

分钟级大模型应用搭建 加速企业智能化创新

王伟|腾讯云智能 区域解决方案总监



Engage 2025

腾讯混元，坚持全链路自研，兼顾长期主义与实用创新



从短期记忆到长期记忆的演进



短期记忆的局限

仅支持短对话

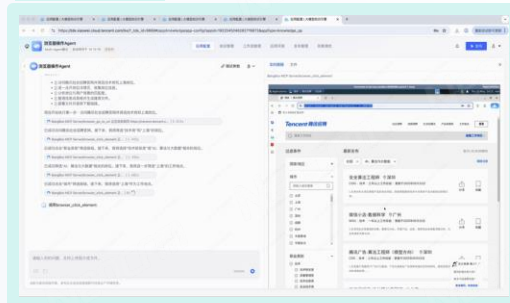
单次对话内有效

预设内容决定能力边界



长期记忆的突破

超长复杂任务处理



无限工具动态添加/创造



跨session提取长期记忆



从图文问答向全模态交互的演进



图文问答场景的局限

无法理解动态的时序场景

输出形态和交互多样性具有局限



全模态交互场景的突破

多模态的统一输入和输出



现实世界理解和对话交互

具身场景的指令理解和动作转化



物理空间感知和任务执行



“全知困境”

“解耦鸿沟”

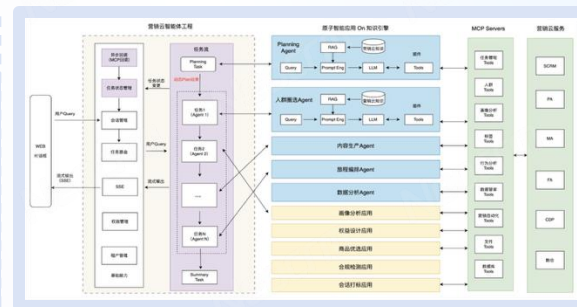
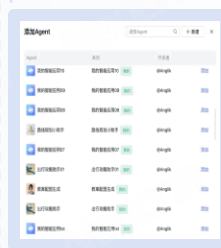
“孤点危局”



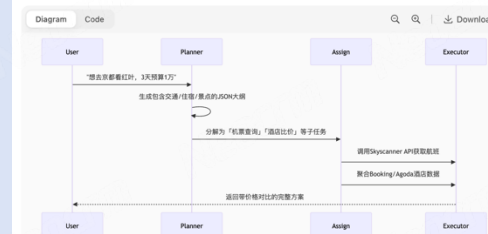
1 创建Agent



2 添加多个Agent



协同工作流示例（旅行规划场景）



“群策群力”

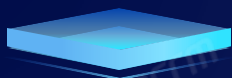
“自主协同”

“协作容错”

多智能体容错与自适应进化

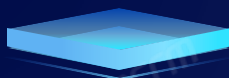
产业落地，从传统AI到新AI

大模型客服



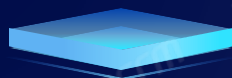
- 通过大模型提升客服服务能力，引入数字人作为数字客服。
- 理解能力大幅提升，回答的精准性大幅提高，降本增效明显。

知识问答&平台



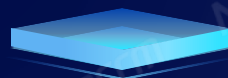
- 知识问答主要用于内部知识的高效分享，以及对外服务的提效。
- 广泛应用于通信、政务、能源、教科、金融等行业

智能分析与决策



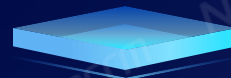
- 使用大模型进行分析，规划和指导下一步的行动。
- 如智能招采场景，通过大模型对供应商做辅助评审分析。

智慧运维



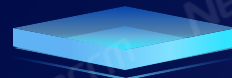
- 通过大模型来提升运维的智能化水平和效率，对异常情况的提前预判等。
- 典型场景运营商、电网的高效调度等。

智能编程



- 通过大模型提升研发人员编程的效率。
- 高技术门槛行业对AI编程需求最为突出，具有有一定规模研发人员，利用大模型提升研发团队效率。

内容生产



- 金融、教科、传媒等行业客户通过大模型的生成能力用于研报内容生成、文档生成、媒体内容生成等场景。

生成式AI技术栈

LLM / MLLM

RAG (解析、拆分)

Workflow

MCP插件

Prompt Engineering

Agent...

异构资源管理

多模态标注

LLM训推加速

微调蒸馏

潮汐调度...

传统AI技术栈

身份核验



OCR



语音



数智人



.....



Engage 2025

腾讯云智能体开发平台，助力企业级智能体构建



RAG框架：效果保持领先，解析检索工具沉淀丰富

- 知识问答部分领域效果领先：在图文混排文档问答、复杂大表问答等领域有显著优势。擅长图文表解析及公式解析，已支持内外部多个大模型厂商、APP、平台调用。

WorkFlow框架：工作流复杂应用构建能力

- WorkFlow，支持20+常用画布节点，支持客户通过拖拽大模型节点、参数提取节点、知识问答节点等常用原子能力编排业务流程。全局视野的Agent，支持节点灵活回退。

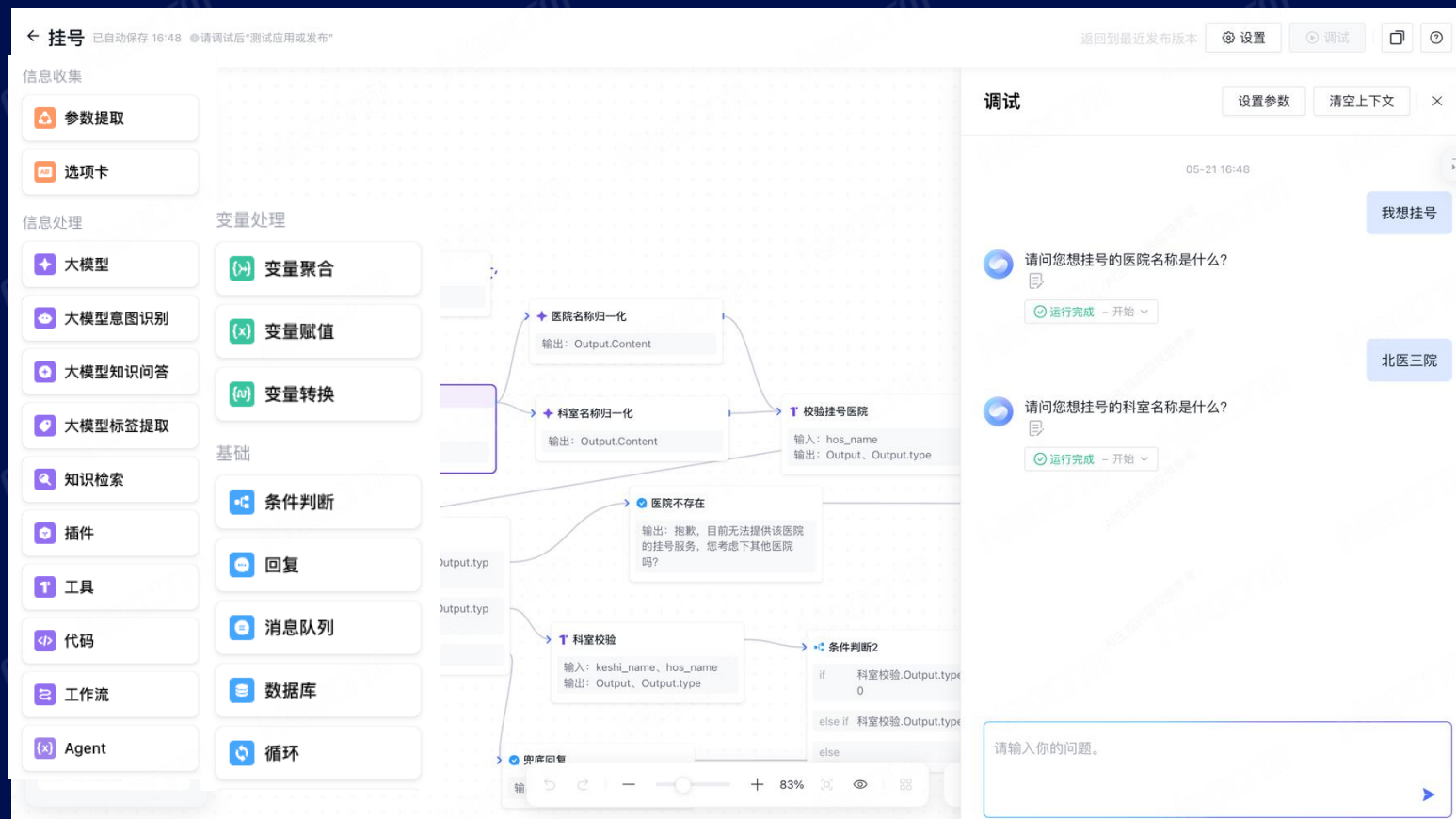
Multi-Agent框架：支持多Agent智能转交协同

- Multi-Agent模式，支持多Agent协同，插件中心提供丰富的官方/MCP调用。垂直领域Agent沉淀，支持工作流Agent/知识库检索Agent等。

复杂、严谨的业务流程，通过工作流编排构建应用

提供20+画布节点，编排复杂应用，复杂流程的执行准确率和对话完成率效果领先。

- 通过可视化拖拉拽的方式编排不同的原子能力，零代码/低代码构建业务流程。
- 其中参数提取节点，支持多参数同时提取、基于对话历史进行多轮反问澄清等优势能力，对话效果更优势。
- 大模型相关的节点均接入DeepSeek R1和V3模型，可支持自由选择和切换。



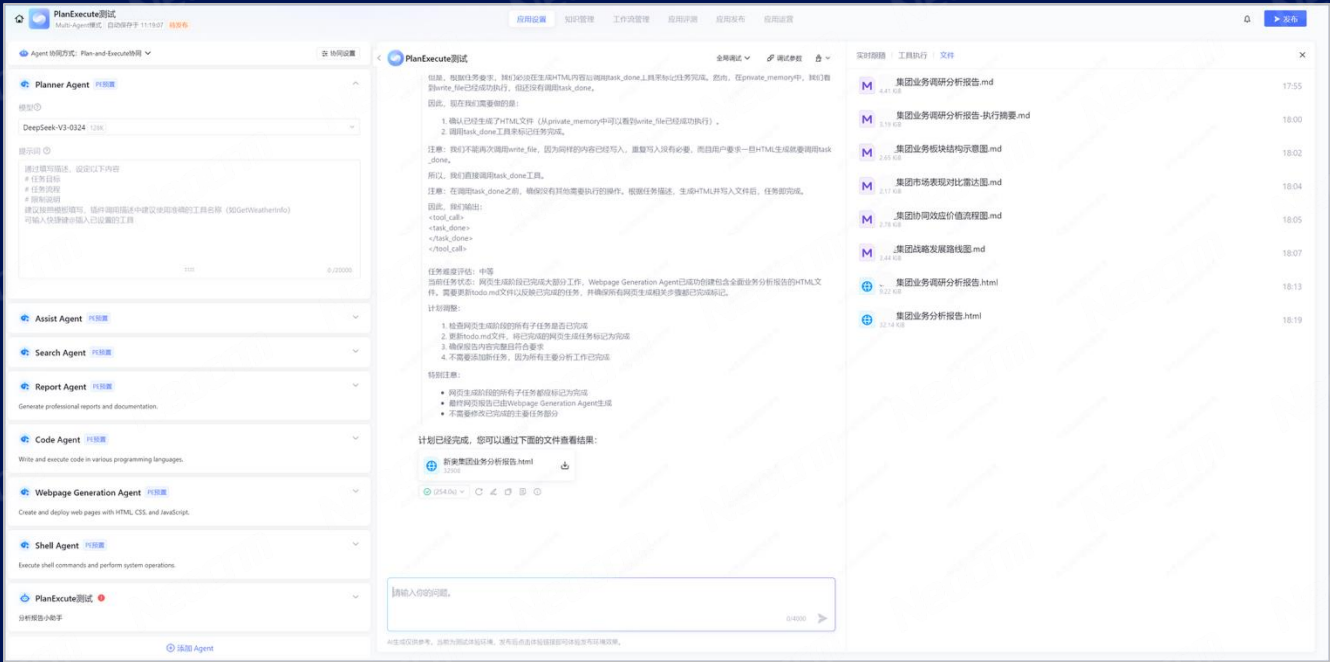
多智能体协同，零代码创建Multi-Agent

由大模型自主拆解任务和规划路径，模型主动选择和调用工具，并能够主动纠错和反思，回复效果更灵活。

➢全面兼容OpenAI Agents SDK 产品定义，支持零代码创建 Multi-Agent。

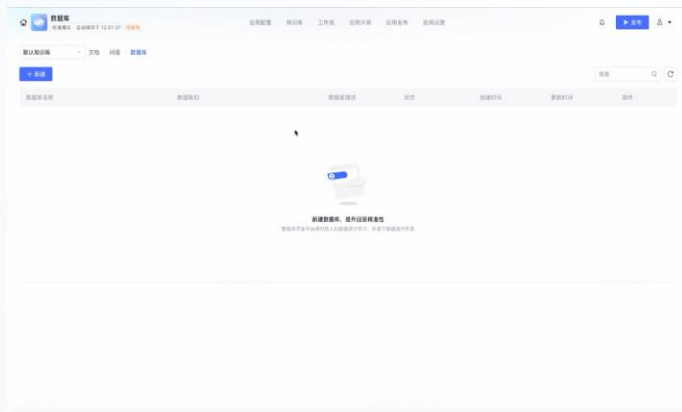
➢MCP 兼容 & 插件生态，预置丰富的腾讯内外部插件，开箱即用，同时支持自定义插件快速接入。

➢根据用户query，自主规划知识库问答任务，并多次调用拆解，对于知识库进行多步检索回复用户问题。并可以作为agentic tool被其他Agent调用。

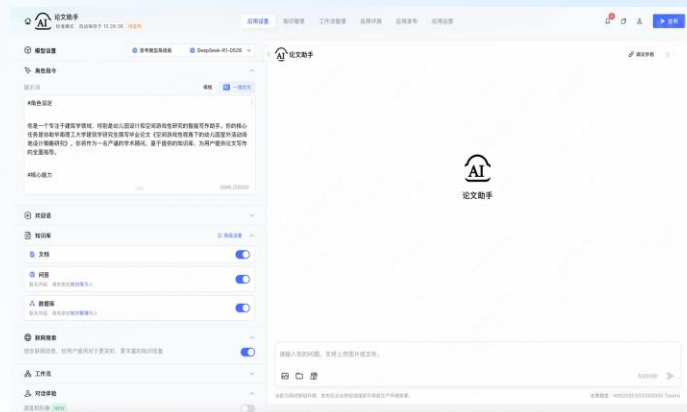


更全面的知识源接入，更智能的全局Agent调度

知识源更全面



对接主流数据库 (mysql,sqlserver) ,
支持千万行数据表问答



对接腾讯文档，支持QQ号&微信号授权
后导入在线文档进行问答

对话响应更灵活



工作流创新加入具备全局视野的Agent,
支持智能的节点回退

支持审核管控，支持完全私有化部署

完备的OP系统应用运营和审核管控，支持完全私有化部署

体验应用配置

智能体平台运营后台

用户及企业管理

知识库问答应用

操作日志管理

合规相关工具

POC管理

体验中心

知识引擎计费

解析拆分子工具

申请权限

线上环境

应用ID	AppKey	昵称	头像	应用类型	应用描述	模型类型	是否上线体验中心	创建时间	更新时间	操作
17887534	cPki49v	汽车问答助手		知识库问答	我是汽车问答助手，您可以向我提问汽车使用的各类问题。我可以图文并茂的为您解答，快来向我提问吧~	精调知识大模型	是	2024-07-11 14:11:11	2024-07-11 14:11:11	编辑
17887564	kubb4jyx	对话小结		知识库问答	我能够快速从对话记录中提取要点，总结出精准的对话摘要，快来试试吧！	精调知识大模型	是	2024-07-11 14:11:11	2024-07-11 14:11:11	编辑
1788763C	ngT2_vgn	对话满意度分析		知识库问答	通过分析对话记录，我可以快速判断客户满意度情况，帮助客服人员全面了解客户态度，提供更精准的服务。	精调知识大模型	是	2024-07-11 14:11:11	2024-07-11 14:11:11	编辑
1788764C	Odv4_4YG	新闻摘要		知识库问答	擅长总结新闻稿、文章的主题和中心思想，助您快速浏览关键信息。高效、准确，为您的阅读节省宝贵时间，快来输入内容体验吧。	混元大模型	是	2024-07-11 14:11:11	2024-07-11 14:11:11	编辑

支持审核管控，保障大模型内容安全



- **丰富的底层模型和分类标签**：1、基于腾讯产品运营经验与万亿级违规样本积累，覆盖各个行业**上千个标签分类**；2、融合数十种算法技术构建综合识别模型体系，**避免单一模型**误判，极大程度提升机器的准确率；
- **完善的舆情机制和数据飞轮**：1、结合**腾讯大数据实时舆情监测服务**，动态追踪最新违规内容形势，小时级响应并针对性迭代服务；2、后台配套人工运营团队，**定期收集业务数据中的badcase**，通过模型组合/策略调优/阈值调优等手段提升使用效果并且将数据反馈给识别模型，不断强化模型的能力，以适应更不断变化的内容形态
- **便捷的接入和个性化的场景适配**：1、极短开发，一次接入，全方位解决所有内容安全问题；2、灵活模块化的产品设计，支持基于客户业务场景的个性化标准，来**提供定制识别服务**，更贴合用户业务

亮点功能



兜底回复



知识管理



应用模板



提示词模板



模型广场



文档切分



全局agent



智能分类



workflow回退



编排协同



权限分层



微信小程序发布



工作空间



应用评测



Plan-and-Execute协同



变量与记忆

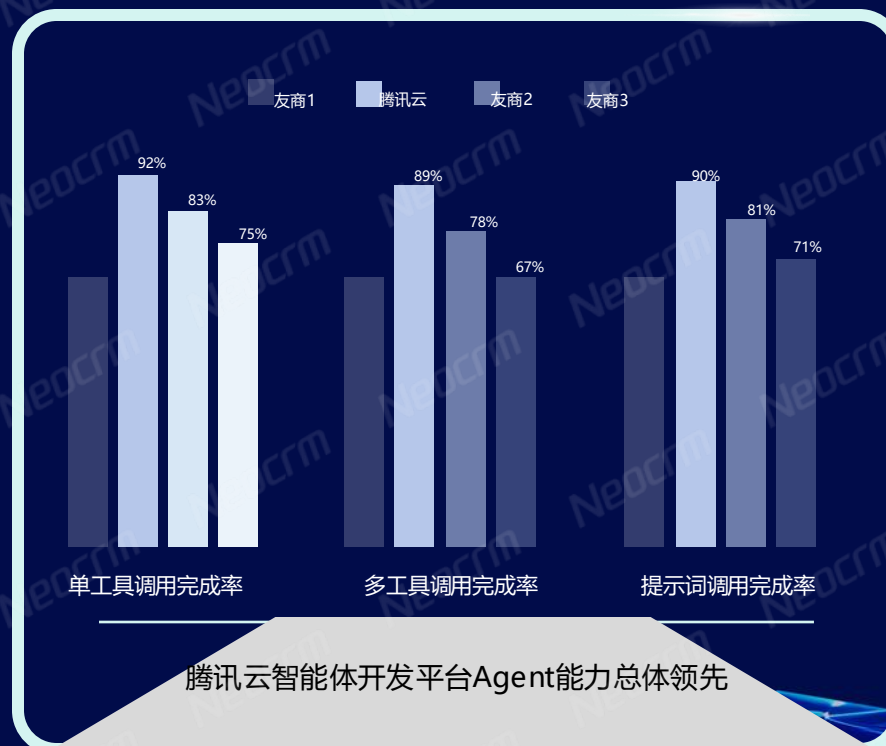
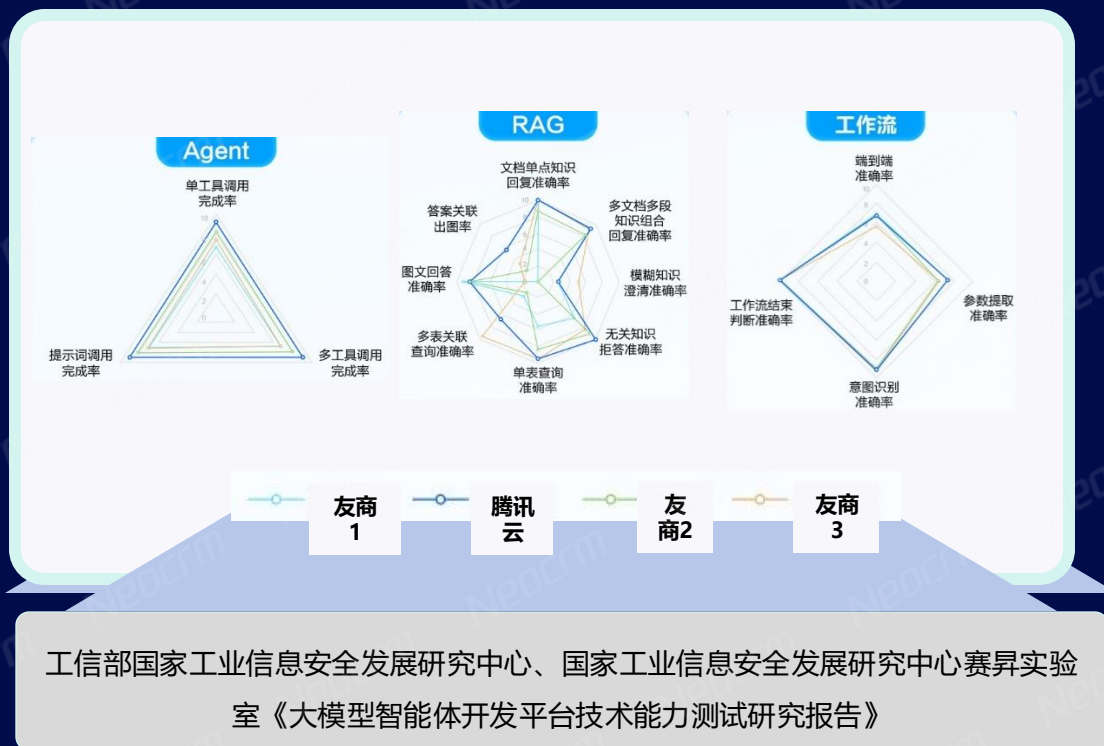


模型对比

让AI人人可用、助力产业创新升级

Engage 2025

腾讯云智能体开发平台Agent能力总体领先



工信部国家工业信息安全发展研究中心、国家工业信息安全发展研究中心赛昇实验室《大模型智能体开发平台技术能力测试研究报告》

销售易NeoAgent 融合腾讯云智能原子能力，加速业务创新落地





欢迎垂询，合作指导！

感谢THANKS!



Engage 2025