

2023年南湖HIT论坛

数据驱动医院运营管理

2023年南湖HIT论坛课件 版权归演讲人所有

2023年11月11日 嘉兴市




扫码观看视频直播

主办:  HIT专家网

承办: 北京和思凯文化传媒有限公司

支持企业:  卫宁健康

 联空网络

 望海康信

 ifmsoft

 B-Soft 创业慧康

 ClinBrain

 美创

 TRANSWARP

南京代码
688051

 白峰科技

 inspur 浪潮



数据智能与医院发展

2023年南湖HIT论坛课件 版权归演讲人所有

孙嘉明

2023.11.11

数据 DATA

源自拉丁语datum的复数形式

// 已知或假定为事实的事物 //

(17世纪40年代)

2023年南湖HIT论坛课件 版权归演讲人所有

最初用于哲学范畴

西方哲学

研究起始

“事物是由什么组成的，以及它们是如何变化的”

主要领域

形而上学 (Metaphysics)

对终极实在的研究

包括世界是由什么构成的 | 世界是否只包含一种实在
(比如物质) | 心灵是什么 | 心灵和身体是如何关联
等

认识论 (epistemology)

对知识的研究

包括知识是什么 | 知识的来源 | 真理是什么 |
我们能否获得有关外部世界的知识等

价值论 (axiology)

对价值的研究

包括价值是什么 | 价值有哪些来源 |
价值判断是客观还是主观的等

逻辑学 (logic)

对正确推理的研究

包括什么叫作一个论证 | 存在哪些种类的论证 |
区分好论证和坏论证的标准等

DATA

哲学范畴

计算机领域

源自拉丁语datum的复数形式

// 已知或假定为事实的事物 //

(17世纪40年代)

第一台通用计算机ENIAC诞生

// 能被计算机操作执行的可传输、存储的信息 //

(1946年)

描述已知

预测未知

DATA

事物相关性

事物因果性

数字空间

未来

组成数字空间的基本材质

构建数字空间的逻辑规则

物理空间

土地要素

劳动力要素

资本要素

技术要素

数据要素成为影响未来全球经济发展的关键

美国

全球数字经济规模最大

持续构建法律法规体系
建立政府数据开放平台
引导企业数据共享开放

DATA.GOV 2009年建立

主要发布联邦政府、州政府和企业的
数据、工具和API接口

数据标准规范

元数据标准 | 统一开放数据格式 |
开放平台目录清单等

一套完整的数据开放评估系统
开放数据仪表盘

(截至2022年12月, 累计免费开放数据集25万余个、数据工具300余个、
众多数据API, 覆盖医疗、能源、财政、教育、农业与气候等数十个领域)

欧盟

“一总四分”的数字战略和法律体系

“一” 战略

2020《欧洲数据战略》(顶层数据发展战略)
“塑造欧洲数字化未来”

建立统一治理框架 | 加强数据基础设施投资 |
提升个体数据权利和技能 | 打造公共欧洲数据空间

“四” 立法

《数据治理法案》(DGA) | 《数据法案》(Data Act)
《数字市场法案》(DMA) | 《数字服务法案》
(DSA)

共同数据空间 (IDS)

全球第一个国家级单一数据流通平台
独创的IDS数据连接器和机构产品双重认证机制

中国

全球数字经济规模第二

2020.4《关于构建更加完善的要素市场化
配置体制机制的意见》

首次明确将数据要素列为第五大生产要素

2021.1《“十四五”数字经济发展规划》

从顶层设计上明确了数字经济及其重点领域发展的
总体思路、发展目标、重点任务和重大举措

2022.12《关于构建数据基础制度更好
发挥数据要素作用的意见》

“数据二十条”, 统筹推进数据产权、流通交易、
收益分配、安全治理, 加快构建数据基础制度体系

2023.10 国家数据局正式揭牌

北京: 公共数据
专区授权运营管理

上海: 促进浦东
新区数据流通交易

杭州: 公共数据
授权运营实施

广东: 健康医疗
数据合规流通标准

医疗健康数据

特征趋势与发展挑战

2023年南湖HIB论坛课件 版权归演讲人所有

全域数据

感知，数字医疗新空间



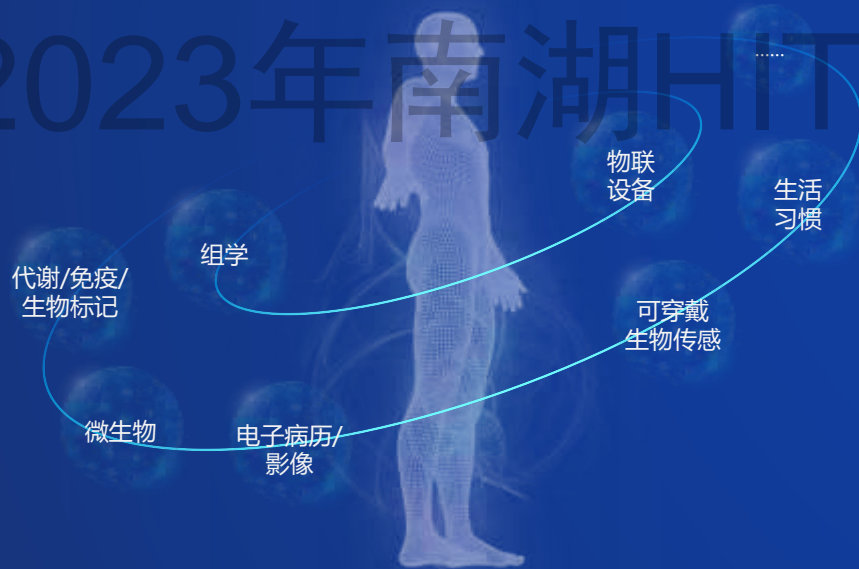
高质量发展需要基于完整、客观、优质的全域数据来评估与促进，而医疗健康数据的复杂多样性是主要挑战

多模态

个人健康及临床诊疗数据

全方位

医院运营管理所需的数据



2023年南湖HIT论坛课件 版权归演讲人所有

符号主义 Symbolism

基于规则、知识库的
专家系统技术路线

本质是关系型数据库，收敛、简化

专家系统



连接主义 Connectionism

基于深度神经网络
更接近于人脑的思维模式

结构不同，非目标导向

神经网络



反向传播算法
Backpropagation

一种在神经网络中
优化权重参数的算法



卷积神经网络
Convolutional Neural Networks, CNNs

图像处理与识别



循环神经网络
Recurrent Neural Networks, RNNs
自然语言处理、语音识别
和时间序列分析预测



生成对抗网络
Generative Adversarial Networks, GANs

图像生成、图像到图像
转换、数据增强



Transformer

引入自注意力机制 (Self-Attention)，自然语言处理

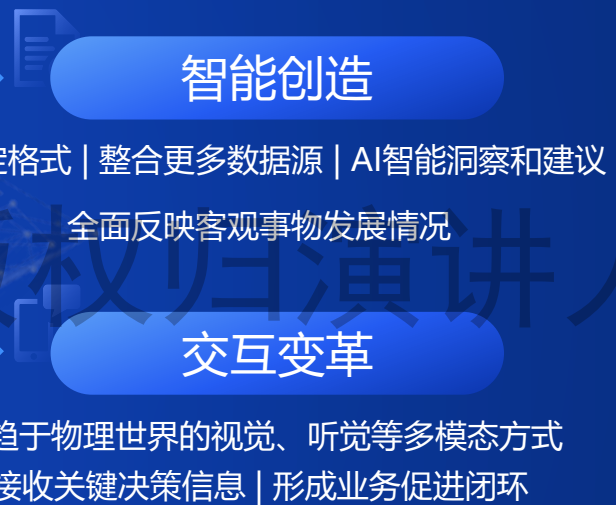
AI First已是当下数字化企业的共识，AI正在催化医学创新变革，
并重新挖掘、定义医疗健康事件之间的相关性、因果性

没有数据，我们将寸步难行

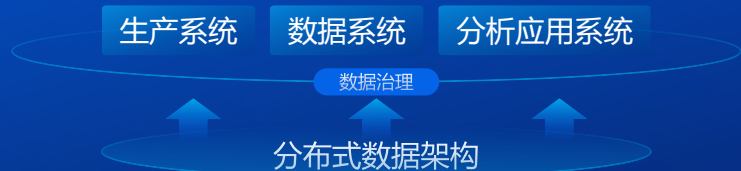
当前数据分析形态



未来智能决策形态



面向分析的主题库应能灵活、高效地构建



面向应用的数据封装实现服务化、组件化



我们需要一个

全新的架构

2023年南湖HIT论坛课件版权归演讲人所有

Be **U**nited

一体

顶层规划 | 一体设计
全生命周期管理等



U

Be **S**mart

智

规则 | 模型 |
预测 | 预警等



S

Be **E**fficient

高效

低代码 | 可视化 |
快速生成应用等

E

2023年南湖HIT论坛课件 版权归演讲人所有

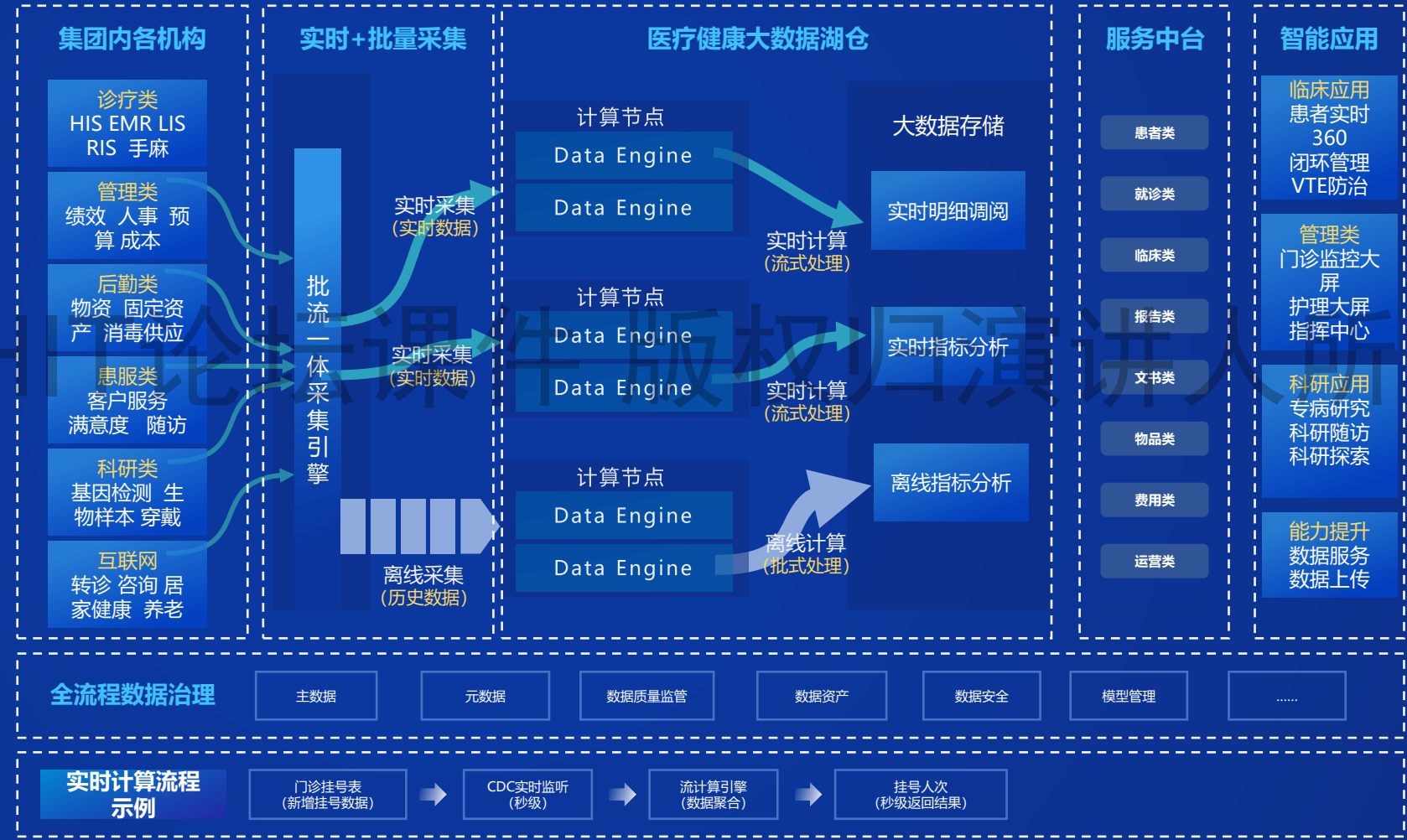
技术 | 更全更快的数据集成引擎

数据来源 全域、全量

数据时效 实时、动态

批流一体 集成引擎

- 集成各种数据类型
- 分布式，灵活扩容
- 批流模式、动态优选



技术 | 精益高效的数据工厂

数

数据要素全生命周期管理

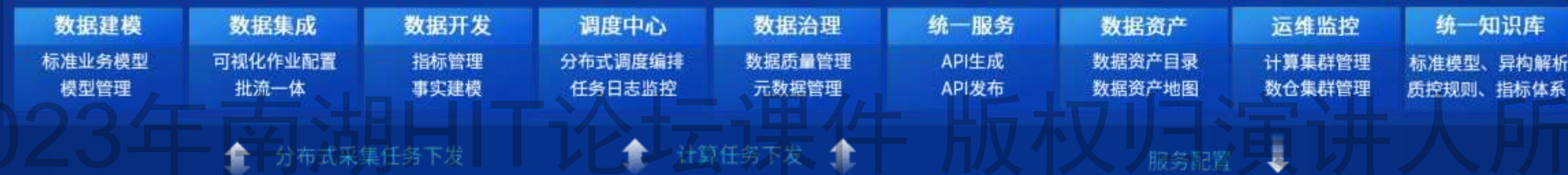
智

模型规则等知识沉淀&AI及智能引擎

器

低代码·可视化工具，快速响应需求

一站式可视化平台 (DataFactory 数据工厂)



Data Factory

2023年南湖HIT论坛课件 版权归演讲人所有

治理 | 持续精进数据标准建设

数据标准、模型、医学知识等内置到数据中台
融入业务全过程，促进数据应用更加精准、快捷

- 术语增加2倍
目前术语体系8000+

- 扩展专病知识规则
覆盖盆底专病、肺结核、甲亢、糖尿病等

- 推动诊疗路径落地
覆盖大三甲常见门诊临床科室，约900+个高频诊断

模型数 900+ 标准词根 1600+
高频字段 3800+ 业务事件 320+

1000+ 诊断



基础医学库

- 系统解剖学/局部解剖学 5,000+
- 实用诊断学/诊断学 100,000+
- 新编药理学/药理学/中国药典 100,000+
- 医学影像学/核医学 2000+

临床病种库

内科	4813	中医	400
外科	2809	儿科	248
妇产科	784	综合	2421
五官科	1341		

医嘱规则库

医嘱规则	300
用药规则	450
医保规则	150
病历质控	150
医保规则	80

治理 | 源头控制，链路闭环

前治理，数据源头控制

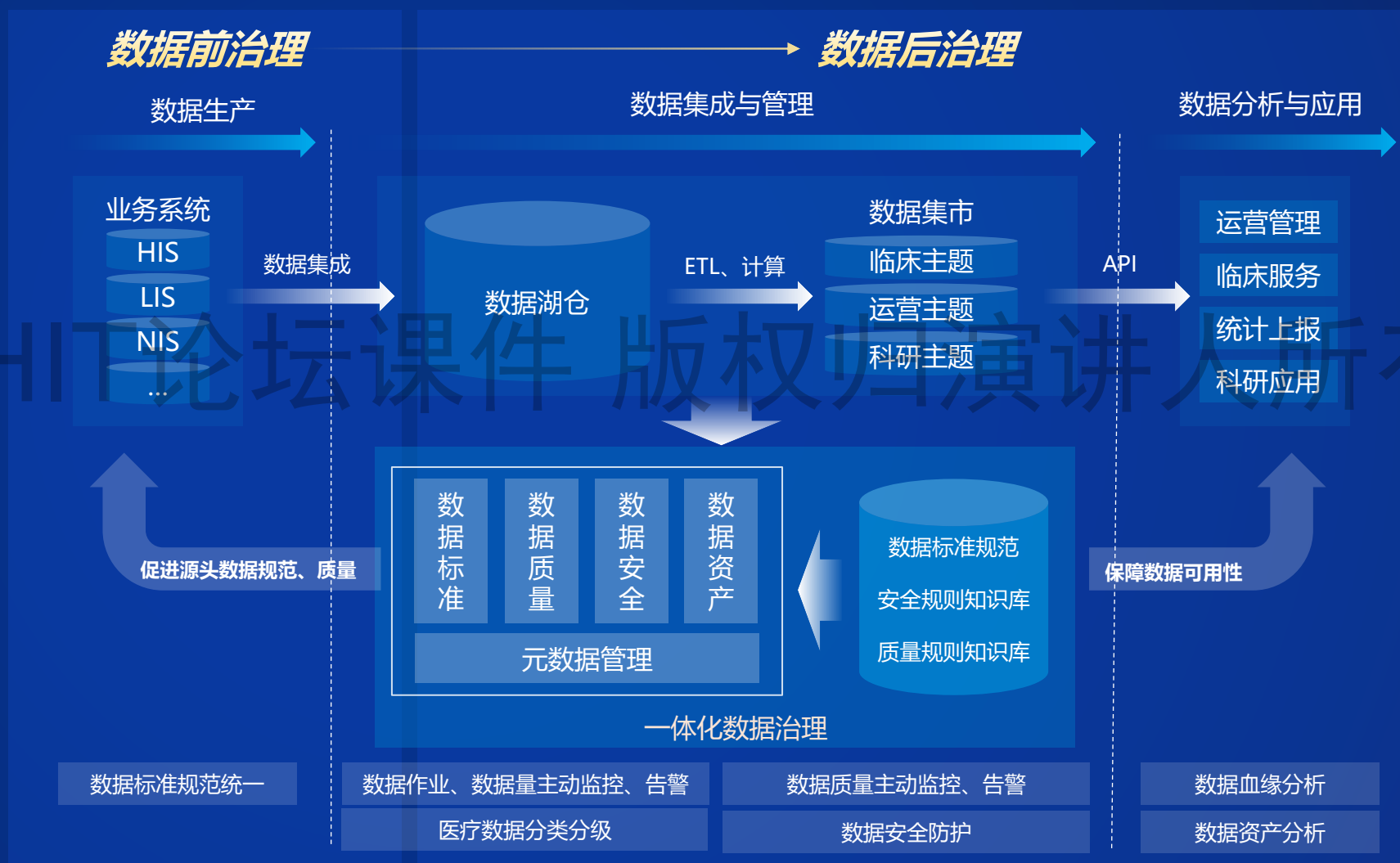
基于主数据统一管理，数据标准及一体化建模，贯通业务系统及新一代平台

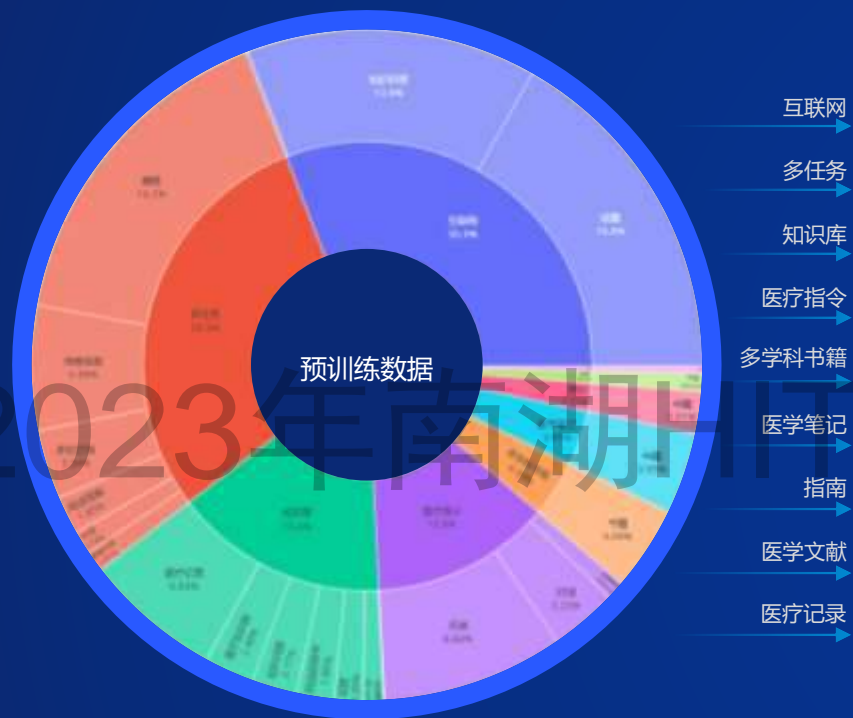
全流程，数据血缘分析

基于统一的元数据管理体系，提供数据&指标追溯、数据血缘分析等全链路追溯

知识化，内置标准规则

- 标准数据元：国标、行标等
- 质控知识：数据域、业务域、应用域规则2000+
- 数据分类分级：按5级敏感等级
- 数据模型：沉淀标准模型500+
- 主数据：资源、组织、术语标准





面向大模型的数据语料的治理

01 数据清洗

统一分类：多源异构数据采集归类
文本格式：统一转化为文本
个人信息去隐私：去隐私化模型

02 低质量数据过滤

主要针对海量多源的数据
人工审核 | 主动学习

03 数据去重

哈希：大量文本数据快速去重
Embedding：问答和指令类型的数据

04 数据采样

根据数据集的质量与分布需求
筛选采样高质量数据
提升大模型回答能力

大模型预训练&评估

WiNGPT

服务 | 数据资产的管理与应用

面向医疗全域数据资产，提供统一的多维度、可视化资产地图展示

如不同资产类型使用分析、数据标准、质量、服务、安全等动态统计，实现数据资产持续精细化运营及深度挖掘

数据标准

当前使用术语、数据元、值域等标准的数量统计以及参考各级标准概览

数据质量

显示数据质控规则及模型的数据统计，开展多维度数据质量综合评估

数据服务

目前已规范开放的数据服务开放数量及类型分析

数据安全

根据数据分级分类5级模型，明确数据资产分级分布，以及敏感数据使用情况



业务系统&数据中心&不同主题场景数据资产

全面掌握了解不同区域数据资产情况，以及临床、科研、运营具体应用场景使用分析

服务 | 数据即服务，兼顾灵活与效率

所有数据都以服务方式提供

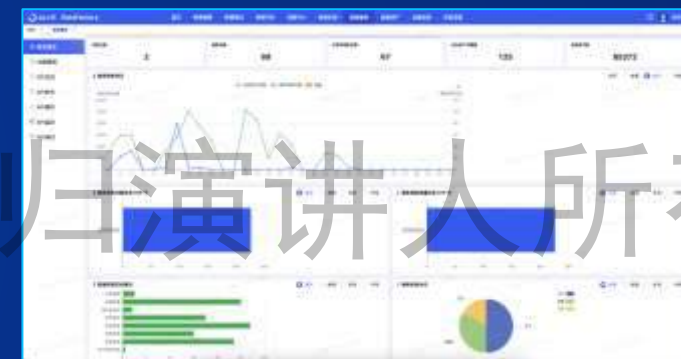
将来所有规范化数据资产，均可封装为可便捷访问的服务API
对外面向各类业务提供数据服务

服务全生命周期管理

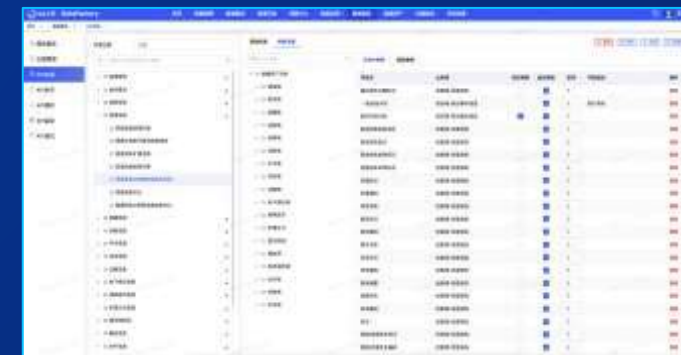
API生成、调试、发布、授权等，实现服务统一的安全管控、
异常监控，保障服务正常运行

灵活支撑
多元应用

统一稳定的
服务中台



数据服务可视化管理，零代码监控管理



统一授权监控，快速响应各类场景

服务 | 加快数据应用，反哺优化迭新

场景示例：可视化快速生成报表

快速支撑数据应用需求

面向不同数据状态分级
(已发布服务、已在数据平台未有服务、未在数据平台)
建立数据服务提供的周期
实现数据服务高效供给



依托报表工具快速创建报表流程



指标体系3000+

三大业务域主题

运营

门诊 | 住院
药品 | 手术
医保等

质控

患者安全 | 出院质控
病历质控 | 医技质控
等

资源

人力 | 财务
物资 | 科研
等



2023年南湖HIT论坛课件 版权归演讲人所有

数据赋能场景

助力高质量发展

以患者为中心 | 叠加知识的专科视图

基于知识的专项评估

疾病诊断及分型模型 | 疾病阶段

多维动态视图展现

不同病种&阶段展现 | 指标动态变化

融入专科应用过程

就诊路径 | 医嘱开立 | 量表评估



感染内科视图



妇科盆底肌专病视图



内分泌专科视图



产科就诊视图

专科应用

专科评估量表嵌入 | 专科处置流程 | 专科病历 | 专科用户界面 | 专科数据展示

专病分析

专项临床指标分析

盆底障碍性诊疗分析

糖化血红蛋白、空腹血糖值等其他临床指标分析

专项评估

感染原因、胸片、分型

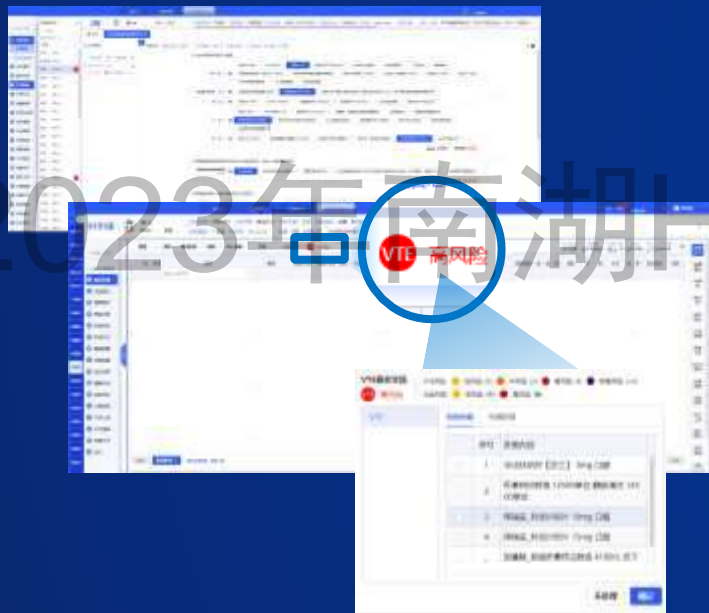
POP-Q检测
PFDI-20/PISQ-12评估

HBA1c 7.0% 指标基线

以患者为中心 | 一体数据下的场景联动

VTE评估预警

实时评估高风险预警
辅助医生及时调整物理及药物预防措施



护理任务驱动

实时消息通知
危急值快速反馈



重症评分

格拉斯哥昏迷评分 (GCS) 为例
GCS作为重症患者的核心数据
实现APACHEII、SOFA等评分的自动引用



以患者为中心的数据联动

数据定义

数据服务

数据中台

知识规则

任务引擎

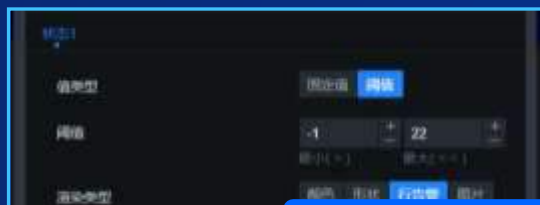
全程贯通闭环 | 管理与业务打通

基于事件驱动的实时监控

依托规则联动业务预警

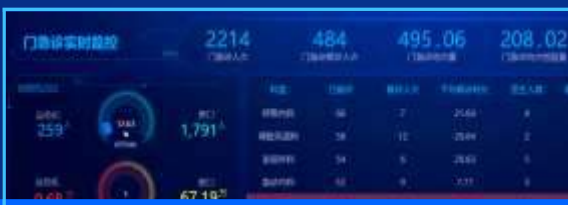
业务评价与持续改进

门诊候诊 | 药房排队 | 检查排队 | 检验（采血）排队 | 急诊&输液排队



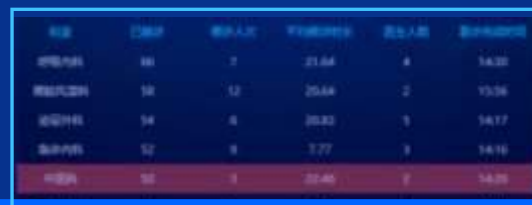
预先配置定义

设置患者候诊时长预警规则，涉及候诊时长、就诊时长、患者数、医生数等多指标综合计算



实时监控展示

实时测算各科室预估看诊完成时间、候诊时长等



动态触发预警

候诊时间超过预设时长，触发红色预警提醒



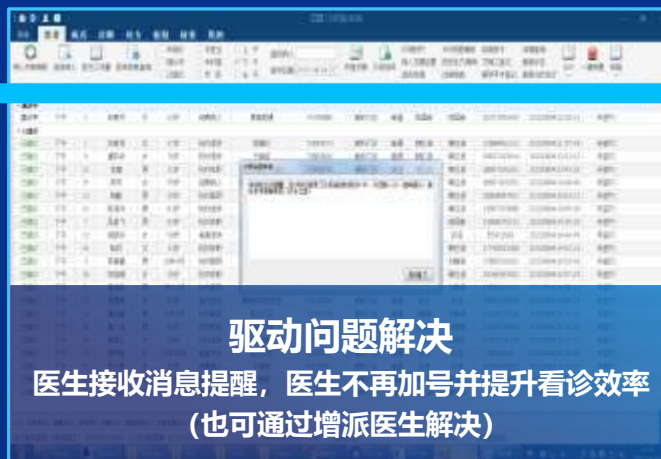
问题下钻分析

选择科室下钻，可查看具体医生看诊人次、看诊时间、上班时间等详情，定位到某医生可能当前候诊患者过多

科室	已就诊	候诊人次	平均候诊时长	医生人数	看诊完成时间
呼吸内科	66	21	21.64	4	16:40
肾内科	58	12	20.64	2	18:06
泌尿外科	54	6	20.83	5	16:27
急诊内科	52	9	7.77	3	16:26
中医科	50	3	19.3	2	16:36
急诊外科	45	4	8.73	4	16:11
口腔科	45	4	17.00	4	17:00
耳鼻喉科	44	1	17.39	1	17:39
产科	43	2	17.39	2	17:39
神经内科	39	1	16:47	1	16:47

业务优化改进

门办主任查看实时大屏，数值显示正常，预警解除



驱动问题解决

医生接收消息提醒，医生不再加号并提升看诊效率（也可通过增派医生解决）



动态联动提醒

BI与业务系统联通，点击医生头像，发送提醒消息

全程贯通闭环 | 院内业务互通

输血闭环示例



数据驱动
闭环各环节的有序规范开展

全院级、全过程知识&规则
实现全闭环数据核查与管控

联动多方系统
统一构建多闭环协作生态



全程贯通闭环 | 院外生态联通

智能中药材追溯及监管示例



实现中药材从种植到患者使用的全闭环监管与质控，可一键直达各个监管环节，支撑中药安全使用

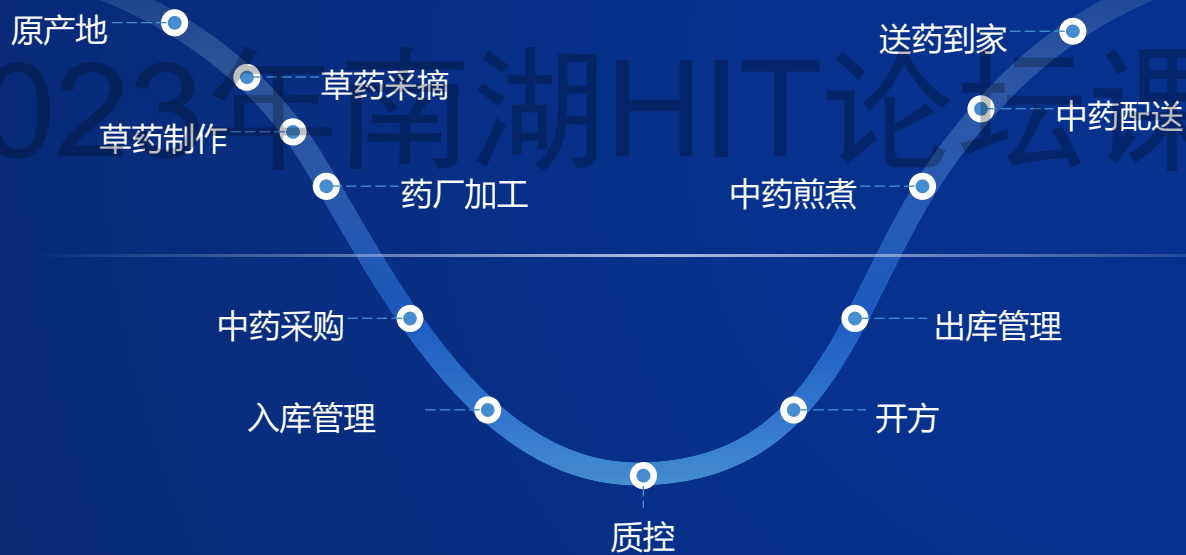
从原产地到医院入库

溯源中药材种源及种植环境

从中药代煎跟踪到配送到家

打通中药材服务“最后一公里”

中药材溯源多维度展示及统计分析



从医院入库到出库质控

实现院内中药材统一库存管理



药材&饮片溯源情况 | 患者煎药情况 | 溯源药材排名 | 煎煮监控等

聚焦专科发展与能力提升

构建管理中心-市级医院-专科联盟应用体系
满足不同层级分析需求

不断拓展其他质量评价体系（如国家单病种等）
多维度评价专科发展

建立病种字典及治理规范
持续扩展丰富专科病种

33类专科 274个病种 6000+质量指标

论证122个病种的个性化指标（平台已展示33个病种143个的个性化指标），分析采集取值定位和计算规则

医院管理机构（申康）
综合评价市级医院整体
专科专病发展情况

36家市级医院
所在医院的专科病种
展示分析

26个专科联盟
该专科在全市市级医院的
综合分析

专科病种质量评价体系

1. 选定目标病种

2. 确定落组规则

3. 论证评价指标

4. 取值口径定义

5. 数据验证核对

通用性指标

体层性指标

为了解工作总量而设

费用性指标

为了解医疗支出而设

效率性指标

为了解工作效率而设

安全性指标

医疗一般性风险
+ 专科诊疗特殊风险

个性化指标

技术规范性指标

专科特有的诊疗要求、特殊技术/设备

临床研究前沿性指标

体现学科发展、专病诊疗研究的优势，
促进市级医院紧跟国际最新临床理念与技术

专科病种质量评价模型

专科综合得分 & 排名

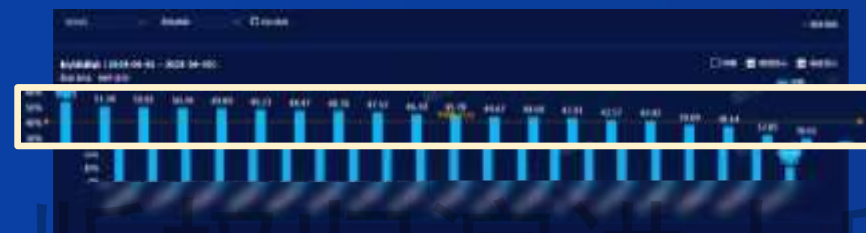
通识+个性指标分析展示

↑
通识性
指标
↓
个性化
指标

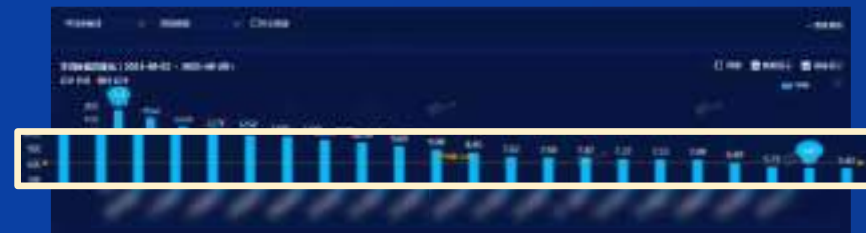


基于基准值的院间比较

材占比排名



平均住院日排名



业务促进建议

专科联盟

发挥专科优势医院的示范作用，提升病种诊疗能力

专项干预

开展处方&耗材合理性使用分析及干预

综合监测

持续监测该类病种患者药品&耗材费用、住院时间等，评价干预措施的效果

模式创新

兼顾安全前提下，推行日间手术模式，提高医疗资源利用率

专科深化发展 | 肿瘤综合诊治 (COC) 模式创新

建设有上海特色的肿瘤综合诊治中心 (COC)

德国综合肿瘤中心 (CCC)

2003年，非营利性组织德国癌症协会
基于MDT理念开展认证工作

经认证的肿瘤中心须达到基于循证指南、
多学科诊治、全病人覆盖、全过程管理、
医院信息化、院间可比较等多项标准

欧洲癌症中心 (ECC) 认证计划

建立一个泛欧数据库 | 跨学科&多专业接受治疗

健康中国“癌症防治行动”

- 落实常见癌症诊疗规范和临床路径
- 推行“单病种、多学科”诊疗模式
- 提高早发现、早诊断比例和五年生存率

以数据为抓手

强化肿瘤临床诊疗规范化路径及质量
深化肿瘤数据科研分析，提升肿瘤科研综合能力

临床
路径

多学科
诊疗

临床
研究

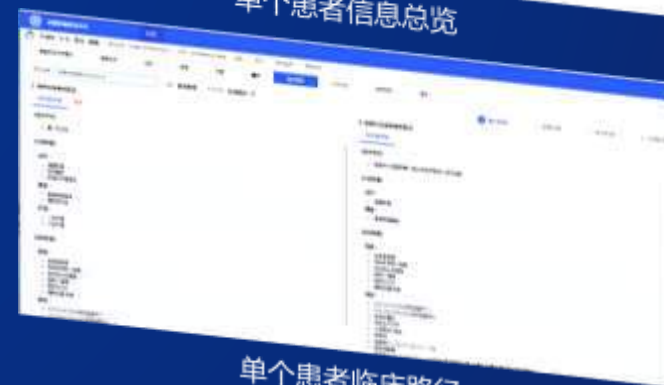
个性
随访

...

肿瘤综合诊治

市级医院肿瘤综合诊治中心(COC)

全面开展肿瘤综合诊治服务试点应用
上海市范围内已遴选2批试点 (第1批: 4家医院&6个瘤种;
第二批: 12家医院&18瘤种&1个中医舒缓)



AI智能变革

重塑医疗价值链



生成式AI



决策

医护临床决策 | 医学科研分析
医疗运营管理



互动

更广泛 | 更紧密 | 长周期
构建医患互动新模式



内容

医嘱、病历记录等
医疗文案书写

2023年南湖HIT论坛课件 版权归演讲人所有

WiNEX Copilot+ 患者服务助手



患者应答，提供专业医疗知识服务，提升就医体验。



分诊预约

提供医疗专业知识
支撑快速精准分诊

预问诊病历

快速生成预问诊病历
让医生及时了解信息



诊前开单

减少候诊次数
节约就诊排队时间

诊前咨询

提供24小时
专业化的诊疗咨询服务



2023年南湖HIT论坛课件 版权归作者所有

WiNEX Copilot+临床辅助决策助手 (CDSS+)



基于大模型增强CDSS全业务场景的集成能力。

强化一致体验

Copilot+CDSS赋能全系统、全流程、多业务场景的数据处理能力统一交互，使各产品有一致的使用体验

增强专项决策

支持专病 (COPD、ASCVD)、专项 (输血闭环、康复治疗) 等模型
WiNGPT+CDSS使得相关产品具备内容智能生成的能力

知识内容升级

知识管理闭环支持知识的全生命周期管理，满足知识库的可持续发展
以PDCA方法论指导，监控分析与改善CDSS运行情况
内置人卫社知识库，专业、权威、及时更新、更适用国内场景

* 满足卫健委《医疗机构临床决策支持系统应用管理规范》

CDSS 新增功能

应用交互

GPT+CDSS服务

智能知识助手

质量与合理性监控

诊断与治疗方案

知识查询

评估表单

模型增强

生成式模型

Langchain

ASCVD预警模型

知识模型

规则引擎

统计算法模型

知识升级

知识运维监控

知识管理闭环

人卫知识库

规则知识

知识图谱

临床指南、文献

WiNEX Copilot+ 质量控制助手

适用多医疗场景的质控模型，使传统基于繁琐规则和模板维护的构建方式，全面走向模型自学习的方式，实现如：病历内涵质控、不良事件、非计划再次手术等质控场景应用。

质量内涵管理

过程透明可信

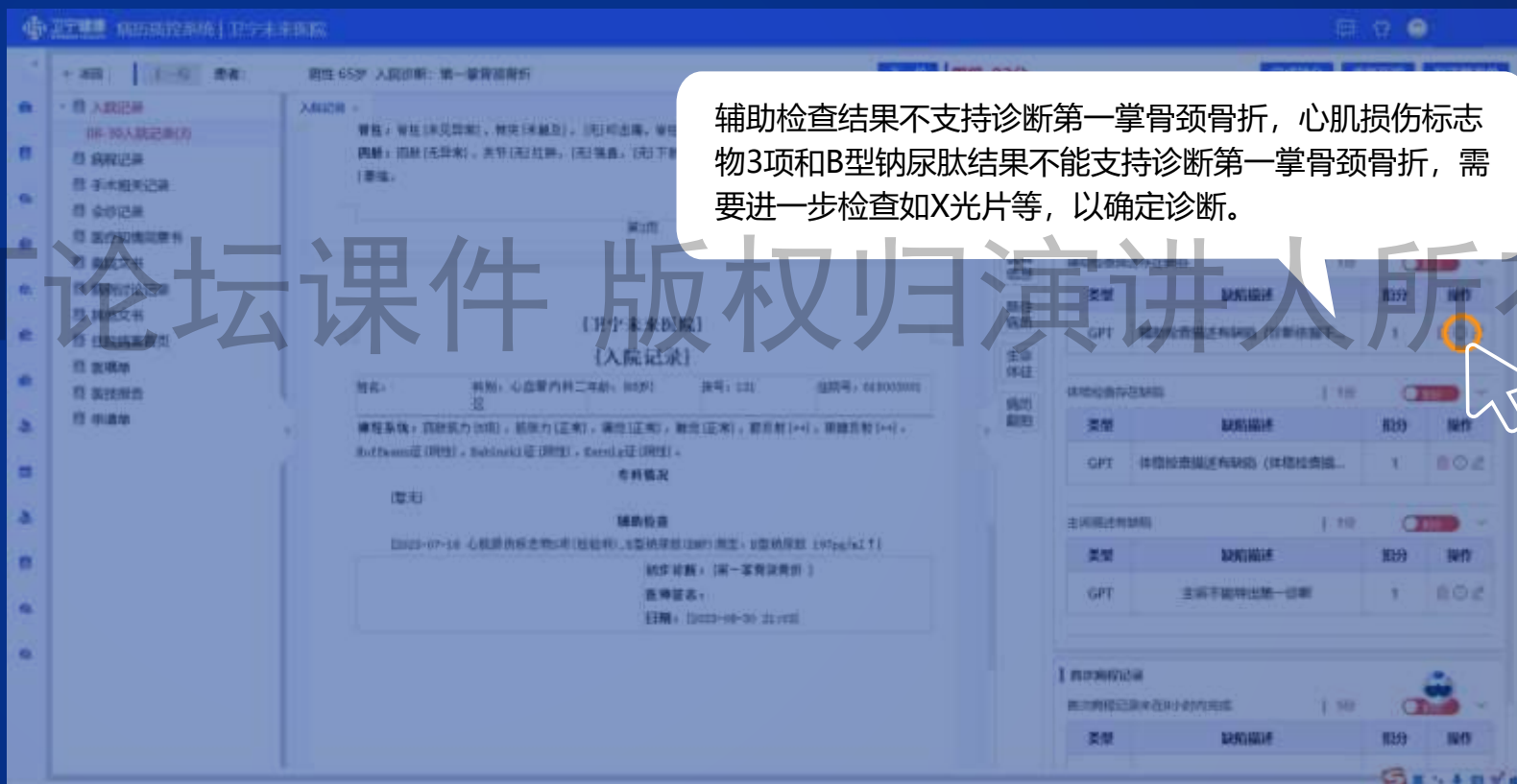
Copilot给出质控打分结果时，可同步浮现依据说明，把黑盒打开，让用户理解和信赖

上下文补全

对不良事件上报等填报内容的上下文语义理解，随时“//”补全填报内容

规则化繁为简

质控规则维护高效简单
模型自学习代替人工维护



2023年南湖HIT论坛课件 版权归演讲人所有

WiNEX Copilot+ 影像报告助手



基于影像报告数据训练后的专业模型，结合患者病情信息快速生成影像医学报告，提高影像报告的生成速度和准确性。

引导式结论

通过引导式结论，自动生成报告，极大提高医生工作效能。

对比式分析

结合患者历史报告资料，辅助对比分析，支持审核工作。

内涵式质控

自动检测报告中的潜在错误或矛盾，保障医疗质量与安全。

上海某医院试运行情况：

300+

医生人均每周调用次数趋于稳定

20%

报告书写效率提升

50%

影像报告书写质控提升

结论：稳定运行，影像诊断结论生成能力已达到中级影像医生水平。



2023年南湖HIT论坛课件 版权归演讲人所有

卫宁构建医疗健康大语言模型WiNGPT



WiNGPT是基于通用模型的医疗垂直领域大模型。结合高质量医疗数据，针对医疗场景需求专门优化和定制，通过将专业的医学知识、医疗数据融会贯通，支持医疗行业各种场景的智能知识服务，提高诊疗效率和医疗服务质量。

易交付

可交付7B/13B

高质量

医疗预训练数据和指令集

可定制

个性化、特色模型定制

场景化

丰富的业务场景对接

知识库

对接不同知识库，能力增强

可优化

持续更新语料，不断迭代优化



WiNGPT特点：“小而专”，大模型上的“小模型”



● WiNGPT1

模型大小：6.2B
训练数据：6.5G
主要功能：3大类能力4种主要医疗场景

● WiNGPT2

模型大小：7B /13B
训练数据：20G
更新功能：适配更丰富的医疗协作场景，接入外部插件，模型能力增强

● WiNGPT3

模型大小：30B /65B
训练数据：>1T
更新功能：加入医学影像多模态处理能力，模型能力更强泛化性更好

2023年南湖AI论坛课件 版权归演讲人所有

数据工程

获得高质量训练数据强化模型能力

数据分类

根据训练阶段和任务场景进行分类

数据清洗

去除无关信息，更正数据中的拼写错误，提取关键信息以及去隐私处理

去重及采样

采用embedding方法去重，根据数据集的质量与分布需求进行有针对性的采样

预训练及指令微调

扩充模型知识库，构建丰富的医疗指令集

持续预训练

扩充模型医疗知识库

自动化构建指令集

从书籍、指南、病例、医疗报告、知识图谱中自动化构建

指令优化

采用人工标注+Self-instruct、Evol-Instruct 等方案，进一步扩展指令集，丰富指令多样性

WiNGPT特点：支持可定制的私有化部署



数据安全

容器化部署在医院本地，保护医疗数据不出医院，避免数据泄露。

可定制化

允许医院根据自身的需求和条件对模型进行定制和优化，以实现更好的性能和效果。

稳定可靠

不依赖于外部网络和服务，可提供更高的系统稳定性和可靠性。

响应快速

减少数据传输的延迟，提高AI模型的响应速度。

成本可控

可通过优化硬件资源的使用来降低成本，例如对内存和 CPU 的充分利用。



WiNGPT

医疗领域大模型

2023年南湖HIT论坛课件 版权归演讲人所有

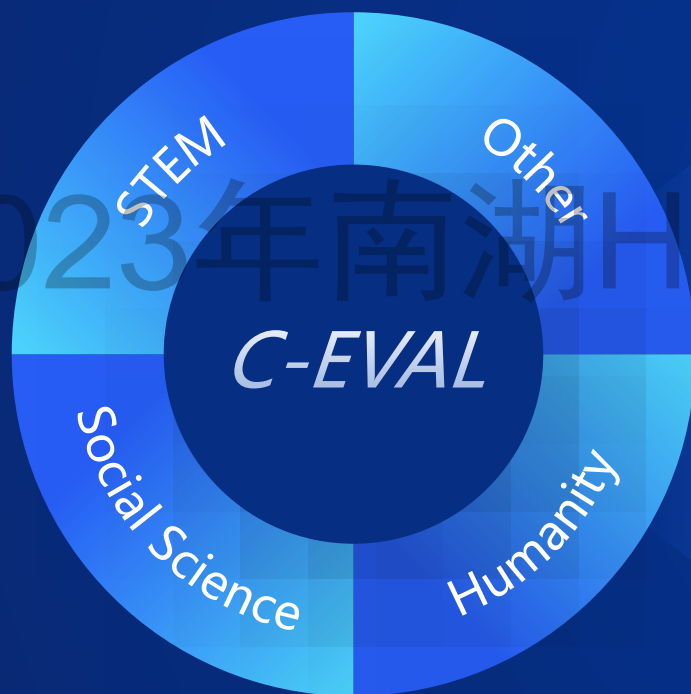
WiNGPT2-7B模型：公开测试集的表现

WiNGPT2-7B与主流中文基础大模型相比，医学相关学科领域平均分表现排名第一

中文基础模型评估 C-Eval (Zero-shot/Few-shot)

C-Eval评估包含52项不同学科和四个难度级别，

我们对四个医疗学科平均分及分数、52项学科平均分、Hard级别学科平均分进行结果展示



	平均 (医学)	兽医学	基础 医学	临床 医学	医师 资格	平均	平均 (Hard)
Bloomz-7b1-mt	48.9	53.3	50.3	42	50.1	44.3	30.8
Chinese LLaMA-13B	33.6	34.8	32	32.5	35	33.3	27.3
ChatGLM-6B*	37.8	35.2	39.4	37	39.5	38.9	29.2
Baichuan-7B	42.4	44.3	43.4	35	47	42.8	31.5
Baichuan-13B	55.6	57.1	56	52	57.3	53.6	36.7
Qwen-7B	59.5	61	57.1	56	63.9	59.6	41
WiNGPT2-7B-Base	69.6	61.4	70.9	69.5	76.7	57.4	42.7

正在推进的WiNGPT2-13B版本，评估表现分更优（医学平均：70.1）

数据

DATA

算法

生成式AI大模型计算跃迁突破

场景

丰富且有挑战的医疗场景交互

算力

数字空间基础设施的重要支撑

数据作为新型生产要素
为医疗健康行业发展注入更多确定性

数据智能助力医院锚定未来