

# 从“按天”到“秒级”大模型 重构医疗理赔审核落地实 践

演讲人：刘升平

云知声 / 研发副总裁

**AiCon**  
全球人工智能开发与应用大会

# 目录

01

保险理赔医疗审核场景分析

02

智能文档处理

03

理赔医疗审核

04

总结与启发

# 极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

北京

1200人

**QCon**

全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

深圳

1000人

**AiCon**

全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

北京

1000人

**AiCon**

全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

4月

6月

8月

10月

12月

**AiCon**

全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

上海

1000人

**QCon**

全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

上海

1200人

# 01 保险医疗理赔场景分析

# 医疗相关险种



保障合同约定的重大疾病  
(如癌症、心脏病、脑中  
风等)

重大疾病险



百万医疗险, 防癌险, 惠  
民保, 团体补充医疗险  
等。保障住院医疗费用、  
特殊门诊、住院前后门急  
诊等, 可报销医保目录外  
费用

医疗险



车险案件中出现人伤情  
况, 保障本次事故中产生  
的医疗费用.

车险人伤



保障因意外伤害产生的医  
疗费用, 包括意外身故/伤  
残、意外医疗

意外险

# ■ 保险理赔-材料准备

## 理赔材料

**基础材料：**理赔申请书、身份证件、银行卡、保险合同

**医疗材料：**病历、诊断证明、医疗费用发票、费用清单、出院小结等

**其他材料：**意外事故证明（如交通事故认定书）、受益人关系证明等

## 理赔条款

**免责条款：**保单中约定的免责条款，如：  
牙科保健与牙科治疗

**既往症约定：**指在第一次投保本保险前罹患的被保险人已知或应该知道的有关疾病或症状。

**理算规则：**医保内赔付比例、医保外赔付比例、起付线、封顶线

# ■ 保险理赔-医疗审核

## 一致性

- 清单收费明细和医嘱不一致
- 检查/检验收费和报告不一致
- 检查
- 手术收费和手术记录不一致
- 无会诊单的中医用药
- 辅助项目无依赖主项目

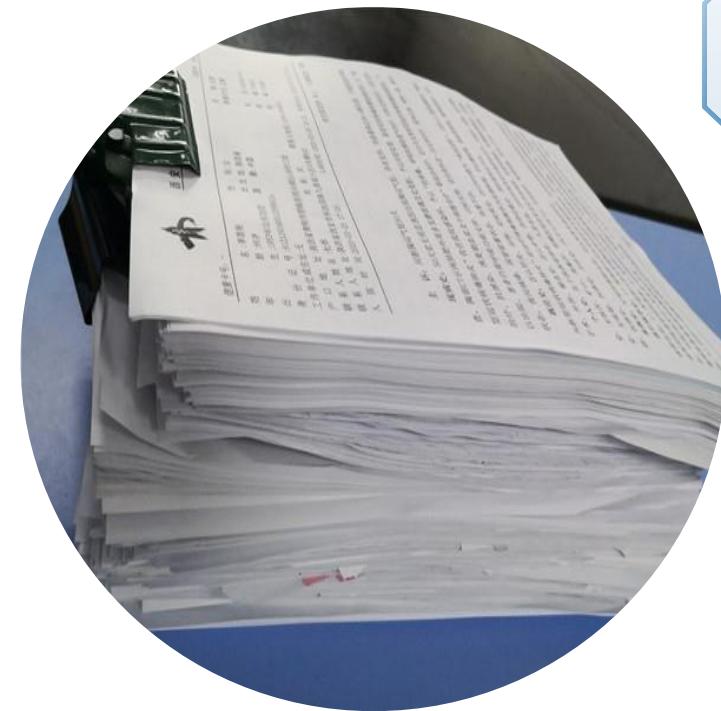
## 合规性

- 限制性用药违规使用
- 诊疗违反限制规范
- 同次多部位影像学检查未执行折价政策
- 手术收费未执行折价政策
- 限人群/年龄项目违规使用
- 收费数量超疗程规定
- 违规带药

## 合理性

- 护理记录不合理
- 定量检查不合理
- 无指征检查/检验
- 项目适应性别和病人不符
- 过度检查/检验
- 分解住院
- 低标准住院
- 收费项目频次超常规

# 现状剖析-医疗审核标准化及降本增效需求



医疗文书**数量大**，  
很难审核全面，且  
标准统一。

医疗**专业性强**，  
需了解疾病、症  
状、药品知识进  
行判断。

医疗文书**内容繁  
杂**，需理解全文  
书进行审核判断。

人工操作

审核  
标准化



核赔  
准确性

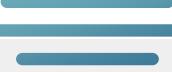


审核  
效率



AI自动

投入  
成本



# 第三方人工审核服务 VS. 医疗AI审核



上传及录入



OCR识别



名称标准化及对码



医疗审核

## 第三方人工审核服务

- 所有文书混在一起
- 主要看费用清单和发票单据

- 通过逐个定制模板进行文字识别及抽取，通用性差

- 人工逐个进行名称标准化录入

- 主要审核费用清单文书，依靠人员经验进行审核

## 医疗AI审核系统

- 智能文书分类
- 判断是否缺少关键文书
- 智能判断重复上传文书并剔除

- 基于模型实现识别图片中的关键信息及关系，完成信息抽取

- 用对码引擎直接把医院项目名称改成医保标准名称，并标记甲乙丙类型

- 理解全病历文书，对用药品、诊疗合理性、一致性等进行判断

# 医疗理赔自动化(AI)的技术挑战(1)- 文档处理

## OCR与版面理解

- 低清/倾斜/噪声
- 印章/水印遮挡;
- 表格/自由文本混排;
- 手写体
- 跨页表格与页眉页脚干扰
- 页码错乱与旋转。

## 信息抽取与标准化

- 实体与术语多样: 诊断/术式/指标/药品/剂量, 缩写与同义写法多
- 版式复杂与噪声: 表格+自由文本+手写+盖章/涂黑, 跨页断裂
- 数值与单位异构: mg、IU、片、次/日等混用; 字段命名不统一

# 医疗理赔自动化(AI)的技术挑战(2)- 自动审核

## 时间线与一致性推理

- 时间表达多样：绝对/相对/区间/模糊/跨机构，记录可能演变或冲突
- 条款口径差异：阈值、时间窗、最短间隔、复查频次因指南不同而变
- 证据分散：同一对象线索散落在出院小结、影像/检验、手术/病理、处方等

## 条款推理与可解释性

- 条款含必要/排除/阈值/时限等复合逻辑，且跨文书取证
- 黑盒模型难以支撑审计可解释与合规追责

# 医疗AI审核SaaS平台系统架构图

## 商险理赔医疗费用审核

核心系统对接

案件创建

影像识别

影像分类

质量检测

关键信息抽取

发票审核

智能审核

报告下载

统计分析



医疗大模型

影像件分类  
及OCR

项目名称  
对码

医疗文书  
阅读理解

医疗知识  
推理及应用



医疗知识图谱

多种数据格式、全量医疗文书数据

# 02 智能文档处理

# 文档质量检测与分类

影像材料 (总计30张)

调整分类

\* 出院记录 (1张) 是指经治医师对患者此次住院期间诊疗情况的总结性记录, 也叫出院小结。



\* 门急诊发票 (13张) 是指完成门诊诊疗后由医院开具的收费票据。



门急诊病历 (1张) 是指门(急)诊诊疗工作中形成的病历资料的总和。



序号	报案号	姓名	就诊类型	总费用	总扣除金额	总扣除比例	案件状态	案件类型	问题说明	住院诊断
1	CS20001627	张三	住院	0	0	0%	人工未审核	人工必审案件	1.未识别到费用清单 2.总费用异常 (存在≤0)	软组织疾患

## 影像自动分类

对上传的影像件进行文书自动分类, 支持医疗单证、票据、证卡等44种文书类型。

## 多页、一页多图自动拆分

连续多页PDF文件自动拆分;  
一页内多张文书内容自动拆分、识别。

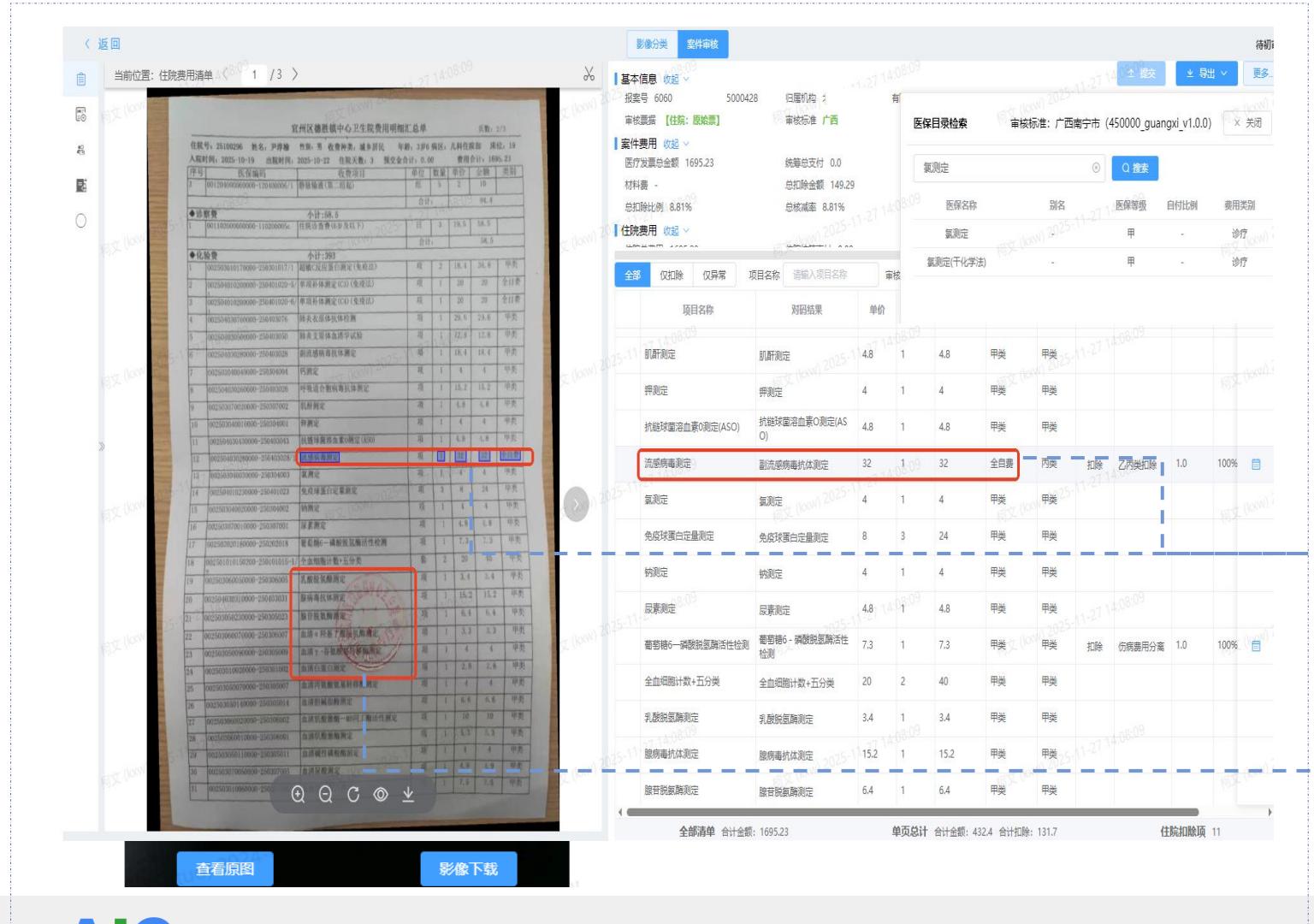
## 重复文书判断

重复文书进行判断并去除, 避免重复上传导致的扣减金额重复计算。通过md5值和识别内容重复度进行双重判断。

## 内容质量、完整性判断

对上传的模糊票据、字体过小等情况进行判断和质量提示, 并支持效果增强、图像摆正;  
辅助判断必要票据的完整性。

## 文档内容识别及抽取



## 医疗通用版式识别

全国万余家医院医疗文书差别很大，费用清单文书有表格、无表格、无表头等多种情况，排版也大不相同，通过通用版式模型识别。

## 便捷审核工具

支持人工增删改，抽取结果原图定位；识别置信度低的进行额外提示。

## 印章、水印消除

对印章部分、水印部分进行消除处理，降低对名称及数字识别的影响。

# 文档溯源，OCR低置信标红，局部修正

将“重录整张票面”变为“只改错字”，单票复核时间从平均3分钟缩短至30秒

当前位置：住院费用清单 <504> 2 / 4 >

影像分类 案件审核

影像模糊

基本信息 收起 报案号 60600114511232025000197 归属机构 陈高瑞 审核类型 【住院：电子票】 伤/害姓名 陈高瑞 审核标准 广西贺州市 乙丙扣除方式 【住院：按项目】

案件费用 收起 医疗发票总金额 2906.73 纳管总支付 0.0 药品费 585.74

材料费 46.99 总扣除金额 489.2 医保外扣除总金额 192.47

总扣比例 16.83% 总核减率 16.83%

住院费用 收起

全部 仅扣除 异常 项目名称 请输入项目名称 审核规则 请选择审核规则 审核

项目名称	单价	数量	金额	影像	审核等级	扣除标记	审核
幽门螺杆菌测定	19.00	1	19.00		甲类		
病毒学试验	20.00	1	20.00	甲	甲类		
病毒学试验(柯萨奇病毒)	20.00	1	20.00	甲	甲类		
肺炎支原体血清学试验	16.00	1	16.00	甲	甲类		
肺炎衣原体血清学试验	37.00	1	37.00	甲	甲类		
一般细菌培养及鉴定	40.00	1	40.00		甲类		
血培养及鉴定	90.00	1	90.00		甲类		
雾化吸入(6岁以下)	9.60	14	134.40				
Ⅱ级护理	20.00	7	140.00	甲	甲类		
机械辅助排痰(6岁以下)	23.00	13	299.00				
硬性耳内镜检查	58.70	1	58.70				
硬性鼻内镜检查	63.50	1	63.50	甲	甲类		
指脉氧监测	3.40	7	23.80		乙类		

全部清单 合计金额：2780.6 单页总计 合计金额：1207.63 合计扣除：0

智能标红：

低置信字段自动标红，集中展示于侧边复核栏。



局部修正：

点击可疑字段，原图局部放大，支持快速修正。



记录留存：

修改记录实时保存，生成复核日志，便于追溯。



# 名称标记化及对码



基于国家医保局要求, 使用国家统一的医保信息业务编码标准是医疗审核最基础要求

## 文书识别结果

常规心电图检查(十二通道)  
眼压检查  
静脉穿刺置管术  
血清直接胆红素测定1  
尿素测定1  
多排螺旋CT平扫  
一次性使用无针密闭输液接头保护器  
下肢动脉血管彩色多普勒超声  
血清高密度脂蛋白胆固醇测定

## 对码名称

十二通道常规心电图检查  
主眼检查  
经皮穿刺股静脉置管术  
直接胆红素(D-Bil)测定  
尿素(Urea)测定  
骶髂关节X线计算机体层(CT)平扫  
无针密闭输液接头  
四肢血管彩色多普勒超声  
高密度脂蛋白3胆固醇测定

## 收费等级

甲类  
甲类  
甲类  
甲类  
甲类  
甲类  
乙类  
乙类  
乙类  
丙类

## 名称自动映射医保目录名称

医院数据繁杂多样, 表达习惯、书写颗粒度差别很大, 结合医疗大语言模型丰富的专业知识, 通过对码引擎自动映射到地市医保目录标准名称。

## 按地市要求判断甲乙丙信息

费用清单上有甲乙丙信息的优先采用, 没有标注的按地市三目规范处理。

已整理超300个地市级三目规范。定期从官网获取更新。

INSG 标普传媒

# OCR技术范式演进

## 多模型 工作流

方案：通过版面矫正、分析、OCR检测、识别、表格识别、KIE信息抽取等多个小模型组合完成任务

缺点：

- 错误传播严重：
- 无法解决疑难场景：套打单据、印章干扰
- 泛化性差：特殊版面易出错

优点：资源占用少，泛化性差

## 通用VL大模型

方案：使用通用VL大模型，通过Prompt工程进行信息提取

缺点：

- 资源占用极高（小参数量模型效果差）
- 无法理解复杂抽取规则
- 幻觉严重
- 抽取结果回溯效果极差

优点：泛化性强

## 端到端专用大模型

方案：采用小参数量通用VL大模型架构，根据业务场景，定制训练任务及数据，微调训练端到端模型

优点：

- 资源占用少（1~3B模型即可取得较好效果）
- 能有效理解复杂抽取规则，满足业务场景需求
- 精准定位和回溯抽取结果在图片中的位置
- 具有较强的泛化性

# OmniDocBench 评测结果

Model Type	Methods	Parameters	Overall↑	Text <sup>Edit↓</sup>	Formula <sup>CDM↑</sup>	Table <sup>TEDS↑</sup>	Table <sup>TEDS-S↑</sup>	Reading Order <sup>Edit↓</sup>
Pipeline Tools	Marker-1.8.2 [45]	-	71.30	0.206	76.66	57.88	71.17	0.250
	Mineru2-pipeline [14]	-	75.51	0.209	76.55	70.90	79.11	0.225
	PP-StructureV3 [10]	-	86.73	0.073	85.79	81.68	89.48	0.073
General VLMs	GPT-4o [7]	-	75.02	0.217	79.70	67.07	76.09	0.148
	InternVL3-76B [46]	76B	80.33	0.131	83.42	70.64	77.74	0.113
	InternVL3.5-241B [47]	241B	82.67	0.142	87.23	75.00	81.28	0.125
	Qwen2.5-VL-72B [24]	72B	87.02	0.094	88.27	82.15	86.22	0.102
	Gemini-2.5 Pro [48]	-	88.03	0.075	85.82	85.71	90.29	0.097
Specialized VLMs	Dolphin [3]	322M	74.67	0.125	67.85	68.70	77.77	0.124
	OCRFlux-3B [49]	3B	74.82	0.193	68.03	75.75	80.23	0.202
	Mistral OCR [50]	-	78.83	0.164	82.84	70.03	78.04	0.144
	POINTS-Reader [4]	3B	80.98	0.134	79.20	77.13	81.66	0.145
	olmOCR-7B [12]	7B	81.79	0.096	86.04	68.92	74.77	0.121
	MinerU2-VLM [14]	0.9B	85.56	0.078	80.95	83.54	87.66	0.086
	Nanonets-OCR-s [51]	3B	85.59	0.093	85.90	80.14	85.57	0.108
	DeepSeek-OCR-Gundam-M	3B	86.46	0.081	89.45	78.02	81.55	0.093
	MonkeyOCR-pro-1.2B [1]	1.9B	86.96	0.084	85.02	84.24	89.02	0.130
	MonkeyOCR-3B [1]	3.7B	87.13	0.075	87.45	81.39	85.92	0.129
	dots.ocr [52]	3B	88.41	0.048	83.22	86.78	90.62	0.053
	MonkeyOCR-pro-3B [1]	3.7B	88.85	0.075	87.25	86.78	90.63	0.128
	MinerU2.5 [2]	1.2B	90.67	0.047	88.46	88.22	92.33	0.047
	PaddleOCR-VL	0.9B	92.56	0.035	91.43	89.76	93.52	0.043

# 产业化技术方案

## 双模型OCR技术：突破传统识别瓶颈

OCR小模型 + 多模态大模型双引擎架构，保证OCR识别的准确性



### 理赔文书全覆盖

支持44种主流理赔文书的智能分类与识别，真正做到覆盖广，包括医疗票据、病历文书、卡证、理赔单据等，信息识别全。



### 双模型交叉验证

传统OCR与大模型双引擎并行处理，交叉验证结果，显著提升复杂版面、干扰与低质量图像的识别准确率。



### 智能图像处理

集质量检测、几何矫正、图像增强于一体，自动修复旋转、倾斜、褶皱等问题。

### 语义纠错与置信度可视化



基于上下文语义的智能推理、纠错，并提供可视化的置信度得分，确保每一条数据的可靠性。



### 复杂场景高精度提取

基于全国万余家医院票据验证，无论是发票、清单、病历还是理赔单证，都能实现高精度的结构化信息提取。

# 03 智能理赔医疗审核



## 基于大模型的全病历理解

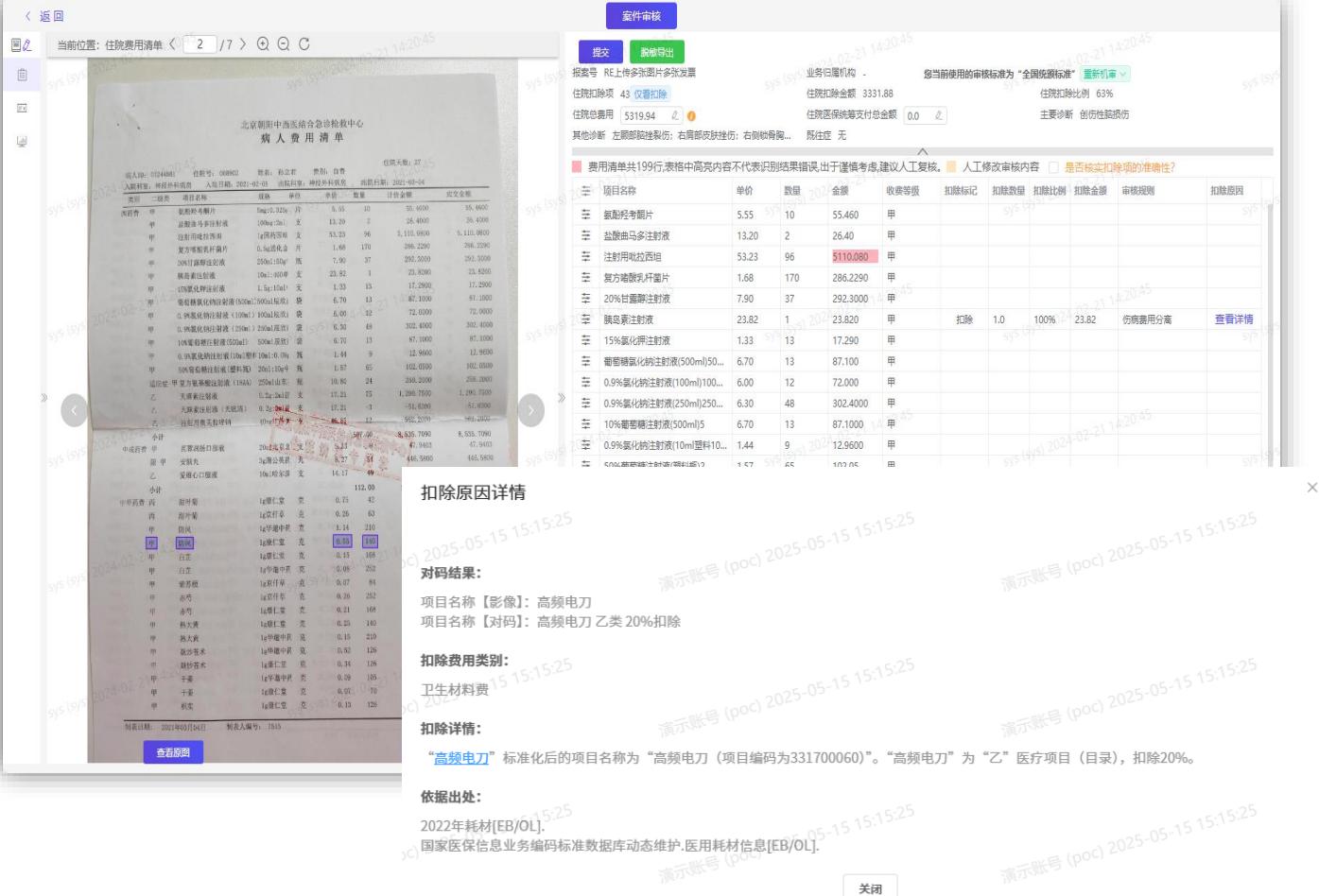
全医疗文书病历理解，包括费用清单、发票、病案首页、入院记录、出院记录、检查检验等。做合理性、一致性判断审核。

## 医疗知识图谱保证专业

基于疾病、症状、用药、体征、部位、检查、手术等实体间的关系，对诊疗合理性、用药合理性等进行判断。

## 审核结果溯源保证审核正确

针对每个审核点进行审核结果的解释溯源，为审核结果提供可解释的依据。



## 多险种、可配置审核规则

内置17类，万余条审核规则；其中政策类标准占比58.27%，医学知识41.73%



## 规范性、一致性、逻辑性、合理性

## 海量审核规则、个性化配置管理

# 伤病分离审核

## 伤病分离审核

基于大模型的强理解能力，基于病历理解和医疗知识判断外伤和疾病治疗的相关性

**主诉：**头晕伴视物旋转 1 月余。

**现病史：**男，65岁，患者于 1 个月前开始无明显诱因出现头晕并伴有视物旋转，严重时伴恶心症状且行走困难，有踩棉花感，同时左上肢麻木，平躺休息后上述症状能较快得到缓解。症状间歇性出现，无明显好转，现患者为求进一步诊治，就诊我院。患者目前精神一般，体力下降，食欲下降，睡眠差，大便正常，排尿正常，体重近期无明显减轻。

**既往史：**平素身体一般，否认“肝炎、结核、疟疾”等传染病史，**糖尿病 2 年余** 饮食、运动控制，未服用药物。否认“肾炎、血液病”等病史。否认外伤史，否认输血史，否认药物、食物过敏史，预防接种随当地进行。

**神经系统：**浅感觉正常，深感觉正常，复合感觉正常，浅反射正常，双侧肱二头肌腱反射、膝反射亢进，生理反射存在，病理反射未引出，Kernig 征正常。生理反射：无异常。

病理反射：未引出。

脑膜刺激征：未引出。

**专科检查：**无

**门诊及院外重要辅助检查：**头颅 CT：未见明显异常；颈部 MRA 造影结果显示：左侧锁骨下动脉近端闭塞。

**初步诊断：**1、**左锁骨下动脉闭塞**；2、糖尿病

提取关键信息及名称标化：

现病史：[{"症状": ""}]

既往史：[{"疾病": "糖尿病"}]

该患者诊断：

主要诊断：左锁骨下动脉闭塞

其他诊断：糖尿病

根据既往史，可得知患者患糖尿病 2 年余，未服用药物。

主要诊断左锁骨下动脉闭塞不会导致糖尿病，因此用于治疗糖尿病的医疗费用（如使用降糖药物西格列汀、二甲双胍、甘精胰岛素等）不予赔付。

# 伤病分离审核

## 伤病分离审核

基于大模型的强理解能力，基于病历理解判断是否外伤引起病情加重

### 主治首次查房

#### 一般情况

记录时间:2021-02-27 17:34

#### 查房记录

今日主任医师代主治医师查房，汇报患者病史、体格检查、诊疗计划及异常的辅助检查：2021-02-27 CT检查1:颈椎间盘(平扫),胸椎椎体(平扫),胸部(平扫)：1、颈椎、胸椎骨折内固定术后改变，请结合病史。2、两肺支气管炎，两肺下叶背段炎性病变，请结合临床。3、胸椎椎体及颈椎平扫未见明显异常。4、C3-4骨性椎管稍狭窄。2021-02-27 09:29 生化常规+全程CRP：总蛋白 58.70 g/L、白蛋白 34.80 g/L、碱性磷酸酶 162.00 U/L、前白蛋白 81.00 mg/L、a-L-岩藻糖苷酶 9.00 U/L、高密度脂蛋白 0.89 mmol/L、载脂蛋白A1 0.72 g/L、铁 3.18  $\mu$ mol/L、钠 129 mmol/L、氯 95.0 mmol/L、钙 2.03 mmol/L、全程C反应蛋白 147.43 mg/L。2021-02-27 08:40 尿常规及沉渣分析：白细胞 6.00 个/uL、上皮细胞 6.00 个/uL。2021-02-27 08:26 美常规(仪器法)：隐血 阳性。2021-02-27 08:19 血凝全套(含D二聚体)：D-二聚体 1.12 ug/mL。2021-02-27 08:15 血常规：白细胞  $11.40 \times 10^9/L$ 、中性粒细胞绝对值  $9.70 \times 10^9/L$ 、嗜酸粒细胞绝对值  $0.00 \times 10^9/L$ 、中性粒细胞百分比 84.90%、淋巴细胞百分比 7.10%、嗜酸粒细胞百分比 0.20%、红细胞  $3.23 \times 10^12/L$ 、血红蛋白 102.00 g/L、红细胞压积 28.9%、血小板  $163.00 \times 10^9/L$ 、平均血小板体积 12.50 fL、血小板分布宽度 17.20。根据患者目前情况，樊有胜副主任医师代主治医师分析如下：患者诊断明确1.脊髓损伤 四肢瘫 2.颈椎骨折术后 3.胸椎骨折术后 4.高血压病II级(中危) 5.冠心病 5.肺部感染。患者CT示有肺炎，血常规白细胞、中性粒细胞、C反应均偏高，临时给予乳酸左氧氟沙星氯化钠注射液1瓶每次100ml，静脉输液 ONCE:痰培养鉴定及药敏(痰液)，明天做头孢西丁钠皮试，如果阴性，可用头孢西丁钠2.0g BID+左氧氟沙星注射液0.5g QD；电解质偏低，特别是钠 129 mmol/L，临时给予0.9%氯化钠注射液400ml+浓氯化钠注射液100ml，静脉输液 ONCE。注意复查电解质；有咳嗽咳痰，痰液粘稠，给予吸入用乙酰半胱氨酸溶液(富露施)每次3ml,0.9%氯化钠注射液每次3ml，雾化吸入 BID稀释痰液，使痰液容易咳出；血小板计数偏高，D-二聚体 1.12 ug/mL偏高，加用硫酸氢氯吡格雷+阿司匹林肠溶片(自备)抗血小板聚集，防止血栓形成，密切观察患者病情变化，根据病情变化调整治疗方案。

#### 既往史

【既往史】既往史：有高血压病史5年，间断服用苯磺酸氨氯地平片 5mg QD。有冠心病史，平素服用阿司匹林肠溶片，具体控制效果不详。否认糖尿病史，否认肺结核、肝炎等传染病史，无家族遗传病史，无疫区接触史，有烟酒史25年，2021-02-09有外伤手术史，颈前路切开复位+ACDF术；2021-02-18颈椎后路椎管减压+钉棒内固定术。【疾病史】疾病史：有有高血压病史5年【传染病史】传染病史：无【预防接种史】预防接种史：不详【手术外伤史】手术外伤史：手术：有2021-02-09颈前路切开复位+ACDF术。2021-02-18颈椎后路椎管减压+钉棒内固定术。外伤：无2021-02-09车祸外伤史。

### 出院诊断

主要诊断: 脊髓损伤。其他诊断: 冠心病。其他诊断: 四肢瘫痪。其他诊断: 颈椎骨折。其他诊断: 胸椎骨折。其他诊断: 高血压病2级(中危)。其他诊断: 肺部感染。

药	工别/仪	录/内/工别/仪	(人衣) /衣				
西成药	硫酸氢氯吡格雷片	硫酸氢氯吡格雷片(波立维)	片	75mg*7片/盒	4.28	7	29.95
西成	吸入用乙酰半胱氨酸溶液	吸入用乙酰半胱氨酸溶液(富露施)	支	3ml*0.3g/支	14.80	2	29.60

硫酸氢氯吡格雷片虽然关联到冠心病，但病历显示，其用以预防手术引起的血栓，可关联到“脊髓损伤”，故予以赔付。

# 用药合理性审核

## 药品超范围审核

长期医嘱 患者信息: 45岁 女 住院 违规金额: 257.07元

一级护理 持续性 日	2021-01-20 11:07	2021-01-21 10:01	2021-01-20 11:07	一级护理	已执行
普食 持续性	2021-01-20 11:07	2021-01-22 09:37	2021-01-20 11:07	普食	已执行
心电监护 持续性 24小时	2021-01-20 11:07	2021-01-21 09:14	2021-01-20 11:07	心电监护	已执行

住院病案首页 患者信息: 45岁 女 住院 违规金额: 257.07元

入院诊断名称: 头皮血肿					
出院诊断	疾病编码	入院	出院		
主要诊断: 头皮血肿	S00.004	1	2		
其他诊断: 上颌窦骨折	S02.400x005	1	2		
头皮挫伤	S00.001	1	2		
多处挫伤	T00.901	1	2		
高血压病3级 (高危)	I10.x00x031	1	2		

医疗费用记录

检验 移酶测定	7.00	7.00	项	1
检验 5.00	5.00	项	1	
检验 血清胆碱酯酶测定	10.00	10.00	项	1
检验 血清单胺氧化酶测定	10.00	10.00	项	1
检验 血清a-L-岩藻糖苷酶测定	25.00	25.00	项	1
西成药 0.9%氯化钠注射液	5.19	5.19	袋	1
西成药 5%葡萄糖注射液 软袋	5.19	5.19	袋	1
西成药 注射用奥美拉唑钠	29.92	29.92	支	1
西成药 硝苯地平缓释片 I(利培)	9.98	0.42	片	24
西成药 维生素B6注射液	1.48	0.74	支	2
西成药 维生素C注射液	1.26	0.63	支	2
检验 腺苷脱氨酶测定	15.00	15.00	项	1
检验 血清肌酸激酶测定	10.00	10.00	项	1
检验 血清肌酸激酶-MB同工酶活性测定	15.00	15.00	项	1
检验 乳酸脱氢酶测定	10.00	10.00	项	1
检验 血清α羟基丁酸脱氢酶测定	10.00	10.00	项	1

住院病案首页 患者信息: 45岁 女 住院 违规金额: 309.07元

应用医疗大模型进行病历理解和关键信息抽取  
利用医疗知识图谱对用药合理性等进行判断。

根据长期医嘱, 患者可以正常饮食, 无特殊说明。  
但药品中的奥美拉唑钠适用限有说明书标明的疾病诊断且有禁食医嘱或吞咽困难的患者。故不赔付。

# 检查化验合理性审核

## 检查化验指征不明确

文书列表	缺陷列表
住院病案首页	
新出院小结	
入院记录	
病程记录	
科间会诊通知单	
医嘱	
检查记录	
检验记录	
医疗费用记录	

医疗费用记录 患者信息: 0000233338 郭书义 男 78岁 神经内科西区二病区 病历评分:0 丙...						
床位	间	40.00		床口	1.00	
诊疗	等速肌力测定	32.00		次	1.00	
其他	医疗废物处置费	2.00		床日	1.00	
检查	室壁运动分析	40.00		次	1.00	
检查	心脏彩色多普勒超声含各心腔等	100.00		次	1.00	

住院病案首页 患者信息: 0000233338 郭书义 男 78岁 神经内科西区二病区 病历评分:0 丙...			
出院诊断	确诊日期	出院情况	治疗天数
主要诊断: 多发性脑梗死	2020-01-03	好转	15
其他诊断: 高血压III级	2020-01-03	好转	15
前列腺增生	2020-01-03	好转	15
慢性硬膜下血肿	2020-01-03	好转	15

设置	
医嘱数量不一致的药品: 阿托伐他汀钙片 (10mg*4+7")	<input type="checkbox"/> 误报备注
-0.78	
医嘱数量不一致的药品: 脉管复康胶囊	<input type="checkbox"/> 误报备注
-6.12	
无医嘱的诊疗: 口腔局部止血	<input type="checkbox"/> 误报备注
-10	
无医嘱的诊疗: 门诊手术室麻醉 (药品除外)	<input type="checkbox"/> 误报备注
-20	
无指征的检查: 心脏彩色多普勒超声含各心腔等	<input type="checkbox"/> 误报备注
-100	
无指征的检查: 血栓弹力图试验 (TEG)	<input type="checkbox"/> 误报备注
-248.4	
无指征的检查: B型钠尿肽 (BNP)	<input type="checkbox"/> 误报备注
1	

入院记录 患者信息: 0000233338 郭书义 男 78岁 神经内科西区二病区 病历评分:0 丙级病...	
年龄: 78岁	入院时间: 2020-01-02, 11:09
婚姻: 已婚	病史记录时间: 2020-01-02, 15:33
民族: 汉	病史陈述者: 患者本人 (可靠)
主诉: 头晕、左下肢行走迟缓1周。	
现病史: 患者近1周来无诱因出现头晕、左下肢行走迟缓, 表现为头部昏沉, 呈头重脚轻感, 无视物旋转、恶心、耳鸣、言语不清及大、小便失禁等症状。未予处理, 上述症状无好转。今日为求诊治入院, 门诊以“脑梗死”为诊断收入科。患者目前神志清, 精神一般, 体力正常, 食欲正常, 睡眠正常, 体重无明显变化, 大便正常, 排尿正常。	
既往史: 既往“高血压”病史20年, (口服维拉帕米片2片2/日、心痛定片2片2/日); “脑梗死”病史13年, 遗留反应迟钝; “前列腺增生”2年。否认有“糖尿病、冠心病”史, 亦否认有“肝炎、结核”等传染性疾病史, 2006年6月因疝气曾行手术治疗, 2000年曾患头面部外伤; 无药物过敏史, 预防接种史随社会。	
个人史: 生长于原籍, 退休职工, 否认有疫水、疫区接触史。无放射性物质接触史, 生活规律, 无烟、酒等不良嗜好。	

患者做了心脏彩超, 但出入院诊断无心脏疾病, 心电图检查无明显异常, 既往史提出无冠心病, 病程中未记录做心脏彩色多普勒的目的及结果分析, 因此系统认定为是无指征检查, 不予报销。

# ■ 医疗审核问题的技术定义

## □ 基本符号：

- ✓ 标准guideline:  $G$  (基于自然语言的标准描述)
- ✓ 待审核文档:  $D$ ,
- ✓ Review policy:  $\pi_\theta$
- ✓ Review trajectory:  $\tau = (a_1, \textcolor{red}{a}_2, \textcolor{red}{o}_2, a_3, \textcolor{blue}{m}_3 \dots)$ ,
  - $a_t$ -review action (推理/工具调用/记忆读写等)
  - $a_{t+1} = \varphi_\theta(\tau_{<t}, G, D)$
- ✓ Review report:  $y = f_\theta(\tau, G, D)$ 
  - $f_\theta(\cdot)$ 为输出函数, 基于review trajectory输出报告(规则判定+对应证据、评分等)
  - 同一个LLM  $\pi_\theta$ 执行审核和输出 (不同prompt) :  $\varphi(\cdot)$ 和 $f(\cdot)$ 共享参数 $\theta$

## □ 目标函数：

$$J(\theta) = \mathbb{E}_{(G,D) \sim \mathcal{P}} \mathbb{E}_{\tau \sim \phi_\theta(G,D)} [R(G, D, \tau, f_\theta(\tau, G, D))]$$

# 医疗AI审核：提示工程

**医保规定：**言语训练限于器质性病变导致的中、重度语言障碍。一个疾病过程支付不超过3个月；每日支付不超过1次。

## Prompt: 【病历信息】

主诉：右侧肢体偏瘫，言语不清1小时

现病史：患者男性、68岁，系“右侧肢体偏瘫，言语不清1小时”入院。

既往史：有高血压病史，未服药；否认“糖尿病”病史，无头颅外伤史，无药物过敏史，饮酒，吸烟20支/天。

查体：体温36.3℃、脉搏100次/分...

诊疗经过：入院后予以神经内科常规护理，一级护理，心电监护，低盐低脂饮食。予以阿替普酶静脉溶栓，川芎嗪、罂粟碱活血、改善微循环.....

出院情况：患者诉下肢恢复较佳，右上肢恢复较差，活动不利，无肢体抽搐，无肢体麻木，无咳量锻炼；

**【任务说明】**请根据以上【病历信息】，判定病人是否存在器质性病变导致的中、重度语言障碍。

## 【背景知识】

“器质性病变”是一个医学概念，指的是....

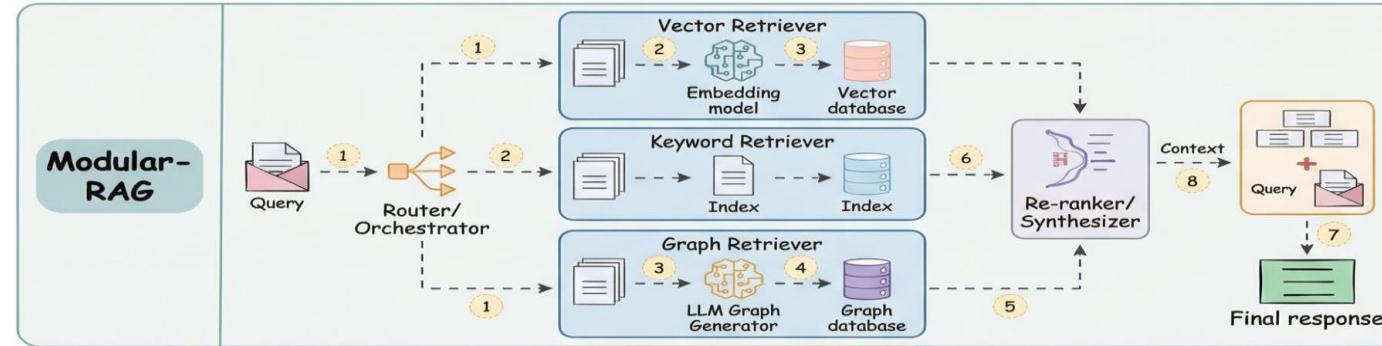
语言障碍的分级及判断主要依据障碍程度、功能影响及临床表现，通常分为轻度、中度、重度三级...

**【输出格式要求】**以json格式返回判定结果，示例如下：

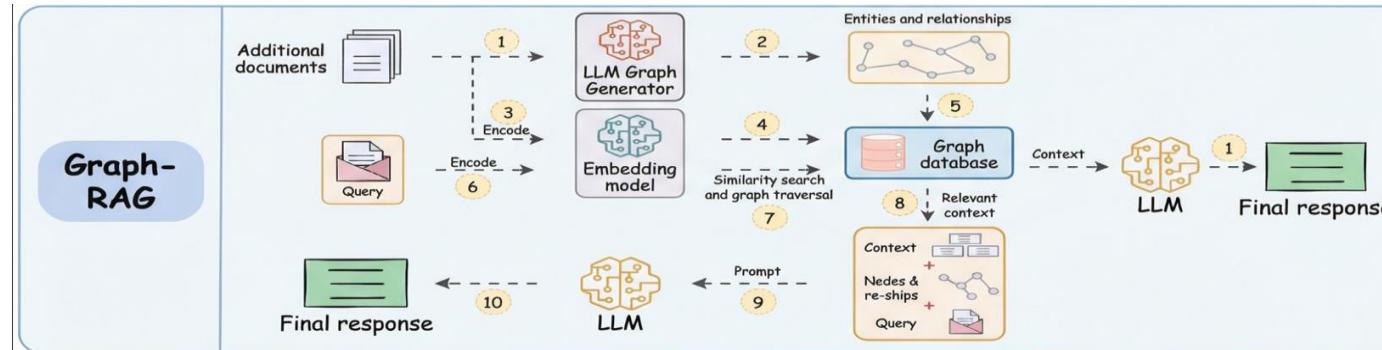
```
{"answer": "true/false", "detail": "简要说明判断依据"}
```

# 医疗AI审核：文本和结构化知识注入

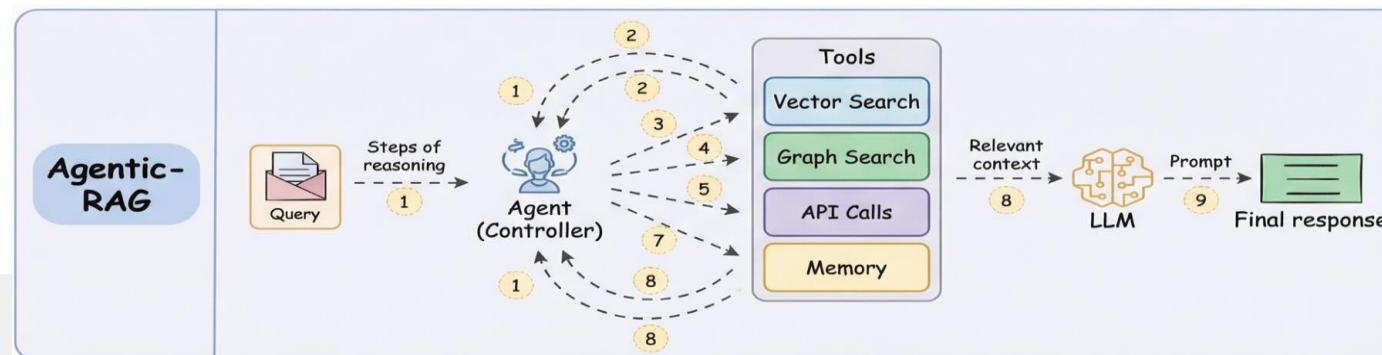
Modular-RAG



Graph-RAG



Agentic RAG



# 医疗AI审核：预编排的工作流

医保政策

言语训练限于器质性病变导致的中、重度语言障碍。一个疾病过程支付不超过3个月；每日支付不超过1次。

条件拆解

条件1（语义）：

限于器质性病变导致的中、重度语言障碍。

条件2（函数）：一个疾病过程支付不超过3个月

```
limitQuality(String ItemName, int quality)
```

函数3（函数）：一个疾病过程支付不超过3个月

```
limitFrequency(String ItemName, float frequency_per_day)
```

结果认定

jsonLogic执行条件逻辑判定，认定最终结果：

```
{  
  "and": [  
    "条件1",  
    "条件2",  
    "条件3"  
  ]  
}
```

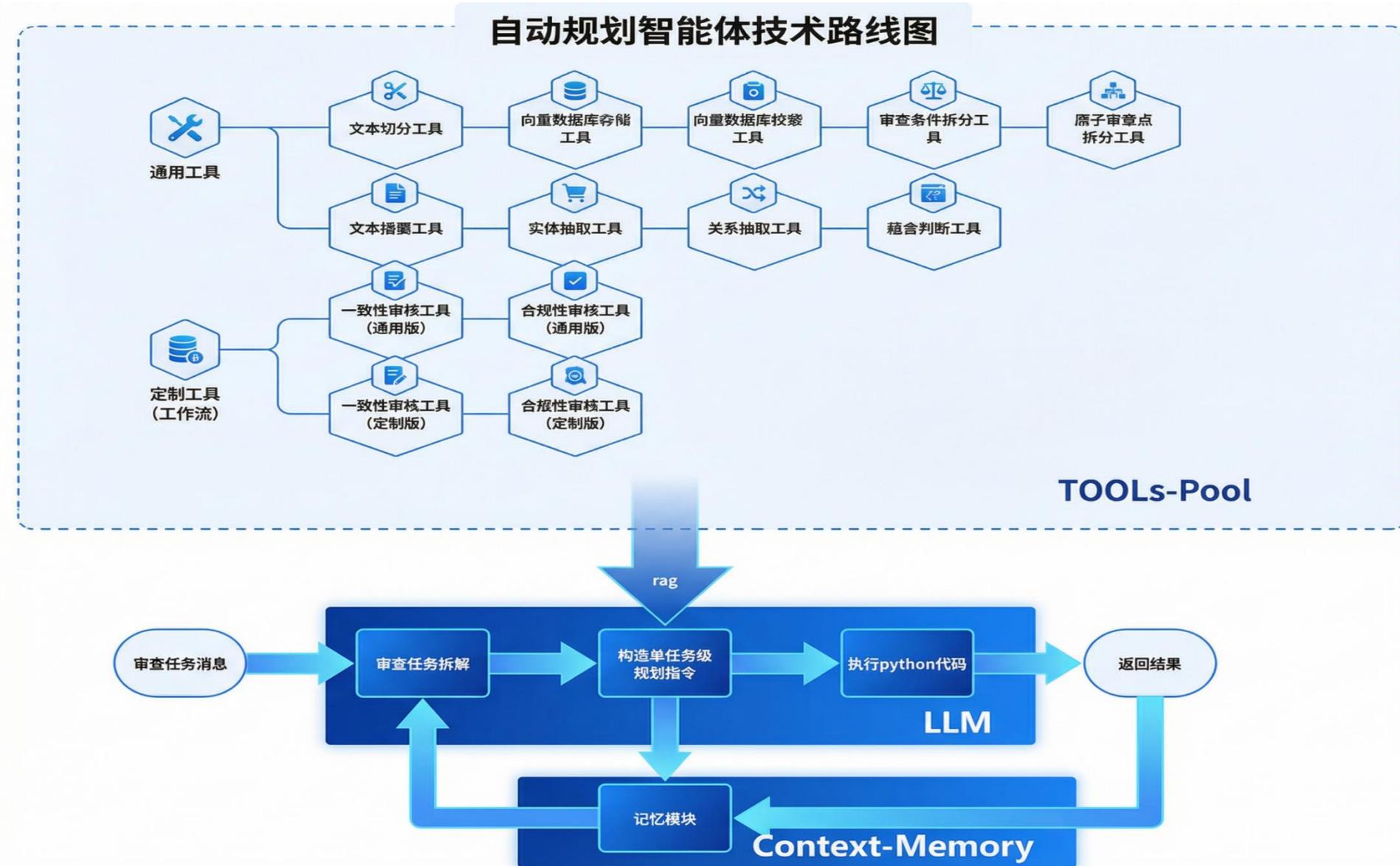
外部工具

医疗知识库

病历读取API

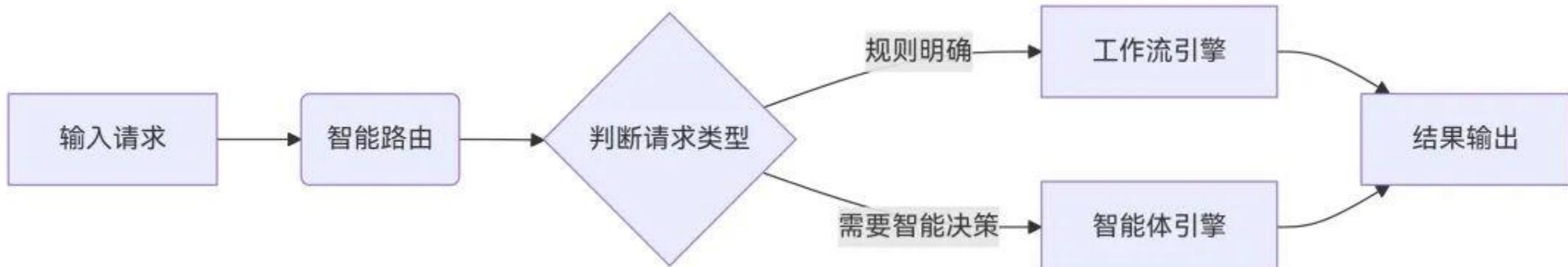
专用Functions

# 医疗AI审核：Agentic workflow (Agent)

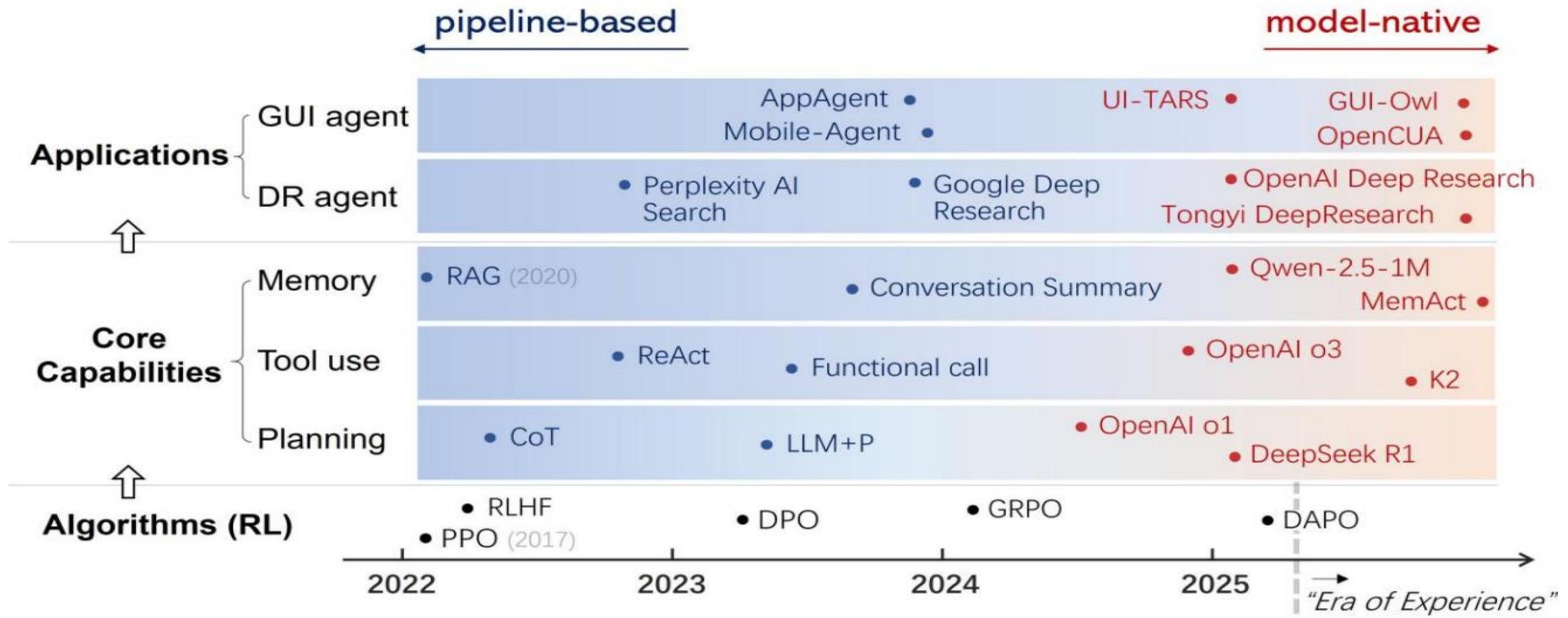


# 医疗AI审核：工作流+Agent 混合

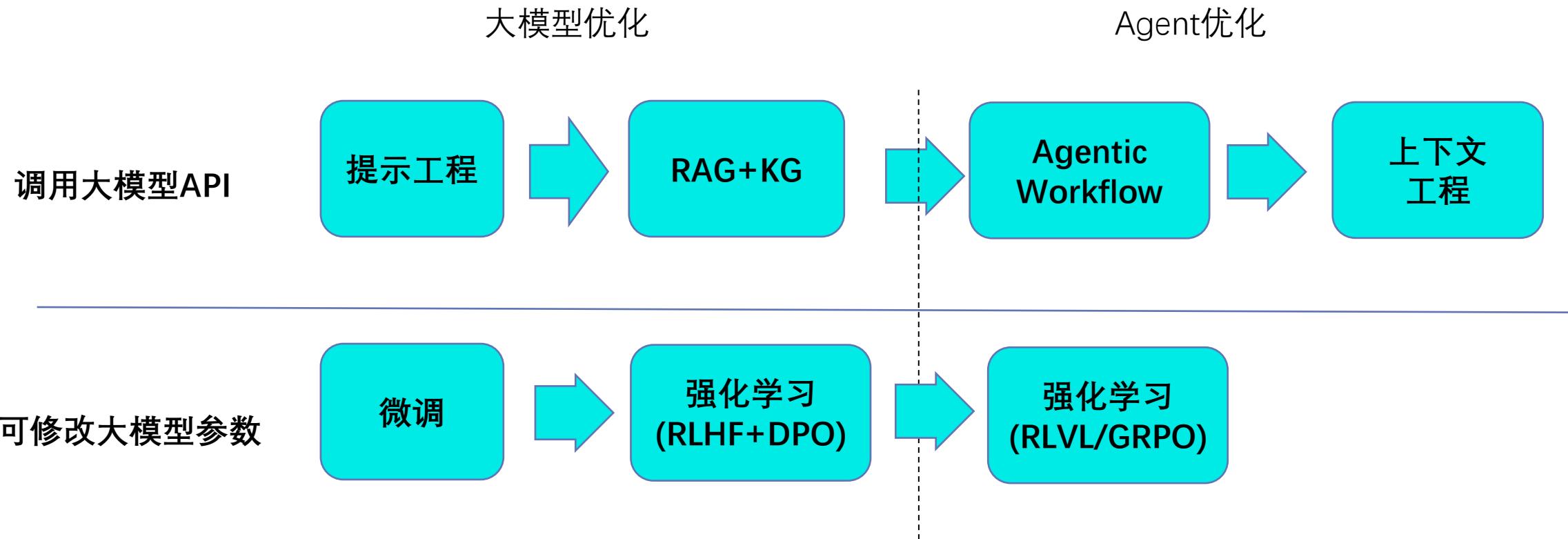
	智能体 (Agent)	工作流 (Workflow)
定义	具备感知、决策、行动能力的智能实体，能自主与环境（包括人类、工具、系统）交互，完成复杂任务	预先定义的任务执行流程，按固定步骤 / 规则自动化完成一系列操作
核心能力	自主决策、动态规划、多工具协同、上下文理解、持续学习	流程自动化、步骤串联、条件判断（如“如果 A 则执行 B”）
自主性	高自主性，可根据环境变化调整策略，无需人类逐步骤干预	低自主性，严格遵循预设流程，无自主决策能力
应用场景	复杂开放场景（如即时性需求）	标准化重复任务（如财务报销流程、订单处理流程、数据填报流程）



# Agent 技术范式演进：从工作流到模型原生

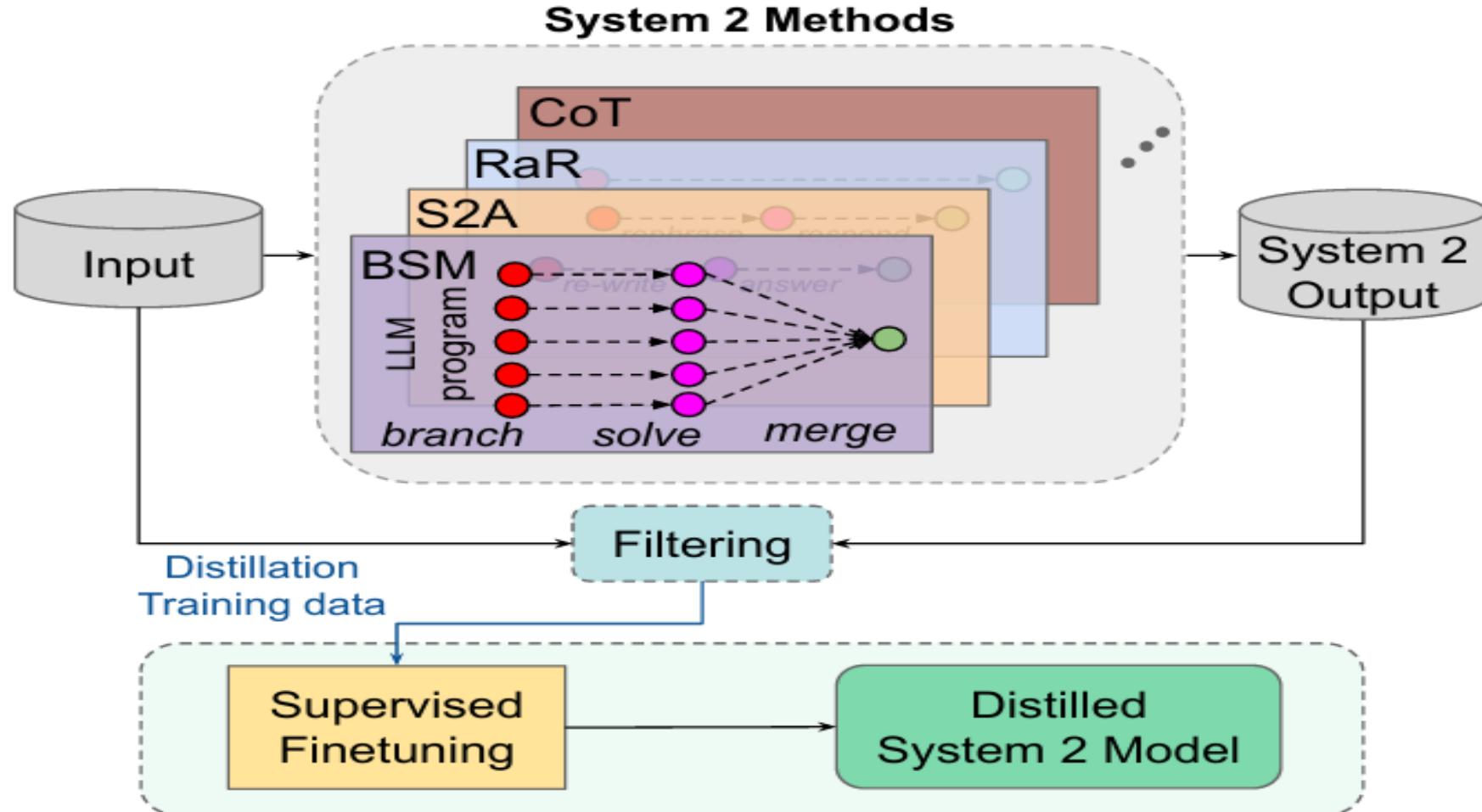


# 面向审核场景的大模型优化技术路线图

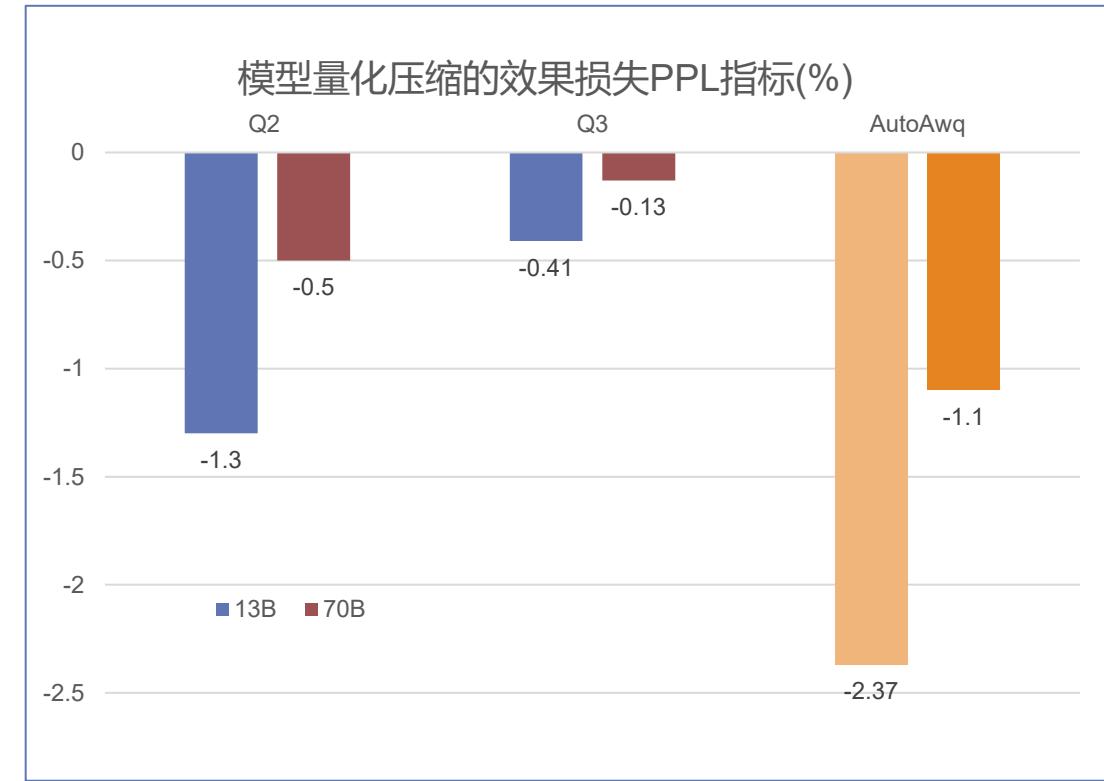
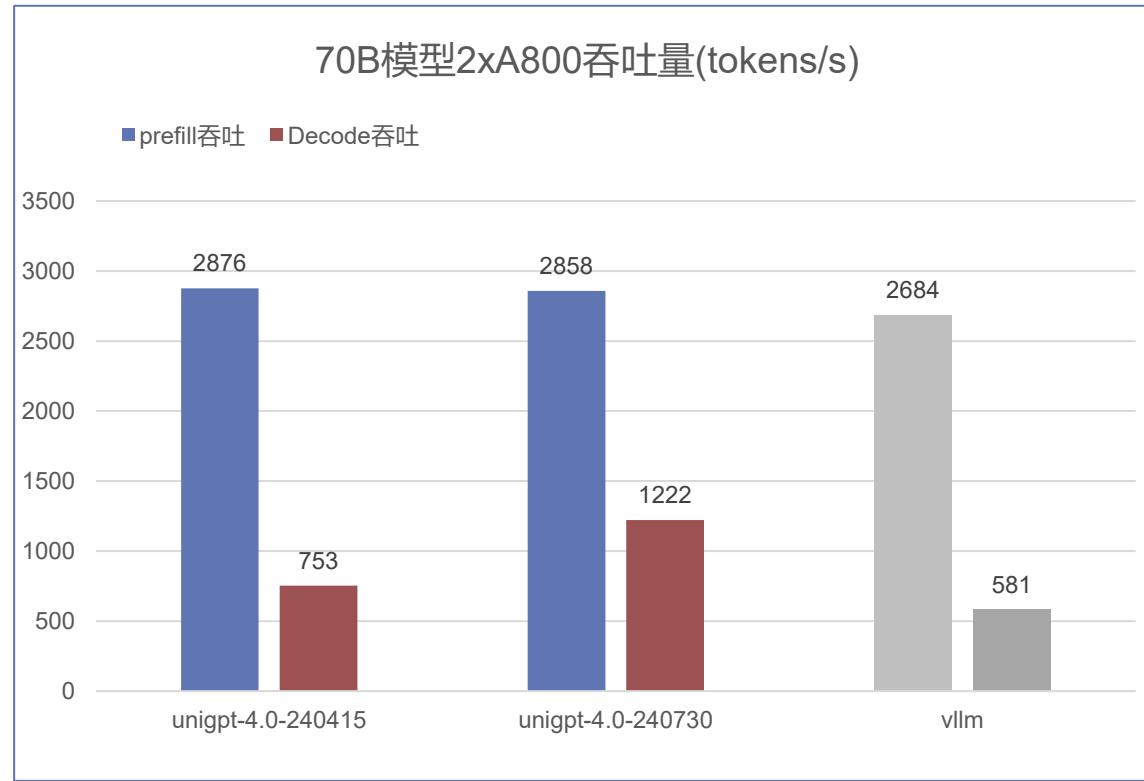


# 工程优化：模型蒸馏

数据蒸馏  
(黑盒蒸馏)



# 工程优化：模型量化



**推理生成速度进一步提升：**

Token生成速度是主流开源工具vLLM的**2.1倍**

**量化压缩效果损失进一步降低：**

70B量化损失0.5%->0.13%，13B量化损失1.3%->0.41%，  
分别是主流开源工具AutoAwq的**17.5%， 11.8%**

# 落地实践效果

## 基金监管系统



传统人工审核模式已不足以应对日益增长的审核需求。同时随着医保制度不断完善规范，医保审核不断细化，医疗费用不合理增长问题亟待解决，医院医保管理需要逐步精准化。借助AI技术，构建基金监管系统。

### 巢湖市医疗保障局文件

巢湖市医疗保障局文件  
巢医保〔2021〕27号  
关于巢湖市试点开展电子病案AI审核的请示  
合肥市医疗保障局：

近年来，巢湖市医保局在上级部门的领导下，深化医保基金全过程监管，依法严厉打击欺诈骗保违法行为，坚持做好医保基金“守护员”，针对医保领域欺诈骗保行为的特点，巢湖市医保局经仔细研究决定对定点医疗机构住院治疗行为进行引导和审核，创新融合运用AI技术领域中自然语言处理（NLP）和知识图谱（KG）两项技术，逐步实现医保基金的流动上，同时“试点电子病案AI审核”。

该方案项目拟通过医保智能审核系统的费用审核审核功能，又通过对完善电子病历（会诊栏、医嘱、检查化验、护理记录等）的结构化信息识别，做到对医保项目的分类审核，且审核结果准确率高，现拟在巢湖市宋庆龄爱心医院完成项目的准确性和可行性，经审核同意后，于2021年1季度2348份住院电子病案，涉及医疗费用1798.92万元。

### 合肥市医疗保障局

关于开展病案AI审核系统试点工作的函  
巢湖市医疗保障局：

你局《关于开展病案AI审核系统试点工作的请示》已收悉，现将有关情况函复如下：

一、同意你局开展病案AI审核试点工作。数据传输应严格按照信息安全保护有关规定，相关系统互联互通要做好数据保护，进行数据审核，创新融合运用AI技术领域中自然语言处理（NLP）和知识图谱（KG）两项技术，逐步实现医保基金的流动上，同时“试点电子病案AI审核”。

该方案项目拟通过医保智能审核系统的费用审核审核功能，又通过对完善电子病历（会诊栏、医嘱、检查化验、护理记录等）的结构化信息识别，做到对医保项目的分类审核，且审核结果准确率高，现拟在巢湖市宋庆龄爱心医院完成项目的准确性和可行性，经审核同意后，于2021年1季度2348份住院电子病案，涉及医疗费用1798.92万元。

### 绵阳市涪城区医疗保障局

绵阳市涪城区医疗保障局  
关于对人保财险绵阳市分公司完成医保基金  
监督检查工作的验收说明  
人保财险绵阳市分公司：

本次分公司在我局“医保基金第三方监督检查项目”中，严格按照相关规定要求具体实施，充分发挥专业团队优势，提升服务质量，人保财险绵阳市分公司完成医保基金监督检查工作，现将有关情况说明如下：

一、严格按照相关法律法规，数据传输应严格按照信息安全保护有关规定，相关系统互联互通要做好数据保护，进行数据审核，创新融合运用AI技术领域中自然语言处理（NLP）和知识图谱（KG）两项技术，逐步实现医保基金的流动上，同时“试点电子病案AI审核”。

该方案项目拟通过医保智能审核系统的费用审核审核功能，又通过对完善电子病历（会诊栏、医嘱、检查化验、护理记录等）的结构化信息识别，做到对医保项目的分类审核，且审核结果准确率高，现拟在巢湖市宋庆龄爱心医院完成项目的准确性和可行性，经审核同意后，于2021年1季度2348份住院电子病案，涉及医疗费用1798.92万元。

示范性落地应用，国内首创

上线时间：2021年-至今

上线范围：已支持全国25个省市地区基金监管相关审核需求

上线系统：病案AI审核、常规基金监管、飞行检查、慢病认定和审核

建设成效：

- ✓ 病历实现**100%全覆盖审核**
- ✓ 审核工作效率提升**11.5倍**
- ✓ 2024年，审核总金额**98亿**，扣除金额**3.4亿**，核减率**3.48%**
- ✓ 上线系统获得**多市医疗保障局高度肯定**，并加大推广力度

# ■ 总结与启发

1. 利用大模型的综合能力，寻找需要多学科交叉知识的应用场景，如医疗+保险，医疗+法律
2. 医保理赔医疗审核的自动化技术：文档智能大模型 + 基于医疗审核大模型的智能体
3. 审核技术的广泛应用：法律，医疗，金融领域的文档审核
4. 审核技术的根本性问题：解决大模型生成与评价能力的不均衡

# 极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

北京

1200人

**QCon**

全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

深圳

1000人

**AiCon**

全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

北京

1000人

**AiCon**

全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

4月

6月

8月

10月

12月

**AiCon**

全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

上海

1000人

**QCon**

全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

上海

1200人

# THANKS

探索 AI 应用边界

Explore the limits of AI applications

AiCon

全球人工智能开发与应用大会