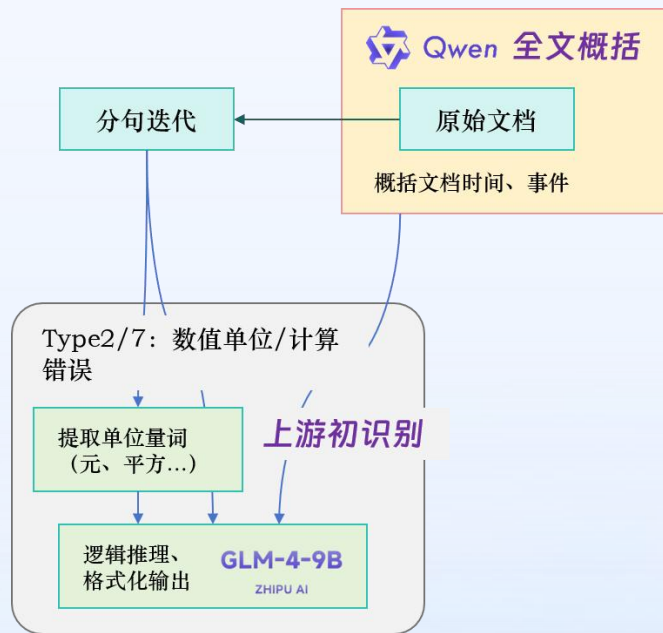
【正确判断】“error_sentence”:[2024年5月17日0时至2024年5月16日23时59分59秒]

【正确判断】“error_sentence”:[2024年2月30日09时00分(北京时间)]

【正确判断】“error_sentence”:[2024年05月29日至2024年05月07日内]

【正确判断】“error_sentence”:[截至2024年5月55日, 营业影院数1.25万家, 较2023年下降1.46%]

【误判】“error_sentence”:[承诺近三年(自2021年03月01日至投标截止时间), 未发生重大及以上质量事故]

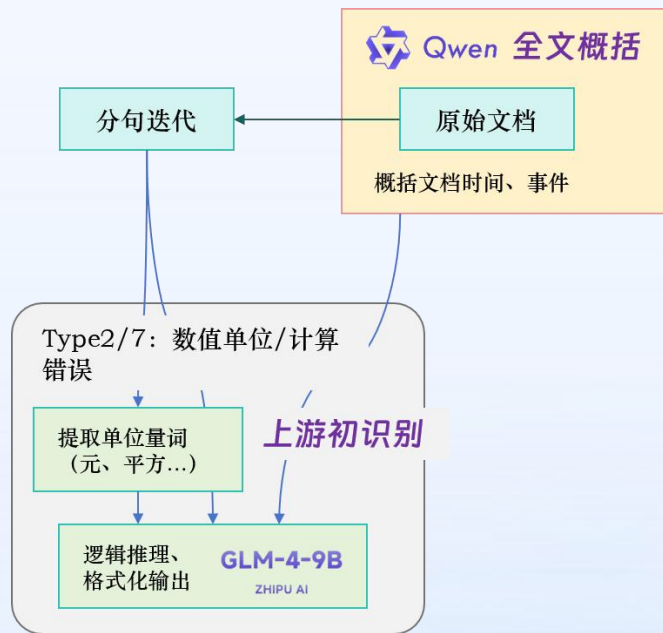


定义、解决方案

- 我们把这种错误定义为，文本中存在前后不一致或违背常理的计量单位，比如采购、招标合同金额往往是以万元甚至亿元计价，而房屋数量、面积等则往往是基本单位，违反常规的计量单位就是存在错误，这种错误也包含了数值前后矛盾，也不再单列。
- 我们对分句后的文档进行逐句迭代，并且抓取如下关键词：
❑ `errs = ['千', '万', '亿', '元', '米', '平方', '支', '套', '个', '件']`

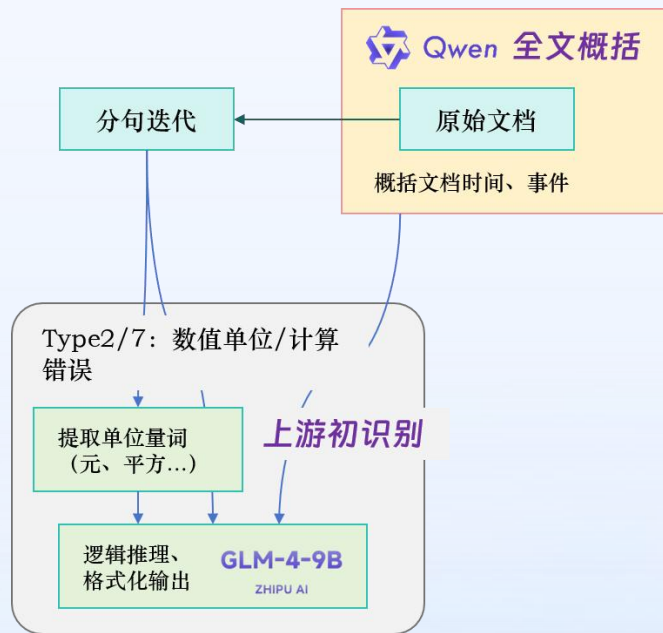
定义、解决方案

- 我们把这种错误定义为，文本中存在两个以上数值且数值间有计算关系时存在的错误，如两个数值相加应等于另一个却不等于的情况。
- 我们对分句后的文档进行逐句迭代，并且抓取如下关键词：
❑ `errs = ['元', '分钟', '个', '小时']`



Prompt

- 给定的上下文:
 - 上文: {context_upper}; 下文: {context_lower}
- 这段文本来自于研报、招标书或者法律条文, 你需要判断以下这个包含数值信息的句子是否存在错误, 数值大小和文本实际含义不符、数值缺失都属于错误, 比如标书金额往往是万元以上, 时间长度应该适中:
 - 句子: {possible_error_sent}
- 请综合上述信息, 你给出的回复需要包含以下这两个字段:
 - 1. TrueOrNot: 如句子没有数值大小错误或缺失, 字段填为`True`; 如句子有过大、过小或缺失的错误数值, 字段填为`False`, 标书金额过小、时间范围过大、面积单位过大等。
 - 2. sentence: 如果句子没有数值大小错误或缺失, 这个字段留空; 如果这个句子有过大、过小或缺失的错误数值, 输出包含错误部分的最小粒度分句, 请用 markdown 格式。
- 请按照以下JSON格式来回答: ...
- 警告: 你的回复将直接用于javascript的JSON.parse解析...



Prompt

- 给定的上下文:
 - 上文: {context_upper}; 下文: {context_lower}
- 这段文本来自于研报、招标书或者法律条文, 你需要判断以下这个包含数值信息的句子的数值是否计算正确, 你可以尝试检查句子中是否存在数值计算关系, 检查是否存在几个数值相加应该等于另一个但却不等于的情况:
 - 句子: {possible_error_sent}
- 请综合上述信息, 你给出的回复需要包含以下这两个字段:
 - 1. TrueOrNot: 如果句子没有数值计算错误, 字段填为 True; 如果句子有数值计算错误, 字段填为 False。
 - 2. sentence: 如没有数值计算错误, 该字段输出为空; 如有计算错误, 输出句子中包含计算错误之处的最小粒度分句。
- 请按照以下JSON格式来回答: ...
- 警告: 你的回复将直接用于javascript的JSON.parse解析...



【正确判断】“error_sentence”:[时间: 0个工作日]

【正确判断】“error_sentence”:[预算金额: 2822.07元]

【正确判断】“error_sentence”:[三年总限价为 1449元]

【正确判断】“error_sentence”:[本周（12月30日-1月5日）电影票房18.65元, 环比增长144.68%]

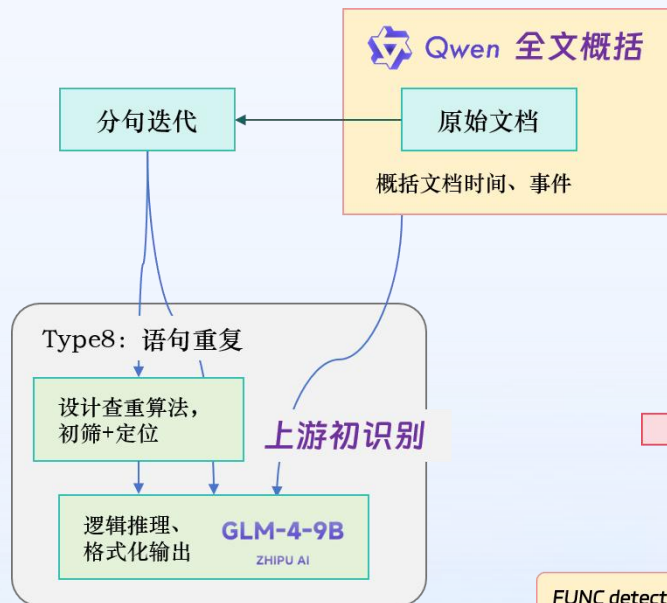
【正确判断】“error_sentence”:[最高限价: 第1包:230万元;第2包:1169元;]

【正确判断】“error_sentence”:[最高限价: 第1包:228.6778元]

【正确判断】“error_sentence”:[《维和防暴队》以3.88元的票房位居榜首]

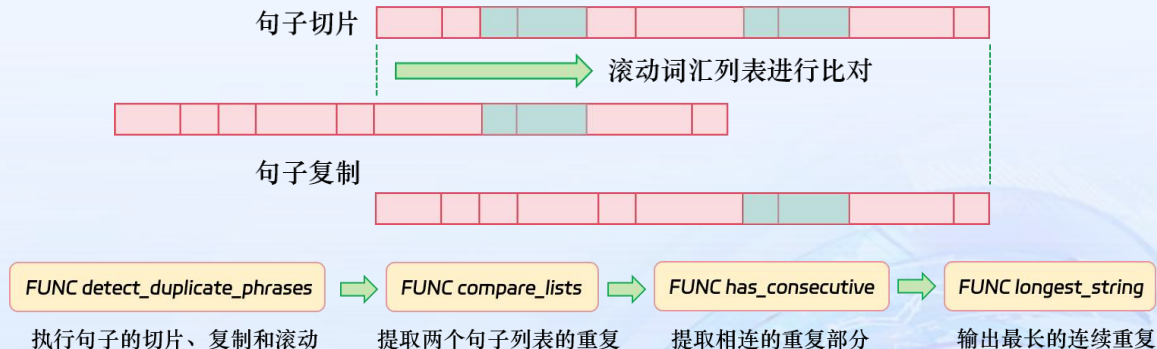
【正确判断】“error_sentence”:[发电设备利用方面, 2024年3月全国发电设备平均利用小时数844万小时, 同比降低2.76%]

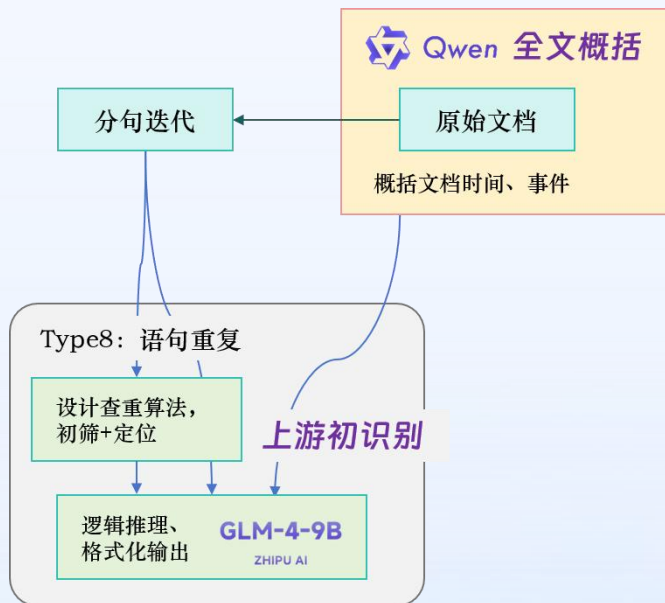
【正确判断】“error_sentence”:[髌关节产品系统年度采购需求量285995个（陶瓷-陶瓷类髌关节102264个、陶瓷聚乙烯类173303个、合金-聚乙烯类1042个）]



定义、解决方案

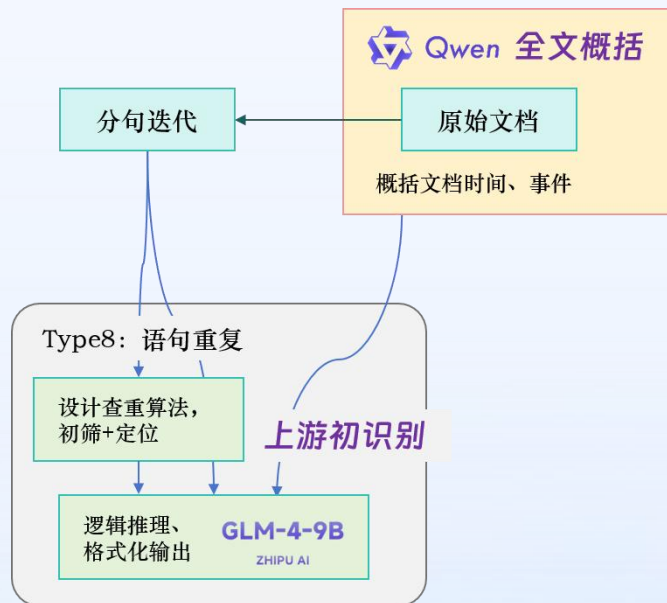
- 我们把这种错误定义为，文本中存在两个完全一致的数值、短语或句子。
- 我们设计了重复识别算法进行初筛，其主要函数如下所示：





Prompt

- 这段文本可能包含人为插入的重复语句作为陷阱，判断这些重复是否多余。以下是段落和可能的重复部分：
 - 段落: {check_sentence}
 - 重复部分: {output}
- 请根据以下标准判断：
 - 如果重复部分对理解内容有帮助，且在段落中有其必要性，则不是重复陷阱；如果重复部分显得突兀且不必要，则可能是重复陷阱。
- 请综合上述信息，你给出的回复需要包含以下这两个字段：
 - 1. TrueOrNot: 如果段落没有重复陷阱，填写 `True`；如果有重复陷阱，填写 `False`。
 - 2. error_sentence: 如没有重复陷阱，该字段输出为空；如有重复陷阱，输出这个句子中包含重复处的最小粒度分句。
- 请按照以下JSON格式来回答：...
- 警告：你的回复将直接用于javascript的JSON.parse解析...



Answer

【正确判断】“error_sentence”:[准入产品物资分类码: 16010418、16012702、16019917、16011507、16011339、16011346、16010418]

【正确判断】“error_sentence”:[依法必须进行招标的项目的投标人有前款所列行为尚未构成犯罪的, 处中标项目金额千分之五以上千分之十以下的罚款, 处千分之五以上千分之十以下的罚款, 对单位直接负责]

【正确判断】“error_sentence”:[26030114氢气压缩机、26030114氢气压缩机之一正常状态。]

【正确判断】“error_sentence”:[降低作业成本30%以上]

【正确判断】“error_sentence”:[携程集团-S]

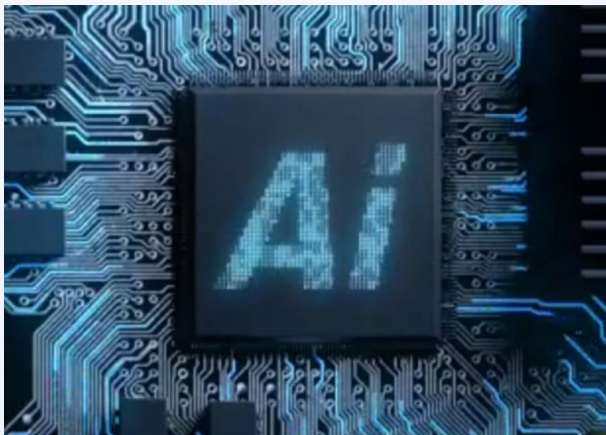
【正确判断】“error_sentence”:[若中标人为2名, 原则上排名第1的分配60%左右的工作量, 排名第1的分配60%左右的工作量, 招标人可根据中标人服务质量情况调整工作量分配。]



04 方案成效

逻辑筛选+

文本概括+金融知识补充+推理（指令跟随）



通过机制判别初筛错误类型和错误句子，进而给出针对性的Prompt，提高了模型识别错误的准确性。



在无微调数据的情况下，充分利用不同LLM的属性，分工执行任务，提高了识别性能。



三LLM分别执行概括、知识补充和逻辑推理任务，保证结构化输出、提高任务执行效率。



05 技术创新

逻辑筛选+

文本概括+金融知识补充+推理（指令跟随）

01

上游初识别

采用pysbd以及模式识别匹配进行分句，建立针对不同类型错误的逻辑判断模型，从而初步筛选问题段落，精细化数据输入，提高模型精度；

02

三LLM交互

三 L L M 交 互： 分 别 利 用 Qwen-Finance-14B的金融属性和 GLM4-9B的理解能力，设计 3LLM分工工作，在有限显存和无外源数据的条件下提高识别能力。

03

下游后处理

针对反复出现的误判进行剔除、对结果进行去重、针对“标书”类文档存在的特征相似的“不、未”和时间错误进行复查。



AFAC2024

感谢您的聆听

队名：不知道叫啥

答辩人：顾国勤