

第二届中新数字经济与人工智能高峰论坛

健康医疗数据的确权与流通

高阳

南京大学智能科学与技术学院书记兼执行院长

第二届中新数字经济与人工智能高峰论坛
THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL
ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

报告4-高阳教授-健康医疗数据的确权与流通

文件 | C:/Users/jerry/OneDrive/桌面/4.29%20Conference/报告4-高阳教授-健康医疗数据的确权与流通...

2 / 63

健康医疗数据确权与流通

高阳

南京大学健康医疗大数据国家研究院

31°

搜索

11:19
2024-04-29

第二届中新数字经济与人工智能高峰论坛
THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL
ECONOMY AND ARTIFICIAL INTELLIGENCE



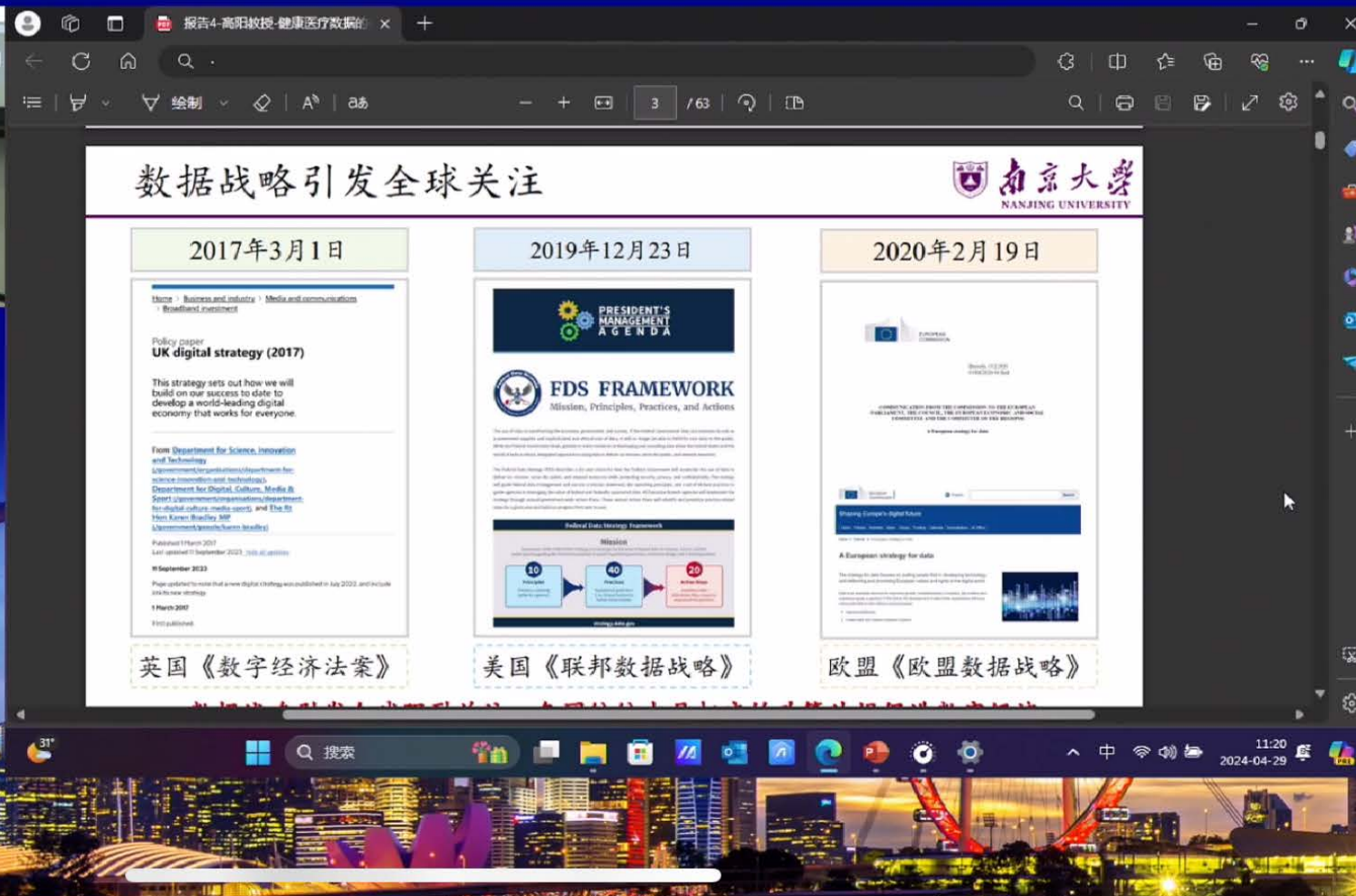
中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

数据确权与流通背景



中共中央全面深化改革委员会第二十六次会议（2022年6月22日）

加快构建数据基础制度

习近平强调，“促进数据高效流通使用、赋能实体经济，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”



我国政府高度重视数据要素并颁布多项政策

- 2017年11月《十九届中共中央政治局第二次集体学习》首次提出数据要素概念。
- 2022年6月22日中央深改委第26次会



我国政府高度重视数据要素市场培育

- 2023年3月7日，十四届全国人大一次会议公布了《国务院关于提请审议国务院机构改革方案的议案》，其中特别提出要组建国家数据局。
- 2023年12月，国家数据局等17部门联合发布《“数据要素X”三年行动计划（2024-2026年）》重点强调加强医疗数据融合创新，推动智慧



国内多地已先行成立大数据局

- 浙江、广东、贵州的这些大数据局已经先行成立了好几年时间。
- 用数据确权推进数据在市场的交易，

31°

搜索

11:20
2024-04-29

第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

报告4-高阳教授-健康医疗数据的

数据确权与流通

□ 健全的数据确权与流通机制将为金融、政务、工业、医疗等领域带来关键支撑，能提高信息透明度和安全性、增强业务协同和服务水平、**促进创新驱动**。

金融领域	政务领域	工业领域	医疗领域
《金融统计报表制度》 《个人金融信息保护试行办法》 《金融数据共享监管试行办法》	《政府信息公开条例》 《国务院关于加快建设全国一体化在线政务服务平台的指导意见》	《数据安全法》 《工业和信息化部关于加强工业互联网安全工作的指导意见》	《个人医疗健康信息安全管理暂行办法》 《关于促进 医疗卫生健康大数据 有序共享的指导意见》
金融数据确权与共享机制研究	政府数据开放与共享的关键技术研究	工业数据确权与共享机制研究	医疗大数据营私保护与共享机制研究

在医疗领域中的数据确权与流通尤为关键

5

第二届中新数字经济与人工智能高峰论坛
THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL
ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

数据要素-“双码”

部门的“监管权”滥用

“擅自对不符合赋码条件的人员赋红码，严重损害健康码管理使用规定的严肃性，造成严重不良社会影响，是典型的乱作为”

——XXX新闻网

个人的“**隐私权**”泄露

“健康码在完成其使命后，其收集、存储、流转的数据是否真正做到销毁，是后疫情时代的重要命题”

——21世纪经济报道



健康医疗数据要素具有体证明显、隐私性强等特点，对社会发展与稳定至关重要

第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



南京大学
NANJING UNIVERSITY

中美欧：个人信息保护的立法和保护要求

□ 强化个人信息保护方面的立法和执法



《个人信息保护法》（2021年11月）

- 管辖范围大（境内、境外）、违法后果严重（5%；责任人100万元人民币）。
- 《关于开展银行保险机构侵害个人信息权益乱象专项整治工作的通知》（2022年8月）
- 银行、保险公司深入全面查找2021年以来个人信息保护方面存在的问题，确保整改到位。



《美国数据隐私和保护法（草案）》ADPPA（尚未生效）

- 在联邦层面建立统一的数据隐私保护法律，加强消费者隐私权利。
- 《金融服务现代化法案》GLBA（2000年3月）
- 对金融信息的收集、使用、披露进行规制。
- 《加州隐私权法案》CPRA 《弗吉尼亚州消费者数据保护法》CDPA（2023年1月）
- 保护各州消费者对个人信息的行使权利。



《通用数据保护条例》GDPR（2018年5月）

- 适用范围广泛（欧盟境内、境外），处罚金额高昂（最高2%或2000万欧元）。
- 《电子隐私条例》ePrivacy Regulation（2021年2月）
- 通过让用户更好地控制自己的个人数据来加强电子通信领域的隐私保护。



2022年7月

违法收集、过度收集个人信息等。
80.26亿（5%）；100万元。



2022年12月

违法收集13岁以下儿童个人信息。
与FTC达成和解后，合计支付
5.2亿美元。



2021年6月

卢森堡国家数据保护委员会（CNDP）
广告定位功能未经适当同意。
7.46亿欧元。

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



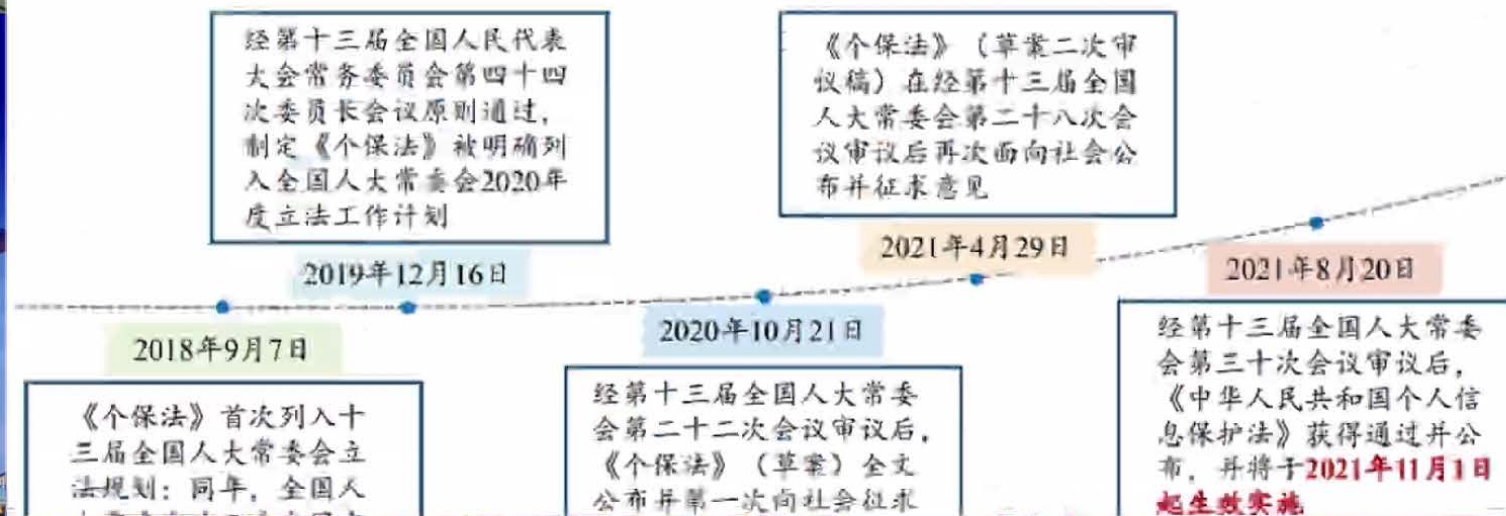
中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

个人信息保护法发展历程



数据要素的定义

□ 数据要素的定义：

数据要素指的是根据特定生产需求汇聚、整理、加工而成的计算机数据及其衍生形态，投入于生产的原始数据集、标准化数据集、各类数据产品及以数据为基础产生的系统、信息和知识均可纳入数据要素讨论的范畴。

□ 数据要素的特点：



容易复制

复制速度快，成本低，难度小

确权困难

数据所有权、控制权、处置权的识别与其他要素有显著差异

非排他性

复制给A并不妨碍复制给B，A在使用时并不妨碍B也同时使用

传播迅速

基于信息化手段，传播无距离限制、规模大

难以估值

价值异质、不确定性和“阿罗信息悖论”

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

健康医疗数据要素



健康医疗数据要素，例如：患者健康信息(个人体征...)、医疗检查结果(血常规...)、药物信息(阿莫西林...)、诊疗记录(病历...)等。



健康医疗数据要素丰富多样，个体数据及机构数据之间关系紧密

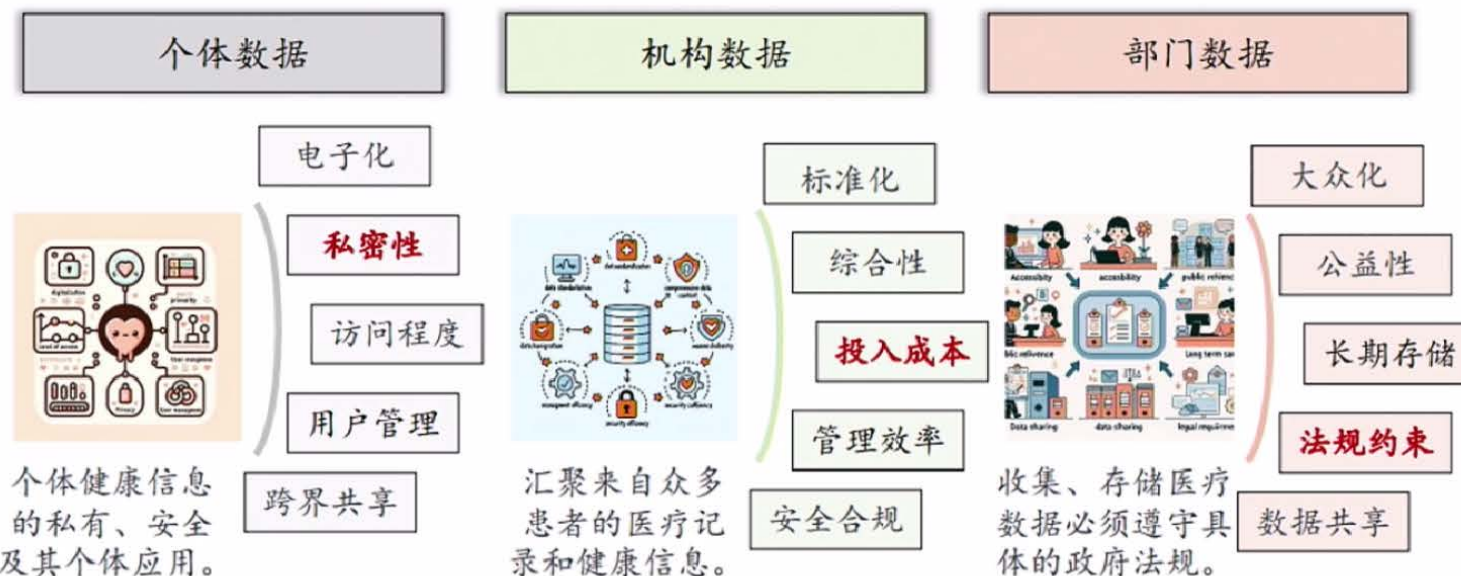
高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

健康医疗数据特点

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



健康医疗数据私密性强、成本高、数量大、法规约束强、潜在价值高

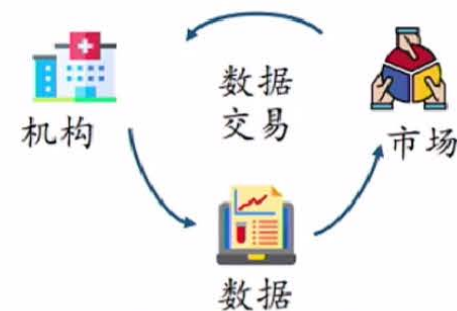
健康医疗数据掌握主体



医院数据包含多个个体数据，数据**同时**由个体和机构**掌握**

有什么策略能够将个体和机构对医疗数据**所属解耦**吗？

机构对解耦去敏后的数据有更好的处置权**便于市场流通**



- 健康医疗数据掌握的主体主要由：个体、机构、部门三大主体构成。部分健康医疗数据三者同时拥有，**对于健康医疗数据的归属和边界没有明确的界定**

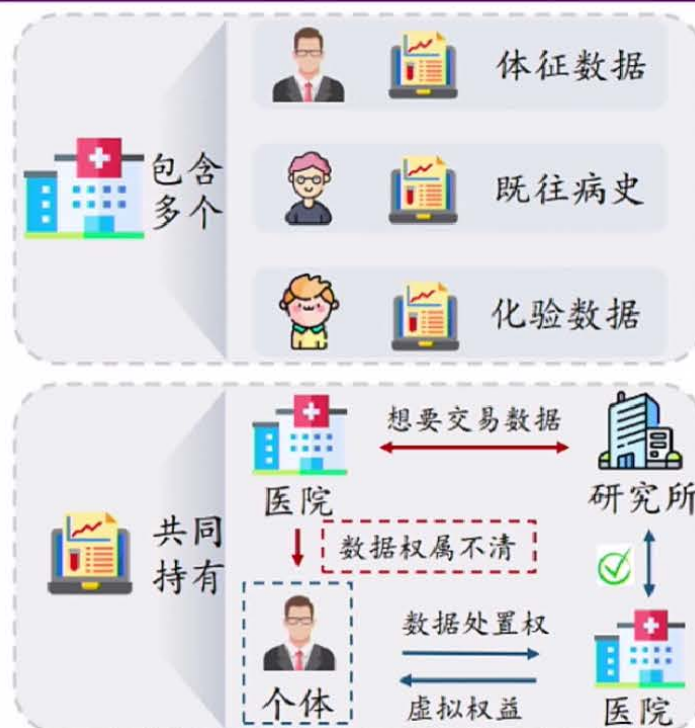
高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

个体与机构数据归属

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



- 医院数据包含多个个体数据。
- 医院机构数据个体数据之间权属界地不清。

如何确权?

- 医院通过虚拟权益向个体申请获得处置权。
- 机构之间流通交易医疗数据。

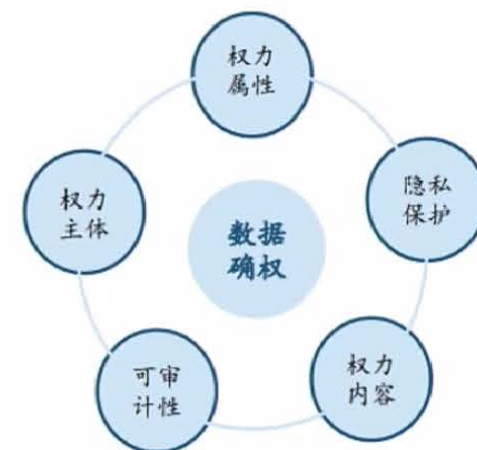
如何流通?

数据确权

□ **数据确权**是指在法律和技术层面对数据资产确立所有权、使用权及其相关权益的过程。

□ **数据确权**的主要目标包括：

- **确定数据权利属性** - 明确数据应受到何种法律保护
- **明确数据权利主体** - 决定数据所有权归属
- **规定权利内容** - 界定各主体对数据的具体权益范围，以及如何分享数据产生的经济效益
- **实现可审计性** - 支持第三方对数据所有权进行有效验证和审计
- **保护数据隐私** - 在不披露原始数据内容情况下，证明数据所有权属性



数据确权是健康医疗数据高效流通的基础

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



NANJING UNIVERSITY

国际数据确权



《数据策略》



《通用数据保护条例》



欧盟数据要素市场化的四种做法

- 严格保护数据产权，形成完整的数据权益法律体系
- 建立“共同数据空间”，创新数据跨领域共享交换工具
- 增加数据供应规模，提高公共数据和其他数据共享程度
- 提高国际交流门槛，提升数据全球流通话语权 and 影响力



《数据保护和隐私机构》



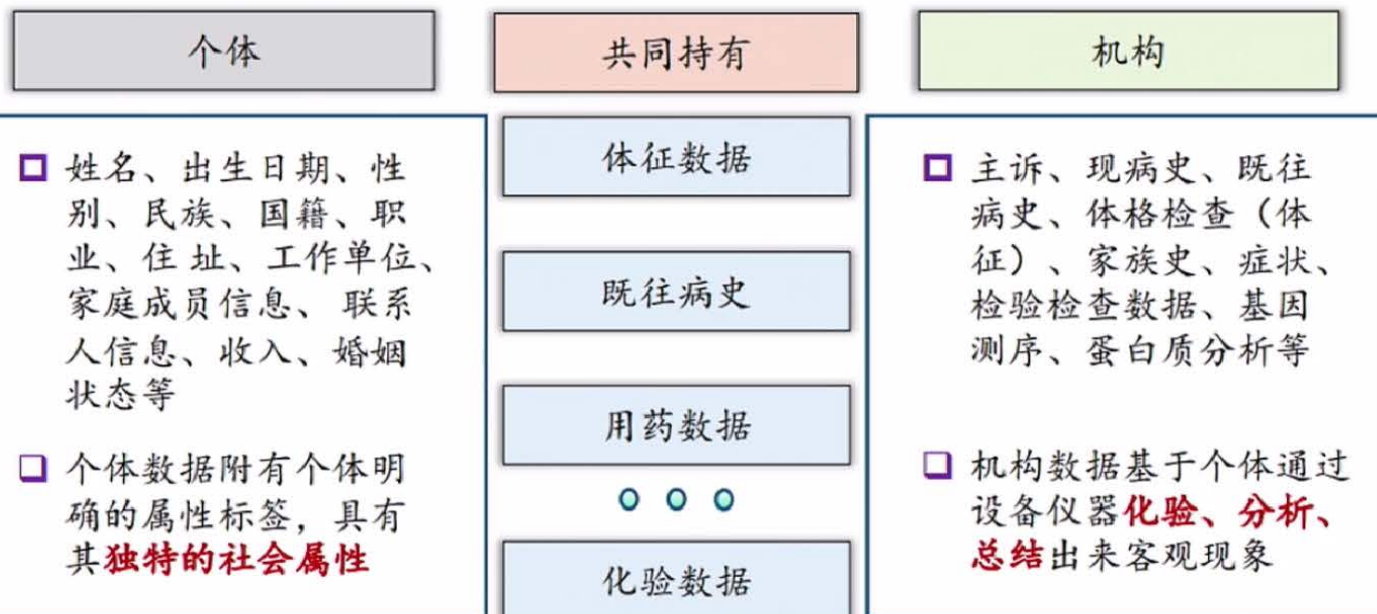
The G7
Data Protection
and Privacy
Authorities

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

各国尤其重视医疗数据，急需明确定义医疗数据的确权与权属边界

个体与机构的权属边界



个体拥有其独特的社会属性，不同于机构化验设备、医师等的总结现象数据

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

虚拟权益机制

虚拟权益

- 可依托医保为权益载体，医院在交易所进行数据交易，所获得的经济回报一部分通过医保的形式从个体手中交换个体数据的处置权，个体获得相应的虚拟回报后，可以通过医保支付其在医院所形成的费用。



- 虚拟权益可以有多种不同的载体，医保是个体在医院问诊时常用的一个载体，医院通过医保支付虚拟权益的方式也能够更快的被个体所接受并使用。

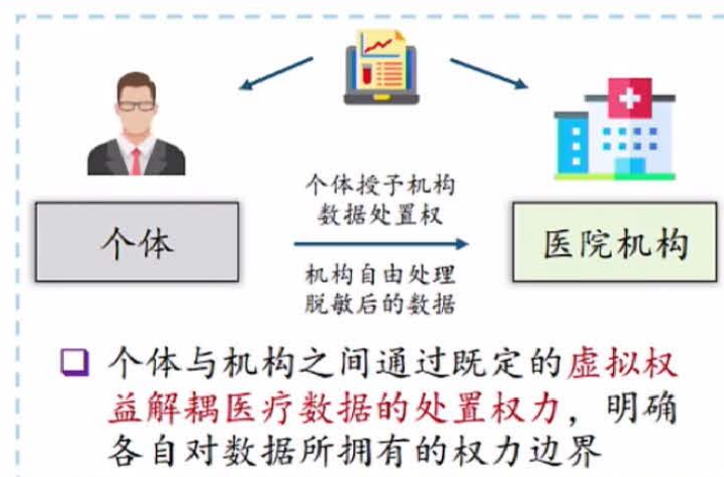
高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

健康医疗数据确权

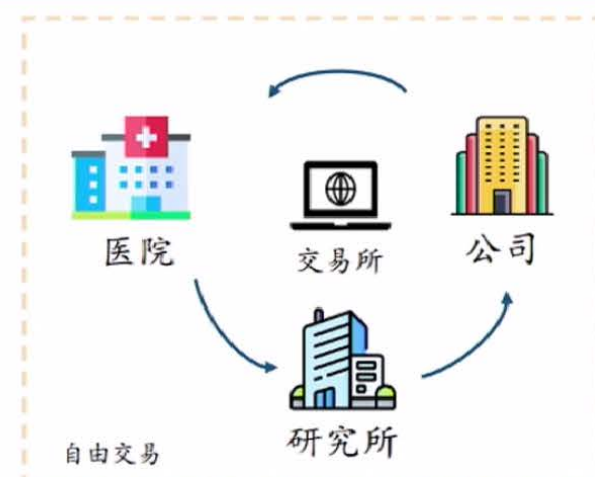
高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



政府部门拥有数据的所有权力，通过税收等其他形式回馈给社会个体、机构

个体与医院机构数据所属解耦

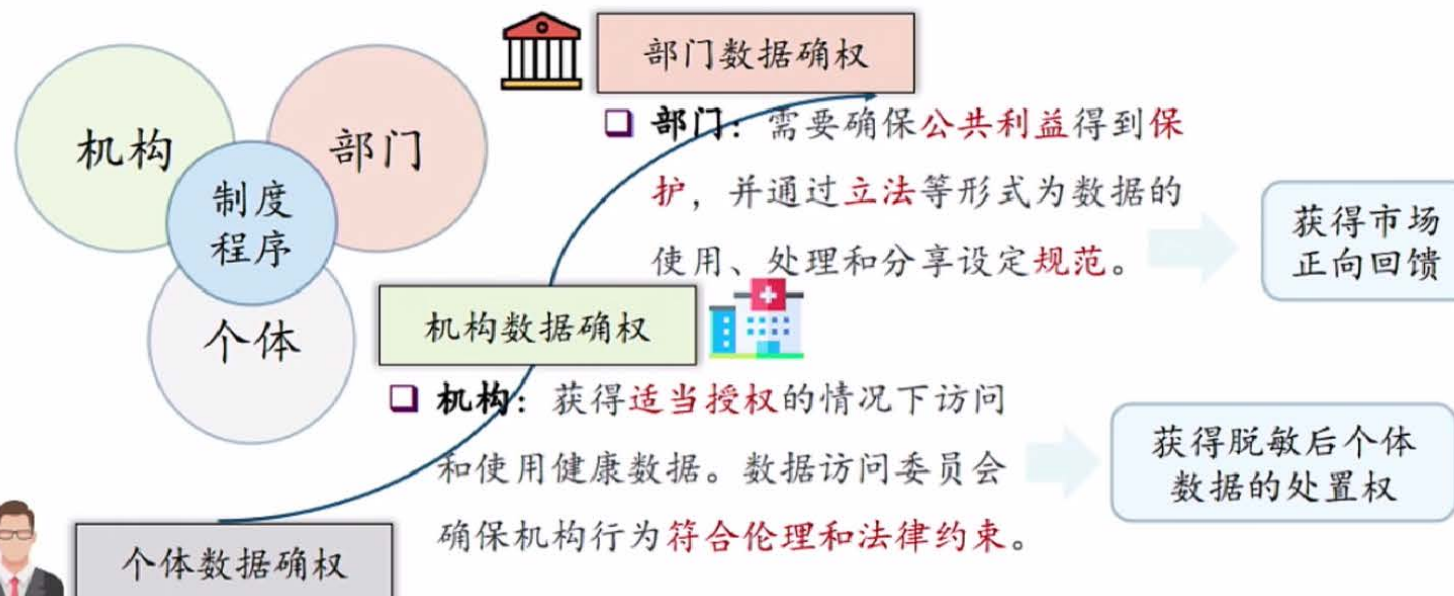


政府部门负责监管，监督，维护保障

机构之间在市场自由交易

健康医疗数据边界的明确促进健康医疗数据在交易市场上更好的流通

健康医疗数据确权



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



健康医疗数据要素具有巨大的经济价值

体对自己的数据有知情同意的权利。

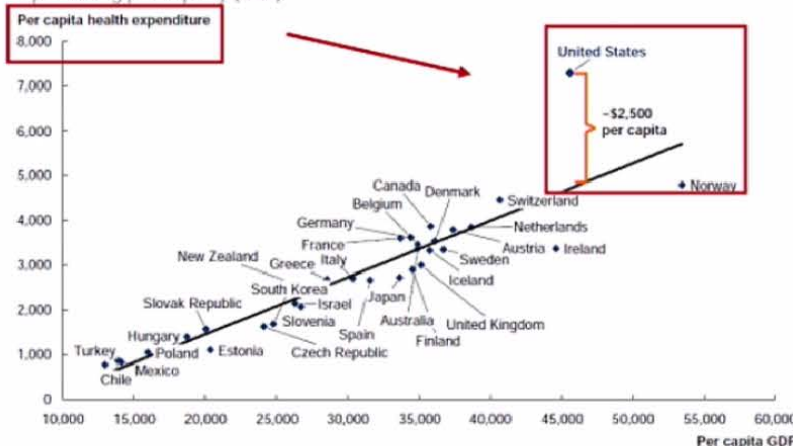
健康医疗数据确权后的流通憧憬



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

A comparison with OECD countries suggests that the total economic potential for efficiency improvements is about \$750 billion
Per capita health expenditure and per capita GDP, OECD countries, 2007
\$ purchasing power parity (PPP)



Kinsey & Company

Healthcare Practice

2024 healthcare services outlook: Challenges and opportunities

The healthcare services sector has seen rapid growth over recent years. Despite the challenges organizations may face in 2024, opportunities for the sector still abound.

麦肯锡，2024 年医疗保健服务展望：挑战与机遇。

健康医疗数据要素具有巨大的经济价值，那健康医疗数据要素应该如何流通呢？

数据流通 - IDV服务

- 作为全球最大的交易所集团之一，ICE下属数据业务部门ICE Data Services携手亚马逊推出 ICE DataVault (IDV) 服务，驱使数据在全球流通开来。

PB级一次性历史数据传输

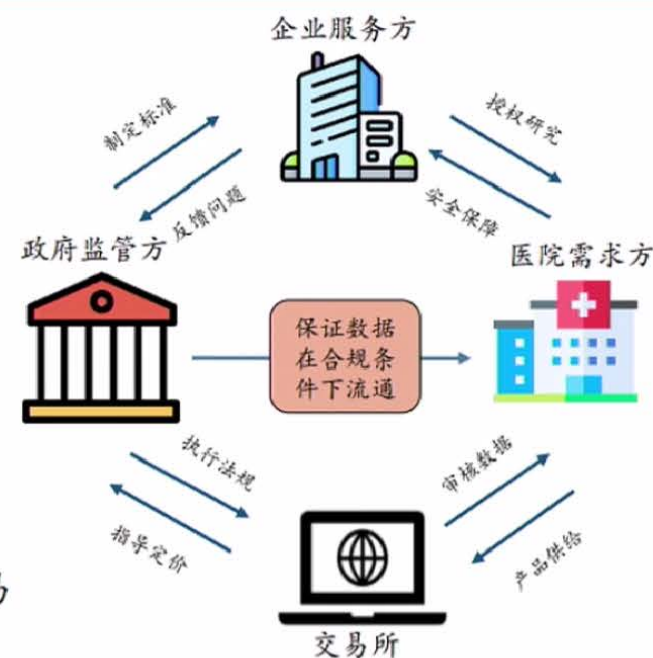


全球用户都可以通过连接访问数据，享受来自不同交易所报价的数据服务

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

健康医疗数据要素流通主线



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

- 健康医疗数据要素只允许场内交易
(有标准的合约并被监管)
- 不允许任何场外的健康医疗数据要素交易
(不允许任何私下协商成交)

健康医疗数据流通模式

健康医疗数据流通模式

模式一 数据开放

- 提供方无偿提供数据
- 需求方免费获取数据
- 无一般等价物参与

定价难

❑ 数据价值难评估，
定价规则不统一

模式二 数据共享

- 同为供给方和需求方
- 相互提供数据
- 无一般等价物参与

互信难

❑ 缺乏统一数据可信
流通基础环境

模式三 数据交易

- 提供方有偿提供数据
- 需求方支付获取费用
- 货币作为媒介

控制难

❑ 规范化流通市场框
架不清晰

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

传统定价方法



数据来源复杂且可
以零成本复制，**难**
以评估成本



数据应用场景复杂
多样导致价格需求
曲线拟合困难，**难**
以评估收益



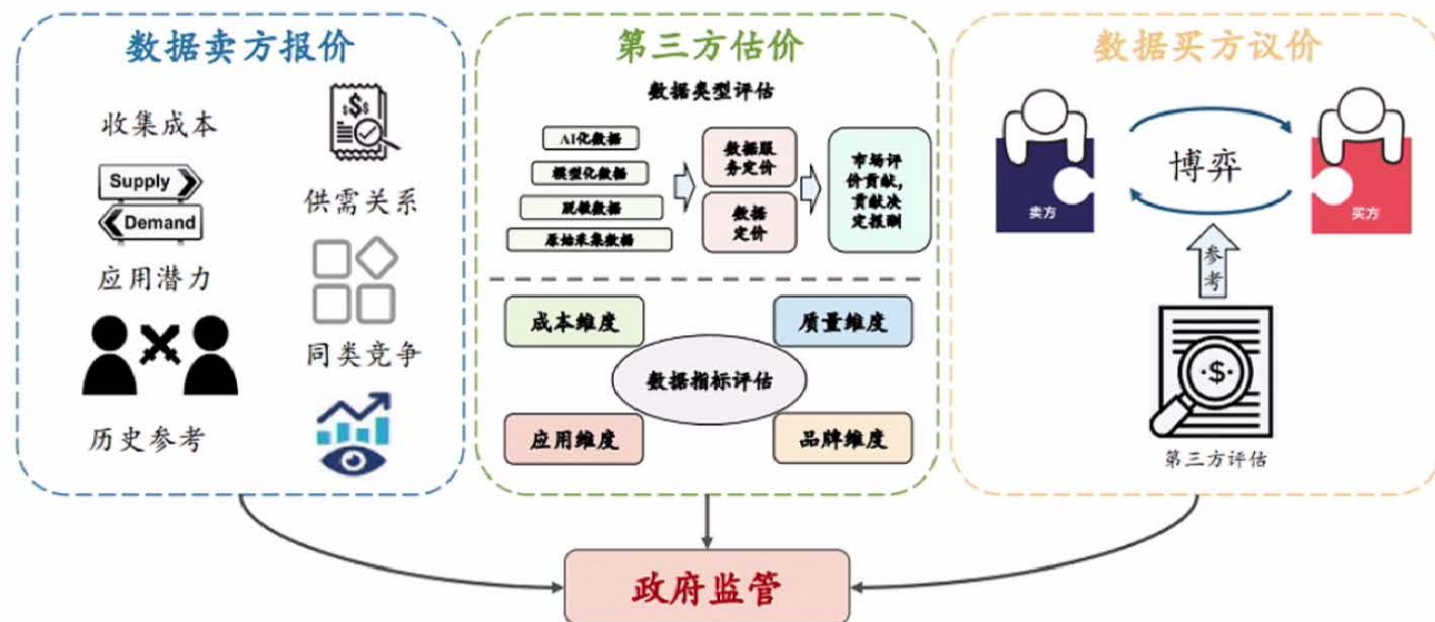
无相同数据产品的
成熟定价，**难以提**
供参考

传统定价方法无法满足健康医疗数据市场的定价需求

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

健康医疗数据要素定价流通



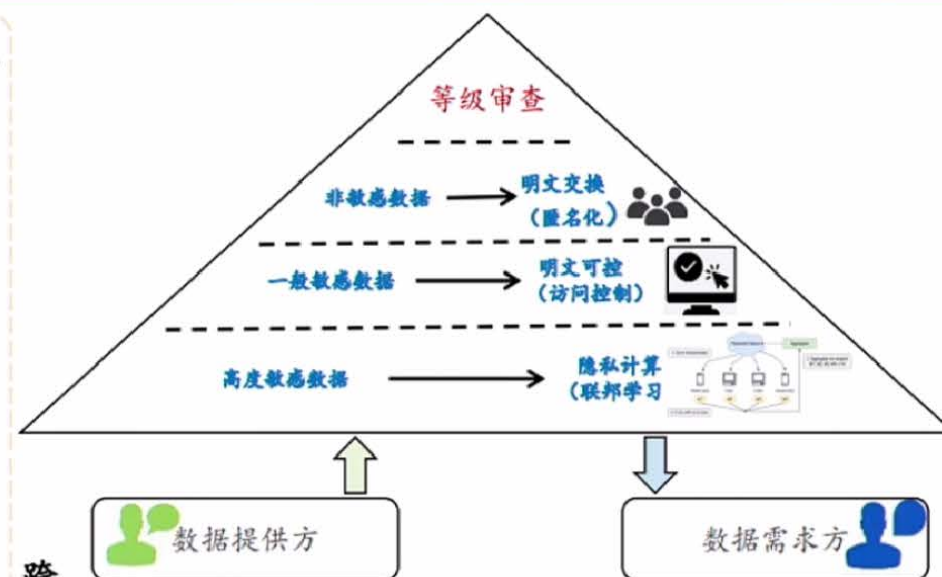
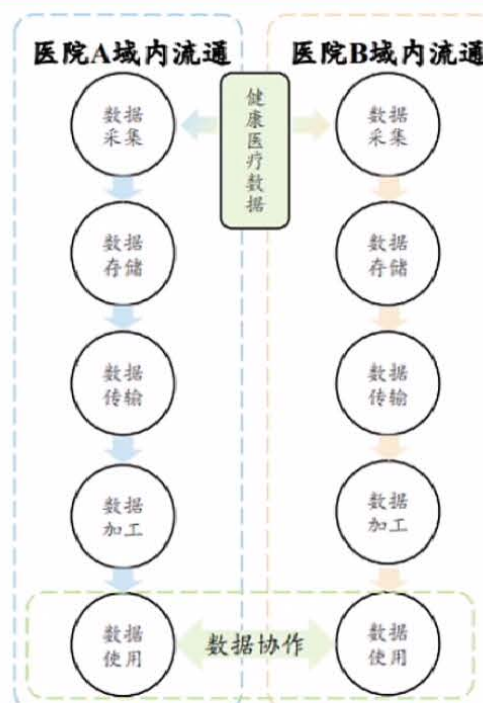
高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

有效解决健康医疗数据价值难以度量、价格共识难以达成的问题

Zhang M, Beltrán F, Liu J. A survey of data pricing for data marketplaces[J]. IEEE Transactions on Big Data, 2023.

健康医疗数据要素流通保护



有效解决健康医疗数据价值
安全保障、隐私泄露的问题

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

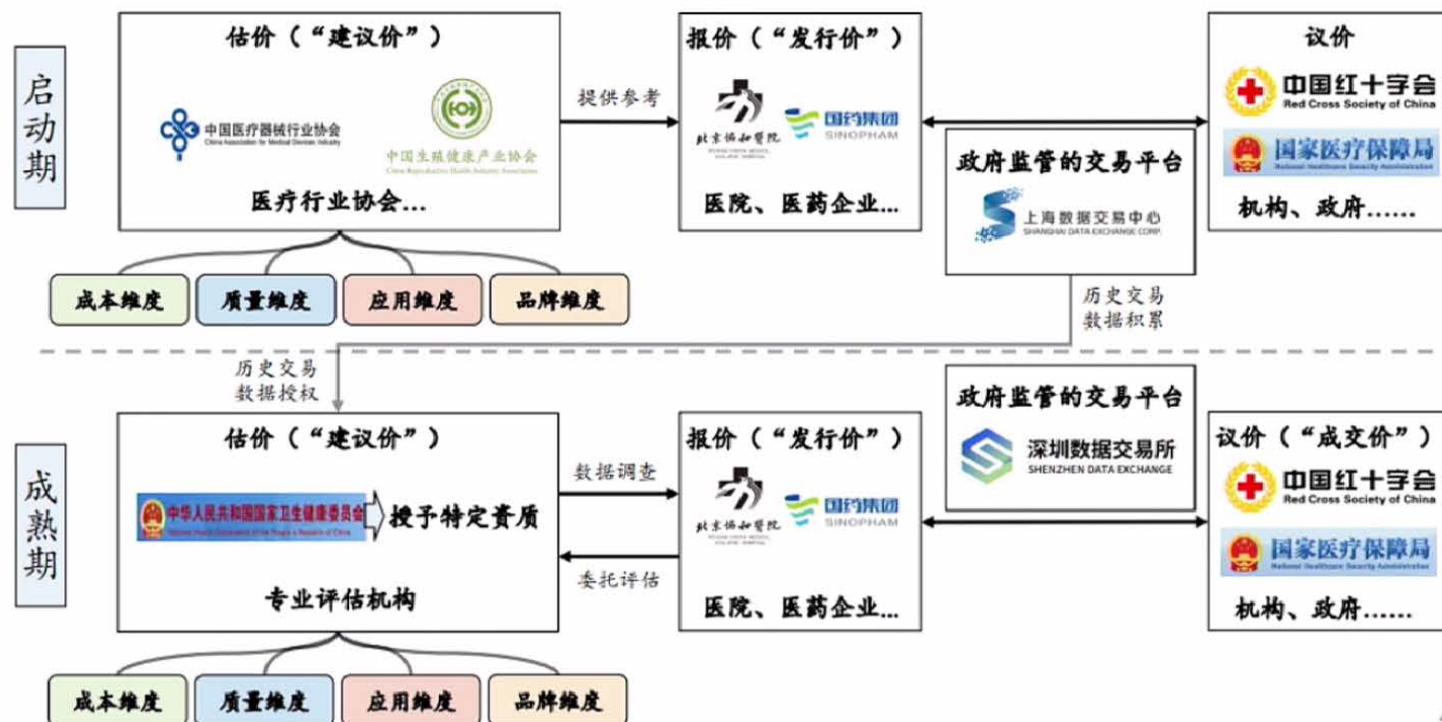
第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558

健康医疗数据要素流通体系



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

第二届中新数字经济与人工智能高峰论坛 THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558



健康医疗大模型

国外医疗语言大模型

stanford-crfm/
BioMedLM

<https://github.com/stanford-crfm/BioMedLM>



uf-hobi-informatics-lab/
GatorTron

all scripts used in gatortron project

<https://github.com/uf-hobi-informatics-lab/GatorTron>



microsoft/BioGPT

<https://github.com/microsoft/BioGPT>



KentOn-Li/
ChatDoctor

<https://github.com/KentOn-Li/ChatDoctor>



中文医疗语言大模型

SCIR-HI/Huatuo-Llama-
Med-Chinese

Repo for Benfiao [original name: Huatuo (华佗)],
Instruction-tuning Large Language Models with
Chinese Medical Knowledge. 本草 (原名: 华佗) 模
型仓库。基于中文医学知识的大语言模型指令微调

<https://github.com/SCIR-HI/Huatuo-Llama-Med-Chinese>



scutecyr/BianQue

中文医疗对话模型(黄晓/BianQue)

<https://github.com/scutecyr/BianQue>



CMKRG/
QiZhenGPT

QZhenGPT: An Open Source Chinese Medical
Large Language Model | 一个开源的中文医疗大语言
模型

<https://github.com/CMKRG/QiZhenGPT>



michael-wzhu/
ChatMed

ChatMed: 中文医疗大模型。善于回答患者/用户
的日常医疗相关问题！

<https://github.com/michael-wzhu/ChatMed>



产业界医疗语言大模型



华为云AI辅助药物设计服务
基于盘古药物分子大模型训练

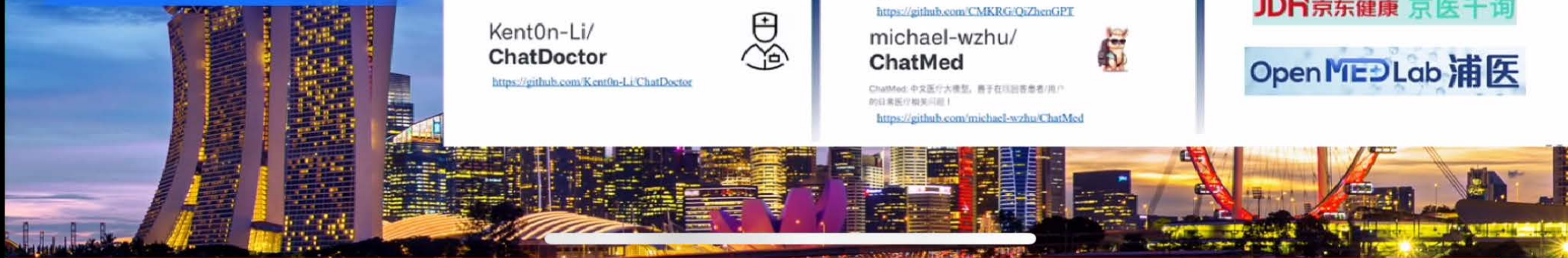


JDH 京东健康 京医千询

Open MED Lab 浦医

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



健康医疗大模型定价

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

泛化性



对新数据的适应能力

模型在从未见过的数据上能表现出良好的性能能力

通用性



解决多个任务的能力

模型能应用于不同的数据集或任务

实用性



应用时的可用性和效率

模型能以合理的时间和资源，快速处理数据并做出决策

人工智能驱动定价

价格数据



定价模型

提供参考

健康医疗数据



健康医疗大模型



下游任务

海量数据-精准高效预测市场需求

先进算法-准确计算分析数据成本

灵活接口-高效监控调整定价策略

围绕大模型的内外特性进行不同策略定价

第二届中新数字经济与人工智能高峰论坛
THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL
ECONOMY AND ARTIFICIAL INTELLIGENCE



中国人工智能学会
Chinese Association for Artificial Intelligence

百度直播 | ID 9178918558

总结与思考：健康医疗数据的流通时代发展



高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长



总结与思考

健康医疗数据更加**需要有效的确权机制**，去有效赋予/区分个体及机构对健康医疗数据的处置权益。



机构需要给予个体合适的**虚拟回报**，以置换个体脱敏后的健康医疗数据的处置权。



搭建健康医疗数据**多科室、多病症状的综合性平台**，采集更多有价值健康医疗数据，发展数据加工后的AI大小模型。

高 阳

CAAI 智能服务专委会副主任
南京大学智能科学与技术学院
书记兼执行院长

