

全球架构师峰会·深圳站

ArchSummit 10th Anniversary

主办方: **InfoQ**

为一线互联网公司核心技术 人员提供优质内容

☑ 每日要闻

☑ 技术干货

☑ 大咖访谈

☑ 行业趋势



关注 InfoQ 官方微博

InfoQ
LIVE

促进软件开发及相关领域知识与创新的传播

InfoQ
LIVE

ArchSummit
LIVE



全球架构师峰会

InfoQ
LIVE

大模型认知时代的新架构

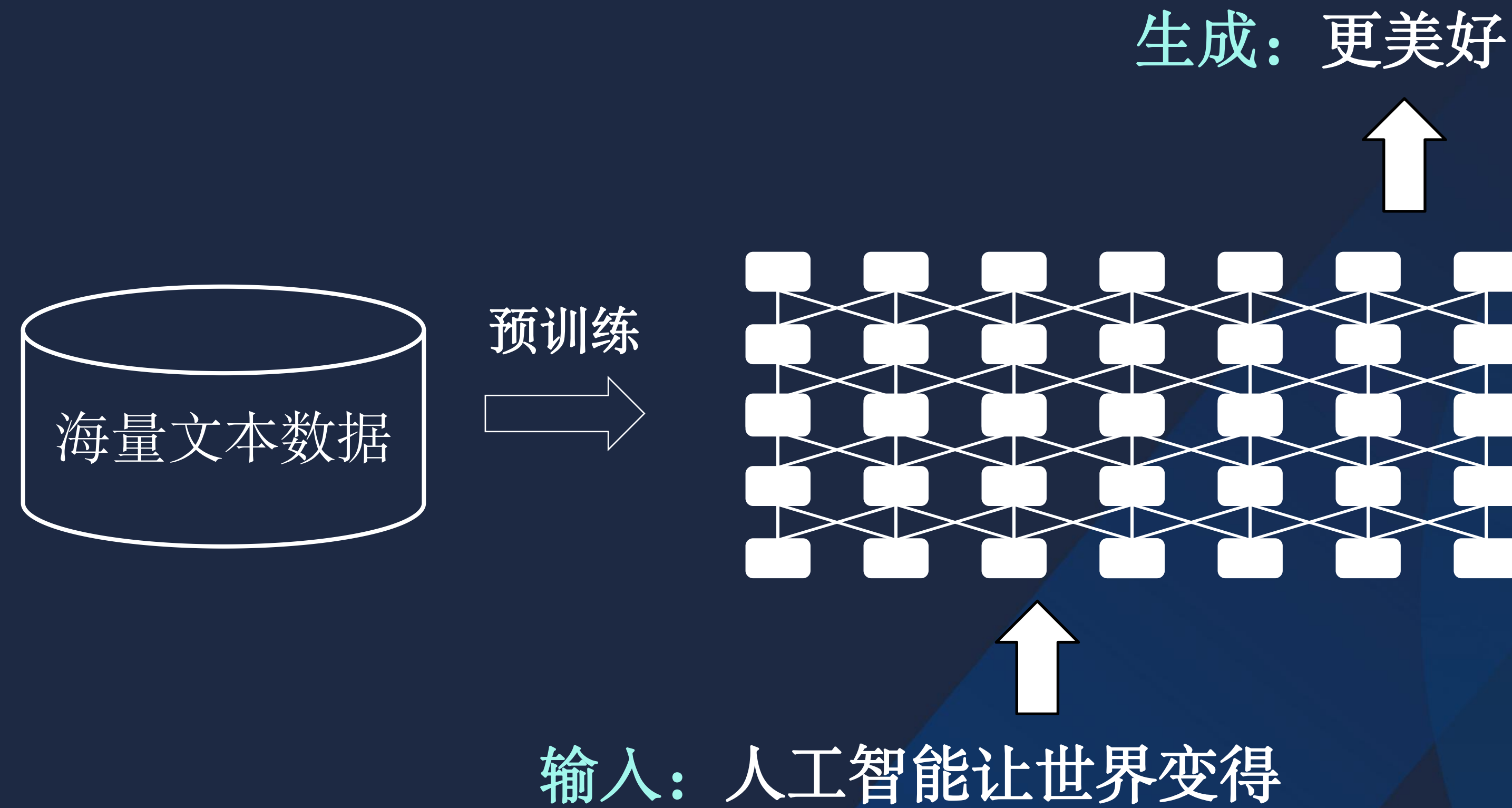
粤港澳大湾区数字经济研究院
(IDEA Research)

张家兴
讲席科学家

认知智能范式演变



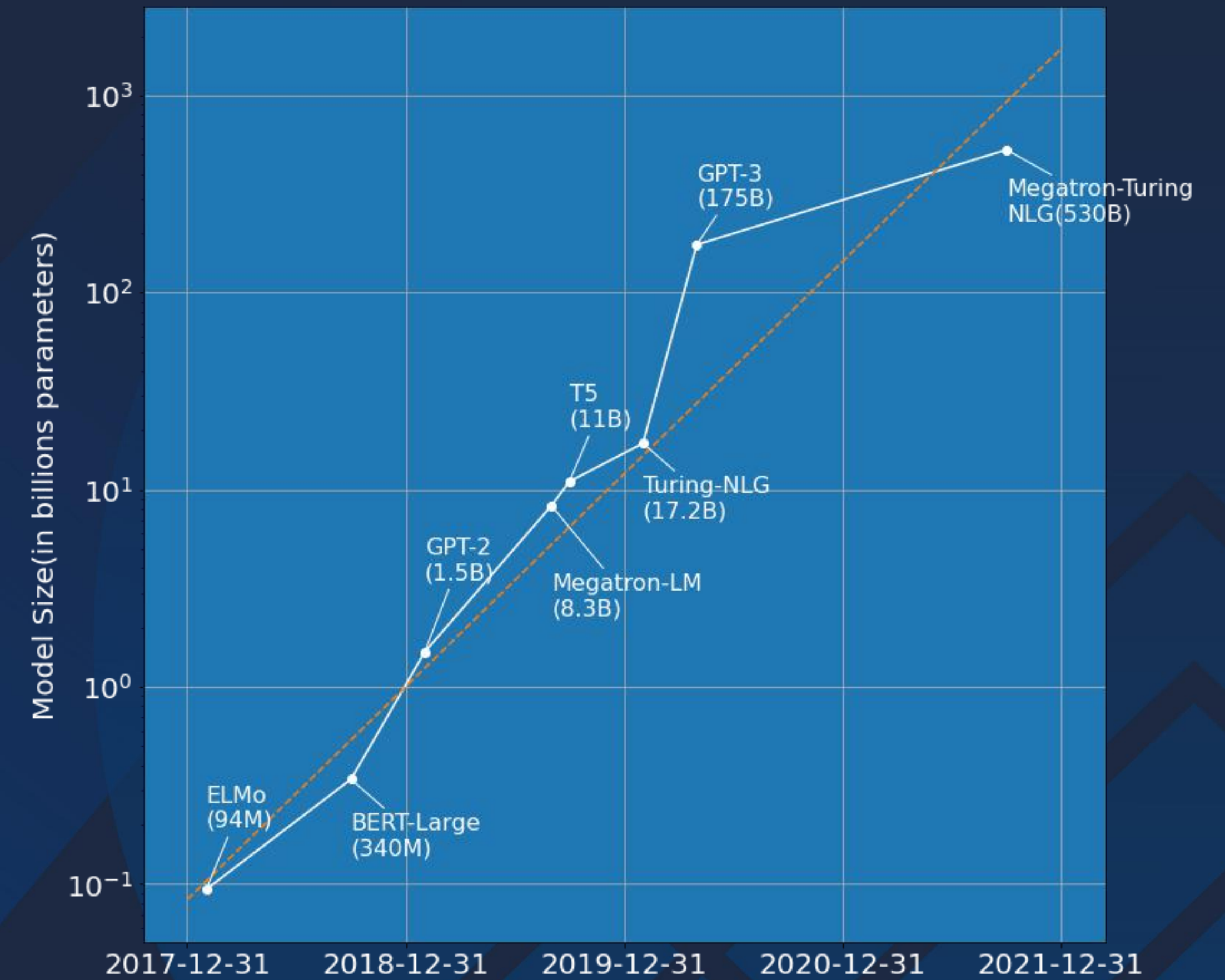
预训练大模型



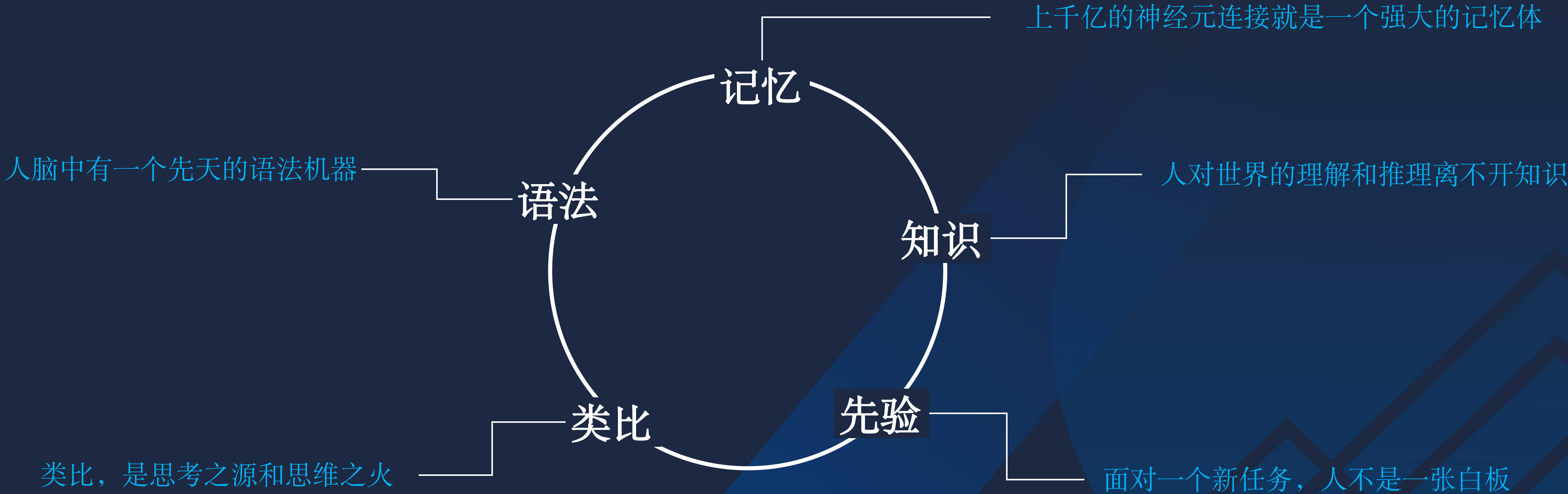
大模型时代



大模型的参数规模每年增加 **10** 倍

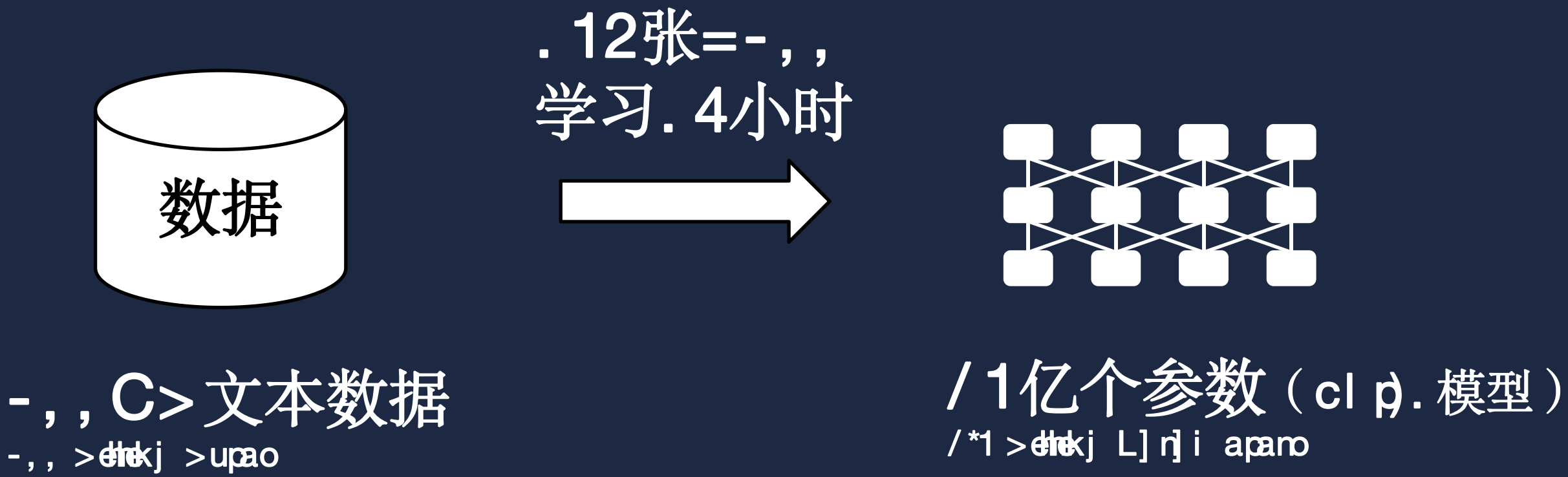


大模型给机器认知智能带来了什么



大模型给AI架构带来了新的思考

记忆：模型即数据



原数据

利率:
1年月费率0.78%, 年费率9.38%;
2年月费率0.77%, 年费率9.26%;
3年月费率0.78%, 年费率9.38%;
4年月费率0.80%, 年费率9.55%;

给定起始文本

利率:
1年月费率0.78%, 年费率9.38%;
2年月费率0.77%, 年费率9.26%;
3年月费率0.78%, 年费率9.38%;
4年月费率0.80%, 年费率9.55%;
5年月费率0.82%, 年费率9.61%;
6年月费率0.83%, 年费率9.72%;

模型生成文本

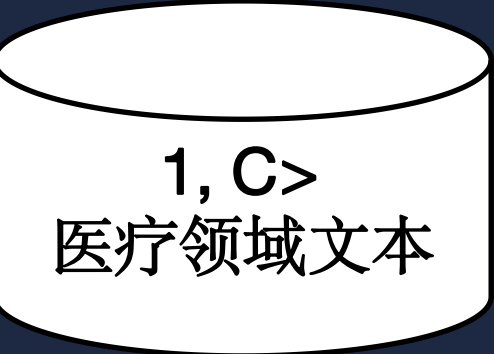
什么是数据

可查询可枚举的结构就是数据

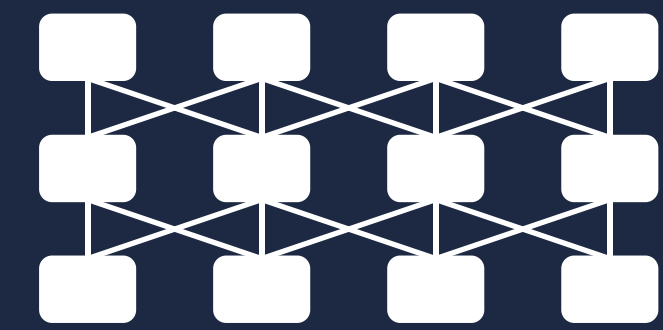
I k` ah] o @] p

存算一体 模糊 检索高效
E l naooekj Opkra)) 印象存储

知识：模型即知识



12张=, ,
学习3天



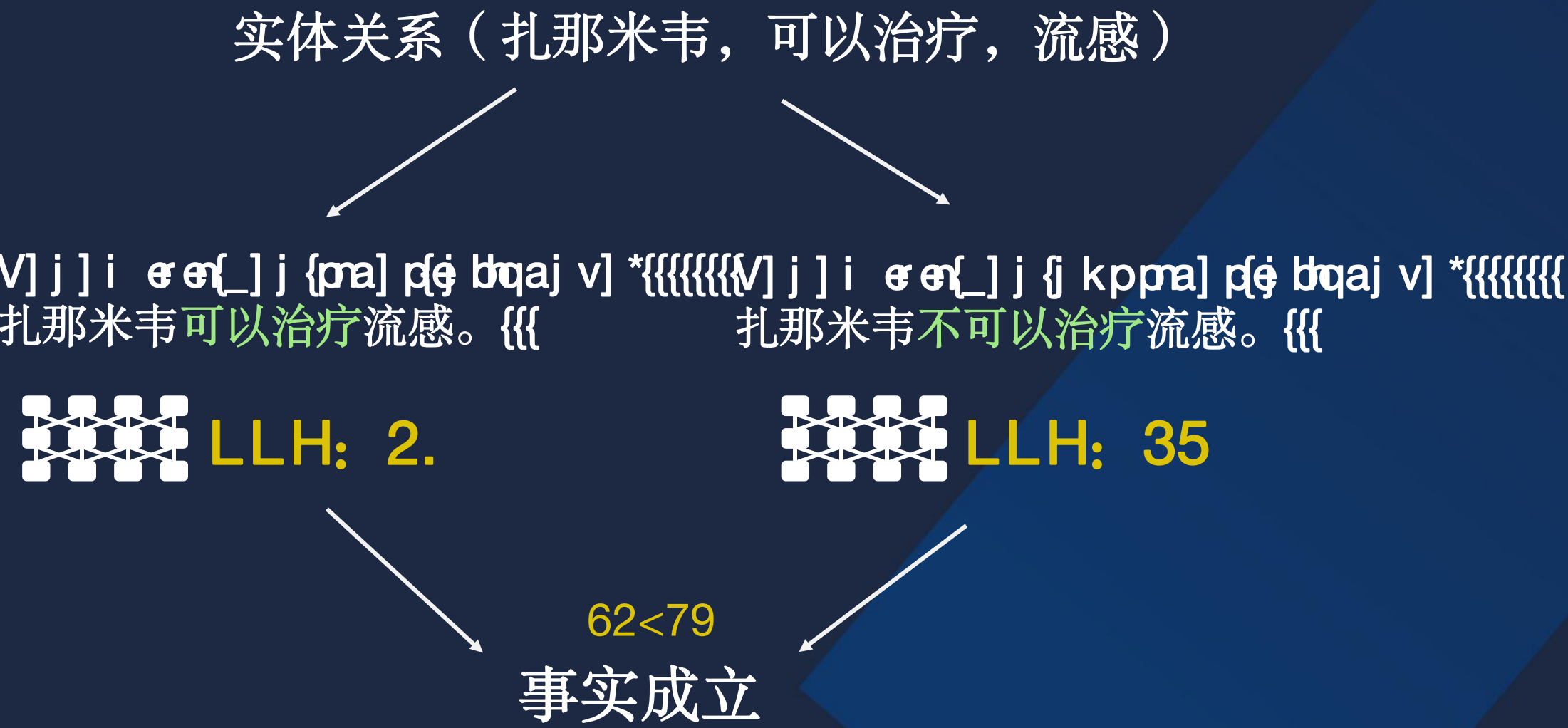
/1亿参数 CLP). 模型

问答

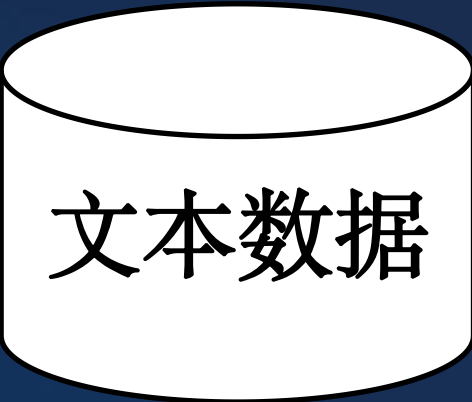
实体关系判定

输入：我得了流感，症状是
输出：发烧、咳嗽、头痛、肌痛、喉咙痛、鼻塞和头痛。

输入：流感的治疗方法是
输出：抗病毒药物扎那米韦、奥司他韦和拉那米韦。



Gj ks læ` ca] o Pat p



算法' 人力

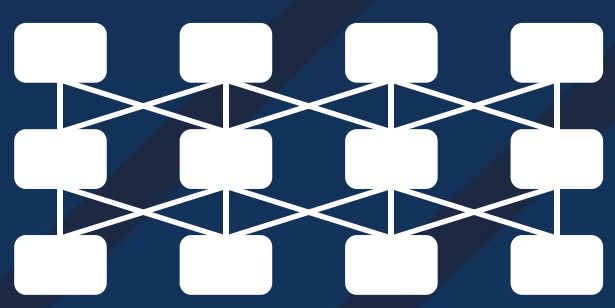
Gj ks læ` ca] o Cr] l d



知识图谱

算力

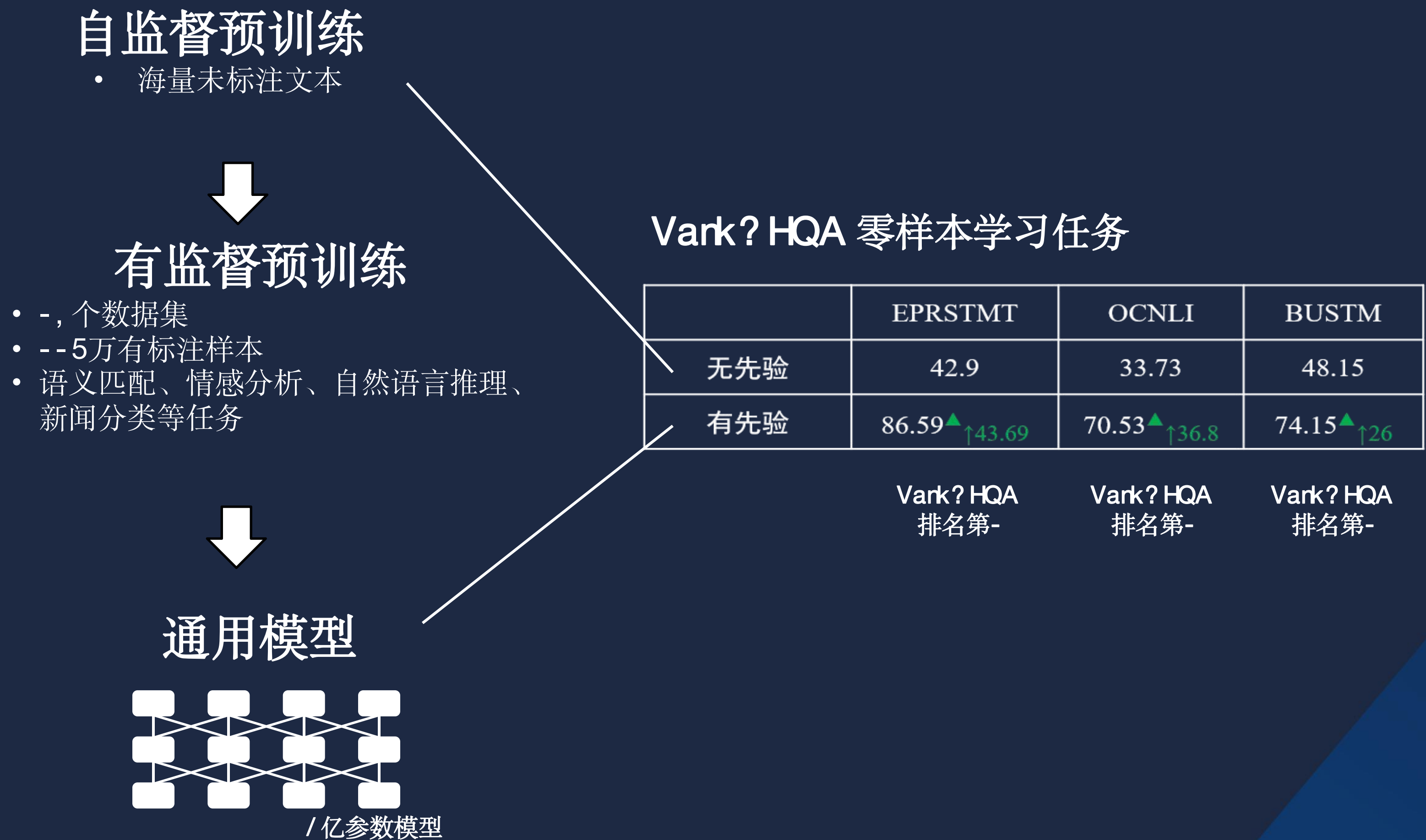
Gj ks læ` ca] o l k` ah



大模型

存储' 表示' 查询' 推理

先验：用计算追上几百万年的差距



为什么要先天的计算

后天任务学习样本更少
后天任务学习速度更快
减少整体计算量

IDEA研究院13亿参数预训练大模型“二郎神”

CLUE社区														
FewCLUE小样本学习榜 Github地址 提交样例 测评规则 小样本榜-提交多份 , 新														
模型描述需包含关键词“FewCLUE”; 提交需实名, 即: 队伍名称、模型名称、Url/Github、模型描述, 需有真实有效。无意义的提交将被移除。														
2021-05-24: 更新了iflytek的测试集(test.json, 长度2600), 请重新拉取一下, 并在这个测试集上做预测; 有问题发邮件: CLUEbenchmark@163.com														
排行	模型	研究机构	测评时间	Score	认证	EPRSTMT	CSLDCP	TNEWSF	IFLYTEKF	OCNLIF	BUSTM	CHIDF	CSLF	CLUEWSCF
1	Human	CLUE	21-05-07	83.934	已认证	90.0	68.0	71	66.0	90.3	88.0	87.1	84.0	98.0
2	二郎神	IDEA研究院	21-11-10	73.816	待认证	87.65	63.89	75.13	48.5	76.21	74.3	89.0	74	76.9
3	RobustPrompt	微信 AI	21-11-02	72.322	待认证	88.45	59.39	73.87	47.58	72.02	76.7	80.55	76.57	77.24
4	GTS	idea_ccnl	21-10-28	71.589	待认证	88.84	62.92	74.33	49.27	68.98	75.9	86.85	68.5	71.38
5	源1.0	浪潮人工智能研究院	21-09-28	71.451	待认证	88.45	62.75	68.33	47.31	45.12	72.05	87.4	80.53	87.93
6	large_0	微信 AI	21-10-27	70.934	待认证	88.45	59.39	73.87	47.58	72.02	76.7	69.45	76.57	77.24

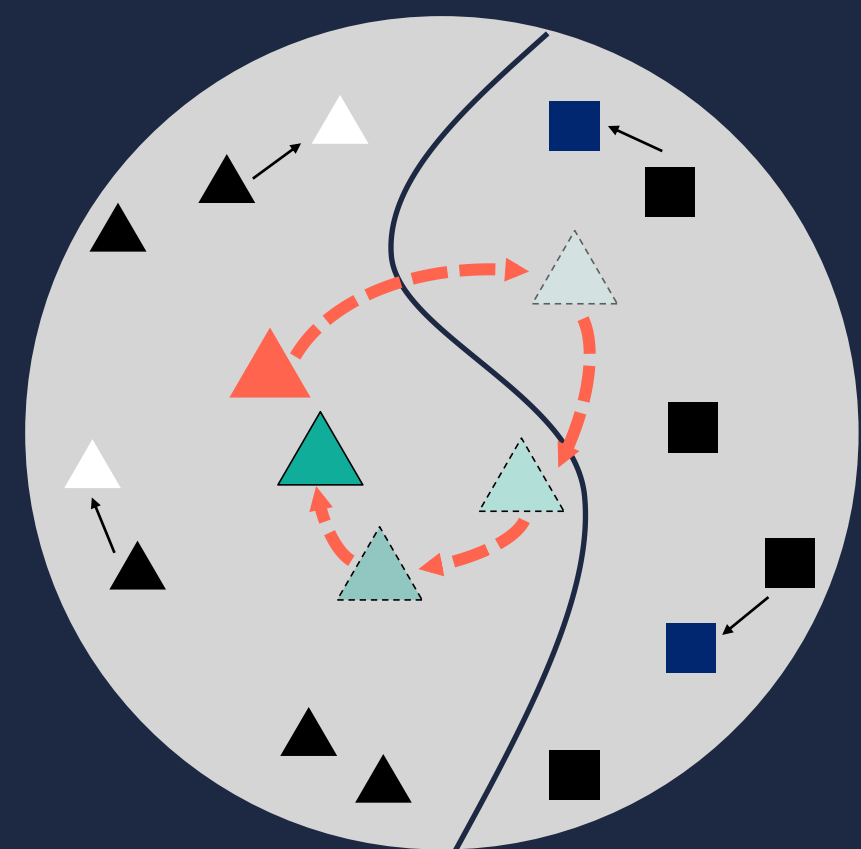
中文小样本学习Bas ? HQA榜单第一名

两项任务超越人类

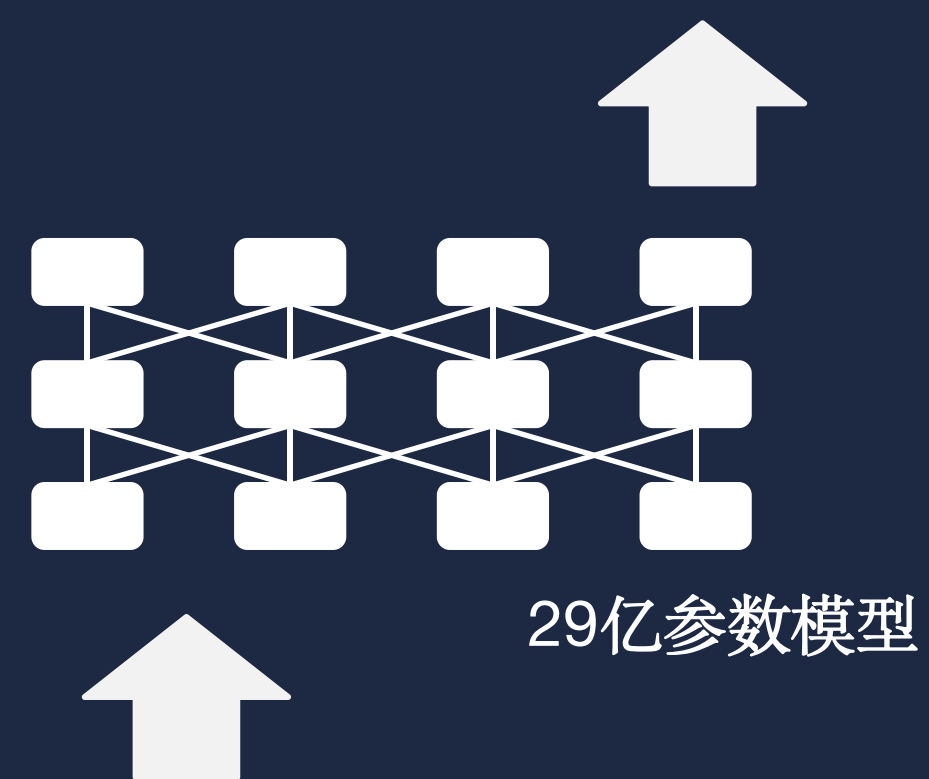
三项任务刷新纪录

类比：突破数据稀缺瓶颈

句子空间的蒙特卡洛树搜索

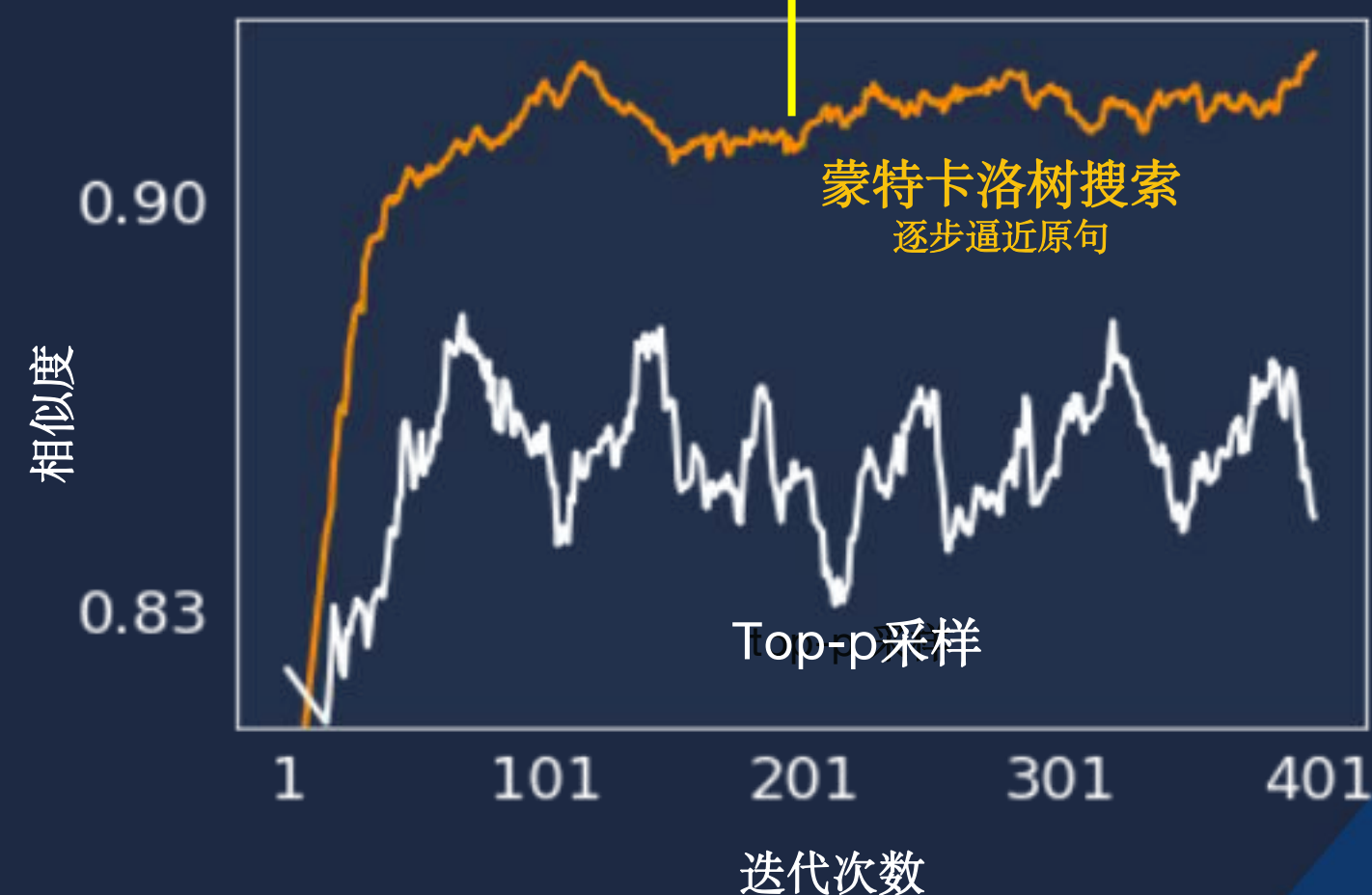


- 1. 大城市这个问题需要引起重视
- 51. 小升初,去农村还是县城?
- 101. 择校,是市区农村学校好还是县乡村小学的好?
- 301. 中国年轻老师选择去农村小学, 还是县城?

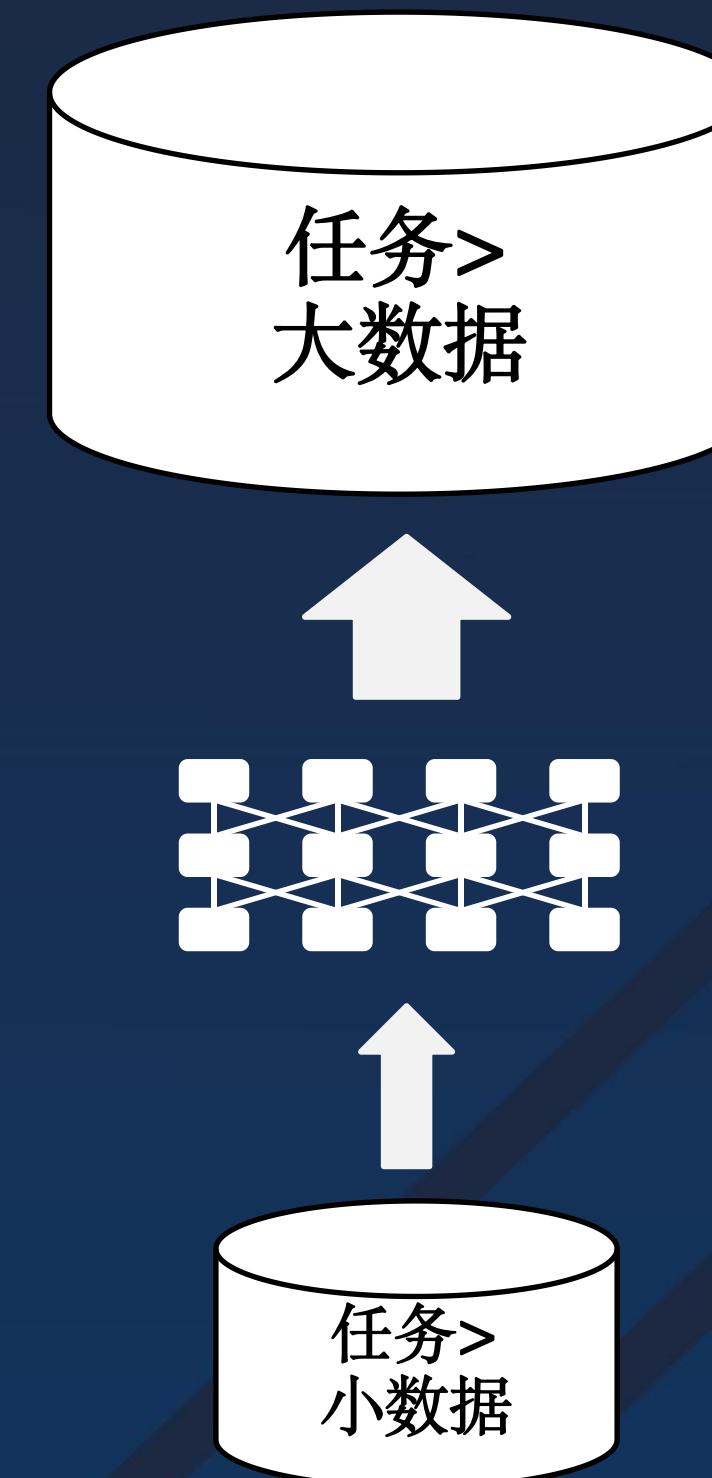
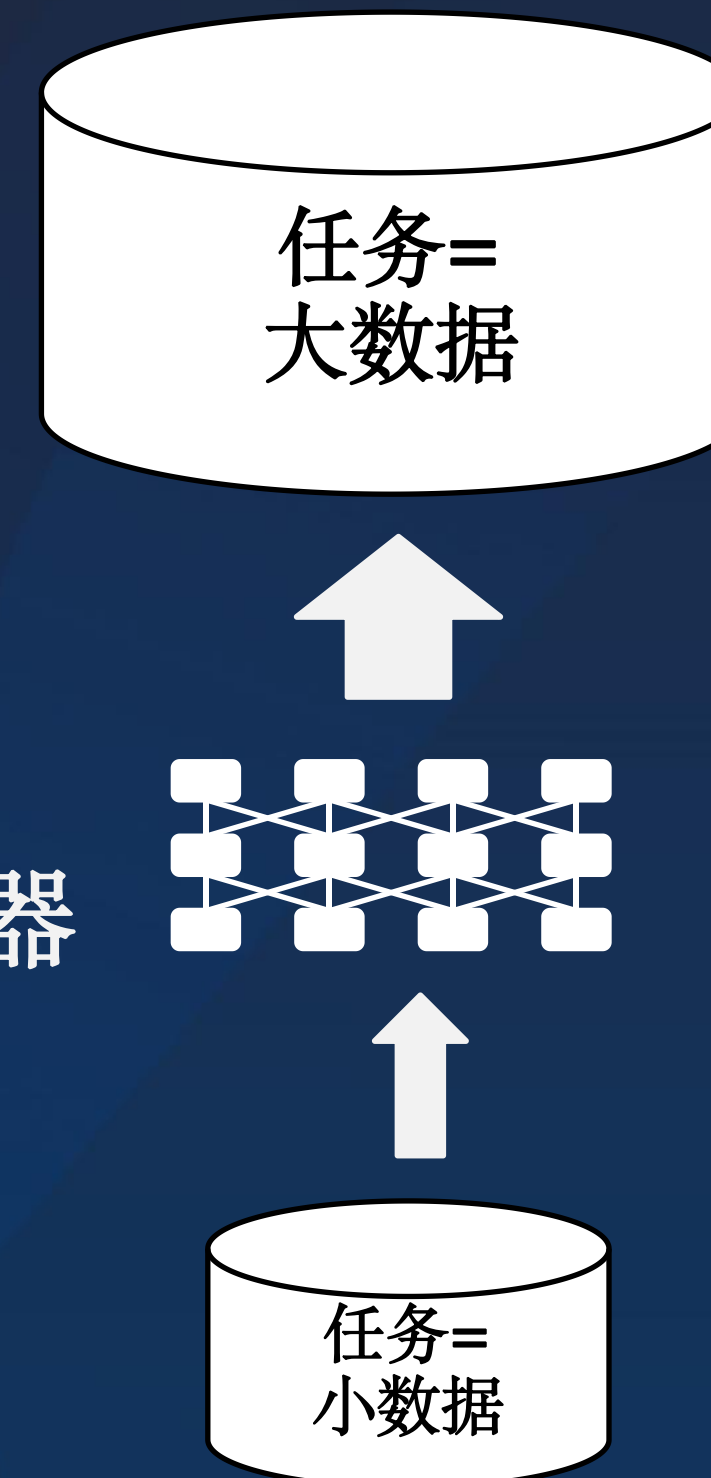


29亿参数模型

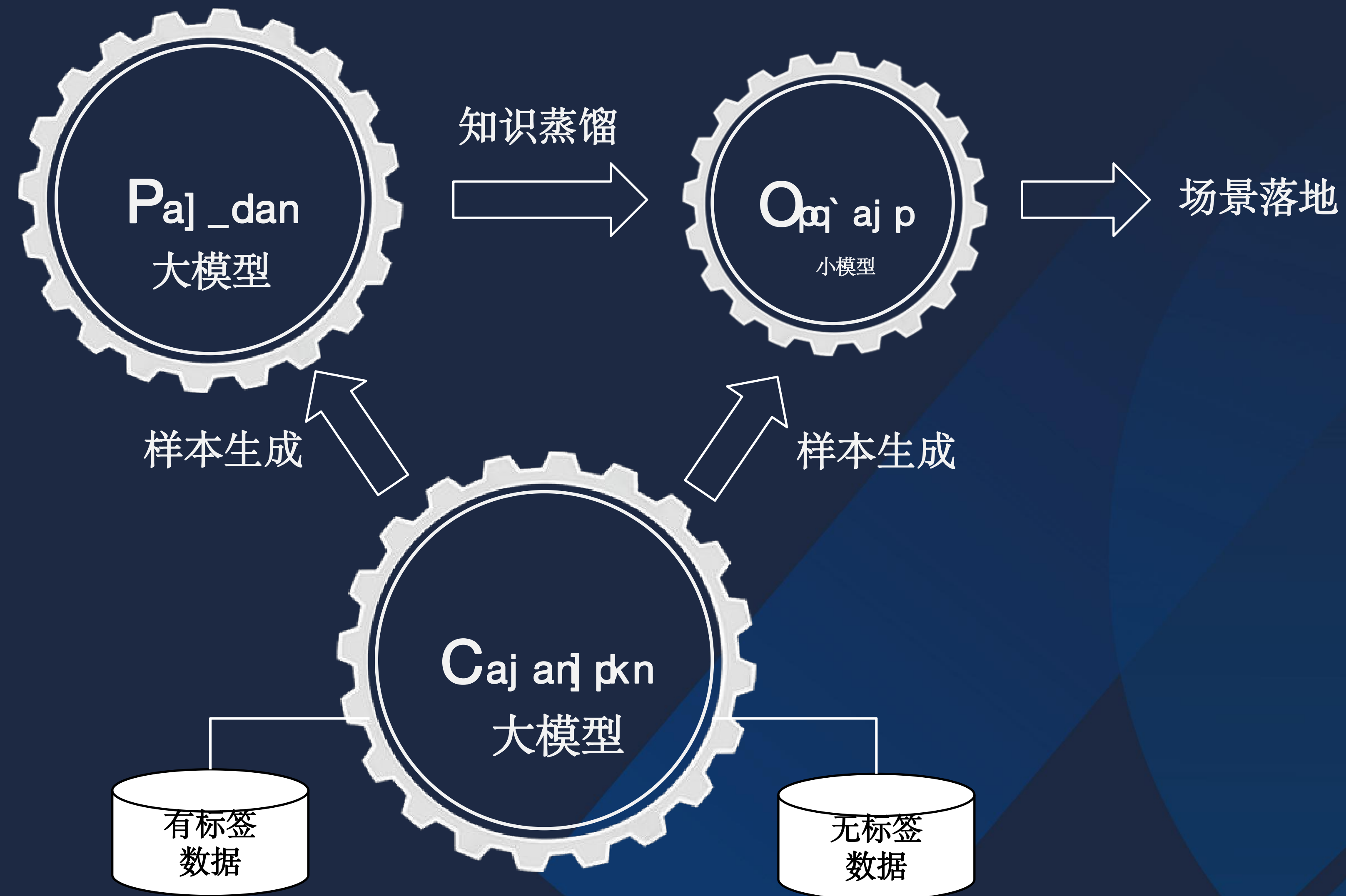
“年轻教师选择农村学校，还是县城学校？”的相似句式



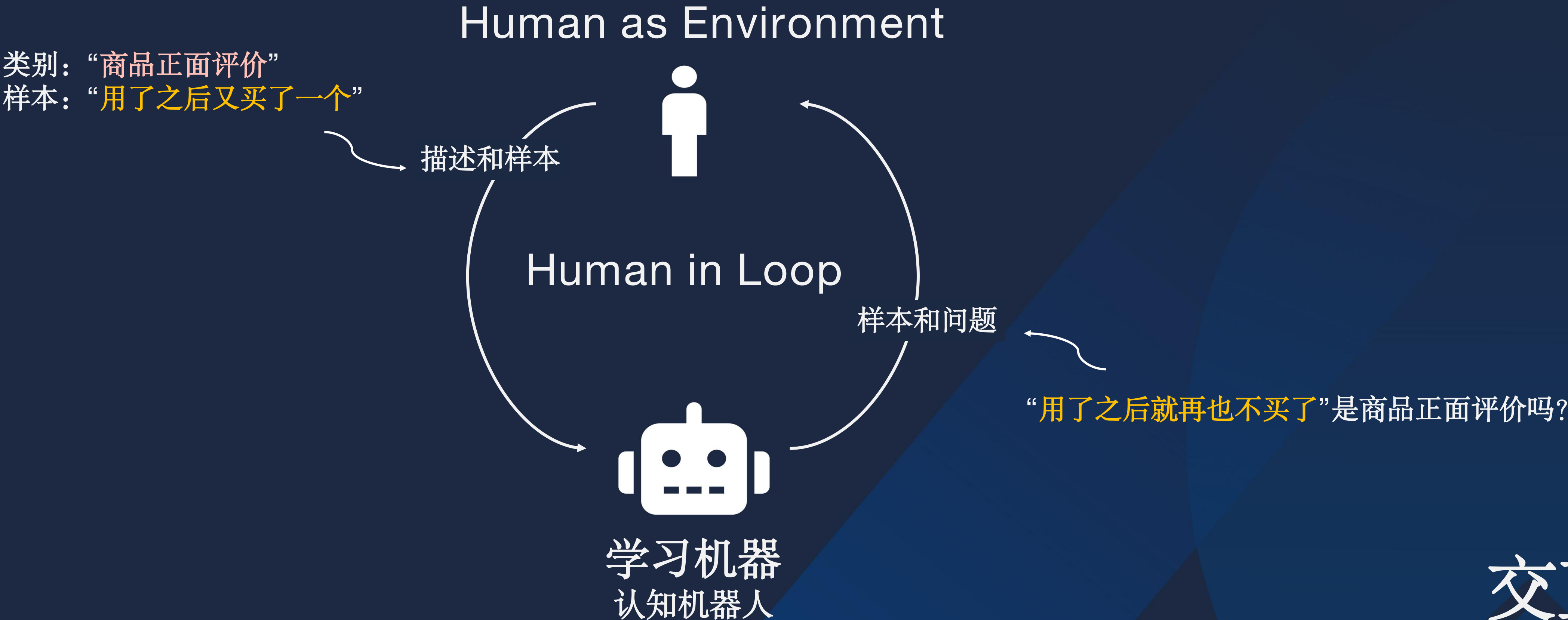
大模型
数据放大器



GTS: 大模型时代的机器学习新框架



从机器学习到学习机器

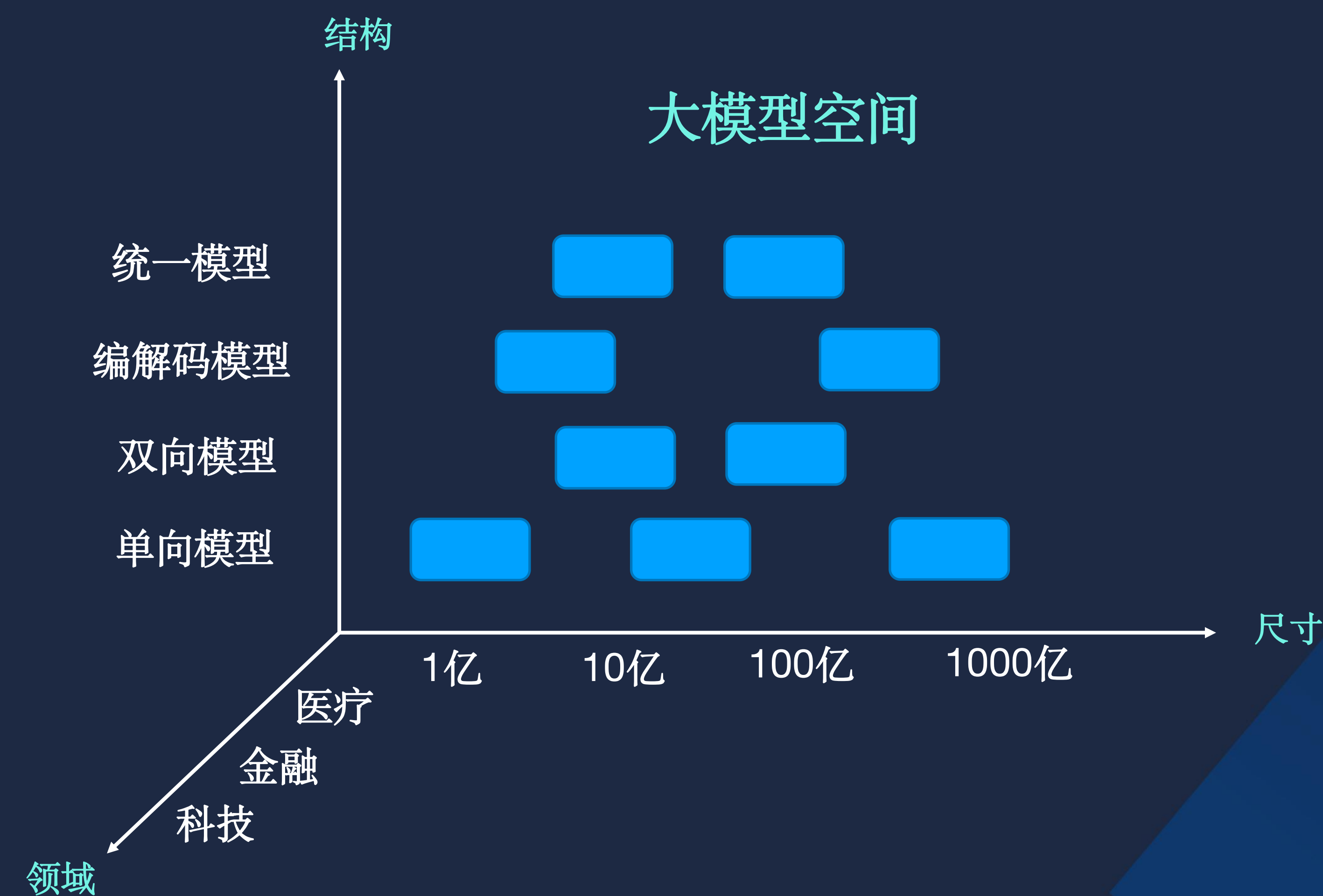


交互式学习
让机器像人一样去学习

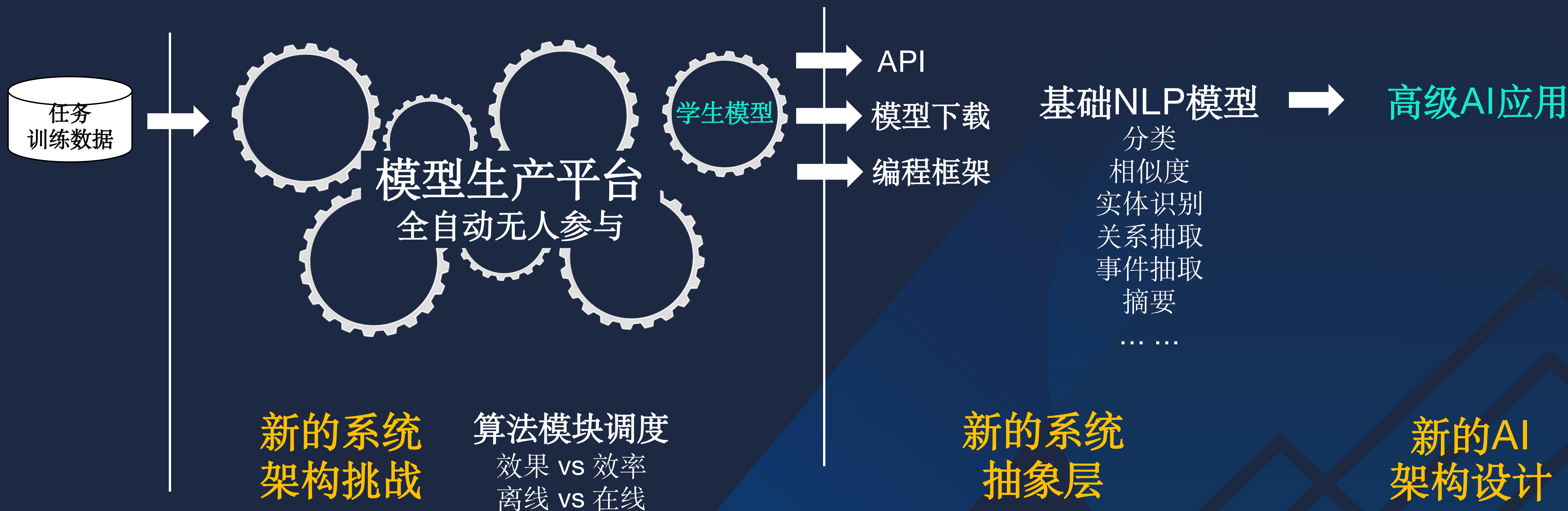
大模型落地

本质性瓶颈
算力 人才

模型开源：合作共赢



自动化模型生产：取代算法工程师



AI应用系统与算法



IDEA研究院
让未来发生

QCon+ 案例研习社



扫码学习大厂案例

学习前沿案例，向行业领先迈进

40个

热门专题

—
行业专家把关内容筹备，
助你快速掌握最新技术发展趋势

200个

实战案例

—
了解大厂前沿实战案例，
为 200 个真问题找到最优解

40场

直播答疑

—
40 位技术大咖，每周分享最新
技术认知，互动答疑

365天

持续学习

—
视频结合配套 PPT
畅学 365 天



THANKS

全球架构师峰会·深圳站

ArchSummit 10th Anniversary

主办方: **InfoQ**
ucue

