



DeepSeek 应用经验分享

王志勇

2025年5月25日

目录

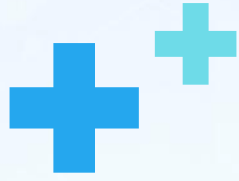
CONTENTS

01 为什么选择DeepSeek?

02 DeepSeek本地化部署

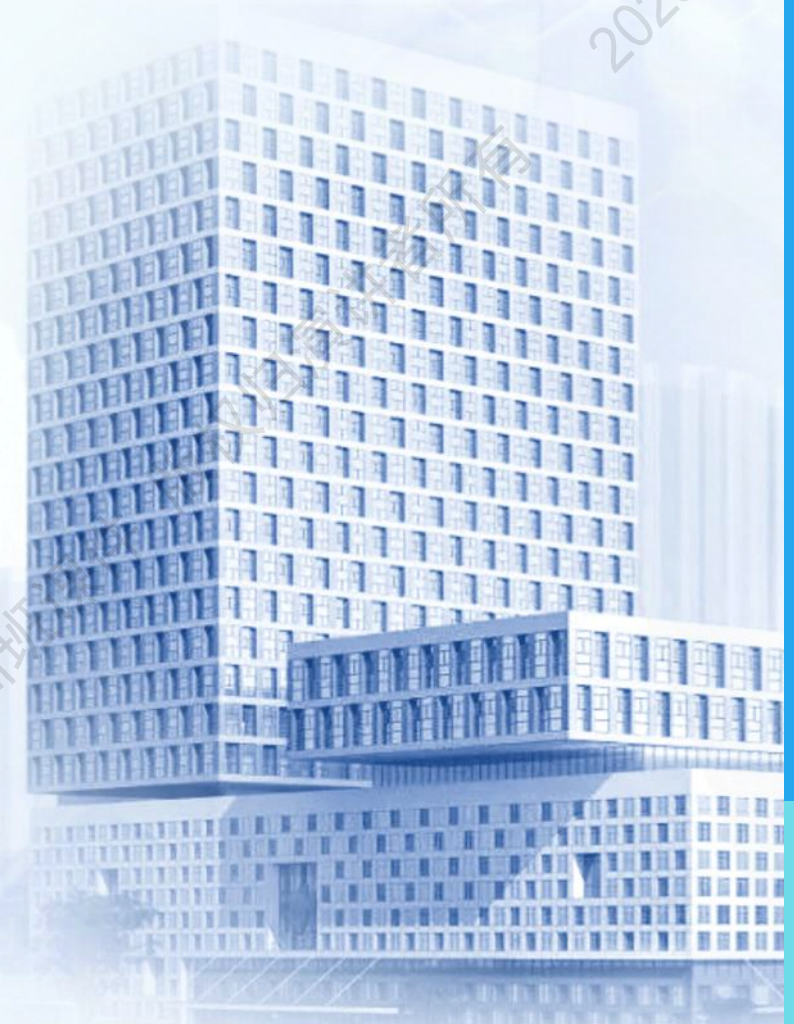
03 落地场景介绍

04 几点体会



PART 01

为什么选择DeepSeek?



2025年医院

演讲者所有

2025年医院人工智能应用专题培训班

版权归作者所有

版权归作者所有

版权归演讲者所有

2025

AI发展的“三大支柱”



左边的齿轮图表示了三者目前的关系：

- 算法的齿轮虽小（表示技术还在发展之中），但却撬动了对于算力和数据的巨大需求
- 这就是今天人工智能的现状，所以要学好用好人工智能，最重要的是要对技术有正确的了解

DeepSeek的相对优势

ChatGPT

DeepSeek

完全开源

高性价比

GPT-4o/o1

DeepSeek-V3/R1

训练成本降低90%+

推理成本降低30%-50%

DeepSeek架构基础及关键技术

架构基础

- **Transformer 模型**：依赖自注意力机制（Self-Attention）处理长距离语义依赖，替代传统RNN/CNN的序列处理限制
- **预训练范式**：自监督学习，通过掩码语言建模（如BERT）、自回归预测（如GPT）等方式从无标注数据中学习通用语言模式
- **大规模多模态语料库**：训练数据通常涵盖书籍、网页、学术论文等多领域文本

关键技术

- **微调（Fine-Tuning）**：在预训练模型基础上，用领域特定数据（如医学文献）调整参数以适应垂直场景
- **提示工程（Prompt Engineering）**：通过设计输入提示词，引导模型输出符合需求的响应
- **检索增强（RAG）**：结合外部知识库动态补充信息，减少模型“幻觉”



DEEPSEEK适合的应用场景

- ✓ **“问答”**：如：智能问答机器人、报告解读、辅助诊断
- ✓ **“审核”**：如：病历质控、论文（标书）、合同
- ✓ **“概括”**：如：病历生成、个性化健康管理方案
- ✓ **“预警”**：VTE风险、不良事件、阳性检查结果



需求方与需求

- ✓ **院方**：需降低运营成本、提供管理效率（如病历质控、资源调度）
- ✓ **医生**：需辅助诊断、病历自动生成等工具**减轻负担**
- ✓ **患者**：需更便捷的智能问答、**持续的**健康管理服务
- ✓ **信息部门**：需易集成的低门槛技术方案



PART 02

DeepSeek本地化部署



为什么要本地化部署?

01

数据不出院，满足合规要求

02

泛化能力强（通过构建不同知识库以满足不同应用场景需求）

03

辅助信息源的完全可控，可获得更加精准的输出



本地化部署可保证医疗数据安全

数据合规性

DeepSeek在设计之初就充分考虑了国内的数据安全法律法规，如《网络安全法》和《个人信息保护法》，确保在数据处理和存储过程中符合相关要求，避免法律风险。



本地化部署

Deepseek支持本地化部署，医院可以将敏感数据存储在本地上服务器上，避免数据跨境传输带来的安全隐患，同时满足国家对医疗数据本地化的监管要求。

加密与审计

Deepseek采用先进的加密技术保护数据传输和存储的安全性，并支持全面的审计功能，确保医疗数据的访问和使用可追溯，进一步提升数据安全水平。

满足定制化需求



● 满足个性化需求

医院通常有特定的业务流程和需求，DeepSeek的开源特性允许医院根据自身情况进行深度定制，例如**可自行建立医院专属知识库**或优化诊断算法，以满足个性化需求。

● 开发门槛低

DeepSeek提供本地**API接口**，只需**简单技术培训**，信息中心工程师就能够与医院现有的HIS(医院信息系统)、EMR(电子病历系统)等**无缝集成**，**对专业AI公司依赖程度低**。

● 持续优化支持

DeepSeek的**开源社区**和开发团队能够为医院提供持续的技术支持和更新服务，帮助医院根据实际使用反馈不断优化系统功能，确保其始终处于最佳运行状态。

针对医院的技术路线

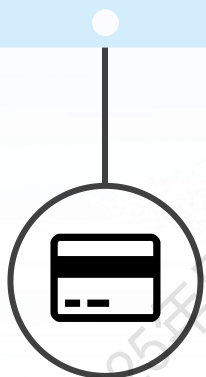
最优组合：DeepSeek-R1 32b/70b+本地RAG知识库

中等规模参数设计（320亿、700亿）在保障诊疗决策精准性的同时，对本地服务器算力需求适中。

模型版本	参数量	显存需求 (FP16)	推荐 GPU (单卡)	多卡支持	量化支持	适用场景
DeepSeek-R1-1.5B	15亿	3GB	GTX 1650 (4GB显存)	无需	支持	低资源设备部署 (树莓派、旧款笔记本)、实时文本生成、嵌入式系统
DeepSeek-R1-7B	70亿	14GB	RTX 3070/4060 (8GB显存)	可选	支持	中等复杂度任务 (文本摘要、翻译)、轻量级多轮对话系统
DeepSeek-R1-8B	80亿	16GB	RTX 4070 (12GB显存)	可选	支持	需更高精度的轻量级任务 (代码生成、逻辑推理)
DeepSeek-R1-14B	140亿	32GB	RTX 4090/A5000 (16GB显存)	推荐	支持	企业级复杂任务 (合同分析、报告生成)、长文本理解与生成
DeepSeek-R1-32B	320亿	64GB	A100 40GB (24GB显存)	推荐	支持	高精度专业领域任务 (医疗/法律咨询)、多模态任务预处理
DeepSeek-R1-70B	700亿	140GB	2x A100 80GB/4x RTX 4090 (多卡并行)	必需	支持	科研机构/大型企业 (金融预测、大规模数据分析)、高复杂度生成任务
DeepSeek-671B	6710亿	512GB+ (单卡显存需求极高, 通常需多节点分布式训练)	8x A100/H100 (服务器集群)	必需	支持	国家级/超大规模 AI 研究 (气候建模、基因组分析)、通用人工智能 (AGI) 探索

部署过程

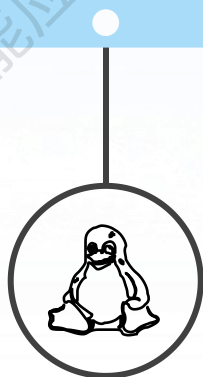
硬件



服务器部署

32b: 单算力卡, 40G显存
70b: 双算力卡, 96G显存

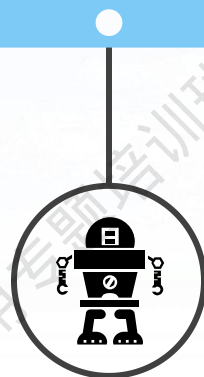
软件环境



系统、平台

操作系统: ubuntu22.04
框架: vLLM
部署方式: Docker或Python 直接部署

模型



deepSeek

1、通过在线源拉取;
2、手动下载后放入指定路径手工启动。

工具



集成工具

知识库: RagFlow & AI 数据库
工作流: Dify
搜索引擎: 多尺度树图 (MSTG) 向量引擎
语音模型: Speech2text

接口



业务系统API

- 业务系统=>Dify的API 工作流(实现鉴权和分流)
- 工作流=>**检索引擎**API(实现向量预过滤及快速索引构建)
- 检索引擎=>DeepSeek(实现定向知识库的结果推理)
- **后续接入MCPServer (实现系统自动操作)**

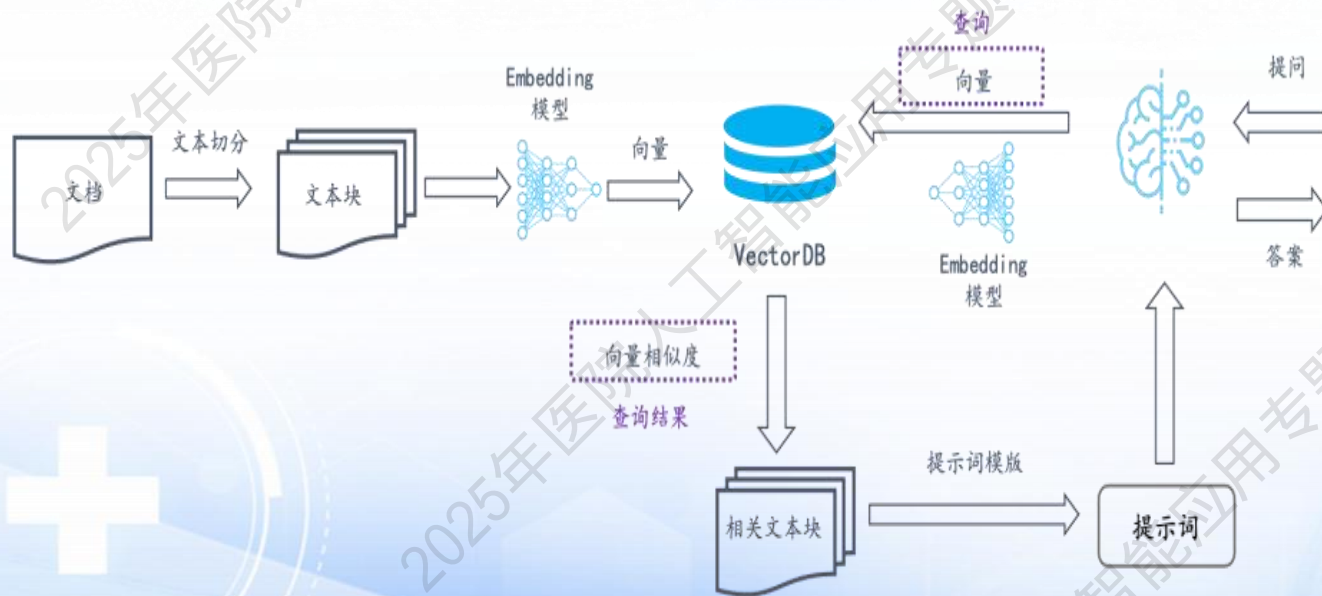
知识库工具：RAGFlow



RAGFlow

@Aiders

RAGFlow 是一个基于深度文档理解的开源 RAG（检索增强生成）引擎。它为任何规模的企业提供了简化的 RAG workflow，结合 LLM（大型语言模型）提供真实的问答能力，并以各种复杂格式数据的可靠引用为支持。



搜索



推荐



个人助理



对话机器人

知识库工具：RAGFlow

RAGFlow存在的问题：

- 知识库无法进行**分类分级管理**
- 不能实现知识文件的**自动标注**
- 需要大量**人工清理数据**的工作
- 无法实现向量知识库和关系数据库的**混合检索**



DeepSeek



Docker



RAGFlow

搭建本地知识库系统

AI数据治理平台

- **AI 数据治理** 以AI 数据库作为底层支撑，通过**上海算法创新研究院**研发的**Data-Centric AI/LLM 工具**，实现多模态数据的数据管理、评估、去重、生成、抽取，自动标注等一系列**算法和 AI ETL 流程**
- 通过**提高数据利用率，降低数据采集、标注和算力成本**，支撑医疗领域专家、数据和算法工程师更低门槛、更高效、更低成本的训练、调优大模型

领域专家、数据科学家、算法工程师



AI数据治理平台

基于 AI 数据库和数据标签、语义信息实现训练中的动态数据调度，解决**数据配比难和调度难**等问题

- **4 天** 超过算法工程师精心设计 **4 个月** 的最佳配比结果，**实现主动学习、课程学习**
- 针对垂域问题，**自动推荐领域最佳语料数据**，提升模型训练效率





提示词模型库构建

提示词构建格式：角色确认=>任务指向=>规则引导=>示例标准=>备注要求

例如：

#角色：从体检报告审核医师角度出发

#任务：严格按照以下规则处理体检报告中的复诊建议

#处理规则：

输入为JSON数据，需解析"JY"字段内容

识别包含以下任一关键词的科室建议：就诊/诊治/进一步检查/复查/随诊/随访等

提取对应科室名称和复诊建议关键词

输出必须保持原始ID和YWID不变

结果格式必须严格使用：

```
{ "ID": 原始值, "YWID": 原始值, "recommendation": [ { "DEPT": "科室名", "INFO": "复诊建议"}, ... ] }
```

#示例：

输入：

```
{"ID": "20250319024038", "YWID": "250004630", "JY": "【脂肪肝】建议：消化科就诊；【冠状动脉壁钙化】建议心内科随诊"}
```

输出：

```
{"ID": "20250319024038", "YWID": "250004630", "recommendation": [{"DEPT": "消化科", "INFO": "脂肪肝"}, {"DEPT": "心内科", "INFO": "冠状动脉壁钙化"}]}
```

#备注要求：

- 必须严格使用指定输出格式，禁止添加任何解释
- 科室名只需保留标准科室名称(如"心内科"而非"心内科门诊")
- 每个建议单独列出，即使同一科室有多个触发词
- 确保准确识别嵌套在句子中的关键词
- 忽略未包含触发词的建议内容

workflows 构建

通过配置Dify工作流的模型、知识库上下文关联以及提示词编排，实现DeepSeek模型与业务系统的无缝接入，使自然语言交互能力嵌入业务流程。



API接口

通过API接口获取不同应用系统数据以及提交DeepSeek处理结果。



数据处理

编写Python程序对接口数据进行格式调整以满足DeepSeek格式要求，以及对DeepSeek的输出做清洗以满足API接口的要求。

```
PYTHON3:  
1 import json  
2 from typing import List, Dict, Any  
3 def main(arg: str) -> Dict:  
4     response_data = json.loads(arg)  
5  
6     details = response_data.get  
7     ("retDetail", [])  
8  
9     extracted_data = []  
10    extracted_text=""  
11  
12    if not details:  
13        return {"result": []}  
14  
15    #for detail in details:  
16        #创建字典对象  
17        item = {  
18            "ID": details[0].get("ID"),  
19            "YVID": details[0].get("YVID"),  
20            "YV": details[0].get("YV")  
21        }  
22        #append  
23        #extracted_data.append(item)  
24        #extracted_text+="ID:"+detail.get  
25        ("ID")+"; YVID:"+detail.get("YVID")  
26        +"; "+detail.get("YV")+"; "  
27        #extracted_text=extracted_text.  
28        replace("\r\n","")  
29  
30    #return  
31    return /
```

```
PYTHON3:  
1 import re  
2 import json  
3 def main(answer: str) -> dict:  
4     cleaned_answer = re.sub(r'<details  
5     [^>]+>/details', '', answer,  
6     flags=re.DOTALL)  
7     final_answer = re.sub(r'^\n+', '',  
8     cleaned_answer)  
9     final_answer=final_answer.replace  
10    ("json","")  
11    final_answer=final_answer.replace  
12    (" ", "")  
13    final_answer=final_answer.replace  
14    ("\\n","")  
15    final_answer=final_answer.replace  
16    ("\\t","")  
17    final_answer=final_answer.replace  
18    ("no","")  
19    final_answer=final_answer.replace  
20    ("n","")  
21    final_answer=final_answer.replace("!  
22    n;n","")  
23  
24    #json_obj = json.loads(json_obj.  
25    ["query"])  
26    final_answer = final_answer.strip().  
27    replace('\\n', ' ').replace('\\t',  
28    ' ')  
29    # 解析为 JSON 对象  
30    json_obj = json.loads(final_answer)
```



里程碑事件

01

2025年2月6日启动DeepSeek院内本地化部署工作。

03

2025年2月12日DeepSeek-R1 32b本地安装完成、知识库环境配置完成

05

2025年2月18日新闻稿发布在全国引起轰动效应，400余家媒体转载

02

2025年2月10日服务器硬件部署到位（借用）。

04

2025年2月15日第一个院内知识库完成数据清洗、导入工作正式上线



媒体报道

媒体类型	信息来源	发布时间	标题
网媒	搜狐新闻	2025-02-18 13:47	革新医疗服务：上海四院完成DeepSeek大模型本地化部署
APP	上观新闻	2025-02-18 13:47	AI医生来看病？上海四院完成DeepSeek大模型本地化部署
短视频	度小视	2025-02-18 13:58	AI医生来看病？上海四院完成DeepSeek大模型本地化部署
微博	新浪微博	2025-02-18 14:53	【助力教学、科研，#上海多机构完成DeepSeek本地化部署#】近日，上海交通大学、上海市第六人民医院、同济大学附属上海市第四人民医院等机构先后宣布，已完成DeepSeek本地化部署环境，利用人工智能
微信	微信	2025-02-18 22:09	DeepSeek已经开始挤压人类工作岗位
论坛	股吧	2025-02-19 00:29	上海四院完成DeepSeek人工智能大模型本地化部署
报刊	新民晚报	2025-02-19 15:50	快速提供精准决策支持



里程碑事件

01

2025年2月27日接待上海市医疗急救中心领导参观交流

03

2025年3月7日接待东方肝胆外科医院、普陀区妇婴保健院参观交流

05

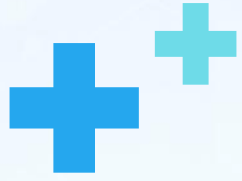
2025年3月24日“关键患者精准追踪系统”正式上线

02

2025年3月3日接待新疆医科大学第一附属医院参观交流

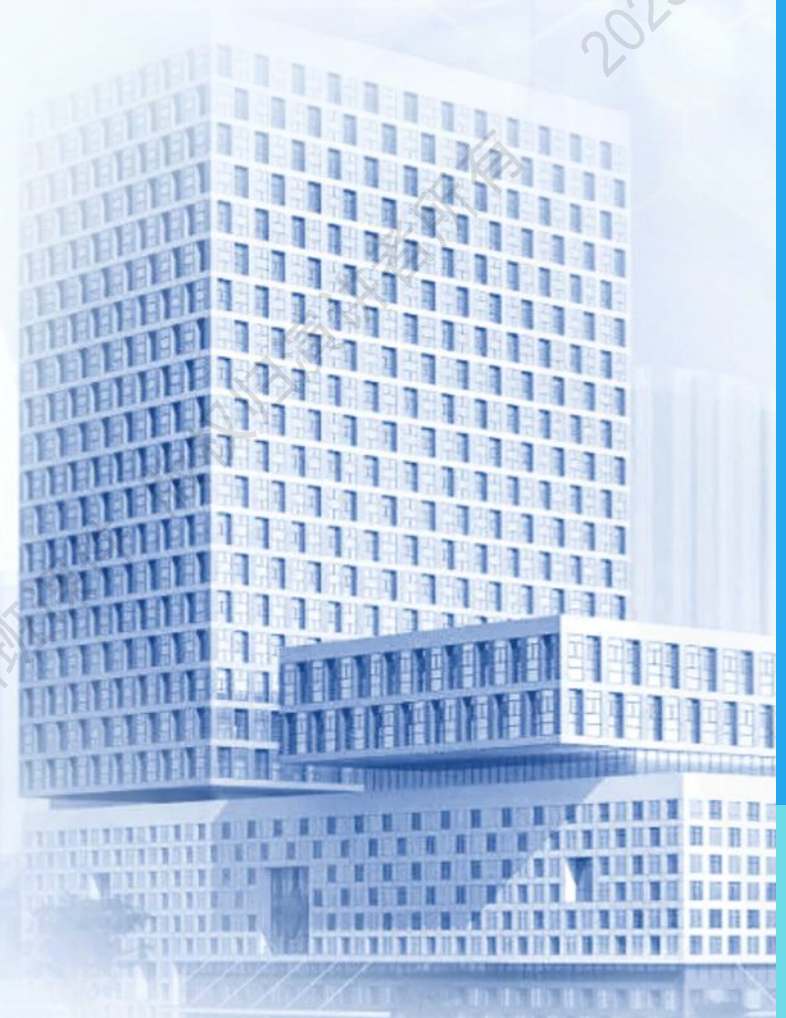
04

2025年3月10日医院办公会通过DeepSeek首批四个场景的立项申请



PART 03

落地场景介绍



问答机器人 (AI客服)

- **文字输入**

提问：痔疮手术出院后需要注意什么？

- **语音输入**

提问：出院后多久复查？



门诊病历生成

诊前：患者通过预问诊二维码，主动填写病情，以减少就诊时间。



07:56 第四人民医院发热门诊患者信息登记

*体温

体温(℃)

*发热开始时间

请选择“发热开始时间”

*症状多选题

干咳	哮喘	呼吸急	胸痛	乏力	盗汗
流涕	眼睛痒或流眼泪	鼻塞	腹泻		
头痛	颈部	四肢酸痛	肌肉	气促	
四肢困倦	呕吐	腹痛	皮疹		

过敏原(白介素) 001105AL

其他症状

请选择“其他症状”

*新冠病毒阳性(单选题)

否 1次 2次 3次 +3次

< 返回 提交

07:54 第四人民医院发热门诊患者信息登记

*近期1次疫苗接种(单选题)

<4个月 3-6个月 6-9个月 9-12个月

>12个月 无

*7日内境外旅居史(单选题)

有 无

*基础性疾病(多选题)

高血压	冠心病	糖尿病	支气管扩张	慢性支气管炎
高血脂	胆心囊	肾脏病	贫血	哮喘
肝硬化	肝硬化	肝硬化	肝硬化	肝硬化
类风湿	类风湿	类风湿	类风湿	类风湿
免疫抑制剂	免疫抑制剂	免疫抑制剂	免疫抑制剂	免疫抑制剂

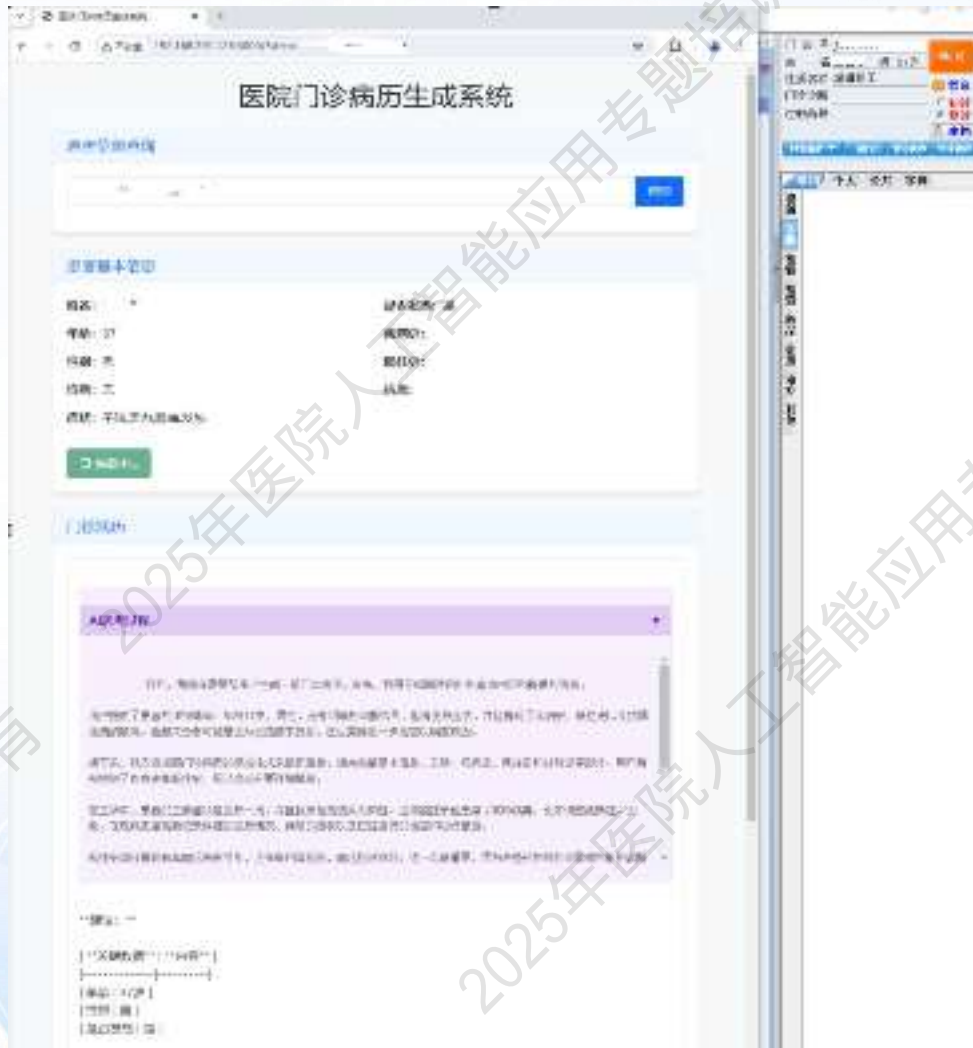
其他基础性疾病

请选择“其他基础性疾病”

< 返回 提交

门诊病历生成

诊中：医生接诊后，系统根据预问诊信息，自动生成诊疗建议及病历文书初稿。



门诊病历生成

诊后：开发患者追踪智能体，筛选需要追踪的病人，自动识别复诊科室、医生和复诊日期。

		主诉：复诊	
2025-03-09 14:22:12	上海市第四人 民医院	现病史：复诊，仍有咳嗽，无发热，有痰血1次，现无痰血，有后背痛	
	呼吸与危重症 医学科	既往史：肠癌	
		药物过敏史：青霉素，克林霉素	
		查体： 神志清楚，无贫血貌，全身皮肤黏膜无黄染，双肺呼吸音清，未闻及干湿啰音，心率72次/分，律齐，无杂音，腹部软，无腹部压痛，无腹部反跳痛，无肌卫，肝脾肋下未及，未触及腰部肿块。	
		辅助检查：拒绝血常规，只查胸部CT及心电图	
		诊断：呼吸道感染	
		处理：患者氧饱和度95%建议急诊，患者拒绝，对症处理，如有不适，即刻就诊，3天后复查，建议耳鼻喉科就诊	
		西药： 美愈伪麻口服溶液(阿贝赛)(10ml*6支/盒) 4盒/20ml tid 口服 头孢呋辛酯干混悬剂(运泰)(0.125g*15包/盒) 1盒/0.25g bid 口服	
		检查： 开单时间:2025-03-09 14:29 检查项目:CT平扫胸部(胸部CT平扫)	



住院病历生成 (含质控)

自动获取患者住院就诊记录、医嘱、检查检验等数据，自动生成出院记录。

上海市第四人民医
姓名: [模糊] 性别: [模糊] 年龄: [模糊] 职业: [模糊]

姓名	性别	年龄	职业	民族	出生地	职业	入院日期	出院日期
[模糊]	[模糊]	[模糊]	[模糊]	[模糊]	[模糊]	[模糊]	[模糊]	[模糊]

出院记录

姓名: [模糊]
性别: 男
年龄: 62岁
入院日期: 2025年03月19日
出院日期: XXXXX (本系统提供, 用XXXXX代替)

一、入院情况

患者因“流涎痰喘1月余”入院。既往多次因复合性溃疡出血、贫血、腰痛、双下肢水肿住院治疗。2024年12月曾行下腔静脉滤器植入术，术后行腹股沟淋巴结术及PTCD术后血红蛋白下降。每日腹股沟引流液约1500ml。入院查体：神清，重度贫血貌，双下肢水肿，移动性浊音(+)，双下肢轻度水肿，辅助检查示血红蛋白显著降低 (Hb 208.3 μmol/L, Dbl 168.3 μmol/L)，白蛋白降低 (33.2 g/L)，C反应蛋白大量升高；双下肢静脉彩超提示：

二、入院诊断

VTE智能辅助决策支持

DeepSeek根据Caprini等问卷信息以及患者的各类诊疗信息，生成辅助治疗及诊断建议。

深静脉血栓 (VTE) 全院防控智能辅助决策系统

住院患者 | 出院患者 | 知识库 | 质量控制 | 管理

患者信息: 16A2(床) 男 75岁 身高: 165 体重: 93 BMI: 34.16 入院时间: 2025-3-28 13:45 手术时间: 2025-3-31 08:55 - 2025-3-31 10:35

风险评估 | 风险趋势 | 检查检验

评估时间 | 开始时间 | 结束时间 | 搜索 | 不给评估 | 同步患者数据 | 更多节点打印

Caprini	高危	手术/后评估	Caprini	高危	常规评估	Caprini	高危	手术/评估	Caprini	高危	入院前评估
未确认的		未确认的	未确认的		未确认的	未确认的		未确认的	未确认的		未确认的
		2025-3-31 11:35			2025-3-31 10:19			2025-3-31 07:43			2025-3-29 08:15

护士审核确认 | 医生审核确认

VTE评估: 11分-高危 | 出血风险: 3分-高危 | wellsDVT: 0分-低度可能 | wellsPE: 0分-低度可能

预防措施: 基础预防 药物预防 物理预防

快速确认

辅助建议: 当前患者VTE风险高危, 出血风险高, 建议血管加压超声明确血栓情况, 密切监测出血风险, 纠正临床出血危险因素, 存在机械预防禁忌症, 酌情药物预防, 或血栓科会诊, 遵医嘱, 1.药物预防建议【低剂量普通肝素】【低分子肝素】【磺达肝癸钠】【直接口服抗凝药】【维生素K拮抗剂】

诊断建议: 当前患者DVT低度可能, 建议监测D-二聚体, 血栓基础预防或遵医嘱; 该患者PE低度可能, 因此, 建议监测D-二聚体, 血栓基础预防或遵医嘱

Caprini (高危) | 外科出血风险评估 (高危) | wells-DVT (低危) | wells-PE (低危) | 机械装置评估 (低危)

物理+药物预防措施
非药物预防或药物预防的相关信息

活动性出血 ×

数据来源: 病程内容 | 编辑

AI急诊哨兵 (研发中)

大模型整合患者语音或文字信息，匹配病历模板生成结构化病历，结合风险评估及检验/检查数据提示风险等级，自动完善生成病史草稿。

智能病史用户界面



HIS病历书写界面

模板选择 公共 我的

胸痛通用, ACS, 主动脉夹层, 急性肺栓塞, 反流性食管炎,
气胸, 肋软骨炎

结构化病历

持续/ 间歇性
(分钟/ 小时/ 天
生命体征 (预检带入)

文字输入病历

持续左胸痛3小时压榨
冷汗气促

语音转TXT **图片资料**



拟诊

ACS

补病史

无

风险评估

心功能III级 (依据)

补病史

端坐 高枕 能平卧

建议处置

心电图、TNT

日程提醒

2小时后复查TNT)

生成病史

风险等级



← 病史草稿, 可修改或删除



最大受益者

医院整体是最大受益者!

➤ 医院

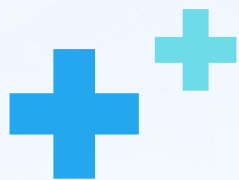
通过精准患者追踪，提升业务量和医疗收入；AI赋能提升诊疗效率、降低误诊率、优化管理流程、实现医疗质量与运营效率双提升。

➤ 医生和患者

受益于精准诊断与个性化服务。

➤ 信息部门

获得技术升级契机。



PART 4

几点体会





几点体会

1. 组织形式：目标驱动、多方配合

➤ 目标驱动

干什么？目标？

➤ 多方协作

临床科室、职能科室、技术厂商密切配合。牵头人很重要（懂临床和

IT）。

➤ 统建算力中心和全科知识库

科室主导建专科知识库



几点体会

2. 注重伦理约束与隐私保护

- **内容安全与合规性**：有害内容过滤、价值观对齐、可解释性
- **防止滥用**：使用条款约束、监控与审计、分级管控
- **隐私保护**：数据收集匿名化处理、数据用途的明确说明
- **传输与安全存储**：端到端加密、数据隔离、定期清理



几点体会

3. 容错率高的场景先行

先选取**相对低风险且容错率高**的场景（如：**患者追踪**）做先行试点，摸索出经验后再持续**优化知识库与 workflows**以适配**高风险**（如：**诊疗决策**）场景的复杂需求。

几点体会

4. 知识库管理与评估工具极其重要

- **科室能自助管理知识库。** 满足自行上传文件、分级分层出知识库原形；通过向量化及数据结构化的模式，存入AI向量及SQL混合模式知识库。
- **科室评估知识库的检索和生成能力。** 对关键指标（答案的准确性、答案的相关性、上下文相关性）的评估。
- **AI收集用户对错误/正确答案的标注、偏好等。** 针对抗噪声能力、拒绝无效回答能力、信息综合能力、反事实稳健性进行评分，构建数据飞轮，用于知识库的修正以及可能的模型微调。

几点体会

5. 付费模式

➤ 医院主导采购

信息化预算或专项资金

➤ 科室主导采购

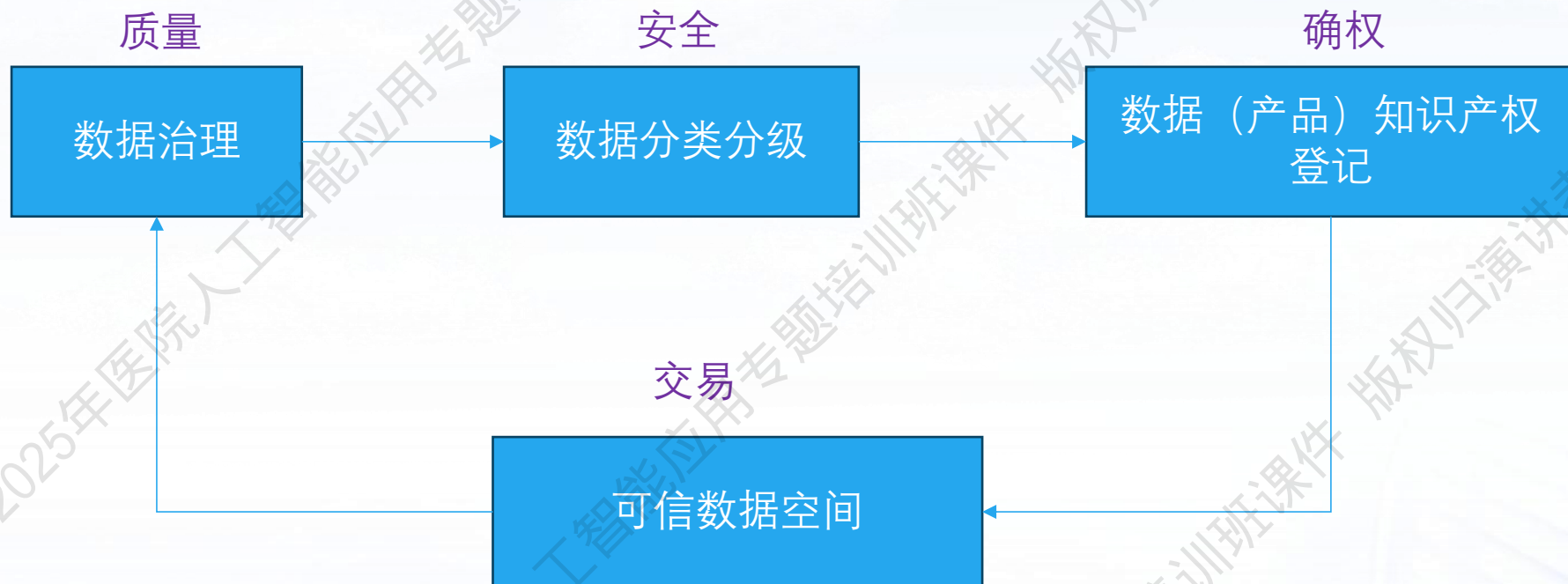
课题经费、学科建设经费等

➤ 其他形式

与医院合作开展ToC业务，利润分成

几点体会

6. 数据治理（资产）与AI应用将逐步走向融合



数据治理，是数据成材的前提！

几点体会

7. 临床科室积累专科数据资产

表1 冠状动脉疾病急诊专病库变量模块

序号	模块	模块内容	变量数量
1	基本信息	研究对象的人口学信息, 就诊信息等	57
2	病史信息	既往病史, 危险因素, 用药史, 家族史等	191
3	体征诊断	体征、诊断和评估等	192
4	检查信息	实验室检验, 影像学检查, 超声检查等	408
5	治疗信息	用药信息, 手术信息, 介入治疗详情等	755
6	护理信息	出入量, 生命体征等	52
7	院内事件	缺血事件、出血时间、并发症等	178
8	随访信息	研究对象随访相关信息	292

- 联合临床科研人员, 信息技术人员, 医疗数据专家等人员, 参考国际上的同类系统, 以及疾病诊疗指南等参考文献, 通过梳理急诊的业务流程, 并且根据冠状动脉疾病领域的科研需求, 建立冠状动脉疾病急诊专病库变量集
- 变量由临床数据、时间信息、序次信息以及其他逻辑计算成专病库变量集, 设定的时间断面包括: 入院前、入院时、住院期间、出院时以及随访
- 专病库变量集一般包含 8大领域数据: 基本信息、病史信息、体征诊断、检查信息、治疗信息、护理信息、院内事件以随访信息等, 结合时间断面, 总共生成的变量超过2000个以上

修库→数据治理→AI应用→数据资产

几点体会

7. 临床科室积累专科数据资产

数据来源：单位自建、多中心联合、共享/公共、...

获取数据的考虑因素：

- ✓ 数据库处于活动状态且可及
- ✓ 符合伦理和数据安全性要求
- ✓ 具有一定的完整性（涵盖主要变量、数据量）
- ✓ 多源异构数据的融合性



几点体会

7. 临床科室积累专科数据资产

• 国外

- 美国住院病人样本数据库 (NIS)
- 全美再入院患者数据库 (NRD)
- 全美急诊患者数据库 (NEDS)
- 重症医学数据库 (MIMIC)
- 肿瘤监测、流行病学和最终结果数据库 (SEER)
- 美国健康与营养数据库 (NHANES)
-

• 国内

- 国家统计直报系统 (所有公立医院的病案首页数据)
- 上海市健康网数据库 (上海市居民健康档案)
- ???

什么是优质数据库？
敢用、能用、好用！



