工单助手--开放性工单稽核

LLM原生稽核框架

全网监控中心 谭小兵

2026-01

培育壮大新动能 --- 技术攻关到成果转化

- 中国共产党第二十届中央委员会第四次全体会议，于2025年10月20日至23日在北京举行。中央委员会总书记习近平作重要讲话。这次会议审议通过了《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》。

- 《建议》强调：要坚定不移推动高质量发展。推动高质量发展是《建议》确定的“十五五”时期经济社会发展的主题。要以新发展理念引领发展，**加快培育新动能**，促进经济结构优化升级，做优增量、盘活存量，保持质的有效提升和量的合理增长，推动经济持续健康发展和社会全面进步。坚持把发展经济的着力点放在实体经济上，优化提升传统产业，**培育壮大新兴产业和未来产业**，促进服务业优质高效发展，构建以先进制造业为骨干的现代化产业体系。进一步全面深化改革，着力破除制约高质量发展的体制机制障碍，努力在重点领域和关键环节取得新突破。发展新质生产力是高质量发展的内在要求。**“十五五”时期要把发展新质生产力摆在更加突出的战略位置**。坚持教育科技人才一体发展，提升国家创新体系整体效能，大幅提高科技自立自强水平，推动科技创新和产业创新深度融合。



目录

背景介绍

场景建模

核心挑战

效果验证

附：工单稽核中的特色功能及演示

目录

背景介绍

场景建模

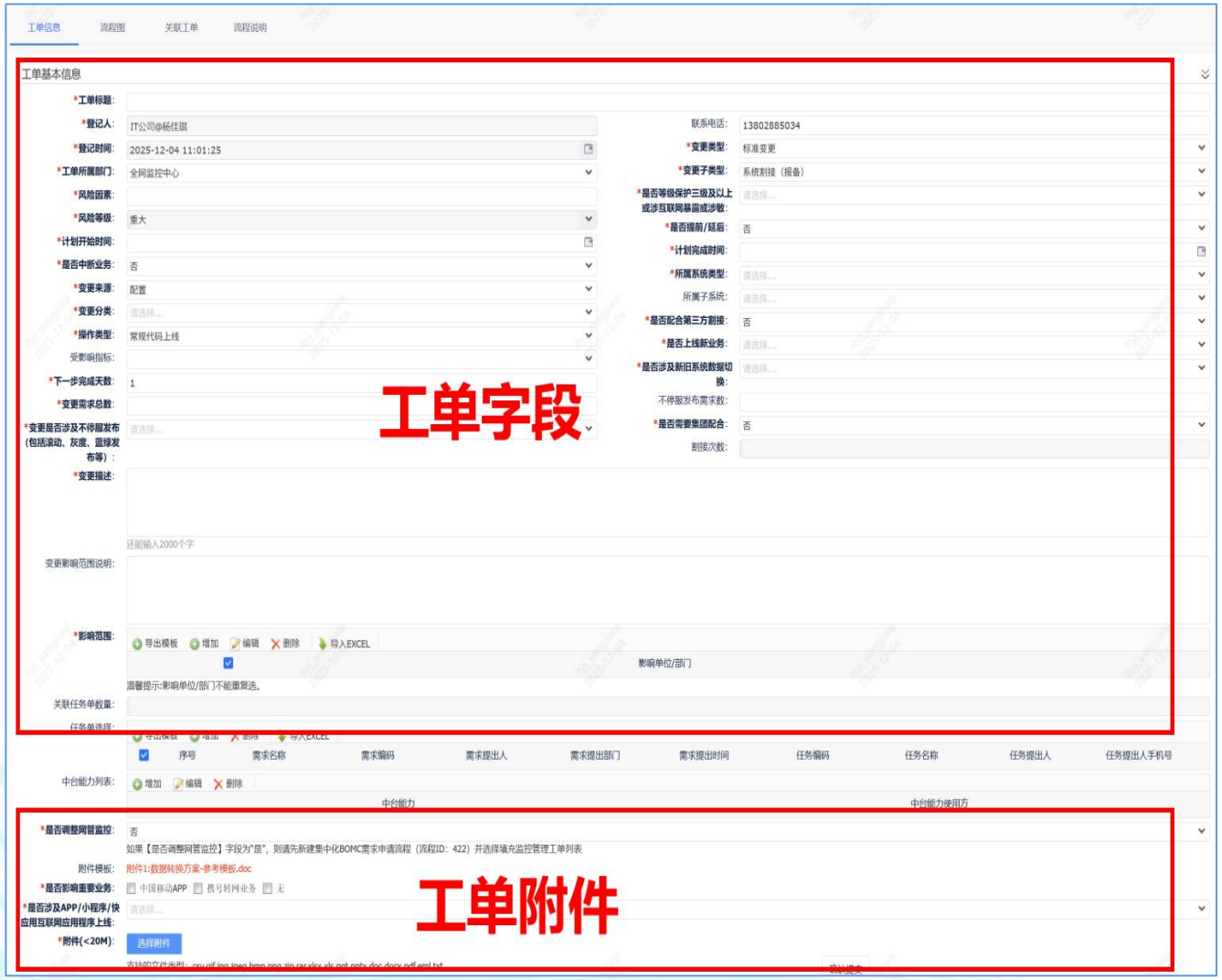
核心挑战

效果验证

附：工单稽核中的特色功能及演示

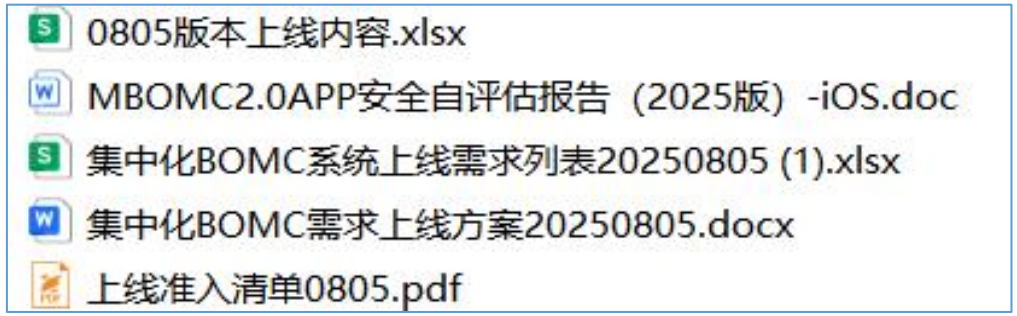
背景介绍 - 什么是开放性工单稽核

工单稽核：判断工单的**内容**是否符合**稽核规则**的要求；工单**内容**：工单附件 & 工单字段；；
开放性：1.**附件内容**是开放性的；2. 工单**某些字段**是开放性的；3. **规则**是开放性的；4. 当然确定性的字段也可稽核；



稽核源	稽核规则说明
字段：变更类型、附件列表 附件内容	稽核点：信息完整性-基本方案 【稽核信息源】工单字段“变更类型”、 “附件列表”，附件内容 【稽核前提】当工单字段“变更类型”为简单变更时，跳过本稽核点，不存在风险。 【稽核标准】 字段“变更类型”为“紧急变更”或“标准变更”时进行如下校验： “附件列表”或附件内容中应包含以下信息点： 大项：变更方案、应急或回退方案； 中项：变更基本信息（内含小项：变更原因、内容、变更类型、风险评估、业务影响评估等）、时间安排、人员安排、变更实施步骤、业务验证、上线测试报告等内容； 缺少小项内容，判定为低风险，并告知用户具体缺失的小项名称。 缺少中项内容，判定为中风险，并告知用户具体缺失的中项名称。 缺少大项内容，判定为高风险，并告知用户具体缺失的大项名称。 若在字段 “附件列表” 中有疑似符合的文件标题但未获取附件内容信息，应按对应级别示警（缺少小项示低风险、中项示中风险、大项示高风险），并告知用户 “因无法判断是否信息完整，请人工介入确认”。 注意点1：“变更”、“割接”、“上线”三词的含义相同，若有割接方案、上线方案等类似表述，也符合要求。 注意点2：应急方案或回退方案一般编写于变更方案内，请着重在变更方案或割接方案里校验是否包含，不要求单独附件。

稽核规则



工单附件实例

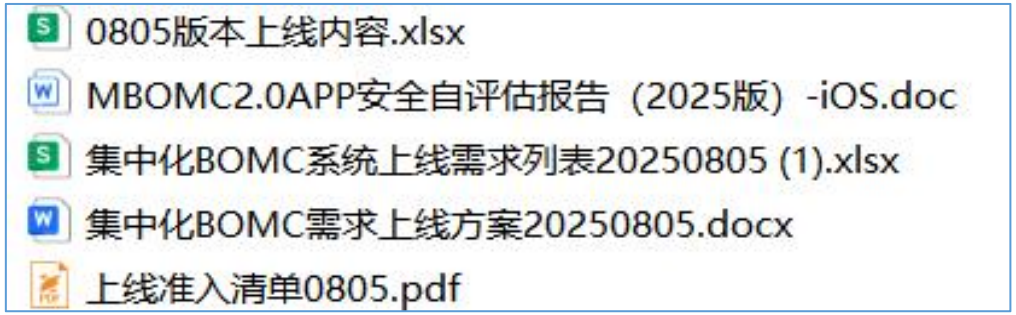
背景介绍 - 业务价值

全网变更单共涉及18个环节，其中**8个环节（45%）**涉及工单信息审核，日常工单审核人员需要比对附件信息与工单填写信息是否一致、附件信息是否完整、正确，各环节审批人员需耗费大量时间（**至少30分钟**）打开附件核查信息。工单稽核在**减少人工核查耗时、降低变更风险**的同时可确保**审计、管理要求**在日常工作中得以**落实**，实现**降本增效**。

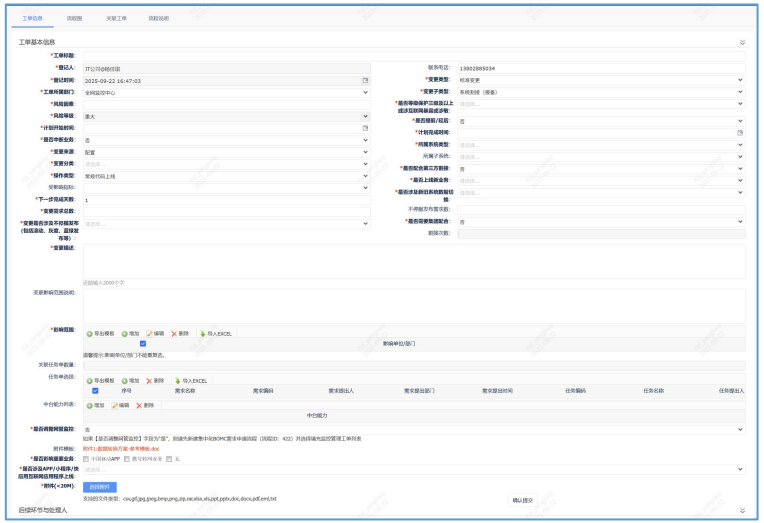
变更申请	计划中
变更主管审批	
上线前测试	
开发总监审批（管信个性化）	等待审批
安全管理员审批	
开源管理员审批	
外部系统会签中	
会审	
变更经理审批	
部门领导审批	
全网监控审批	

审批环节

仅变更单一项，**月均**可节约：
1800张×0.5时÷8时/人天
=**112.5** 人天



工单附件



工单字段

S-GOC涵盖200+种流程，月均10w+张工单，其中全网变更单基于变更管理办法、运维管理办法等要求制定的**统一流程**，各生产部门、各省均通过该流程进行割接报备，月均1800张，具有**强通用性、受众广、可复制性高。**

背景介绍 - 需求分析

功能性稽核需求：基于管理规范与日常运维经验梳理出14条稽核规则，主要包含信息完整性、信息合理性、信息一致性、其他稽核点和个性化稽核点，通过大模型对**14条稽核规则**进行匹配、稽核、反馈最终结果。

非功能性能需求：高稳定性、高准确率、极速响应

信息完整性：

- 通用信息
- 安全信息
- 开源场景
- APP场景
- 超割接时间审批场景

工单字段

工单附件

信息合理性：

- 变更类型
- 变更风险
- 变更步骤
- APP场景

工单字段

工单附件

信息一致性：

- 变更时间
- 需求数量

工单字段

工单附件

其他稽核点：

- 割接次数
- 工单标题规范性

工单字段

个性化稽核点：

- 变更评审表

工单字段

工单附件

- **稳定性**：近幂等性，同样输出给出同样的输出；
- **扩展性**：规则扩展、场景泛化等；
- **性能**：必须在限定时间内完成，越快越好；

- **可信可解释性**：严肃场景，给出理由、证据来源；
- **可维护性**：易于理解、修改、调试和升级：如支持图片；
- **可观测性**：过程变量、报表统计、指标量化分析；

目录

背景介绍

场景建模

核心挑战

效果验证

附：工单稽核中的特色功能及演示

场景建模 – 主要挑战

当前的工单稽核主要挑战来自于两个方面：一是LLM本身的限制；二是业务场景的挑战。

LLM本身限制

LLM本质是概率模型

- LLM本质是一种概率模型，存在幻觉，其输出并不稳定。
- Prompt的改动容易牵一发动全身；尤其当稽核规则变多时，规则间的冲突也增多，稳定性难以控制；

模型上下文长度有限

- 模型的上下文长度有限，附件内容超过上下文长度时，造成附件内容的截断，从而出现”**看不全**“的问题，导致稽核准确率下降。
- 使用过大的上下文会带来**噪声干扰**，也会增大资源消耗，影响效率。

不擅长数值计算

- 稽核场景中，存在大量数据计算：如，间隔不超过7天；金额小于3000等；
- 数据计算要求绝对准确，把数据计算完全交给大模型，会存在准确率低、稳定性差的问题。

业务场景挑战

业务规则复杂

- 规则包含前置条件；
- 规则包含自定义概念；
- 规则包含业务背景知识；
- 规则包含多输入源组合逻辑：字段、附件、知识库；
- 规则要求跨文档稽核；

附件复杂

- 种类多：pdf/word/excel/图片等；
- 模态多：文本、表格、图片等；
- 内容多：一个附件可能包含几百页内容。如6456485这个工单中，同时包含一个244页的附件和一个111页的附件，总长超1000K；

业务要求高

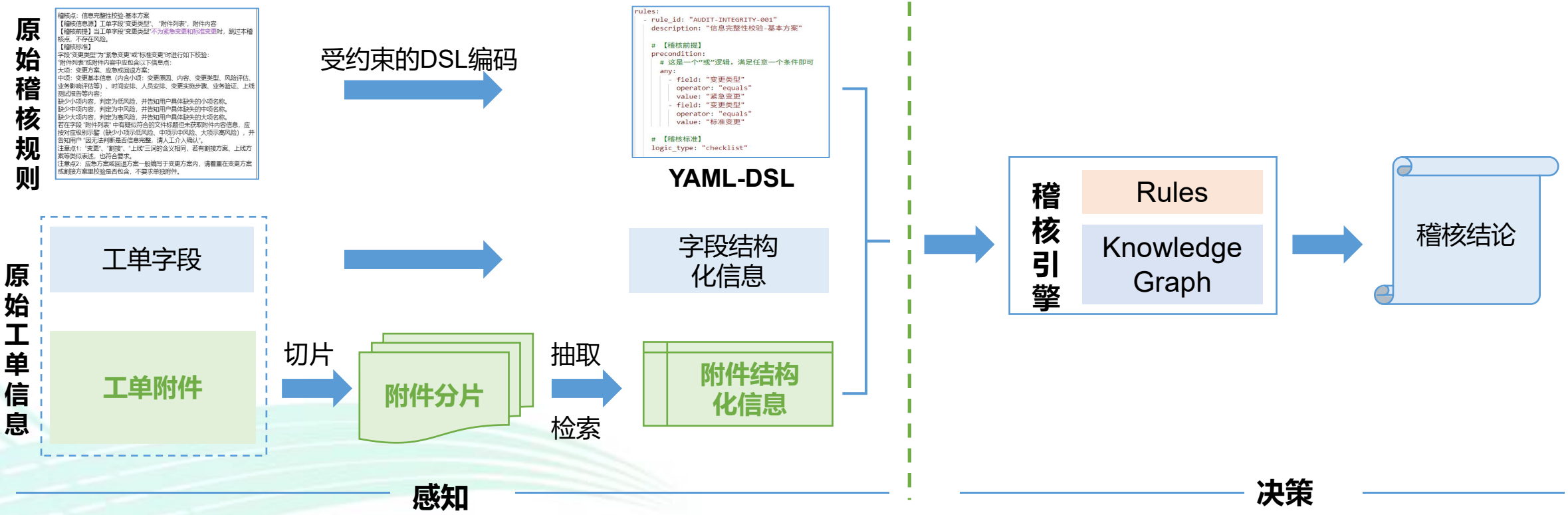
- 审计稽核是严肃场景，要求可信可解释性强；
- 规则变动频繁：要求扩展性强，可维护性强；
- 稳定性要求高：复稽核时，必须保持高一致性；
- 准确率高

场景建模 - 稽核方案 - 神经符号融合稽核系统

- 本方案基于**感知**与**决策**分离的思想，充分利用大模型的优势，规避大模型的劣势。采用 **神经符号混合**（Neuro-Symbolic）架构：左侧以 LLM 为核心的“神经感知系统”，右侧以 YAML-DSL 规则与知识图谱为核心的“符号决策系统”，实现感知与决策解耦。

- **神经** = 神经网络类系统，主要就是大模型 / 深度学习部分。它主要负责工单切片理解、摘要、分类；NL2DSL；非结构化文本处理、跨文档检索等。它具备如下几个特征：
  - **软约束、概率性**：同样的输入有一定随机性（概率分布采样），更擅长“模糊”、“相似”、“语义”层面的判断；
  - **强感知能力**：能看懂长文本、多文档、弱格式化信息；
  - **不透明**：内部是参数和向量，逻辑不易直接追踪。

- **符号** = 显式的规则 / 逻辑 / 知识结构，如DSL规则及引擎、知识图谱等。它具备如下几个特征：
  - **强确定性**：同样的输入 → 一样的输出；
  - **可解释**：每一条命中规则、每一条路径都可以还原给审计/专家看；
  - **可维护**：可以“按条修改”，方便治理和审计。



目录

背景介绍

场景建模

核心挑战

效果验证

附：工单稽核中的特色功能及演示

实验效果

2025年4月开始业务场景调研与筛选；2025年6月发布了v0.1版本；2025年11月发布v0.2内测版本。
当前共有13条稽核规则，54项稽核点（高9中19低26）

	2025-04	2025-06	2025-09	2025-11
版本发布	• 业务背景调研 • 业务场景筛选 • 技术原型开发	• v0.1版本发布 • 问题收集 & 功能优化 • 优化方案预验证	• v0.2版本原型开发 • v0.2版本生产化开发	• v0.2内测版本发布 • 市服和全网监控中心试用
风险维度	/	/	/	• 高风险：**P=92.3%，R=89.7%** • 中风险：P=94.7% • 低风险：P=95.1%
工单维度	/	• Accuracy：72.1% • 工单维度统计 • 以中高风险为准	/	• Accuracy：90.1% • 工单维度统计 • 以中高风险为准

试用效果

V0.2版本2025年11月开始内部试用；期间由于封网的原因，截止1219，真正试用的只有三周。
当前共有13条稽核规则，54项稽核点（高9中19低26）

	评价标准	1124-1130	1201-1205 (-1比例：49.1%)	1215-1221 (-1比例：75.54%)	1223-0104 (-1比例：71.31%)	0105-0118 (-1比例：74.82%)
主要问题	• 一个工单总分100 • 最低得0分	• 压缩包（超量文件） • 时间理解 • 变更方案识别 • 安全评估报告	• 压缩包（超量文件） • 测试报告 • 变更方案识别 • 互联网等级保护	• 压缩包（超量文件） • 变更方案识别 • 测试报告 • 市服评审表	• 压缩包（超量文件） • 变更方案识别 • 时间问题	• 压缩包（超量文件） • 变更方案识别
评价方法一	• 错一个高风险扣100分 • 错一个中风险扣100分 • 错一个低风险扣3分	80.5	88.8	81.3	86.0	94.1
评价方法二	• 错一个高风险扣20分 • 错一个中风险扣10分 • 错一个低风险扣3分	95.3	95.5	92.6	-	-

目录

背景介绍

场景建模

核心挑战

效果验证

附：工单稽核中的特色功能及演示

特色功能1 - LLM-DSL代码生成（Difference Automata）

不同场景的业务稽核规则迥然不同，如何解决场景迁移门槛高、能力复用难、定制化开发多、迁移周期长的问题：

- **打通【自然语言】 --生成--> 【DSL-YAML规则】 的技术路径，为实现零门槛迁移复用奠定了基础。**

LLM-based code generation and system migration in language-driven engineering

Daniel Busch[1] · Alexander Bainczyk[1] · Steven Smyth[1] · Bernhard Steffen[1]

Accepted: 4 April 2025 / Published online: 29 April 2025
© The Author(s) 2025



特色功能1 - LLM-DSL代码生成



稽核结果

稽核报告

原始稽核结果(JSON)


```
[
  {
    "rule_id": "AUTO_差旅报销超过3000元必须有正规发票",
    "rule_name": "差旅报销超过3000元必须有正规发票",
    "status": "FAILED",
    "violations": [
      {
        "risk": "HIGH",
        "message": "该报销单为差旅报销，金额超过 3000 元，但未提供正规发票，属于高风险报销，请人工重点核查。",
        "evidence": {}
      }
    ]
  }
]
```

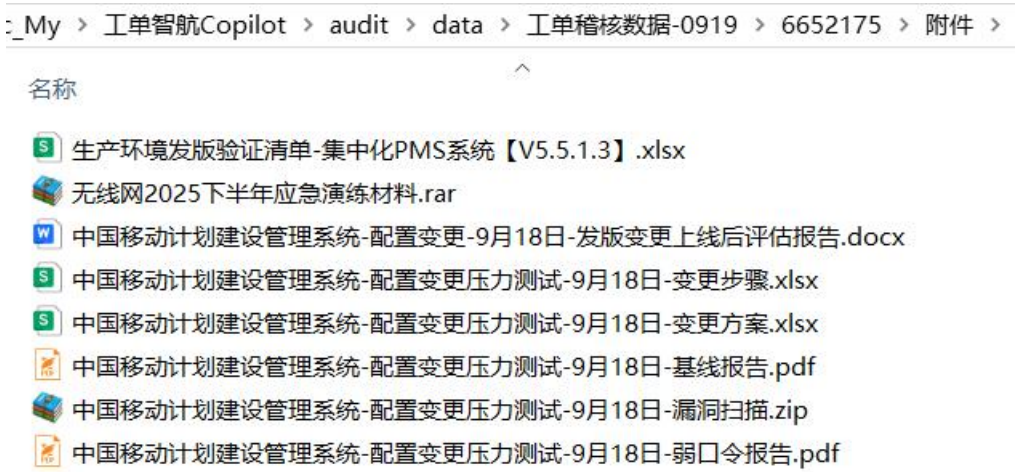
特色功能2 - 跨文档稽核

链式稽核：在工单稽核中，存在一种典型的跨文档稽核场景：在一个文档中通过【详见xxx、见xxx、如xxx】引入别的文档。

- **需要精准语义识别**，如“步骤文档”，其实是指《中国移动计划建设管理系统-配置变更压力测试-9月18日-变更步骤.xlsx》
- **混合文档类型**：在EXCEL中引入WORD、在WORD中引入图片等；加大了稽核难度；
- **链路追踪**：要求稽核系统自动进行多跳文档链路追踪；

【割接紧急程度的评估】	本次变更属于常规版本发版，且已经过功能测试、业务验证，需各微服务同步上线验证
【上线前测试】	已进行测试，详见测试报告
【割接实施步骤】	详见变更步骤文档
【上线验证】	验证步骤： （1） 通知葛智宁、相关微服务产品经理进行测试，进行以下测试验证： http://cpms.hq.cmcc/pms/index.html#/login （2）通知相关业务实施顾问进行测试，进行以下测试验证： http://cpms.hq.cmcc/pms/index.html#/login 冒烟测试验证内容： 1. 生产正常登录、注销； 2. 待办、已办、列表可以正常打开，且速度正常。

工单 号：6652175



特色功能3 - 多模态稽核

多模态稽核：在工单稽核中，存在典型的多模态稽核场景：

- **5G随e签**：各类文件中都可能存在：单个签名、多个签名；如word/excel/pdf/jpg等；
- **邮件截图**：各类文件中都可能存在邮件截图：报备截图、审批通过截图、沟通过程截图等；
- **系统页面**：各类文件中都可能存在各种系统截个图：作业系统、审批系统、命令行系统等；
- **其它类型**：流程图、架构图、拆线图、指标数据图等；

- **位置还原（Layout Recovery）**： 自动解析各类文档（Word/Excel/PDF），精准定位图片插入坐标，还原其在文档流中的物理位置；如合并单元格场景；

- **语义对齐（Semantic Alignment）**： 深度识别图片的业务含义与文本内容，将视觉信息与文档上下文进行逻辑关联，构建图文互证的完整证据链。如：识别图片的类别、用途与内容：“这是一张手写签名，签名人是张三”、“这是一张邮件截图，内容为变更审批，审批结果为通过”、“这是一张邮件截图，内容为变更报备”

- **复杂抗噪（Noise Resistance）**： 具备复杂文档环境下的抗干扰能力，能在多类型图片共存的场景中精准锁定目标图片，剔除冗余信息干扰，输出高置信度的标准化结果。

图片类型	典型示例	用途说明
签名（重要）		判断工单流程是否合规、是否有对应责任人签署。
邮件截图（重要）	 可以。	判断工单是否经过交接、申报或相关沟通流程。
工单流程截图（重要）		判断工单是否已按要求进入流程节点、完成审批链路。
其他	系统操作页面截图、代码/日志截图、架构/拓扑/流程图、PPT 截图、公司标识等	此类图片通常与工单核心判断无直接关系，仅作简单内容描述或不做处理。

特色功能3 - 多模态稽核


提出部门：平台能力中心
项目经理：
总监签字：
日　　期：
2025年09月18日

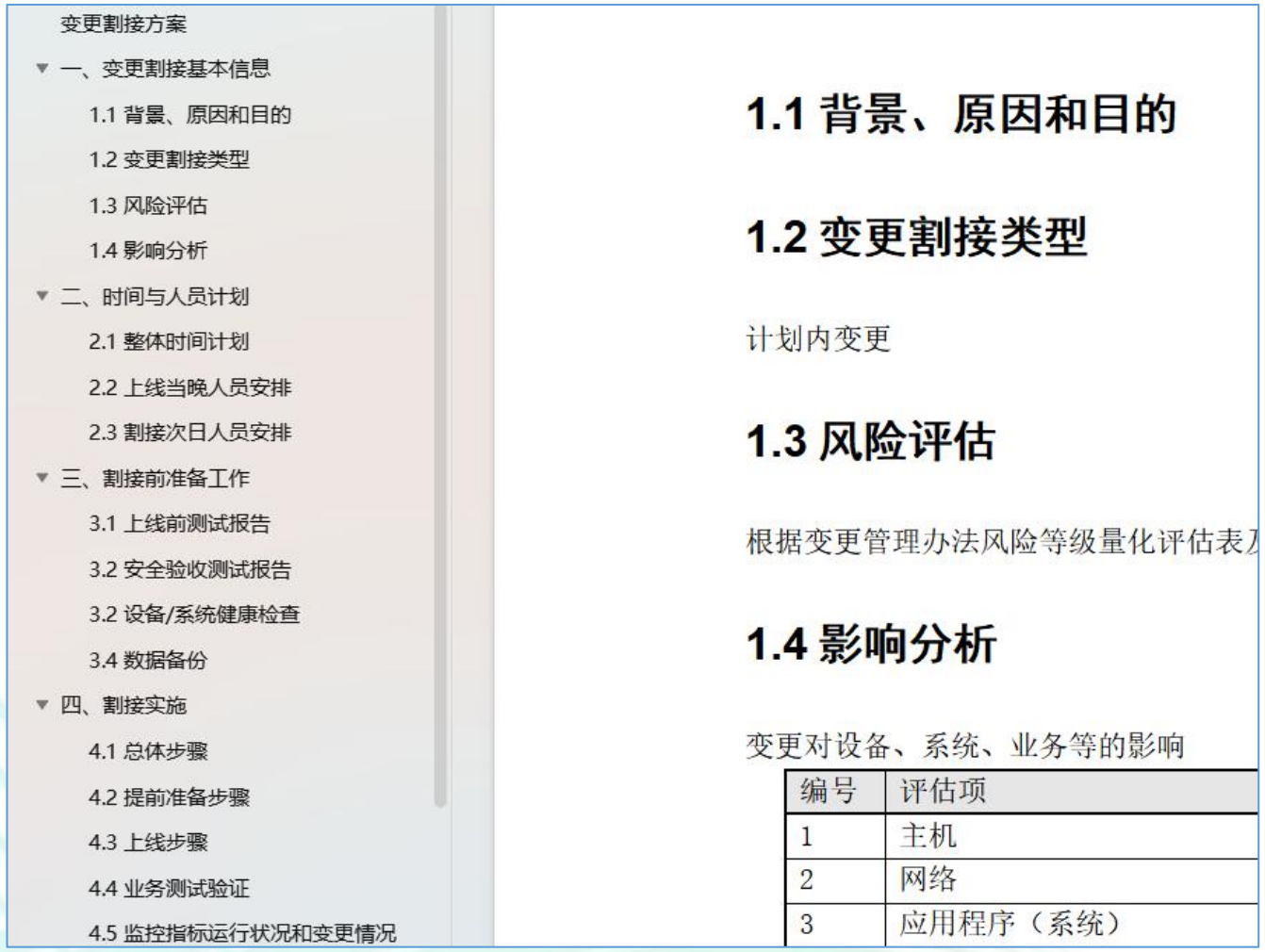

提出部门：平台能力中心
项目经理：
总监签字：
日　　期：
2025年09月18日
【图3-代码/日志页面】
【代码/日志页面】这是一张

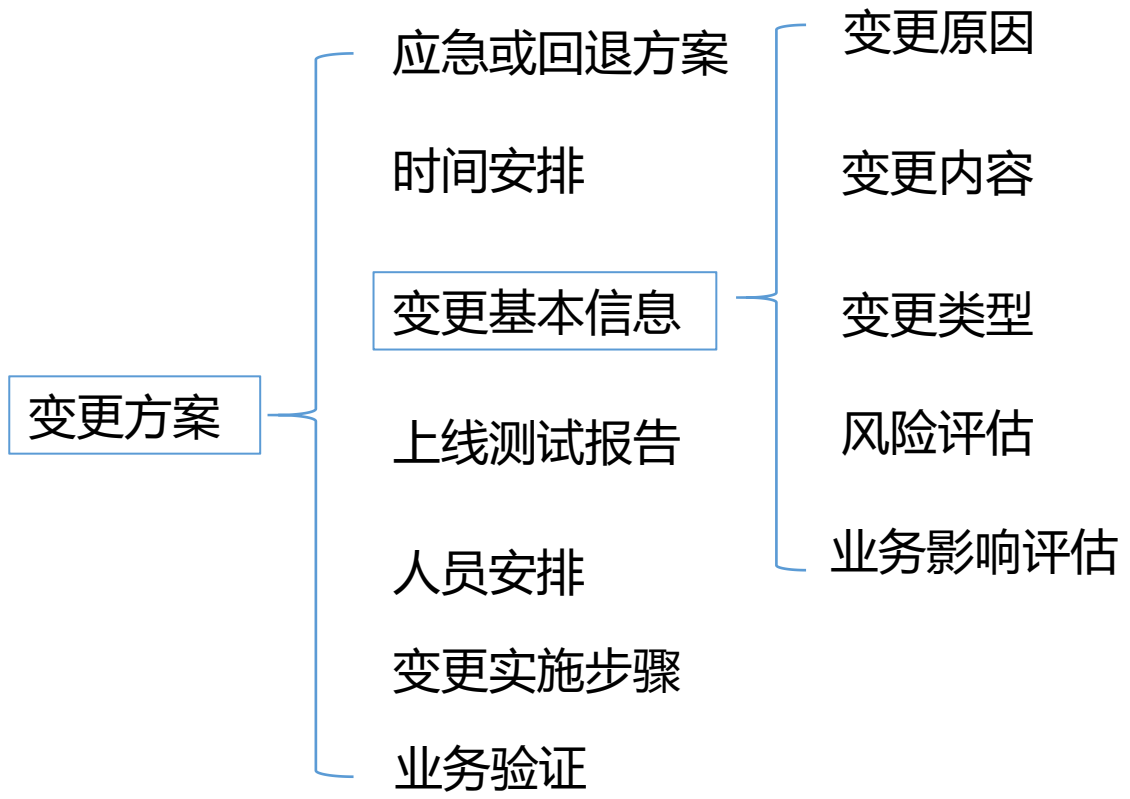

提取文字
更多功能
4. 遗留问题及风险
问答功能由于研究院大模型在维护中，因此暂时无法进行正常的问答操作。
5. 测试结论
集团党建系统测试无缺陷遗留，测试通过。
测试人员：


3.12 办理后待办消失
点击处理跳转文章页
【图26-网页截图】【网页截图】这是一张待描述图片
数据进入已办事项中
【图27-代码/日志页面】【代码/日志页面】这是一张待描述图片
4. 遗留问题及风险
问答功能由于研究院大模型在维护中，因此暂时无法进行正常的问答操作。
5. 测试结论
集团党建系统测试无缺陷遗留，测试通过。
测试人员：【图28-手写签名】【签名识别】

特色功能4 - 复杂逻辑

变更方案完整性稽核，具备一个复杂的”**决策树**“逻辑。先判断变更方案是否存在，再判断各章节是否健全；最后再判断子章节是否完整。





谢谢~