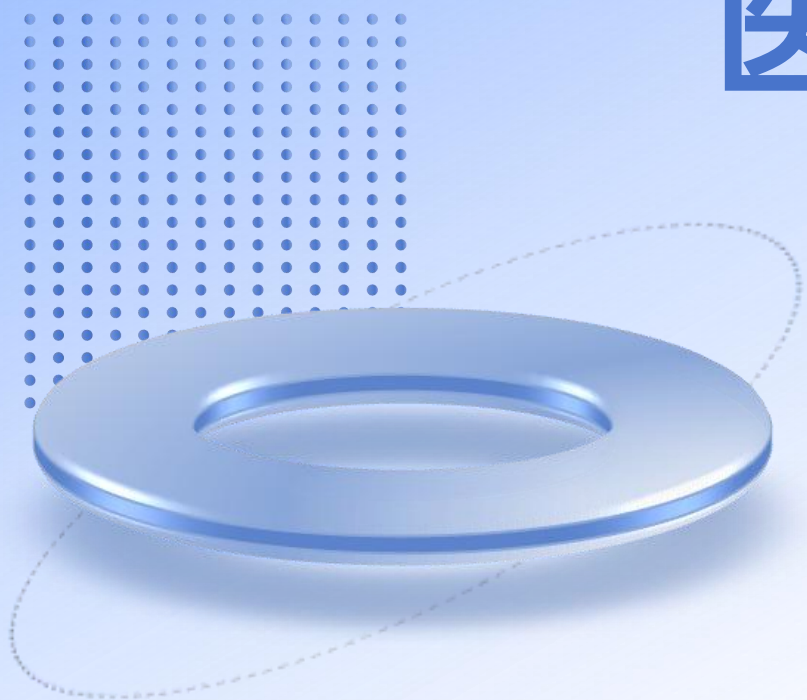


医疗数据质控标准化 处理框架



....

LET'S EMBARK ON TODAY'S SHARING JOURNEY TOGETHER

医数云析

背景与需求：AI医疗时代下做数据质控标准的意义

医疗数据爆炸式增长，质量参差不齐

01

数据存在标准不统一、采集方式多样、设备误差、人为录入错误等问题，导致数据质量低下，直接影响AI模型的训练效果和临床决策的可靠性

AI医疗应用对数据质量的高要求

02

低质量数据会导致AI模型过拟合、泛化能力差、预测结果不可靠，甚至引发医疗风险

数据质控标准化框架的核心价值

03

- (1) 提升AI模型训练效果
- (2) 推动医疗数据标准化与互联互通
- (3) 降低合规与伦理风险

目录 CONTENTS

01 概述

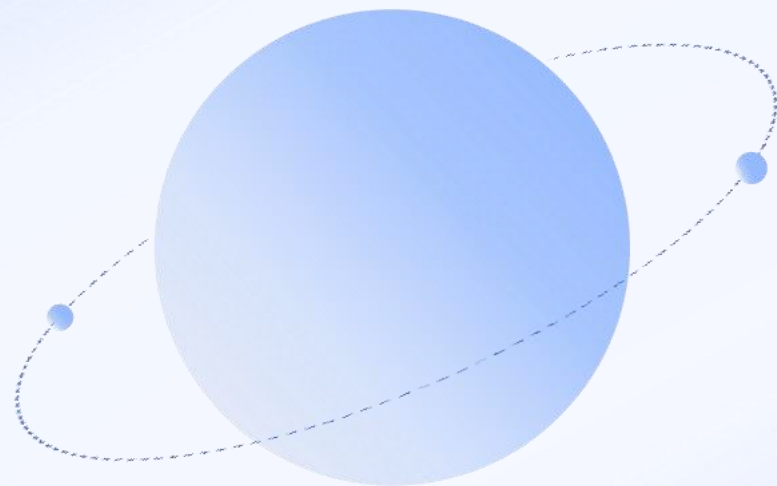
02 规则层

03 治理层

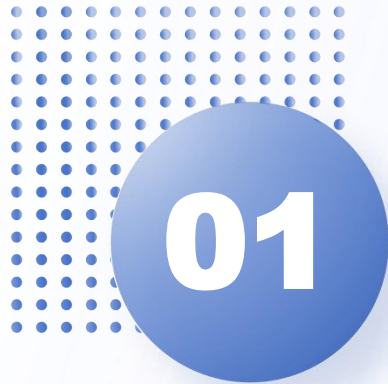
04 清洗层

05 技术优势

06 应用案例



....



PART 01

框架概述

Let's embark on today's journey of sharing and communication together

架构设计

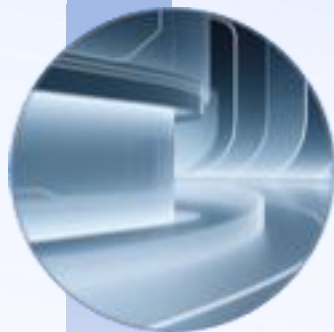


三层架构组成

规则层：按照行业标准，对不同数据的类型特性进行规则化处理。

治理层：对不同设备产生数据的多样复杂性进行纠偏，需定制精细化处理。

清洗层：通用模型识别，枚举字典排查，双层矫正。

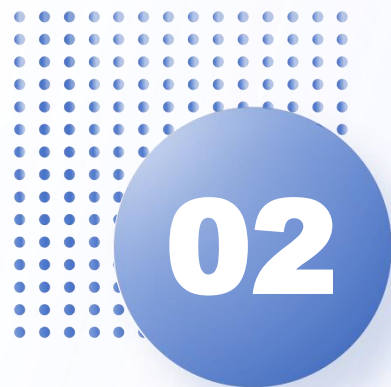


标准化与智能化结合

框架将标准化流程与人工智能技术相结合，既保证了数据处理的规范性，又利用AI的智能特性提高了处理效率和精度。这种结合方式能够有效提升数据质量，为数据分析和决策提供可靠支持。

灵活性与规范性兼顾

在保留最大应用灵活性的原则下，尽量处理到标准化格式，要即不损失信息又能适用各种输入端格式，保证数据的及时、通用可用性



PART 02

规则层

Let's embark on today's journey of sharing and communication together





规则层

01

定义：是一套带参的、按各项目分项建立的一套质控处理规则，分通用规则和本地规则，都有量化模型

意义：按照既定的量化处理法，对数据进行处理，以达到现行行标等。

通用规则

02

定义：是根据每个项目的 方法学不同、单位不同、数据类型不同、参考范围不同而需要的调整，通过行业通用库导入；

意义：‘统一度量衡’，达到可比可合并的统一形式。

本地规则

03

定义：是根据原始数据的产出干预不同（如校准模型、计算方式、可报告范围、干预条件等），做本地化调整，差异消除。

意义：差异消除归类，升维或降维，在尽不损失信息的情况下。



参照现行、原研新生

医疗数据的统一是极其复杂的事，不到用时不能”开启“处理，处理了往往就‘废了’、损失了、失真了。

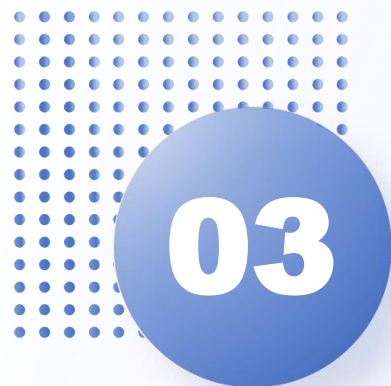
处理原则和难度是：要考虑到所有后续可能的应用场景，已有的和前瞻性的研究。

行业通用库导入

即使有通行的标准库可导，貌似很能标准化的东西，但在大数据量下，往往这些之前被认为可接受的‘容差’也会造成错误。这就貌似对行业标准加上了个适用数据量、维度的限制概念。

本地策略定制

一提本地化策略就是个头疼概念，七门八样，哪个维度区分的都有，按一个模子削出来不‘砍肉’根本不可能。
是对‘园丁’技术、专业度、知识面、分析能力的全面考验。



PART 03

治理层

Let's embark on today's journey of sharing and communication together



处理复杂数据属性



设备固有制约

治理层处理数据采集设备固有制约，如精度、采样频率等，确保数据在设备限制范围内尽可能准确，减少设备因素对数据质量的影响。

01



不可控环境因素

考虑数据采集过程中不可控环境因素，如温度、湿度等，治理层通过算法校正补偿，降低环境因素对数据质量的干扰。

02



人为条件制约类

有可归为标准类的，有不可归的，有配合归的，也有不愿归的，就像各家有各家的生活方式一样。技术也不是万能，人事也要配合。

03

专利技术



体检大数据偏差矫正法

01

专利技术“体检大数据偏差矫正法”通过体检大数据计算群体偏差系数，对数据集进行后质控纠偏，以达到和外部验证集融合的程度。

AI通用大模型+专业知识库训练

02

利用通用语义模型的优势和专业知识库想结合，连接部分数据库处理权限，在可控风险范围内让AI来当专家和管理员，帮我们完成一些工作。

图像描述型数据专项治理

03

针对形态学影像学的图像描述性语义字段，我们根据医学特定的学科范类和研究目的来做分型处理框架。即从数学上重新划分医学场景目的，把医生思维转换为数学AI思维模型，构建分析处理流模型。

医学数据专项治理

描述型字段的通用处理

描述的文字段数据，不能在不确定需求的情况下去治理，会损失。所以我们的方案是：做每一把开启的‘钥匙’，每把钥匙有他特定‘抓手’，去提取其有意义的维度。

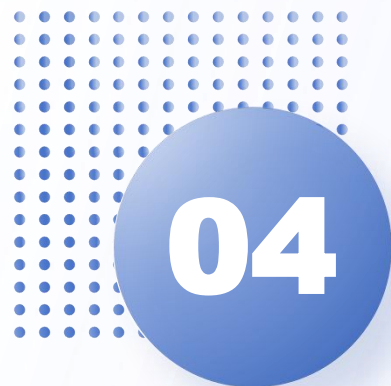
同种描述的合并处理

中国语言博大精深，一种症状十种描述，一个部位八种叫法，明确是一个事物的东西合并就用基础大语言模型。

图像等非语义字符化类型

图像：往往由特定产生设备的厂家去做语义转化处理。非字非图类：如染色体核型结果，是图还是符，报告上是图也要，字也要，数字还要。医生描述更是哪种简单用哪种看心情，此类情况最复杂只能因地制宜。





PART 04

清洗层

Let's embark on today's journey of sharing and communication together

名称/定量/定性数据



枚举字典

对于常见多种名称、错误类型适做枚举，明确归类，方便调整。

逻辑纠错

往往是一些明显的，前端拦截漏洞造成的，入库后‘滤渣’，这样的‘渣’剔出来容易，改正难，采用‘淘米’制，吃多少淘多少。对于多发有一定规律性的来写脚本处理。

数据一致性校验

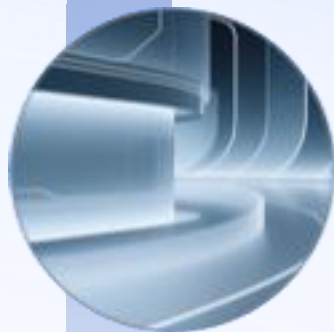
对数据进行一致性校验，确保数据在不同维度和层面的一致性，为数据分析和决策提供可靠支持。

非结构化数据



大模型底座识别修正

文本类，通用大模型底座识别修正，即可达到理想效果。如果有，再加上专业知识库学习后的AI用词更地道。

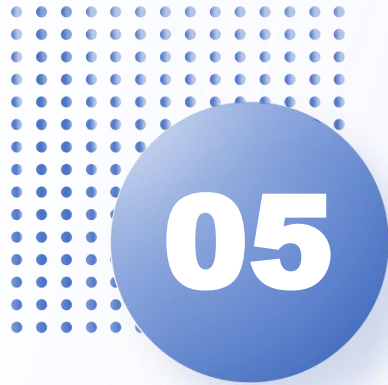


清洗原则

一般只做语义语法字词上的明显错误，如无后需明确，不做结构化处理。

数据精细化处理

一般是只做修正，不做雕琢，防止歧义产生，数据原始味道失真。



PART 05

技术优势

Let's embark on today's journey of sharing and communication together



规范性与智能性提升

01

将标准化流程与人工智能技术相结合，既遵循了行业规范性，又利用AI的智能灵活性，优势互补。

灵活高效，按需调用

02

嫁接在链路层和应用层之间，可常开启服务于特定的应用，也可按需提取，定制输出。

外部标准库更新

03

为达到最新的更好的处理方案，策略比对，定期导入新框架，新技术，选择性更新，保持跟行业前沿接轨。

专利算法、云端质控溯源链



多项专利算法应用

- ★“体检大数据室间偏差矫正法”
- ★“生物学变异动态时序法”
- ★“相关性归一算法”
- ★“不同数据量级别下不同精度项目的统计学指标估算法”

01



云端区域比对，质控溯源证书

有云端众多医疗机构的质控平行数据，可为单家单项目做区域大数据比对，偏差计算，质控溯源定级等。

02



参与标准计划，共享课题研究

web3时代模式，数据贡献者即受益者，鼓参与我们的研究项目，分担横向、纵向课题的研究。

03

行业高度和成绩



参与标准制定

01

参与多项团体标准和企业标准制定，确保符合行业最佳实践，具有广泛适用性和权威性，提升产品的行业认可度。

医学大数据创新平台

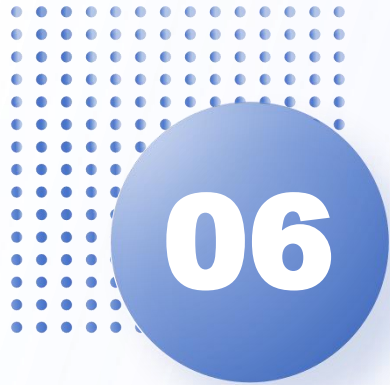
02

在天津总部与检验医师协会协办的“青年医学大数据创新平台”，以创新前沿的AI大数据相关项目为研究，以各成员医疗单位为实践基地，已辐射至10多家京津冀鲁周边的医疗单位，与其联合申报大小课题30+项。

高校产学研

03

在天津总部地区与天津医科大学生物医学工程学院联合建立“医学大数据AI医疗应用项目成果孵化中心”、“医数云析大学生创新、创业实践基地”，并以此发展全国医学高效生信类专业的多中心联合。



PART 06

应用案例

Let's embark on today's journey of sharing and communication together

客户应用案例



★急性缺血性脑卒中临床决策系统

为某家医院神经内科做的脑梗愈后评估模型，有化验指标、行动认知评分指标、年龄病史等，采用神经网络和决策树等几个主要模型，综合预测准确率达82%，并取得了团标、数据资产认定。

★老年体检大数据AI分析系统

采为社区公卫老年查体做的体检报告AI分析系统，采用盘古大模型+本地化训练共用的模式，对报告给出疾病预测和健康分析建议等，并对每个单项指标做多年的趋势性分析。

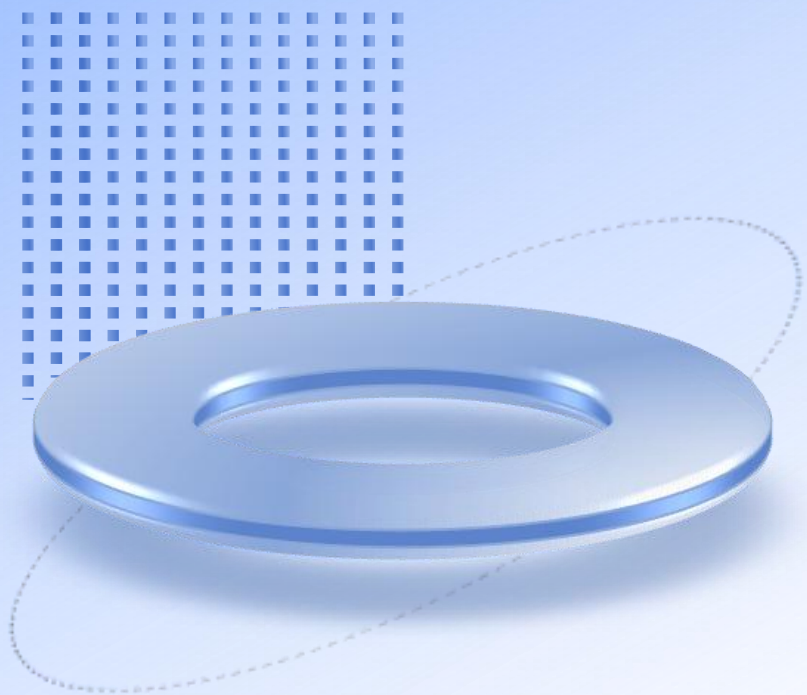
★为某上市公司全国1000家客户医院定制的质控区域管理系统

区域“室内质控室间化”比对系统是我们的核心产品，也是此单机版的质控框架产品的区域版依托和扩充，为其提供外部验证集和模型应用的泛化支持。

总结



在AI医疗时代，数据质量直接决定智能系统的上限。构建标准化、智能化的大数据质控框架，不仅是技术需求，更是保障患者安全、提升医疗效率、推动行业创新的关键基础设施。



感谢观看

THANKS

....

THANK YOU VERY MUCH FOR WATCHING

医数云析