大模型对医院信息系统的影响分析

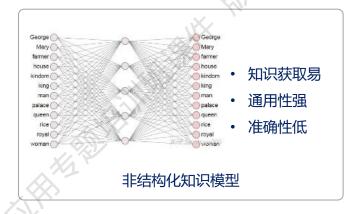
薛万国

才大模型应用能力的分析

大模型给医学AI带来的本质改变

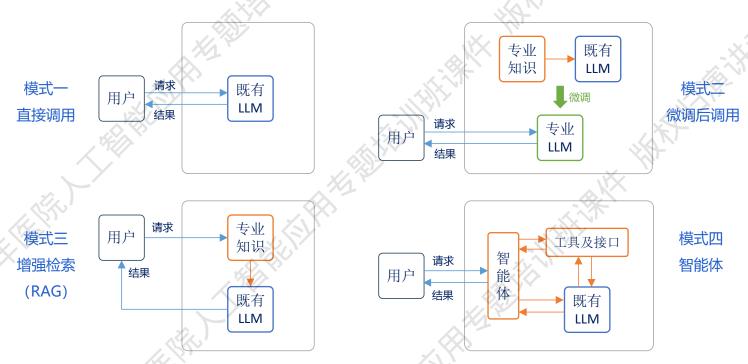
- > 知识表示的改变: 从结构化知识到非结构化知识
 - 传统的医学AI以结构化知识表示和应用为主: 规则知识库、知识图谱
 - 大语言模型以非结构化知识表示和应用为特点
- ▶ 外在能力的改变: 从特定场景到通用场景
 - 从特定场景的AI转为通用型AI,极大地提升了模型适应的多场景能力
 - 无监督为主的知识获取,极大地降低了掌握复杂医学知识的难度





大模型的应用模式

> 大模型在医院的四种应用模式



大模型在技术上的局限与适用性

- ▶ 大模型与生俱来的"幻觉"
 - 大模型学习了大量的文字资料,本质上是打碎了的信息碎片以及关联关系
 - 大模型之所以不同于检索系统,不是"鹦鹉学舌"而是知识加工
 - 在信息碎片基础上再生成的内容必然与"幻觉"相生相伴,其优势在于不确定性,劣势也在于不确定性



大模型在技术上的局限与适用性 (续)

- "幻觉"限制了大模型医学应用的场景
 - 对于严谨的医学应用场景,医学伦理对于机器的错误是"零容忍"的
 - · 若采用RAG等知识库技术"限制"大模型的能力,则会出现与传统医学AI应用类似的局限



大模型在技术上的局限与适用性 (续)

- > 大模型的资源开销和性能
 - 大模型推理所需资源开销大、响应慢,配置8GPU卡的"一体机"可支持数十至上百个并发
 - 对于医院应用场景和典型配置,既有大模型适合于低频使用,难以支撑多用户高频交互
 - 两类场景类似于OLAP与OLTP的特点区别



对信息系统内外在影响的分析

大模型对医院信息系统的潜在影响

> 大模型将给医院信息系统的外在功能和内在架构带来重大影响



大模型对医院信息系统功能的影响

- > 功能重构,降低了智能化的门槛
 - 用大模型替代既有采用自然语言处理、知识库(知识图谱)等智能化功能
 - 可提升功能的准确性、通用性,大大简化原有功能的开发实现
- > 可重构的典型应用
 - · 辅助诊疗决策CDSS
 - 病历质控
 - 患者智能分导诊
 - 临床科研平台病历结构化



大模型对医院信息系统功能的影响 (续)

- > 场景创新
 - 通用化的知识库、强大的自然语言理解与处理、突出的对话能力,激发出新的应用场景
- ▶ 典型应用场景

面向临床医护

- •病历摘要
- •医疗护理文书生成

面向患者服务

- •健康教育
- •健康咨询
- •报告解读
- •预问诊
- •智能随访

面向医院管理

- •运营报告生成
- •运营问题分析
- •医院管理知识问答

面向医学研究

- •研究路径指导
- •病历语义搜索

应用系统功能形态的发展



—引自森亿AI医疗相关报道

应用系统功能形态的发展 (续)

▶ 例: 自然语言指令生成运营报表



应用系统功能形态的发展 (续)

▶ 例: 医患语音交互, 自动生成门诊病历

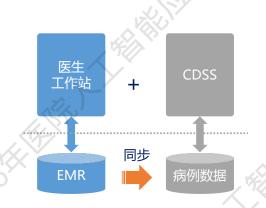




—引自open.zuoshouyisheng.com/dialogue_structure

应用系统与大模型的深度融合

- ▶ 传统的AI功能多以外挂的形式出现
 - 传统AI的实现复杂而专业,通常由第三方开发,再与应用系统界面集成
 - 如传统的CDSS、病案质控助手等

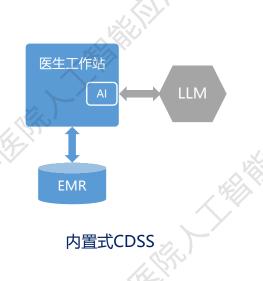


外挂式CDSS



应用系统与大模型的深度融合 (续)

- ▶ 应用系统或智能体直接调用大模型
 - 一方面,利用大模型的通用推理能力,可极大地简化智能化功能的开发
 - 另一方面, 内置智能化功能可以充分获取病历数据, 并免除数据同步的繁琐

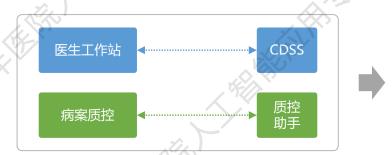




直接调用 大模型

自然语言作为接口语言

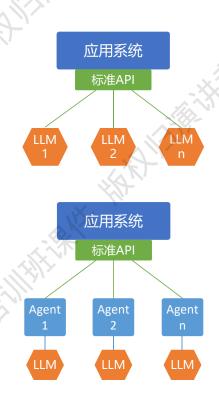
- ▶ 传统AI应用
 - 不同的任务调用不同的接口和AI模块
- ▶ 基于大模型的AI应用
 - 由自然语言提示词指引不同类型的任务
 - 任务不同,同一模型与接口,提示词不同
 - 为保持和规范生成质量,设置提示词模板





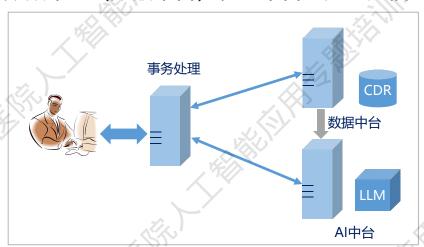
应用系统与大模型的集成问题及解决方法

- > 不同的应用系统绑定了不同的大模型
 - 不同的模型在不同类型任务中的表现有所不同
 - 一些应用系统按照特定用途对特定模型进行了微调训练
 - 医院需要部署不同的大模型
- > 应用系统需要与医院部署的大模型进行适配
 - 不同的大模型的调用接口可能不同
 - 同样功能不同厂商的AI智能体调用接口不同
- > 依靠标准构建应用系统和大模型的和谐生态
 - 建立大模型的标准API:基础AI能力接口标准
 - 建立各类智能体的标准API: 应用间接口标准



对数据的整合度要求更高

- > 智能化应用的实现需要医疗数据和大模型的融合
 - 以界面方式集成的医生工作站,难以适应大模型对完整病例数据的需求
 - 数据集成和集中管理成为大模型应用的基本条件
- ➤ 临床数据中心 (数据中台) 和AI中台成为医院信息系统的基本构件



数据中台包括CDR以及能够供AI 直接调用的加工数据

AI中台包括大模型、智能体以及 以知识库形式构建的智能化功能 部件

用户知识库成为基本构件

> 用户知识库的重要作用

- 通过内容约束,限制大模型的"幻觉"
- 实现大模型输出的本地个性化
- 解决大模型的知识更新时效性问题

> 医院信息系统架构的基础部件

- 向量数据库:医院非结构化知识载体,如临 床指南、医院制度文件
- 知识图谱:医学基础性结构化知识载体,如医学术语系统
- 数据库: 医院个性化知识载体, 如科室设置



医院IT基础设施面临的新挑战

- > 大模型运行资源开销大
 - 以DeepSeek为例,不同规模的模型的推荐硬件配置如下:

模型规模	FP16显存需求	4-bit量化需求	推荐显卡配置
67B	134GB	33.5GB	4×A100 40GB / 2×A100 80GB
236B	20GB	5-7GB	单卡A100 80GB / RTX 4090
236B	640GB (8×H100)	160GB (2×H100)	H100 80GB多卡集群
671B	1,543GB	386GB	16×H100或32×A100
70B	140GB	40GB	4×A100 80GB
7B	14GB	4GB	RTX 3090/4090
70B	140GB (理论)	35-50GB	多卡RTX 4090/A100/H100集群
1.5B	3-5GB	0.8-1GB	单卡RTX 3050或纯CPU
671B	1342GB (理论)	350-436GB	多节点A100/HT00集群 + 云服务支持
	67B 236B 236B 671B 70B 7B 70B 1.5B	67B 134GB 236B 20GB 236B 640GB (8×H100) 671B 1,543GB 70B 140GB 7B 14GB 70B 140GB (理论) 1.5B 3-5GB	67B 134GB 33.5GB 236B 20GB 5-7GB 236B 640GB (8×H100) 160GB (2×H100) 671B 1,543GB 386GB 70B 140GB 40GB 7B 14GB 4GB 70B 140GB (理论) 35-50GB 1.5B 3-5GB 0.8-1GB

• 据报道,基于昇腾的满血DeepSeek一体机,在输入输出长度4096/1024典型场景下,系统输出吞吐为1288Token/s,多用户并发数128

医院IT基础设施面临的新挑战 (续)

- > AI算力集群将成为医院的新需求
 - 特别是在多种模型并存的场景下,对于算力和显存空间要求更高
 - 少数医院因为模型调优的需求,需要更高的AI训练性能
 - AI (GPU) 服务器集群需要高性能的网络互联,如NVLink、InfiniBand网络
- > IT中心机房的供电散热受到挑战
 - 传统服务器机柜的供电为单路32A, 也即6KW/机柜的供电能力
 - 一台8卡AI服务器的功耗在5-10kW,例如NVIDIA的8卡HGX A100超算平台总功耗达6.5kW,其中GPU功耗占3.2kW。AI服务器组成整体机柜后功率密度将达40kW以上
 - 功耗的增长也必然带来散热的大幅增长,对于机房空调带来很大压力

结束语

大模型的出现,既是人工智能的技术进步,也给医院信息系统的发展创造了新的契机,必将给医院信息系统的功能能力、功能形态、数据架构、集成模式、基础设施等诸多方面带来重大变化。我们在关注大模型的能力的同时,需要深入洞察潜在的变革因素,主动应变。

改变才刚刚开始。

2025HHELERIN TREASURE THE STATE OF THE STATE