

51CTO WOT

World Of Tech 2024

WOT全球技术 创新大会

智启新纪
慧创万物



AI大模型时代的业务架构 智能化转型

蔡林51Talk技术副总裁

“

51Talk

51Talk，中国在线英语教育的领导品牌，创立于2011年，从创立之初，便将“让每个人都有能力对话世界”作为自己的使命，于2016年6月10日成功登陆美国纽交所，股票代码COE。成为2016年中国赴美上市第一股、中国在线教育赴美上市第一股、全球首家上市的B2C在线教育企业；同时我们也是第一家深耕菲律宾外教的在线英语教育机构。

”

“



蔡林
51Talk技术副总裁

9年 ToB 经验，8年 ToC 经验；擅长 DDD 、业务架构设计，微服务设计；擅长 Java.PHP 技术栈；带领100人+技术团队，用3个月解决别人8个月解决不了的生死难题；用2周时间完成别人4个月项目上线；善于打造技术驱动业务、技术创新业务，用技术创新给公司带来数亿营收增长；目前在做 AI 及大模型在业务上的落地，兼任 AI研究院院长。

”

目录

- AI的边界：大模型时代的技术趋势和商业趋势
- AI的价值：All-in On AI的价值是什么
- 核心认知：愿景很美好，现实很残酷，能否落地，如何落地
- 案例分享：落地的案例
- 经验总结：落地经验总结

AI的边界

AI在技术层面和商业层面的边界

01



炸裂的GPT4O



Pretraining	<ul style="list-style-type: none"> • Compute cluster scaling • Hardware correctness 	<ul style="list-style-type: none"> • Data • Optimization & architecture 	<ul style="list-style-type: none"> • Distributed training infrastructure • Training run babysitting
Long context	<ul style="list-style-type: none"> • Long context research 	<ul style="list-style-type: none"> • Long context kernels 	
Vision	<ul style="list-style-type: none"> • Architecture research • Compute cluster scaling 	<ul style="list-style-type: none"> • Data • Training run babysitting 	<ul style="list-style-type: none"> • Distributed training infrastructure • Alignment Data • Hardware correctness • Deployment & post-training
RL & Alignment	<ul style="list-style-type: none"> • Dataset contributions • Refusals 	<ul style="list-style-type: none"> • Data infrastructure • Flagship training runs 	<ul style="list-style-type: none"> • ChatML format • Code capability • Model safety • Foundational RLHF and InstructGPT work
Evaluation & Analysis	<ul style="list-style-type: none"> • OpenAI Evals library • Capability evaluations • Instruction following and API evals • Overreliance analysis • System card & broader impacts analysis 	<ul style="list-style-type: none"> • Model-graded evaluation infrastructure • Coding evaluations • Novel capability discovery • Privacy and PII evaluations 	<ul style="list-style-type: none"> • Acceleration forecasting • Real-world use case evaluations • Vision evaluations • Safety and policy evaluations • Non-proliferation, international humanitarian law & national security red teaming • ChatGPT evaluations • Contamination investigations • Economic impact evaluation • OpenAI adversarial testers
Deployment	<ul style="list-style-type: none"> • Inference research • Trust & safety engineering • Reliability engineering 	<ul style="list-style-type: none"> • GPT-4 API & ChatML deployment • Trust & safety monitoring and response • Product management 	<ul style="list-style-type: none"> • GPT-4 web experience • Trust & safety policy • Inference infrastructure • Deployment compute

模型进化



看得懂



听得懂



有手



有脑

大模型涌现出物理世界的理解，更像人

大模型社会化

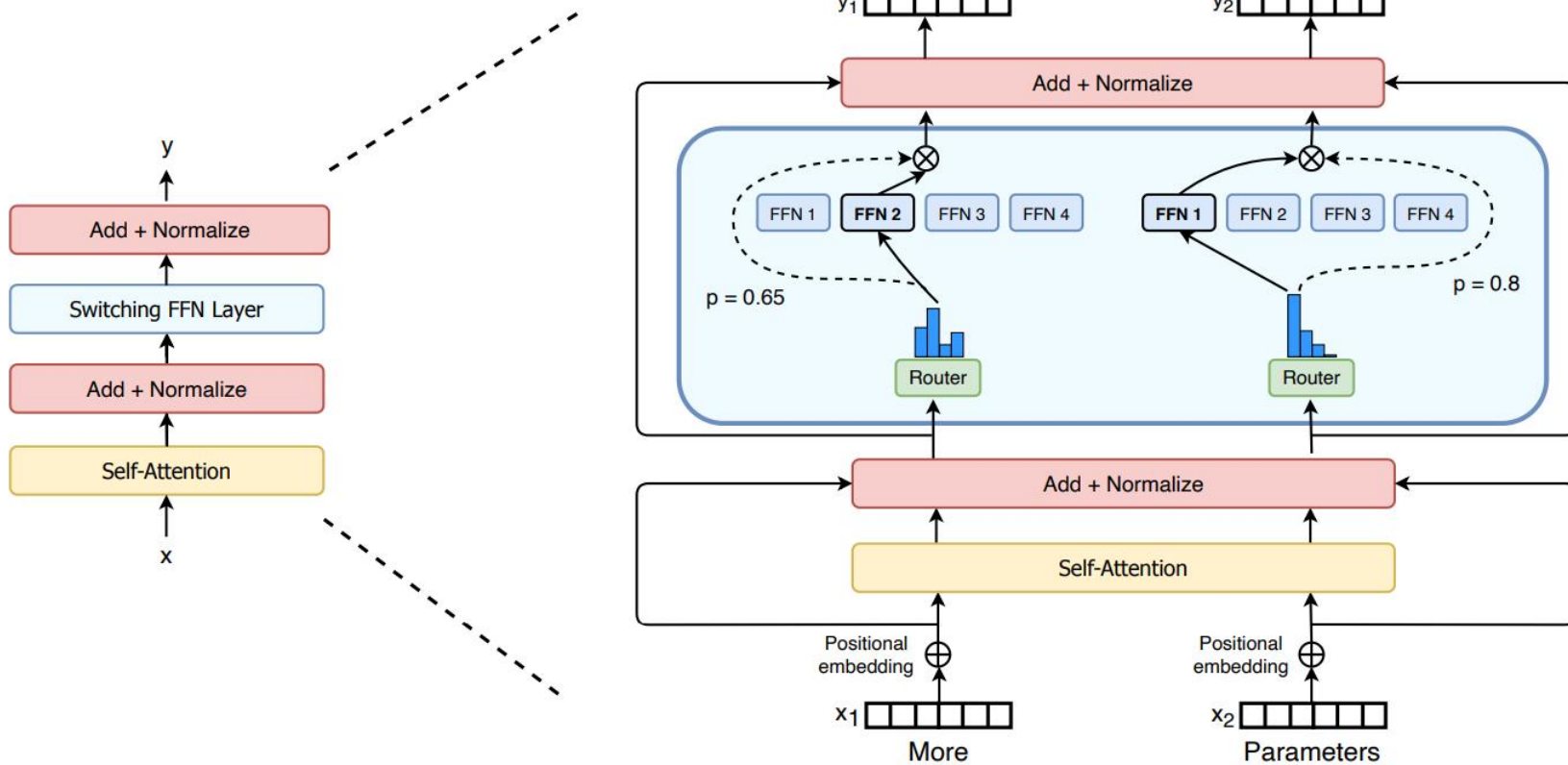
广泛应用和普及

从恐慌到接受

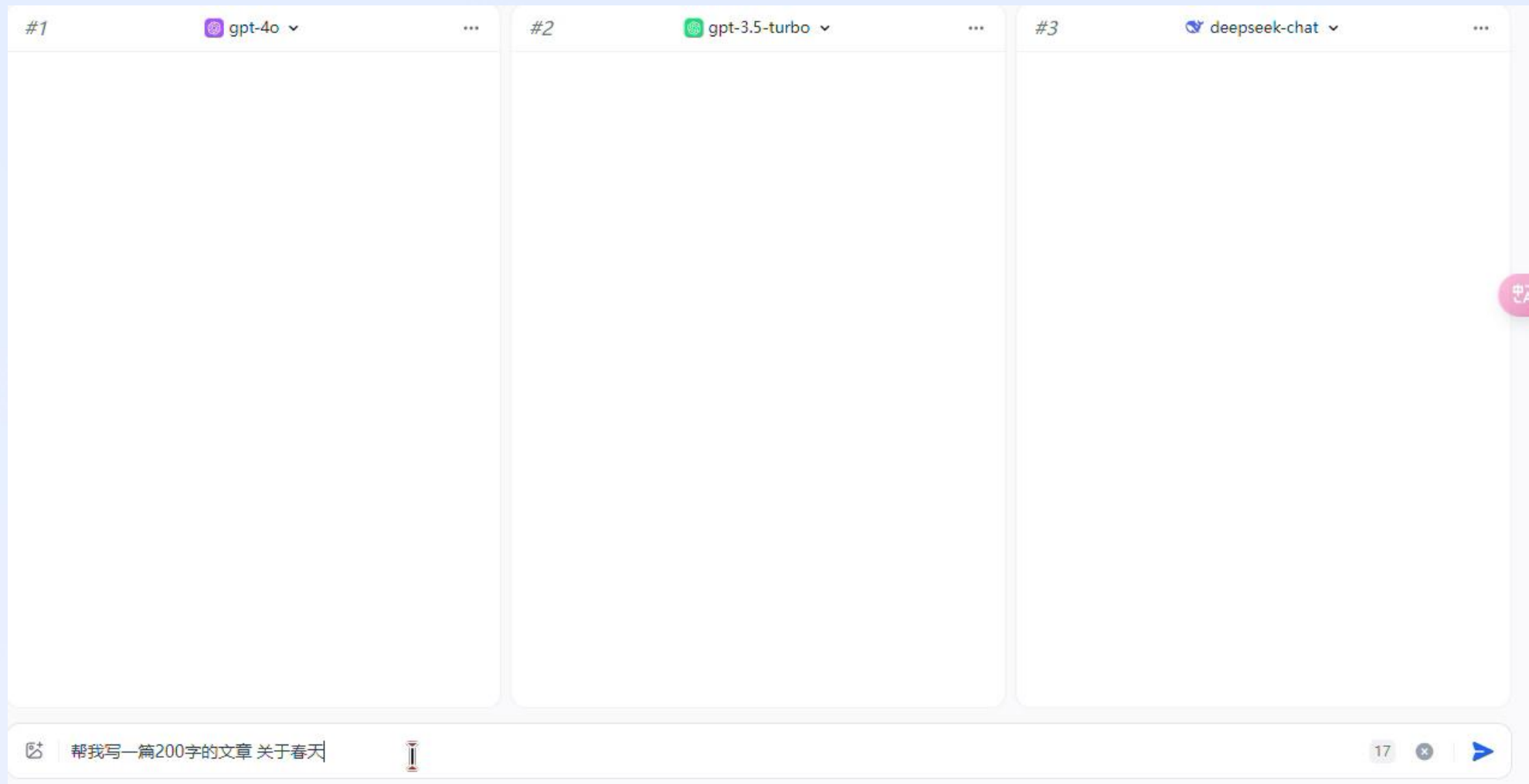
算法公平

文化和社会变革

未来的模型会变小，大模型会拆成小模型，MOE架构

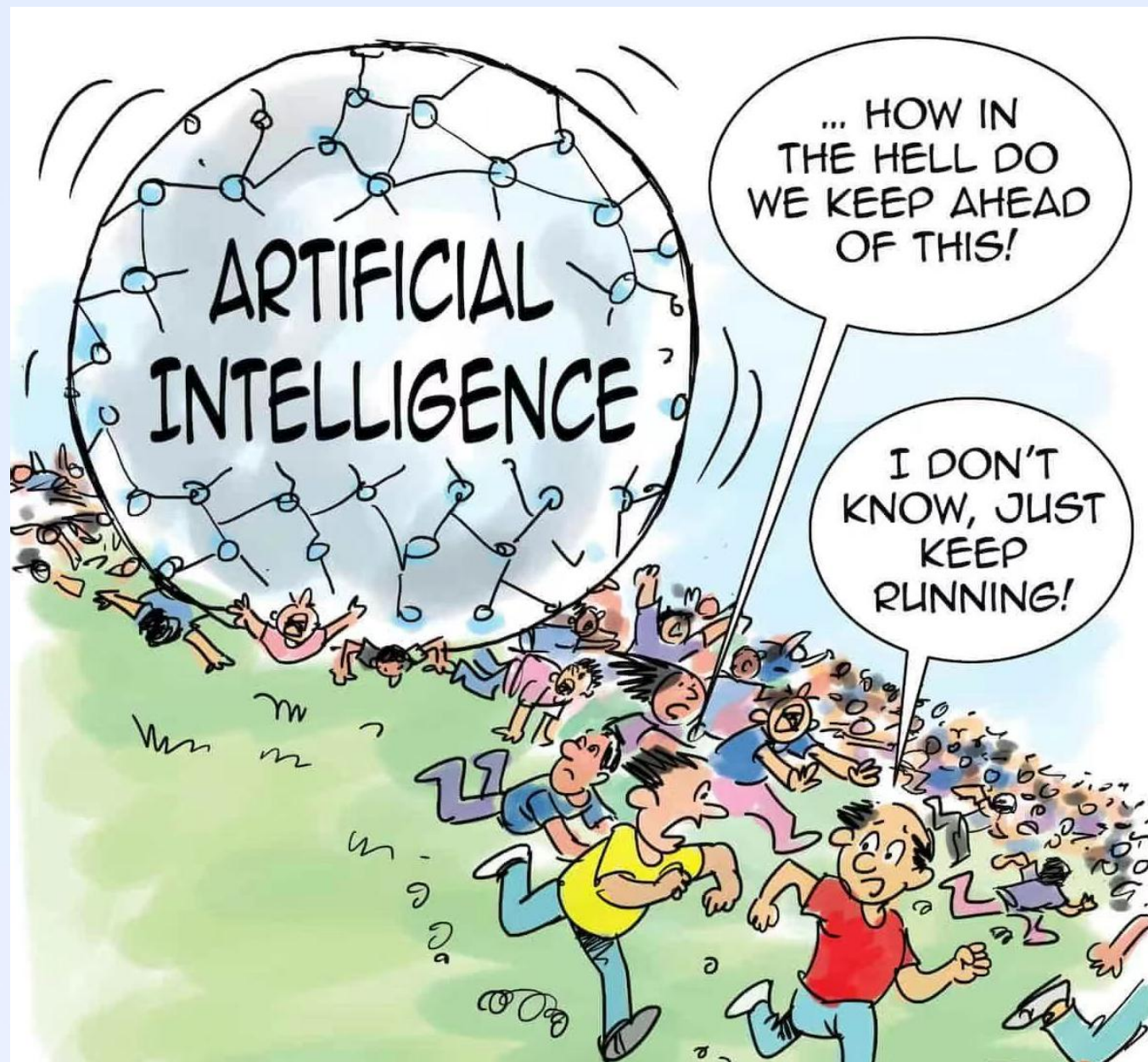


Token免费，技术平权，速度越来越快（Groq的速度每秒300token）



传统AI会和大模型结合的更加紧密

AI技术的类型	技术是如何运行的
统计机器学习	创建基于历史数据训练的预测模型
无监督式机器学习	识别相似物体的特点，预分类
自我监督学习	在数据中寻找隐含信号，逐渐训练
自监督学习	通过实践和最大化回报来学习
强化学习	使用特征的隐蔽层进行新颖分类
神经网络	使用预测模型的多层隐藏层
深度学习	学习从记忆的数据集中识别模式
深度学习图像识别	学习理解并生成语言和文本
深度学习自然语言处理	根据“如果……就……”规则做出简单决策
基于逻辑的AI系统	结合 workflow，激活时间表规则的决策
基于语义的AI	识别人类语言并将其转化为文本



AI的价值

02

- All-in On AI的价值是什么



商业价值

船票

造纸术

iPhone时代

AI 会让组织更敏捷

- 执行速度：应用AI，减少决策和延迟
- 降低成本：业务流程的自动化
- 复杂度降低：连接信息点和预测结果，提高理解和决策能力
- 转变参与方式：改变客户和员工以及智能系统的交互方式
- 驱动创新：应用AI驱动创新

AI 将带来现有商业的重构

- 交互重构：从用户适应产品，到产品是适用用户
- 数据和信息的重构：从结构化数据到非结构数据
- 服务重构：从组织、撮合到直接提供服务
- 反馈机制的重构：从功能迭代，到模型迭代
- 从GUI到LUI交互重构

核心认知

03

愿景很美好，现实很残酷，能否落地，如何落地



蓝领AI

AI关注的不是研究或实验，而是哪些项目可以在最短的时间为业务带来价值

AI不是工具：因为他没有说明书

所有人用起来，让个人生产力的提升成为公司的应用集合

拆解企业 workflow，打造企业专属工具

在产品和流程中找AI的机会，找企业的痛点是什么，而不是找AI能做什么

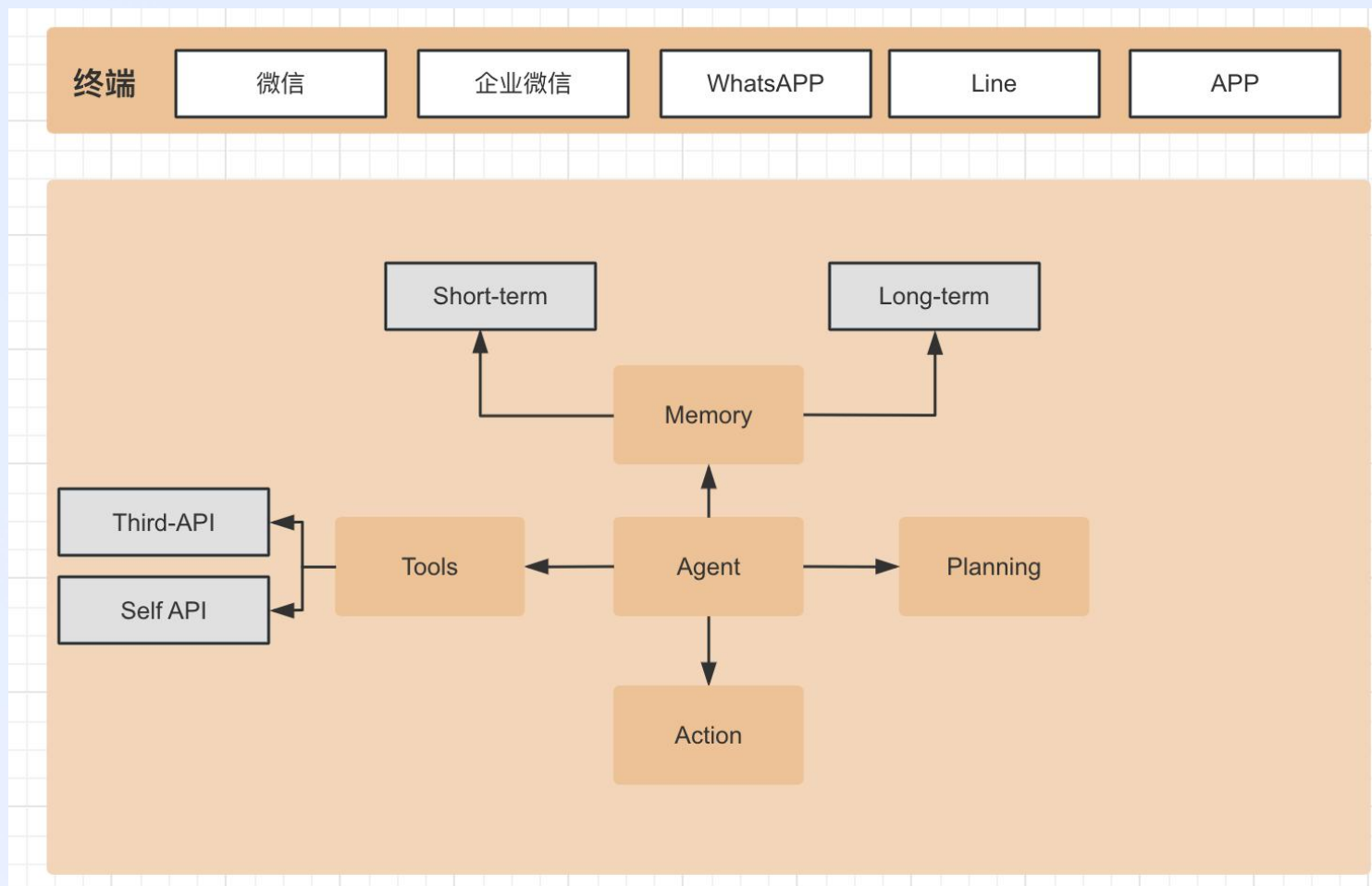
案例分享

落地案例分享

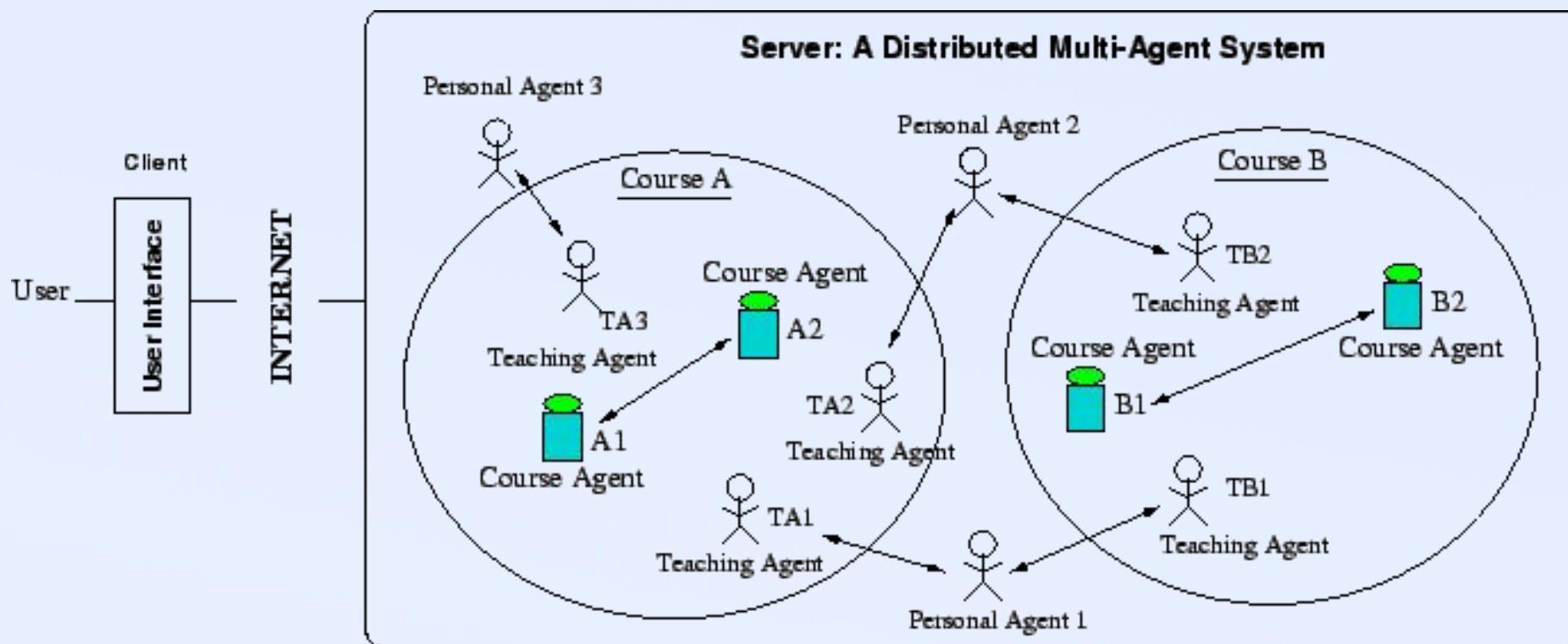
04



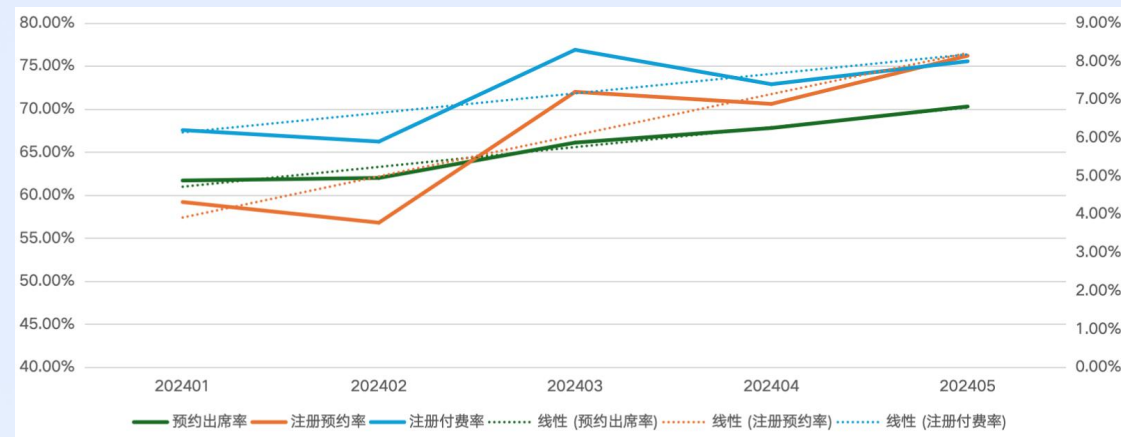
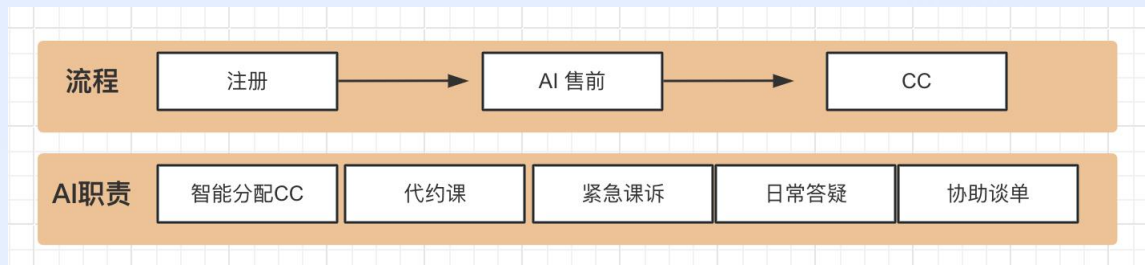
基于LLM的Agent打通内部业务线所有 workflow、数据、以及所有终端



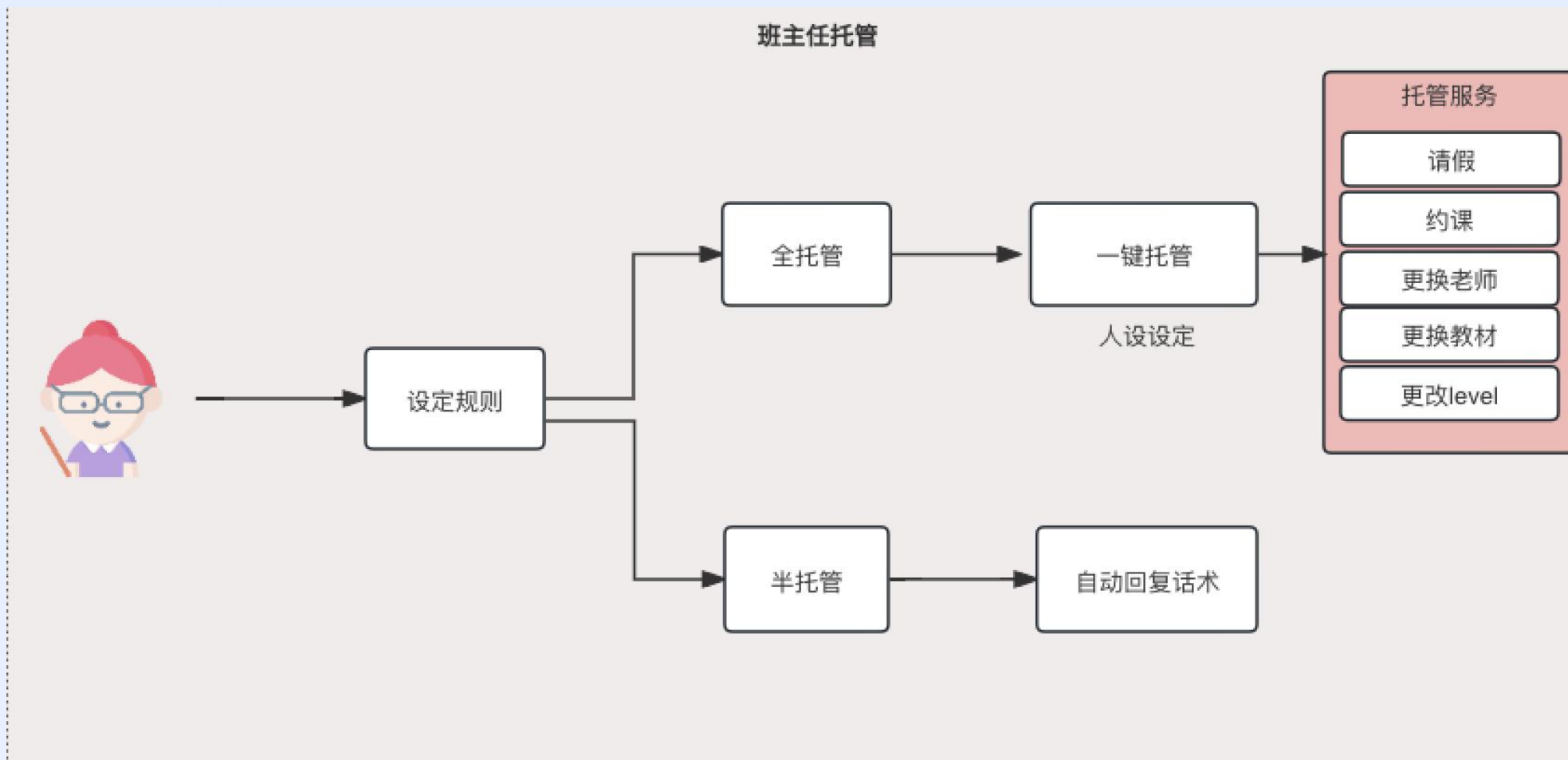
基于多智能体框架，协调客服、TMK、CC、SS、外教等智能体



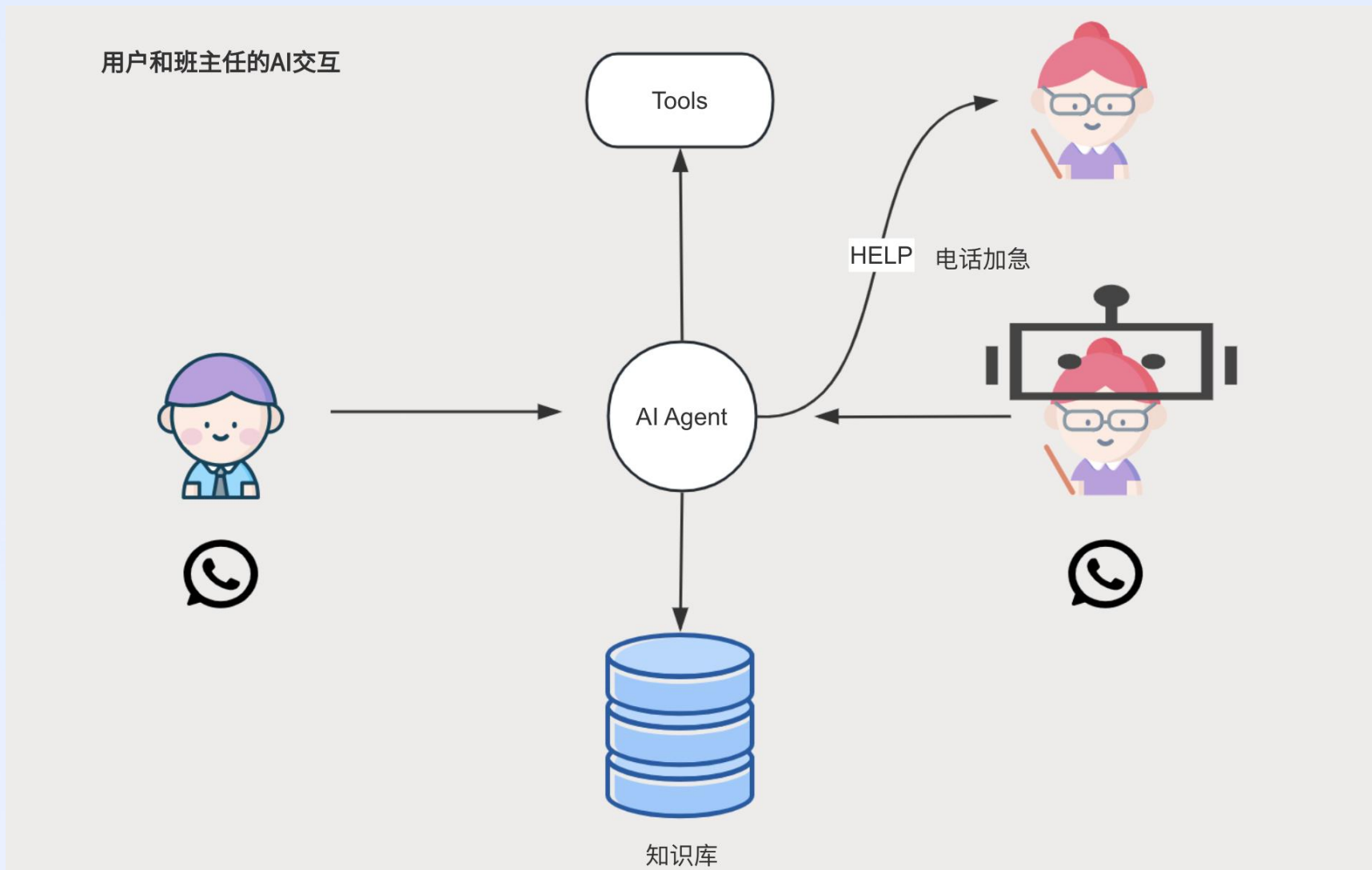
售前智能体，提供7*24小时的售前咨询转化、售后服务提效，提供售前的保温工作



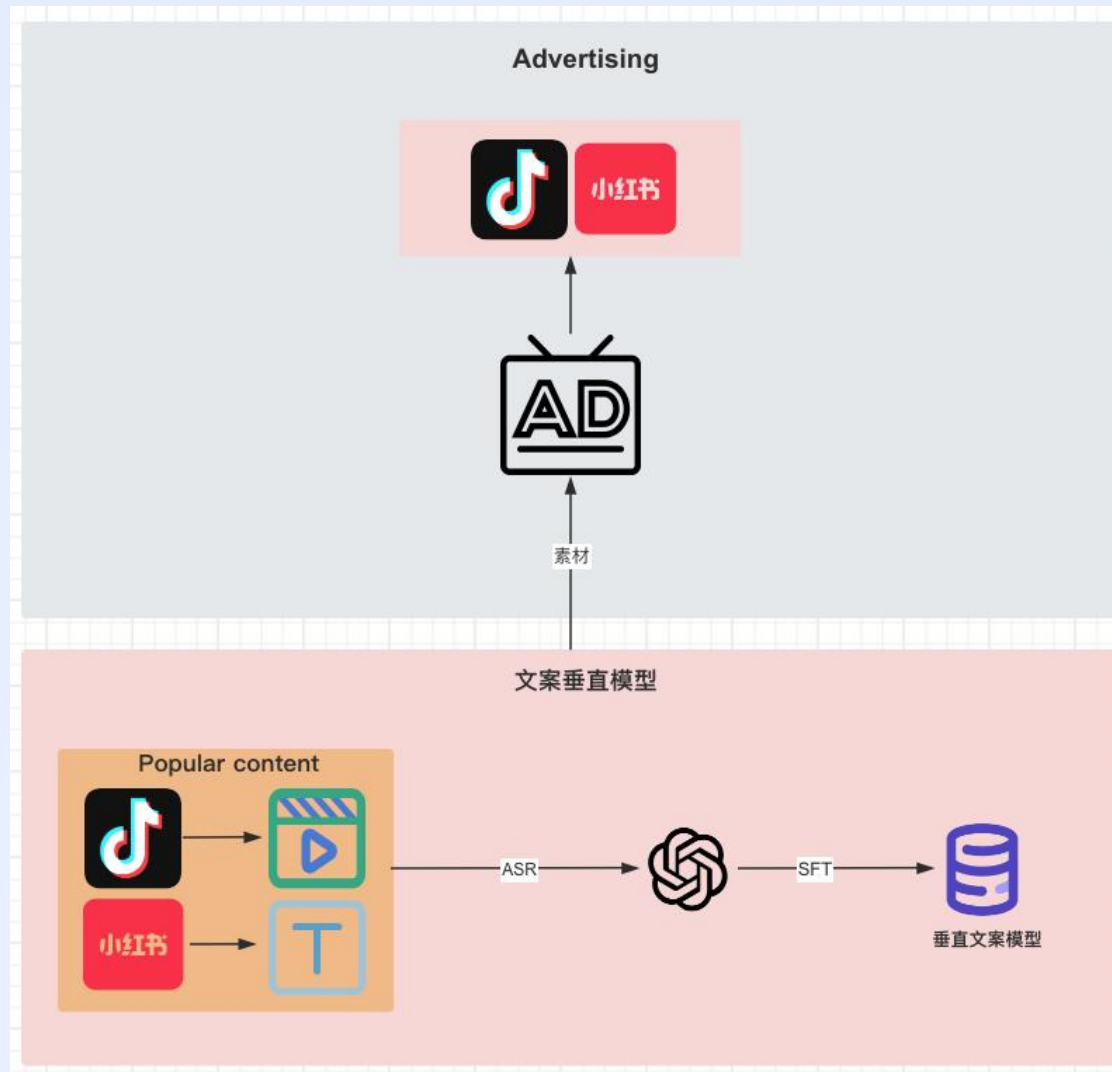
班主任智能体



班主任智能体



投放智能体



经验总结

落地经验总结

05



经验再总结

- 不搞“运动”
- 成立AI领导机构（排头兵）
- 从流程中进行定量分析，找突破点
- 整理自己的数据，数据是企业的核心
- 储备AI的人才：机器学习、NLP的人才
- 拆解工作流，打造专属工具
- 让人专注于人的事情，不是替代人，而是

推动AI转型，核心不是为了替代人类，而是如何重新设计工作岗位，重新培训员工，人和AI结合提高商业效率。

All-in On AI

谢谢观看

THANKS