Tencent 腾讯 | 腾讯云 | WORK BUDDY

AI Agent

企业解决方案及落地实践

2026年



腾讯云互联网行业出品

目录

CONTENTS

01
AI整体解决方案
从超级个人到超级团队
P01-07

02
内部提效
企业研发运维全链路提效

03
技术创新
从模型训练到智能体应用的技术底座
P48-56

04
业务创新—SaaS Agent
AI Agent重塑行业SaaS体验
P57-78

05
业务创新—内容生产
AI驱动规模化内容生产与创意升级
P79-103

06
业务创新—AI社交陪伴
AI角色沉浸陪伴，情感化互动体验
P104-112

更多免费行业报告 并购家 www.ipoipo.net

07

业务创新—游戏内AI玩法

AI驱动NPC智能化与玩法创新

P113-120

08

业务创新—运营/智能客服

AI驱动客户全生命周期智能化运营

P121-130

09

WorkBuddy生态共建

共创超级团队

P131-150

01

AI整体解决方案

从超级个人到超级团队

方案

1.1 腾讯云Agent解决方案全景图 / P01

1.2 腾讯WorkBuddy：企业级智能体解决方案 / P02

1.3 腾讯云ClawPro：企业级AI Agent管控台 / P04

1.4 腾讯云Agent Runtime：Agent原生的新一代基础设施平台 / P06

1.5 腾讯云TokenHub：统一的大模型服务入口 / P07

1.1 腾讯云Agent解决方案全景图

1 打造面向Agent时代的全栈AI引擎

个人提效 LightClaw、QClaw
办公提效 WorkBuddy
知识提效 乐享
营销提效 企点营销云
研发提效 CodeBuddy
企业应用提效 ADP × ClawPro
数据提效 Data Agent

微信 元宝 小程序 QQ 企业微信

全生命周期安全防护

- Lighthouse 原生安全
- AI Agent 安全中心
- AI Agent 安全网关
- CodeBuddy Code Security
- 腾讯 iOA
- 腾讯电脑管家

全栈应用开箱即用

生态开放技能丰富

SkillHub
专属开源技能中心

25000+生态开源Skills
腾讯自研原生Skills
腾讯文档 Skill
QQ浏览器 Skill
腾讯乐享知识库 Skill
腾讯地图 Skill
腾讯云语音 Skill
腾讯会议 Skill
腾讯云CodeBuddy Skill
企业微信 Skill
ima Skill
数据安全（HaS）Skill
...

全系模型灵活调度

腾讯混元大模型
第三方模型

TokenHub
大模型服务平台

- 全栈模型供给 >
  自研模型+行业模型+开源模型
- 全生命周期管理 >
  训练+微调+推理+部署一站式服务
- Agent 专属优化 >
  场景化模型路由与调度

企业级Agent治理底座

Agent Runtime企业级Agent治理方案
执行引擎 | 云沙箱 | 记忆 | 网关 | 可观测

AI Infra：智算底座
通用算力 | 异构算力 | 高性能网络 | 高性能分布式存储 | Agent原生容器环境 | 加速+算力调度

1.2 腾讯WorkBuddy：企业级智能体解决方案

1 企业智能体的三阶段跃迁：从“能对话”到“能干活”，再到“能上岗”

角色从工具升级为流程连接器，再进化为按 KPI 交付的组织新角色

能对话	能干活	能上岗
第一阶段 · 知识问答	第二阶段 · 办公助手	第三阶段 · 数字员工
代表：DeepSeek、ChatGPT	代表：WorkBuddy	形态：Agent 工作台 + 业务系统集成
核心能力：自然语言理解与行业知识检索	核心能力：嵌入工作流调工具、操作文档、协助流程	核心能力：进入系统、操作数据，完成业务闭环并对结果负责
角色定位：问答工具	角色定位：流程连接器	角色定位：按 KPI 交付的组织新角色

当前阶段
办公助手 → 数字员工

从流程连接器，迈向按“目标”交付的组织新角色
系统越复杂、流程越密集的组织，越是数字员工规模化落地的主战场。

2 打造Agent时代的办公新范式：以数字员工为执行主体的人机协同

以 WorkBuddy 为代表：人用自然语言给目标、边界和判断；数字员工理解上下文、调用工具、持续执行并交付结果

第一代	工具办公	第二代	协同办公	第三代	数字员工
人操作软件完成工作		人在平台中协同工作		数字员工直接完成工作	
代表：Office		代表：主流协同办公平台（某钉 / 某书等）		代表：WorkBuddy	
人负责		人负责		人负责	
找功能 · 录内容 · 调格式		沟通 · 分工 · 填报 · 催办		给目标 · 设边界 · 做判断	
软件角色		平台角色		数字员工	
等待操作 · 响应指令		连接人 · 消息 · 文档 · 流程		理解任务 · 调工具 · 持续执行	
交付一份文件		交付一次协作过程		交付一项完成的工作	

为什么第三代必然到来？执行主体正在从人转向数字员工
协同平台解决了“人怎么一起做”，AI Agent 开始解决“谁来把事情做完”

01 工作对象变复杂
任务横跨消息、文档和多个系统
仅靠人协同，执行压力不断上升

02 软件开始会执行
AI Agent 能理解目标、规划步骤
调用工具，并持续推进到结果

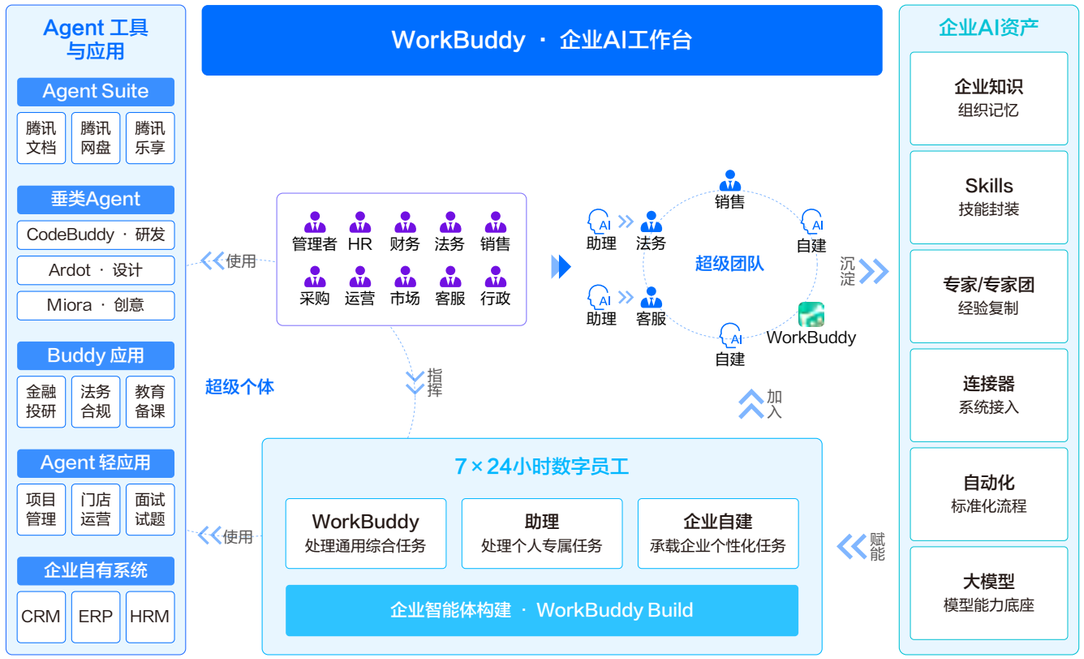
03 组织产能可复制
数字员工可以复制执行力
企业竞争转向“人 + 数字员工”总产能

新一代办公定义
人负责目标、边界与判断；数字员工理解上下文、调用工具、持续执行，直接交付完成的工作。

更多免费行业报告 并购家 www.ipoipo.net

3 让 AI 通过“企业AI工作台”真正进入组织工作

连接工具、岗位与企业资产，把超级个体组织成可持续交付的超级团队



4 第三代办公新范式：既增强每个人，也重构整个组织

狭义办公决定智能能力能否规模化普及；广义办公决定智能能力能否进入经营核心

狭义办公 以个人为中心的通用智能办公		广义办公 以业务目标为中心的组织智能办公
战略位置　普惠入口 · 个人生产力底座		战略位置　价值兑现 · 组织生产力核心
01 进入日常 PPT汇报 | 文档协作 | 表格分析 | 会议沟通	能力延伸	01 进入系统 CRM | ERP | MES | HR / 项目系统
02 增强个人 理解问题 | 创作内容 | 分析数据 | 辅助判断		02 组织执行 调人员 | 跑流程 | 取数据 | 守权限
03 沉淀资产 知识经验 | 统一口径 | 任务上下文 | 可复用方法		03 结果兑现 推进商机 | 处理订单 | 跟踪交付 | 闭环异常
决定「普及广度」 覆盖更多人 × 进入高频工作 × 积累组织知识		决定「价值深度」 进入核心业务 × 改变业务状态 × 创造经营结果

人人可用的AI工作入口 + 深入业务的数字员工执行体系 = 全新一代办公范式

1.3 腾讯云ClawPro：企业级AI Agent管控台

1 腾讯云 ClawPro · 企业级 AI Agent 管控台

产品介绍

开箱即用的企业级AI智能体管控平台

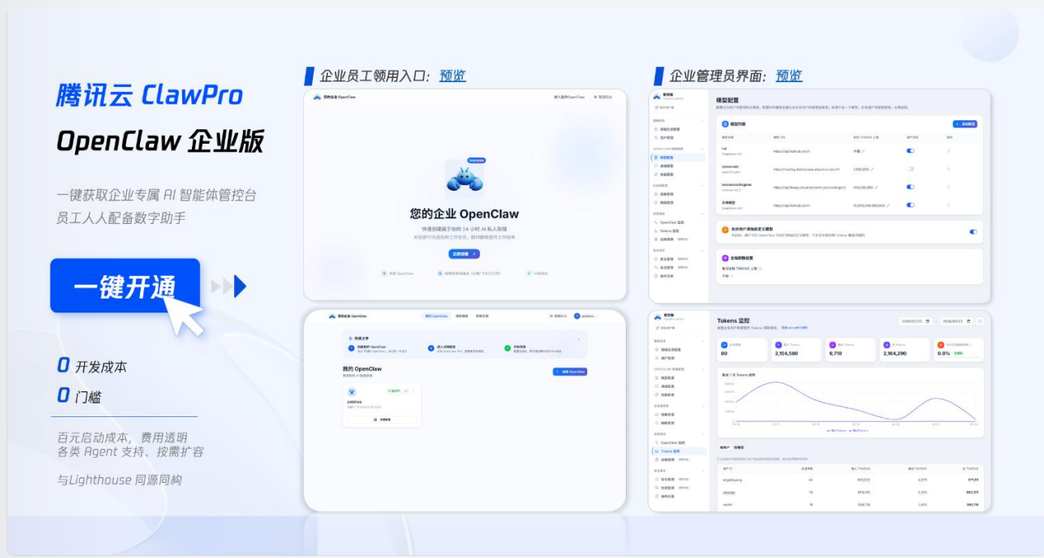
实现员工一键开通，Agent即开即用；企业统一管控，全流程安全可靠，助力企业轻松管理所有Agent实例

产品定位

开源开放，任意Agent的企业管控平台

开源、自研、自定义

管控端与员工端双端界面预览



工信部首批认证

百万级用户实战验证

企业级全链路安全防护

覆盖SaaS服务、金融、互联网、财税等行业

2 腾讯云 ClawPro · 产品能力全景图

用户触点

管控端 全局管控，企业专属，简单易用

模型管理 | 企业Skill库 | IM通道 | 镜像管理 | 审计日志

员工端 一键申请，快速创建，按需使用

一键创建 → 即刻使用 | 用量统计 | 开箱即用，零技术门槛

ClawPro 核心能力

首发 IM	首发 Skill	IAM	Mem	Cost	Net	独家 LH
平滑适配	技能管理	身份权限	记忆体系	成本管控	网络加固	轻量云同构
• 微信 / QQ	• 初始技能包	• IAM登录	• 记忆存档	• Token管控	• 物理级隔离	• 百万用户验证
• 企微 / 元宝	• 龙虾角色 / 工种	• 三级RBAC	• 记忆检索	• 资源管控	• 随机端口	• 同款镜像Skills
• 飞书	• 私有Skill库	• 组织架构	• 记忆管理	• 用量预警	• 安全组管控	• 管控面同源
• 钉钉	• 国内Skill社区	• 资源管控	• 向量桶优化	• 成本优化	• VPC隔离	

安全体系

- **能看见** 资产盘点 风险可视
- **可审计** 全链路溯源 对话I / O 审计
- **强管控** 运行管控 环境隔离
- **全扫描** Skill 供应链 安全扫描

AI基础设施

- **C** 计算系列产品 CVM、TAT、Runtime
- **S** 存储系列产品 CBS、COS、CLS、CFS
- **N** 网络系列产品 GA、VPC、安全组
- **S** 安全系列产品 网络防护、终端防护

3 腾讯云ClawPro · 企业级AI Agent 管控台

核心能力

零门槛分钟级上线
10分钟开通，员工1分钟申领即用

任意 Agent 接入
开源、自研、自定义Agent创建＆统一管控

多层安全体系
看 · 审 · 管 · 扫，高危命令实时拦截

全IM生态融合
微信 / QQ / 企微 / 飞书 / 钉钉一键接入

灵活分组管理
部门 / 项目 / 个人三级，权限分组下发

精细成本管控
Token / Agent三级配额，按组按人核算

Claw Team协同
多Agent团队级任务编排

企业专属技能库
MCP / 插件 / Skill私有库，批量下发

1.4 腾讯云Agent Runtime：Agent原生的新一代基础设施平台

1 产品全景

客户与场景

- Agent强化学习训练 / 评测场景 大模型厂商
- Agent构建与托管 Agent服务商
- 企业内Agent落地 千行百业企业

Agent 不同层级产品能力按需组装

Agent Runtime

治理平台层 — 以Agent为第一视角的统一操作面，面向企业IT运营者，解决成千上万Agent治理问题

Agent Engine 生命周期管理 | 治理与安全 | 资源与成本管控 | 组织级资产沉淀

基础产品层 — 七大原子化能力模块，每个独立可用，按需组合支撑治理平台层和场景化解决方案

Agent Sandbox | Agent Wall | Agent Gateway | Agent Identity

Agent Observability | Agent Skills | Agent Memory

2 产品架构



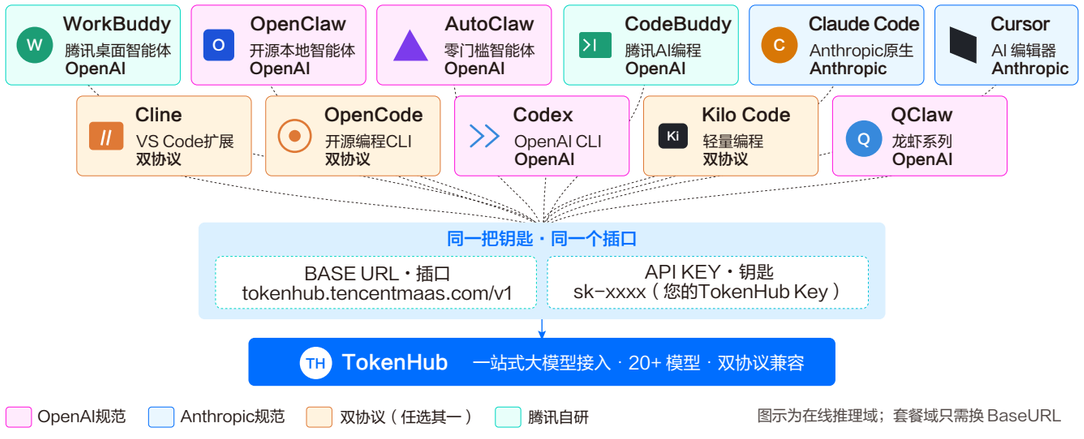
更多免费行业报告 并购家 www.ipoipo.net

1.5 腾讯云TokenHub：统一的大模型服务入口

1 腾讯云TokenHub

大模型服务平台TokenHub致力于为企业和开发者提供统一的大模型服务入口，整合腾讯自研的混元大模型能力，并引入优质第三方模型，覆盖通用对话、深度推理、代码生成、视觉理解、图像生成、视频生成等多类场景。平台支持按量调用、保障型资源、专属部署等多种服务模式，帮助企业简单、高效地获取AI算力。

➤ 主流工具全覆盖：现成工具也一样接



配的东西永远只有两个

BaseURL和API Key。换套餐/按量时两个都要对应地换（同域配对，缺一不可）。

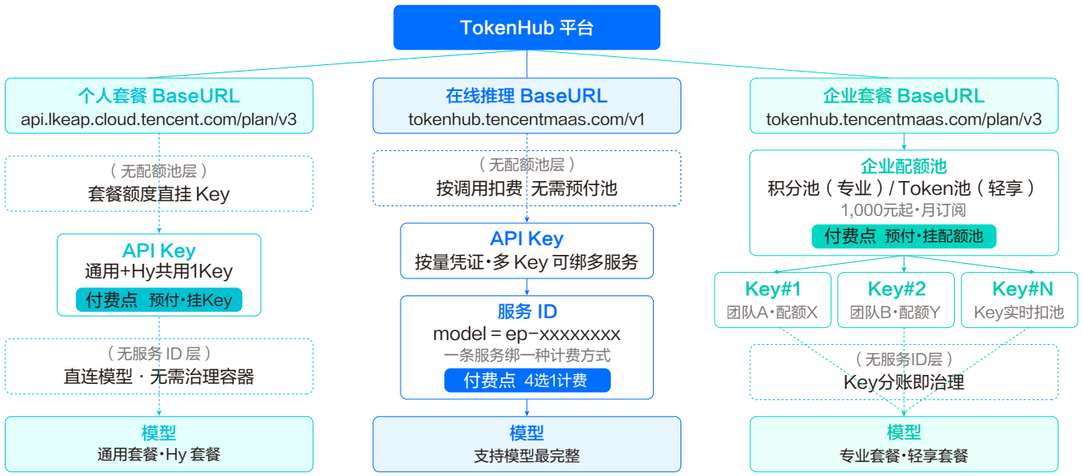
同一产品线下Key复用

一把Key可以让该产品线下所有兼容工具同时转起来；跨部门分账就在控制台多发几把Key。

协议是工具自选的

Claude Code/Cursor用Anthropic规范；WorkBuddy/OpenClaw/Codex用OpenAI规范；Cline/OpenCode/Kilo Code两种都能走。我们都兼容。

2 产品线



02

内部提效

企业研发运维全链路提效

方案

2.1 WorkBuddy 打造第三代办公新范式 / P09
2.2 CodeBuddy赋能企业AI Coding实践转型 / P15
2.3 腾讯云游戏研发AI解决方案 / P20
2.4 腾讯云游戏精细化运营AI Agent解决方案 / P22
2.5 腾讯云AI驱动SRE智能故障定位解决方案 / P24
2.6 腾讯云AI Agent数据库智能运维方案 / P26

案例

标杆案例——盛和 / P29
标杆案例——淘米 / P31
标杆案例——游族 / P33
标杆案例——游卡 / P35
标杆案例——花椒 / P37
标杆案例——搜狐畅游 / P39
标杆案例——紫龙游戏 / P41
标杆案例——马帮 / P45

2.1 WorkBuddy 打造第三代办公新范式

1 广义办公：让企业从“系统各自记录”走向“全局判断、统一行动”

系统只能回答局部问题；业务大脑把分散事实连成全景，并把管理判断持续推进到结果。

一个真实的业务问题

这张订单，能不能按期交付？

答案分散在四个地方

- CRM：客户承诺 · 优先级
- ERP：库存 · 物料 · 采购
- MES：产能 · 工单 · 异常
- 协同：责任人 · 最新进度

过去：靠人逐个查、逐个问、开会拼答案

WorkBuddy · 企业业务大脑

把分散事实变成可执行判断

- 看全局：客户、订单、库存、产能、进度放在一起
- 做判断：识别缺口、冲突、风险与优先顺序
- 推行动：落实到系统、流程、责任人并持续追踪

给出的不是一份报告，而是：结论 + 依据 + 下一步行动

企业得到的组织答案

不再是“每个系统各说一段”

- 同一张全景 | 事实不再分散
- 同一个判断 | 依据能够解释
- 同一套行动 | 决策真正落地

把关键人经验变成组织能力

看问题
各自报数、各说一段 → 全员看同一张业务全景

做决策
会上拼信息、凭经验判断 → 基于完整事实持续判断

推执行
会后靠人催、结果容易断 → 责任、动作、状态持续闭环

关键判断：广义办公的价值，是让企业不只是“有系统”，而是能看清全局、做出判断，并把判断持续变成业务结果。

2 用自然语言，把一个岗位的工作方式配置成数字员工

不是做一个会聊天的助手，而是把目标、上下文、知识、工具、运行与权限配置成一套可审计的岗位系统。

从一句目标，到一个可评测、可授权、可上岗的数字员工

- 定清目标
  - 岗位职责 · 触发条件
  - 交付物 · 完成标准
  - 边界与转人工条件
- 接上资产
  - 企业知识 · 业务数据
  - CRM / ERP / 流程工具
  - 角色身份与权限范围
- 真任务评测
  - 历史任务真实回放
  - 准确性 · 完整性评测
  - 异常与边界场景测试
- 灰度上岗
  - 小范围岗位试用
  - 关键动作人工确认
  - 持续观测并迭代优化

上岗前，必须配置清楚6件事

六项配置共同决定：能否做对、做稳、做得可控

- 目标与任务：职责范围 · 触发条件 · 交付物 · 完成标准 · 禁区与转人工条件
- 上下文管理：客户 / 订单 / 项目状态 · 当前会话与历史记录 · 可见范围与时效边界
- 知识与记忆：产品知识 · SOP · 规则 · 历史案例 · 专家经验 · 哪些可记、哪些需清除
- 工具与流程：CRM · ERP · MES · OA · 查询 · 填单 · 建单 · 回写 · 动作确认与防重复执行
- 运行与恢复：定时 / 事件触发 · 超时重试 · 断点续跑 · 失败升级人工接管
- 权限与管控：角色授权 · 最小权限 · 敏感动作二次确认 · 全过程审计与追溯



真实落地案例

首发 | 磐石小O

- 查客户 · 看数据 · 问流程
- 身份可感知 · 权限可控制
- 关键动作确认 · 结果可追溯

3 连接系统之后，数字员工才真正进入业务

从理解任务，到调用系统、改变业务状态并回写留痕，形成可核验的执行闭环。

① 系统怎么打通 | 数字员工通过鼎捷 MCP 操作真实业务系统 **以鼎捷为例**

场景任务：问数 · 报工 · 委外 → WorkBuddy数字员工：理解目标 · 规划步骤 → 调用 → 鼎捷MCP连接层：调用系统 · 回写结果

鼎捷MCP暴露的真实系统能力

- CRM：客户 · 商机
- ERP：订单 · 库存
- MES：工单 · 设备
- KG：知识 · 关系



真实界面证据 | 4 类MCP 已启用，并可在联调环境调用

执行闭环 | 读取鼎捷数据 → 跨系统判断 → 新建 / 更新单据 → 回写结果并留痕

联调验证的不是“能否回答”，而是能否真正改变 CRM / ERP / MES 中的业务状态

② 连上之后做什么 | 11 类已验证场景，就是上方链路的“场景任务”　上面证明连得上；下面证明有活可干

- 生产制造 | 5 类：ERP / MES | 委外发货回货；智能问数；设备报修；语音报工；智能委外
- 售前洞察 | 1 类：CRM / KG | 竞品分析助手；客户需求 + 对标信息 → 方案判断
- 智能交付 | 5 类：CRM / ERP / KG | Excel 导入；配置评估；流程规划；蓝图助手；操作手册生成

4 企业 AI 资产不是“存起来的资料”，而是可被反复调用的生产力

每一次真实任务都沉淀知识、规则、流程、Skill 与结果反馈；下一次直接复用，并在使用中持续进化

资产形成闭环：做一次，沉淀一次；再用一次，增值一次

① 真实任务：异常 · 决策 · 交付 → ② 结构化沉淀：提炼知识、规则与流程 → ③ 场景调用：岗位中直接使用 → ④ 结果回填：保留结果与反馈 → ⑤ 持续进化：验证 · 修正 · 版本化 → 回到 ①

企业AI资产库

不是归档，而是可调用、可验证、可版本化

- 专家经验
- 业务规则
- 标准流程
- Skill 组件
- 数据口径
- 结果案例

为什么企业 AI 资产非常重要

它把个人能力变成企业可持续拥有的能力

1. 关键经验留在企业：资深骨干、专家与骨干的判断，不再随人员流动而丢失。
2. 新人更快达到岗位标准：在真实任务中直接调用成熟方法，显著缩短培养周期。
3. 交付更稳定，风险更可控：统一规则、口径和流程，减少错漏、返工与结果波动。
4. 优秀实践可以规模复制：从一个人、一个岗位，快速扩展到部门、工厂与集团。

资产复利的结果

越用越多 × 越用越准 × 越用越有价值 → 降低个人依赖；缩短新人培养；减少错误返工；加快跨厂复制

更多免费行业报告 并购家 www.ipoipo.net

5 财务与 HR：合规与举证，缺的从来不是勤奋

同一条链路：数据散在三处 → 判断做不了 → 出事才知道，补上的是可查底表与敞口金额

财务 | FINANCE · 往来与核算

场景描述

应收账款账龄分析与逾期预警

月末结账后老板问「到底谁欠我们钱、欠了多久」。要从销售开票、银行回款、发票三处手工拼账龄。

核心痛点

- 销售台账与财务口径对不上，拼出来的数每次都不一样
- 账上躺着300万应收，老板以为都是钱
- 逾期超过1年的，核销概率高达80% —— 但没人按账龄排过序

以前怎么做

以前	用了WorkBuddy
三处数据手工拼账龄，每月重算一遍	三处原始数据直接拼，口径差异逐笔标注
逾期客户凭印象列催收清单	四档账龄+坏账计提测算自动出
已到账未核销的差异没人逐笔查	催收清单按「金额 × 天数」排序，附历史核销提示

可量化的效果

87万	300万	2条
逾期敞口被摆到老板面前	应收总额首次分出四档账龄	口径固定：货到60天不算逾期

HR | HUMAN RESOURCES · 员工关系与合同合规

场景描述

劳动合同台账重建与到期预警

内审或劳动监察通知一来，要在一周内交出一份能扛住检查的合同台账，并回答有没有漏签。

核心痛点

- 纸柜一摞合同、桌面三个版本台账，谁也说不清哪个是真的
- 合同躺在档案柜没人翻，直到员工申请仲裁才知道 3 份早过期
- 「连续两次固定期限」的无固定期限临界，从来没人算过

以前怎么做

以前	用了WorkBuddy
三个版本台账各自维护，到期靠人记	扫描件与三版Excel一起重建底表
续签临界靠印象，法定顺延常被忽略	无固定期限临界自动计算并标记
归档无标准，该销毁的和该留的混一起	按离职保存二年规则分出归档 / 可销毁

可量化的效果

30/60/90天	3份	1条
三级到期前瞻，每月更新一次即可	已过期合同被翻出，附敞口金额	续签未归档成为必查项

6 法务与销售：条款、时效、资质，错一处就是钱

都要在很短时间里给出能担责的结论，而依据分散在别人手里

法务 | LEGAL AFFAIRS · 合同全周期管理

合同风险分级初审（四栏清单版）

新客户发来24页对方版框架合同，首单46万、年意向120万、要求60天账期，销售催下周就签。

核心痛点

- 业务嫌法务审得慢，一个下午要出审查意见
- 真正的坑不是审漏字眼，是付款条款与违约金的不对等没人算过账

以前 → 用了 WorkBuddy

以前	用了 WorkBuddy
逐条改到业务看不懂，销售拿不去谈	四栏清单 × 高 / 中 / 低三级，先看清哪几条值钱
违约金、账期停在文字描述，没换算成钱	违约金与账期全部换算成金额
审查意见口头传达，出问题时无据可查	每条高风险配改法 + 谈判话术 + 底线建议

可量化的效果

46万 / 120万	167天	3条
首单与年意向金额进入风险测算	瑞泰回款教训固化成检查项	高风险条款 + 谈判要点，销售直接可用

销售 / 售前 | SALES & PRESALES · 投标与资质

资质证书台账与到期预警

技改项目下周招标，要求 ISO9001 + ISO14001 + 安全生产许可证。开始找证书才发现问题。

核心痛点

- 营业执照在保险柜、ISO 电子版在同事两年前邮箱、许可证原件挂在墙上
- ISO14001 翻出来才发现上个月底已过期 22 天

以前 → 用了 WorkBuddy

以前	用了 WorkBuddy
招标公告来了才开始翻抽屉和邮箱	扫描件、手机拍照、邮箱电子版混着扔，台账自己长出来
到期靠人记，换证周期长的证来不及	90/30双提醒线，按换证周期倒推
投标资质够不够用，逐项人工比对	逐项比对本次招标要求，给缺口与补救路径

可量化的效果

28%	22 天	6 栏
资质过期占资格类废标 —— 最高频项有了第一道闸	已过期天数首次被量化暴露	台账含原件位置，从此是必填栏

7 市场与采购：钱花出去了，效果与风险都没人算过

共同点是缺一张能把关键数字放在一起的表 —— 决策才不靠感觉

市场 | MARKETING · 活动与展会

展会 ROI 复盘表

展会结束六周，服务商来问明年早鸟展位预订。老板问「今年这场到底值不值」。

核心痛点

- 手上只有预算表、扫码数、销售一句「有几个客户在跟」，拼不出 ROI
- 行业展会平均投入产出比 1:0.8，多数品牌参展即亏损
- 复盘结论是「起码品牌得到了曝光」—— 无法验证

以前 → 用了 WorkBuddy

以前	用了 WorkBuddy
花费、线索、转化三张表各自躺着，没人合起来看	预算表 + 扫码留资 + 销售跟进状态三源合一
线索质量凭销售一句话定性	成本分项 + 线索成色 + 三口径 ROI + 行业基准对照
明年订不订靠早鸟价压力决定	画出两条参展决策线与回本测算

可量化的效果

18.6万	32万	105条
最大单笔支出第一次有了交代	明年早鸟预订对应两条决策线	线索逐条六周状态回收，争议标注

供应链 / 采购 | SUPPLY CHAIN · 订单执行与跟单

逾期订单风险日报（给老板版）

每天早上采购主管要在管理群报一句「采购都正常」—— 而老板娘要的是扫一眼就知道哪张要出事。

核心痛点

- 23张在途订单里6张已逾期，全在采购脑子里和微信聊天记录里
- 老板娘以为货都好好的，直到产线等料停了半小时
- 供应商的「明天到」是幽灵承诺，跟踪表状态永远滞后

以前 → 用了 WorkBuddy

以前	用了 WorkBuddy
跟踪表状态月更，微信承诺散落不入表	跟踪表 + 微信记录 + 仓库收货三处直接读，状态从对话里提取
报一句「采购都正常」，风险不可见	在途总览 + 风险三单 + 需老板出面栏
哪张会停线靠主管记忆判断	按停线风险排序，附催货话术与承诺记录位

可量化的效果

3张 + 1张	3天	23 → 6
「都正常」变成风险三单 + 一张需您出面	停线风险处理窗口从零天变回三天	在途订单中逾期数被逐日刷新暴露

8 跨境电商：规则、广告、库存、退款相互牵动，单点看指标必错

经营利润异常会诊 — 经营管理

销售额涨9%、利润反降11,000 —— 运营说广告、广告说退款、财务说汇率，开会两小时没结论

以前：三个部门各导自己那块数据，财务花一星期拼成大表，归因靠经验和嗓门

用 WorkBuddy：先校口径能不能比 → 利润桥接分项列金额，合计必须算平

价值：一笔3,200 USD秒杀费被同时计入广告费和促销费 —— 重复计费被当场翻出

广告–库存–利润联动预警 — 广告增长

ROAS 4.2正准备加预算，但可售只够27.8天、海运还要34天 —— 放量等于花钱加速买断货

以前：广告效果好就加预算，库存的事问供应链；交期信采购系统，不和工厂再确认

用 WorkBuddy：库存三态拆分（可售/预留/在途）+ 补货时间线，断货窗口提前算

价值：工厂排产表把断货从账面 7 天修正为 11天；空运/降档/硬断三案各带总账，人拍板

FBA赔偿与异常货件追回 — 商品开发

平台自动赔了31笔，没人核过其中4笔是按￥52估算成本赔的——采购发票写的是￥68

以前：自动赔偿到账就当结束；逐笔核要一整天，排不进日程

用 WorkBuddy：事件–SKU–赔偿三方匹配，按采购发票核对制造成本，少赔差额逐笔算

价值：漏赔7笔1,486 + 少赔4笔728 = 可追回 2,214 USD；时效临期的优先提交

PPC搜索词与真实利润诊断 — 广告增长

ACoS 41%按教科书该停词了 —— 但曝光没降、CTR稳定，是CVR从12.4%跌到8.1%

以前：按 ACoS 排序，高的就否词降价；广告数据只看广告后台，不看成本表

用 WorkBuddy：漏斗定位哪一环坏了，盈亏平衡 ACoS = 34% 是唯一基准

价值：否掉3个高消耗词每周省1,140 USD；而错杀品牌词每月丢2,600 USD自然关联流量

84条FBA库存事件的赔偿核对结果（海拓生活科技）

84 = 31 已赔 + 7 漏赔 + 4 少赔 + 42 无需赔偿（正常流转，算式配平）



已赔付	37%
漏赔 7 笔	8%
少赔 4 笔	5%
无需赔偿	50%

关键洞察

可追回合计 2,214 USD；2025年另有一批 940 USD因超过索赔时效永久归零

9 零售私域：对所有会员一视同仁是最省事的做法，也是最贵的做法

RFM 分层与差异化触达 — 会员运营

60%私域GMV来自20%会员，但这20%是哪20% —— 回答不上来

以前：导3张表做透视 → 手工分层 → 不同人算出的分层还不一样

用 WorkBuddy：只用ID + 日期 + 金额三个字段，映射8–11段人群

价值：冠军段不打折、高价值流失段大额挽回、冬眠段自动化 —— 同样预算打在刀刃上

异常指标归因分析 — 客户策划

试驾预约 320 → 180，环比 −43.8% —— 客户会上要的不是「大概是天气」，是每条归因带多少单

以前：各部门解释自己没问题，归因停在定性描述，补救动作靠猜

用 WorkBuddy：6 因素按可信度降序排列，每条给影响单数与数据支撑，并交叉验证

价值：各因素归因合计 −180 单，略高于实际缺口，说明因素间存在共振与抵扣 —— 这个诚实比凑齐 100% 更可信

试驾预约−43.8%的归因链（按可信度降序 · 每条带影响单数）

6个因素合计解释−180单，与实际−140 单相比略有过归因

竞品官降3万截胡	−58单	82%
公众号打开率崩塌	−42单	68%
信息流素材疲劳	−28单	61%
华东暴雨高温抑制到店	−22单	55%
SDR 离职 SLA 崩了	−18单	47%
新款上市「等等党」	−12单	38%

关键洞察

过归因反而说明因素间有共振：竞品降价放大了 SDR 响应慢的影响 —— 这才是可行动的洞察

沉睡会员识别与唤醒 — 会员运营

微信里躺着买过的人，但谁在沉睡、值不值得投成本挽回，没有分级也没有名单

以前：统一推送统一折扣，预算花在不回头的客户身上

用 WorkBuddy：3级沉睡分层 + 差异化话术 + 可挽回GMV 预估

价值：流失从「掉了GMV才反应」变成滑段时就有信号，预警在流失发生前

周报 / 月报自动化 — 客户策划

给客户 CMO的周报，要的是深度洞察与行动计划，不是4个KPI加一张封面

以前：多源数据人工合并，洞察靠写周报的人当场想，每周重来一遍

用 WorkBuddy：多源清洗→归因洞察→策略建议→客户级 PPT 周报

价值：Agency 员工用同一套逻辑跑出同一份结果 —— 可复现、可审计、可追踪迁移

10 WorkBuddy 构建企业级 Agent 安全体系

WorkBuddy将安全治理贯穿数据流转、权限调用、任务执行与操作审计全过程，支撑 Agent 安全、可控、合规地进入核心业务。

客户端沙箱与防泄漏

本地沙箱隔离执行，敏感信息在企业本地完成识别与脱敏，全程不出网。

- 多平台原生沙箱隔离
- 文件 · 命令 · 网络管控
- 敏感信息本地识别脱敏
- Skill加载前安全检测

AIGC内容安全审核

模型调用前三层防护，生成过程实时审核，违规内容即刻截断输出。

- 提示词注入防御
- 输入 / 输出双向审核
- 流式生成实时截断
- MCP与自定义模型管控

云端沙箱与凭据保护

一任务一沙箱强隔离，最小权限临时令牌，业务密钥不落沙箱内。

- 一任务一沙箱强隔离
- RBAC与临时令牌分离
- SKILL密钥信封加密
- 出站统一网关代理

模型推理与审计留痕

TEE 机密计算隐私推理，全链路加密传输，行为集中留存可追溯。

- TEE机密计算推理
- 全链路TLS加密传输
- 统一审计平台留存
- 高危行为实时告警

个人信息保护影响评估标识

二星级+，全国首个且唯一获评智能体产品。



中国信通院首批安全认证

WorkBuddy首批通过中国信通院的云端智能体基线能力标准评估



更多免费行业报告 并购家 www.ipoipo.net

2.2 CodeBuddy赋能企业AI Coding实践转型

1 AI时代下，企业软件工程的变革

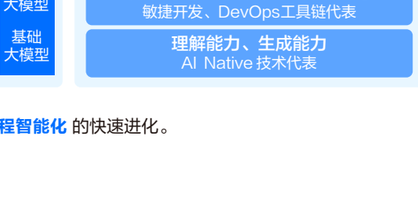
传统软件开发

- 人是核心生产力，人越多，可以支撑的软件项目就越多，产出越多。
- 人才密度大，越复杂的项目，对于人的要求越高。
- 业务的复杂，带来组织的复杂、流程的复杂，最终影响工程的复杂度。
- 基于瀑布、敏捷等研发模式（平台）解决人与人的协作问题。
- 基于平台工程、DevOps平台解决人与工程系统的协作问题。



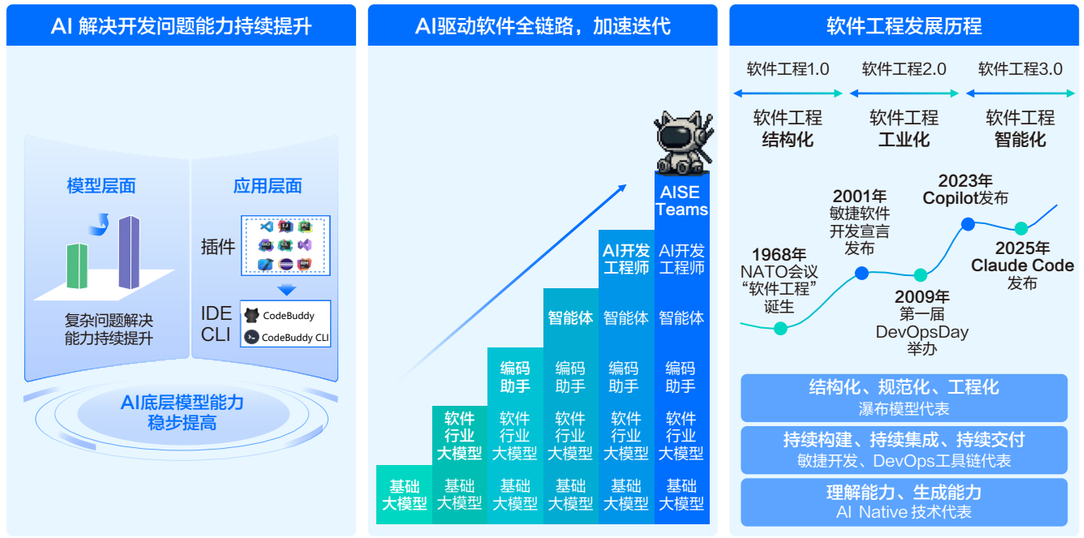
AI Coding模式

- AI是生产力的核心，大部分编程工作从人力向算力迁移。
- 算力可以快速扩容，成本远低于人力。
- 人从直接生产力，变为对于算力的设计、编排与维护。
- 定义研发过程中的各类AI角色，承担各自的责任边界。
- 全新的Multi-Agent协作方式。



2 AI Coding发展趋势 – AI Coding快速发展推动软件工程智能化

随着大语言模型技术的不断发展，AI 解决开发问题的能力持续提升。



软件工程领域正在发生从 **软件工程工业化** 到 **软件工程智能化** 的快速进化。

3 AI Coding实践的三大阶段

AI 辅助开发	AI 驱动开发	AI 原生开发
协作关系	协作关系	协作关系
人主导、AI 辅助	AI 主导、人监督	AI 自主、人验收
阶段特点： AI辅助解决单点问题，开发者仍然是研发流程的控制者与主要生产力	**阶段特点：** AI 开始接管单点环节，可以闭环完成任务，人从生产者逐步变为监督者、规则制定者	**阶段特点：** 通过AI Agent Teams闭环端到端开发过程，人负责验收结果和 遇到问题的修复
研发流程： 对于企业现有研发流程、工具没有冲突性影响，不影响人与人的协作关系	**研发流程：** 可以复用企业大的研发流程，但是企业内部的流程规范、人的知识与技能需要沉淀为 AI 可复用的规则与 Skills	**研发流程：** AI to AI的流程，会衍生出新的 AI 之间协作范式
度量方法： 前期以代码采纳率、代码生成率为代表，后期逐步牵引到 AI 场景覆盖，智能化提升上	**度量方法：** AI 可以独立闭环任务的情况下，带来的是 算力 VS 人力 的价值、效率提升	**度量方法：** 业务价值度量

4 AI辅助：能力建设，场景牵引

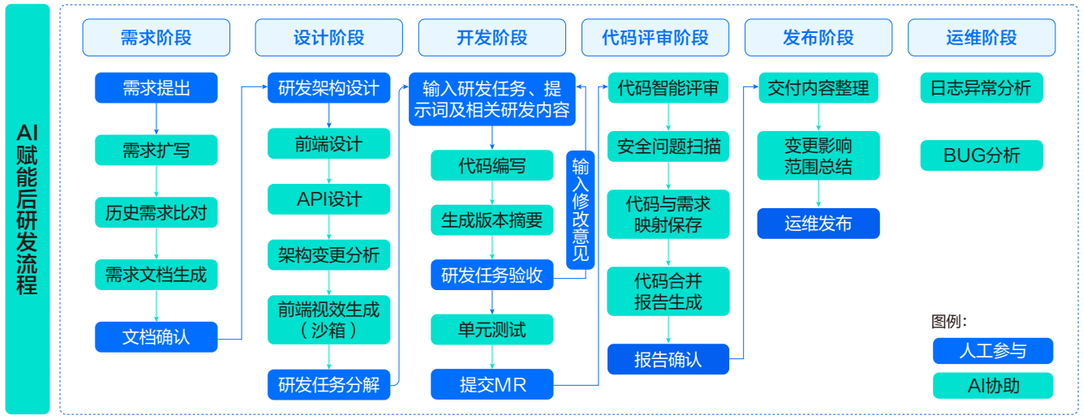


5 AI辅助开发 带来的研发流程变化

研发流程在Agent的加持下，从当前纯人工实施的各节点，转化为由AI完成基础工作、人工审核、修改、验收的模式



6 AI辅助开发过程中的SPEC规约定义

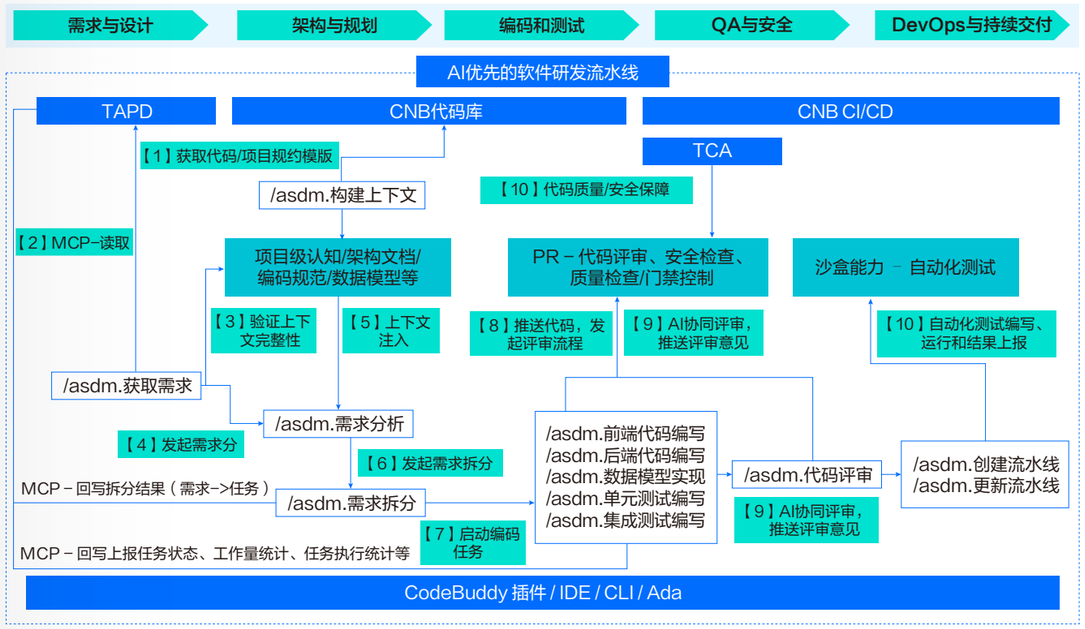


需求规范	架构&设计规范	开发规范	评审规范	交付规范	可观测规范
业务需求模板	前端设计规范	代码风格规范	代码评审规范	资源管理规范	日志管理&查询
需求规格说明	技术选型规范	代码质量规范	安全问题要求	部署&回退规范	监控接入标准
需求描述规范	模块&分层规范	分支管理规范	MR信息规范	业务发布检查规范	调用链接入&查询
历史需求文档/需求树	API 接口规范	版本管理规范	评审报告模板	部署影响分析方法	业务探针接入
需求&代码关系	架构变更说明	单元测试规范	需求&MR映射	部署&验证规范	可观测&代码关联

SPEC是用来指导AI代码生成的一切契约的总和

7 结合腾讯开发者工具，落地AI Coding实践

借助将企业研发策略通过各类规约进行显性化定义与版本化管理，实现传统模式难以企及的研发过程标准化、规范化，以及卓越的质量与安全保障。



备注： asdm是CodeBuddy生态伙伴构建的一套AI规约编码的工具集，可以参考https://asdm.ai/

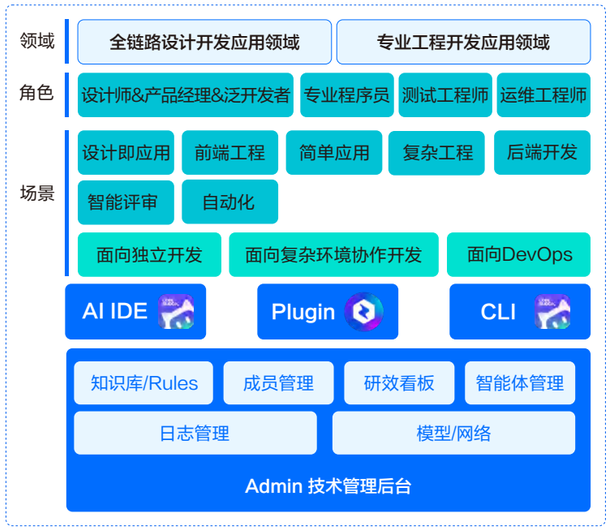
8 CodeBuddy 产品形态

CodeBuddy是AI原生的开发工作台

产品形态上由 Plugin，IDE，CLI 组成，完整覆盖企业开发场景，并为技术管理者提供统一的管理后台

AI Plugin 智能副驾	AI IDE 一站式AI工作台	AI CLI Agent操作系统
• 集成至主流/用户习惯IDE，与IDE为同产品 • 团队内快速落地AI Coding	• “集成开发环境”转为“智能开发环境” • 产设研一体，软件工程全生命周期覆盖	• Unix类命令行交互 • 批处理，多系统异步协作，7×24运行 • 深度融入司内工具链/云+DevOps

Plugin，IDE，CLI 三种前端产品形态



统一接入Admin技术管理后台

更多免费行业报告 并购家 www.ipoipo.net

9 CodeBuddy产品核心功能架构



2.3 腾讯云游戏研发AI解决方案

1 业务场景

游戏行业正往精品化方向发展，玩家对玩法创新与内容质量的要求持续提高，产品能否在核心玩法与内容体验上持续打磨，直接决定其留存与口碑表现。行业研发范式正从“长周期立项、重资源打磨”转向“小步快跑、敏捷验证”：先低成本验证核心玩法，再投入正式研发与运营。该方案适用于需要快速验证核心玩法、持续迭代内容的游戏项目，覆盖从早期立项到正式研发的全过程。

2 业务挑战

- **玩法验证周期长：**从策划文档到可玩 Demo 需多方协同，周期常以周或月计，早期方向风险难评估。
- **跨职能协作成本高：**策划、美术、技术之间存在信息壁垒，需求传递过程中易失真，反复对齐消耗大量成本。
- **引擎内开发门槛高：**进入商业引擎正式开发时需重写大量代码、重建资产管线，技术债与人力成本高昂。
- **快速迭代压力大：**上线后玩法与内容需高频更新，传统开发与美术产能难以匹配玩家消耗速度。

3 解决方案

➤ 阶段 1：策划设计（WorkBuddy）

把创意转成可执行的设计文档。传统方式下，玩法框架依赖策划经验、数值依赖主观判断、规则依赖口头传达，误差在传递链路中逐级放大。WorkBuddy通过策划、数值、架构、QA多专家协作，结合封装好的Skill，产出策划框架、数值模型、关卡定义与里程碑规划。原本需要1-2天完成的策划案，通常可压缩至半天左右，作为下游各阶段的统一输入基准，降低跨职能对齐成本。

➤ 阶段 2：原型验证（WorkBuddy）

验证核心玩法是否成立，同时控制验证成本。传统路径需先搭建商业引擎环境，这一步周期长、投入大；若最终验证结果不理想，前期在环境搭建上的时间与人力投入即被浪费，试错成本高。WorkBuddy的软件开发团队专家将策划文档、任务拆解、代码开发、自动化测试整合至同一对话空间中闭环处理，跳过重型引擎环境，直接基于Web技术栈生成可运行Demo，代码完成即可进行验证。周期由数周压缩至天级，以更低的试错成本换取更快的决策速度。

➤ 阶段 3：开发联调（CodeBuddy）

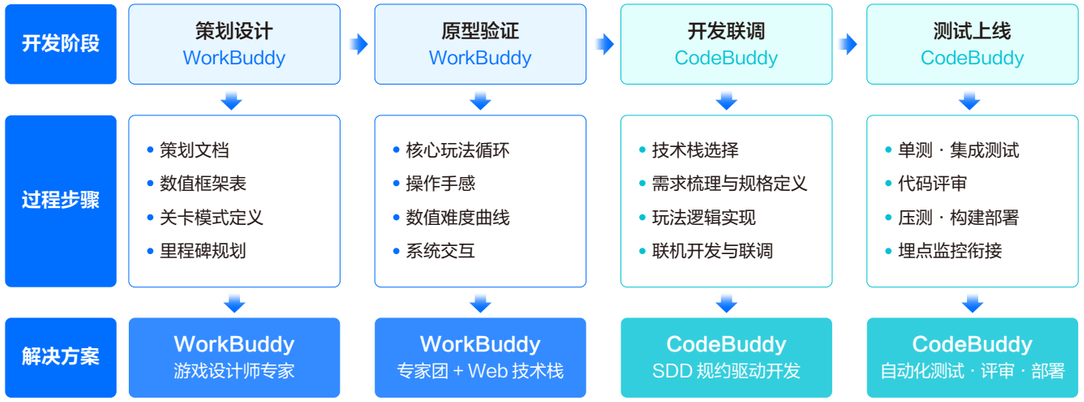
实现全部游戏玩法和业务逻辑。传统通过提示词描述需求、由AI自主生成代码的“vibe coding”方式，面对简单任务效率较高，但应用于游戏引擎级别的复杂工程时容易失控：产出不稳定、难以验证，代码易演变为难以追溯的黑箱。CodeBuddy采用SDD（规约驱动开发）应对这一问题，先明确规格再生成代码：先把技术栈选型、需求定义写成清晰规约，再据此规划实现方案、由AI生成代码，最后验证产出是否符合规约，形成规约→规划→实现→验证的四阶段迭代循环，使AI在明确边界内稳定产出可交付代码。

➤ 阶段 4：测试上线（CodeBuddy）

完成质量保障，衔接上线运营。传统方式中，“发现问题→定位→修复→重新验证”这一循环最为耗时：人工测试覆盖有限，部分问题需等到压力测试甚至上线后才能暴露；代码评审依赖人工审查，项目规模扩大后容易出现遗漏；部署环境配置不一致，也常导致本地运行正常、线上出现异常的情况。CodeBuddy将单元测试、代码评审、压力测试、构建部署、埋点监控整合为自动化闭环，未通过测试将自动定位问题并修复后重新验证，通过后直接衔接埋点监控与运营。周期由数周压缩至数天，代码质量始终可追溯、可验证。

4 架构图

AI游戏开发解决方案总览



5 相关产品

WorkBuddy　CodeBuddy

2.4 腾讯云游戏精细化运营AI Agent解决方案

1 业务场景

游戏行业已从“流量红利”转向“存量经营”——玩家获取成本攀升、产品长线运营周期拉长，谁能把每一位玩家在全生命周期内的体验、付费与活跃经营到极致，谁就能在存量竞争中胜出。而“精细化运营”的落地高度依赖数据洞察与高频决策，传统依赖人工规则与经验的运营模式，在玩家规模、行为复杂度与响应实时性上已逼近天花板。

AI Agent（智能体）凭借“感知—决策—执行”的自主闭环能力，正在重塑游戏运营的各个环节：从新玩家激活、活跃留存、流失预警与召回、付费转化与 LTV 提升，到个性化内容推荐，以及数据驱动的实时经营分析与归因诊断。以下场景是游戏精细化运营中与 Data+AI 最相关的主战场。

2 业务挑战

- **全域数据孤岛，统一玩家画像缺失** —— 玩家行为、交易、客服、日志分散在游戏服、BI、CRM、工单等多套系统，缺乏统一身份与实时画像，Agent 没有可靠的“感知”基础，无法做出准确决策。
- **指标口径不一，AI 分析结果不可信** ——同一指标（如“日活”“付费率”）在多团队、多系统下定义各异，语义口径缺乏统一管理。大模型直接面向原始表做问数与归因时，容易产生条件错配、指标歧义、JOIN 错乱等幻觉错误——运营决策建立在错误数据上，比没有数据更危险。
- **实时特征获取难、“计算之墙”高** —— 精细化运营依赖实时信号（如当前会话行为、近期流失倾向），但海量行为日志的实时采集、计算与特征工程成本高、门槛高，Agent 难以及时拿到决策所需特征。
- **经验主义决策、缺乏预测能力** —— 运营策略依赖人工经验与事后复盘，缺少事前 / 事中预测（流失预测、LTV 预测），错失最佳干预窗口。
- **千人千面运营规模化难** —— 规则运营颗粒度粗，难以在千万级玩家中规模化实现个性化推送、推荐与干预，运营人力成为瓶颈。
- **运营动作响应慢、跨系统执行断点** —— 即便发现了问题，从“分析”到“执行干预”（推送 / 发券 / 拉群）仍需人工在多系统间跳转，链路长、易遗漏，缺乏自动闭环。
- **企业自建 AI Agent 四大陷阱** —— ① 部署复杂（需串联模型 / 知识库 / 工具 / 工作流）；② 成本黑洞（模型调用与算力不可控）；③ 安全隐患（数据泄露 / 越权）；④ 管理缺失（缺乏统一观测与治理）。

3 解决方案

➤ Data+AI 实时数据底座

- **打破孤岛、统一画像：**依托**数据开发治理平台 WeData**，统一接入并建模游戏行为、交易等多源数据，完成标准化画像与数据治理，解决“全域数据孤岛”的挑战。
- **实时计算、化解墙障：**基于**流计算 Oceanus（Flink）**实时清洗高频日志、抽取动态特征，并动态写入 **TCHouse-D 高性能实时数仓**，支撑秒级、高并发的特征检索，跨越“计算之墙”。
- **预测驱动、告别经验：**利用**数据湖计算 DLC** 的联邦查询与分析挖掘能力，结合 WeData 调度流，定期/实时运行流失预测与 LTV 变现计算，将预测结果回写至 TCHouse-D 特征库，为 Agent 注入“预测”维度的感知。

➤ DataBuddy：AI 原生的大数据智能体工作台

围绕精细化运营“用数—建数—治数”三大环节，DataBuddy 提供三个各司其职、协同作业的数据 Agent：

- **智能分析 Agent**

  智能问数：运营与业务人员通过自然语言直接查数，系统自动理解指标与维度，返回数据结果及 AI 解读。

  异动归因：自动监控发现指标异常波动，AI 自主规划多维度拆解、关联因素定位与原因总结，输出可执行建议。

  分析报告与 AI 仪表盘：对话生成完整分析报告，自然语言一键生成看板，支持千人千面策略的分层效果。

更多免费行业报告 并购家 www.ipoipo.net

- **数据工程 Agent**

  对话式链路搭建：通过对话完成多源数据接入、数仓分层方案设计、自然语言生成 SQL、任务依赖识别与工作流 DAG 编排、一键发布调度。

  智能运维：自动巡检任务运行状态，异常智能告警；AI 诊断失败原因，保障流失预警、实时特征等关键链路不掉线。

- **数据治理 Agent**

  语义层唯一口径真相源：每个指标只有一个定义，语义层与治理引擎深度融合、由治理 Agent 持续保鲜，从机制上抑制 NL2SQL 幻觉。

  巡检自愈：治理 Agent 自主巡检元数据、语义一致性、数据质量、安全合规、成本五大维度。

4 架构图



5 相关产品

数据底座

流计算 Oceanus、数据开发治理平台 WeData、数据湖计算 DLC、高性能实时数仓 TCHouse-D

AI 原生大数据智能体工作台

DataBuddy（数据分析 / 数据工程 / 数据治理三场景一体）

2.5 腾讯云AI驱动SRE智能故障定位解决方案

1 业务场景

本方案基于腾讯AI办公助手WorkBuddy与CloudQ云资源纳管能力，将资深SRE的排障方法论沉淀为可协作的智能体专家集群，聚焦“故障定位”核心场景，以“对话即服务”的方式，让业务、研发、运维乃至管理层都能用自然语言驱动分钟级的全链路诊断与远程闭环处置。

2 业务挑战

- **排障链路长、强依赖专家经验：**分布式架构下故障传播路径复杂，从公网接入到业务代码需逐层定界，高度依赖少数资深工程师的经验直觉，知识难以复制、难以规模化，人员流动即能力流失。
- **复杂网络与跨云资源定界难：**公网抖动、运营商路由、CLB健康检查、安全组策略、后端端口连通性等任一环节异常都会导致业务不可用，故障跨越多个云产品，人工逐项核查效率极低、极易遗漏。
- **资源“看似正常”却卡顿——微架构盲区：**常规监控只覆盖利用率 / 饱和度等资源维度，无法洞察 CPU 内部的指令执行效率。当 IPC 偏低、缓存命中率下降或分支预测失败率升高时，资源大盘正常，业务却已劣化，故障长期无法定位。
- **多种异构数据库，日常管理难度大：**专业数据库运维智能体，立于安全可控、专家技能、进化记忆三大支柱：四级分权高危永拦、十万级工单沉淀能力、三层记忆边干边学。
- **告警噪声大、根因关联弱、MTTR高：**海量告警缺乏根因关联，运维团队疲于被动救火，难以从“告警风暴”中快速锁定真正的故障根因并完成处置验证闭环。

3 解决方案

方案以「双引擎诊断 + 五层白盒定界 + 全链路数据底座 + 远程闭环处置」为核心，构建覆盖“发现—定位—处置—验证”全流程的智能故障定位专家团，将资深 SRE 的排障链路自动化、标准化、智能化。

- **双引擎智能诊断**

先以 USE 资源画像对 CPU、内存、磁盘、网络做广度扫描，快速回答“资源够不够用、有没有排队、有没有报错”；再以 PMC 微架构穿透，洞察 CPU 指令执行效率，定位“资源看似正常、体验却在劣化”的隐蔽瓶颈。两套方法论串联，先广度、再深度，避免误判与漏判。

- **五层白盒诊断流水线**

将资深 SRE 的排障链路抽象为自顶向下的五层流水线：公网可达性 → 基础设施（主机、数据库）→ 接入链路 → 业务日志与链路 → 专家研判，从公网入口一路穿透到业务代码，分钟级锁定故障所在层，杜绝盲目排查。

- **全链路可观测数据底座**

统一纳管云拨测、云监控、CLB、CLS、主机性能、DatabaseClaw 及 Prometheus 业务指标等多源数据，为故障定位提供交叉验证支撑，让每一个根因结论都有数据背书。

- **远程一键闭环处置**

基于自动化助手无需登录服务器即可远程执行处置动作，并在处置后自动复测关键指标、确认业务恢复，形成“定位—处置—验证”闭环。所有诊断操作遵循“只读铁律”，处置动作需显式确认，保障生产安全。

- **多智能体编排与经验沉淀**

WorkBuddy 负责意图理解与多智能体协作调度，CloudQ 负责云资源统一纳管；诊断主理人调度多位专科分析智能体并行作业。如：主机性能专业分析，DatabaseClaw 慢查询优化，排障经验持续沉淀，实现故障定位能力可复制、可传承、可规模化。

4 架构图

下图为 AI 驱动的 SRE 智能故障定位整体架构：自然语言入口→WorkBuddy + CloudQ 编排层→双引擎诊断（USE 资源画像 / PMC 微架构）→五层白盒诊断流水线→全链路数据源→根因定位与 TAT 远程闭环处置。

AI驱动的SRE智能运维专家图 · 全链路故障定位架构

自然语言驱动 · 双引擎诊断 · 五层白盒定界 · 远程闭环处置

用户入口：业务 / 研发 / 运维 | 管理层 | 对话即服务 | 自然语言提问

编排层：WorkBuddy + CloudQ
意图理解 · 任务拆解 · 多智能体协作 · 只读安全护栏
记忆沉淀 | 专家知识库 | TAT 远程执行

智能体专家团

网络诊断专家	主机性能分析专家	数据库智能体 DatabaseClaw	业务链路专家	根因研判专家
公网可达 · 路由链路	USE 资源画像+PMC 微架构双引擎穿透	慢查询优化 · 参数调优	日志检索 · Trace 归因	交叉验证 · 优化建议

五层白盒诊断流水线（分钟级逐层定界）

L1 公网可达	L2 基础设施	L3 接入链路	L4 业务日志	L5 专家研判
CAT 云拨测 Ping / MTR / DNS	主机 CVM · 数据库 DatabaseClaw CDB / Redis · 云监控指标	CLB后端健康 端口连通性	CLS 日志检索 Trace 链路	根因归因 优化建议

数据源：CAT拨测 | 云监控Monitor | CLB健康检查 | CLS日志 | 数据库 | Prometheus业务指标 | 慢查询/连接数

定位与处置（持续闭环）

根因定位	远程处置	闭环验证	报告沉淀
网络 / 资源 / 服务 / 业务四类异常	TAT一键重启 / 限流 / 扩容	复测指标→确认恢复	MTTR小时级→分钟级

分钟级	5层	2大引擎	一键处置
故障定位耗时	白盒诊断纵深	资源 + 微架构	远程无需登录

WorkBuddy x CloudQ x DatabaseClaw | 让运维经验可复制、可传承、可规模化

5 方案价值

MTTR小时级→分钟级
自动化逐层定界替代人工串行排查，大幅压缩故障恢复时间。

穿透微架构盲区
双引擎补齐常规监控无法覆盖的CPU指令级性能视角，隐性劣化无所遁形。

根因可解释、有据可依
多源数据交叉验证，每层结论都有指标与日志背书，杜绝拍脑袋定位。

经验可复制、可规模化
资深SRE方法论沉淀为智能体集群，降低对个人经验的依赖，人员流动不再等于能力流失。

对话即服务、全员可用
自然语言驱动，业务、研发、运维、管理层均可低门槛发起故障定位。

6 相关产品

WorkBuddy | CloudQ | 云拨测 CAT | CLS | DatabaseClaw | 可观测平台 TCOP

2.6 腾讯云AI Agent数据库智能运维方案

1 DatabaseClaw · 腾讯云数据库 AI Agent

让AI像资深DBA一样管理和运维你的数据库 · 把 DBA 十 余年经验Skill化

10万级	1600+	14款	3类
真实工单经验沉淀	OpenAPI 接口可调用	覆盖腾讯云数据库全线	Skill 供给 · 官方 / SkillHub / 自建

DatabaseClaw：能诊断、敢执行、会学习 —— 你的专属AI数据库专家

产品定位
生产级 AI Agent 平台
对话式运维 | Skill套件

产品边界
聚焦生产可控的数据库 Agent 能力
X LLM套壳 | X 白屏工具 | X 通用助手

Safety
安全可控·分级防护

建立严格的L1-L4分级防护体系，确保AI行为在生产环境中完全可控、可追责，守住数据与操作的安全底线。VPC不出域 + CAM权限一致 + 全链路审计。

分级防护体系

L4	高危变更 · 永久 Deny
L3	变更操作 · 需二次确认
L2	分析诊断 · 自动执行
L1	只读操作 · 自动执行

Skill
专家技能·海量沉淀

把故障诊断、性能优化、日常运维、变更审核抽象为可调用、可组合、可复用的Skill；一个Agent管14款数据库全系，接入1600+OpenAPI，覆盖实例管理到诊断分析全链路。

核心场景
故障诊断 | 性能调优 | 架构评估 | 容量规划 | 变更审核 | 主动巡检

支持引擎（14 款云数据库全系）
MySQL | Redis | MongoDB | PostgreSQL | TDSQL-C | +More

Memory
进化记忆·边干边学

具备实例级 / 客户级 / 决策链三层记忆架构，构建专属知识库，记住历史问题、处理方式和业务特征，让每次诊断都比上一次更精准，实现 “边干边学” 的持续进化闭环。

三层记忆架构

决策链记忆	逻辑复用
客户级记忆	业务偏好
实例级记忆	运行特征

更多免费行业报告 并购家 www.ipoipo.net

2 产品架构介绍

DatabaseClaw · AI统一入口 · 推理引擎　DBbrain · 性能诊断 · 数据资产底座　DMC · 安全访问 · 变更管控中枢

交互层

DatabaseClaw 统一智能入口
对话交互 | IM 渠道接入（企微 / 飞书 / 钉钉） | API接口 | Agent skill 化接入
主动式服务 · 跨产品数据汇总 · 专家诊断 SOP

DBbrain 工位（保留）
大盘 · 专家经验 Top-Down

DMC 工位（保留）
高频实时交互编辑器

能力层

诊断分析能力
性能分析 | 容量评估 | 慢SQL诊断 | 锁分析

数据操作能力
SQL执行 | 元数据管理 | 导入导出

变更管控能力
工单流程 | SQL审核 | 无锁变更

推理分析层
专家经验 | 记忆能力 | 评测机制

基础设施层

DBbrain 数据库智能管家·实时异常检测、问题数据模板化

数据采集管道	实时计算引擎	数据预聚合	性能监控座舱
指标监控数据	指标异常检测	慢SQL模版	全局实例大盘
PFS内核观测	执行计划分析	指标时序聚合模版	DB Load 视图
慢日志	容量趋势预测	锁等待事件模版	审计聚合视图
审计日志	慢SQL根因分析	审计日志模版	
全量数据基石	实时异常判断	结构化压缩数据	可视化呈现

DMC 数据库管理

审批与工单引擎
多步骤工单流程 | SQL语法审核 | 无锁变更执行
合规变更保障

数据源统一管理
多类型数据源接入 | 连接信息规范管理 | 数据源健康检测
统一接入基础

凭证与权限管理
凭证集中托管 | 权限管控对齐 | 操作最小化授权
写操作安全基础

3 智能运行的起点：让 Agent 安全接管数据库任务

VPC部署，数据不出域　架构安全
Agent 实例运行在客户专属 VPC 内

客户专属 VPC·内网直连
TKE容器集群 Agent 运行环境 → 客户数据库 业务数据源
仅传输脱敏推理请求·业务数据零外发

容器化强隔离
Agent 运行环境独立可用

大模型与VPC仅交互元数据 / 脱敏请求
业务数据不离开VPC边界

支持金融、政务等高合规私有化部署
满足专有云、行业云场景

四级操作分权　AI 行为护栏
权限随风险递增，审批随级别收紧

L1	只读	SELECT / SHOW / EXPLAIN　自动执行
L2	分析	诊断 / 巡检 / 问答　自动执行
L3	变更	DML / Kill Session / 参数更新　需二次确认
L4	高危变更	DDL / TRUNCATE / DROP　永久Deny

权限隔离 + 安全执行　合规审计
对齐 CAM 和数据库访问体系，权限严格一致

Agent 实例 → 用户云上 CAM角色 → API 权限 / 资源权限

Agent 角色扮演
通过用户角色扮演获得所有权限和资源视角

CAM 权限一致性
实例权限与子账号 CAM 1:1 映射，AI 不超出用户授权

全链路操作留痕
完整记录 Agent 每一次决策与执行轨迹，可审计可追溯

Agent 实例部署在用户 VPC 内，通过用户角色扮演获得权限，AI 通过 Soul 进行行为分析，最终执行受云权限体系控制
- AI 可操作范围受云 API 权限和资源权限限制，默认只读（管控平台操作限制）
- AI 访问数据 走云上 DMC 安全访问策略（数据库连接访问）

4 智能运行的关键：用 Skill 封装数据库能力

10万级	1600+	14款	持续进化
真实工单经验沉淀	OpenAPI 接口可调用	覆盖腾讯云数据库全线产品	Skill与线上工单同源迭代

Skill 是 Agent 的能力单元，将复杂的数据库运维能力封装为可调用、可组合、可复用的模块
——十万级真实工单经验产品化，让每个用户都能获得专家级的数据库运维能力。

场景矩阵　应用层

事前预防	事中应急		日常运维
容量突增预测	综合性能诊断	负载异常分析	告警主动分析
智能健康巡检	慢查询分析	主从复制延迟分析	备份策略分析
安全风险扫描	HA 切换分析	内存异常 / OOM 分析	DTS 迁移全流程助手
版本变更性能比对	大 Key · 热 Key 分析	DTS 延迟 / 错误诊断	DTS 数据质量保障

技术底座支撑
DBbrain 数据基座 + 云数据库基础设施

DBbrain 数据基座
PFS性能洞察 · 审计日志SQL分析 · 实时异常检测

云数据库基础设施
统一管控 · 集中日志 · 基础监控告警 · 高可用架构 · 备份恢复

统一能力接口层
1600+ OpenAPI可调用 · 覆盖14款云数据库产品 · 标准化编排

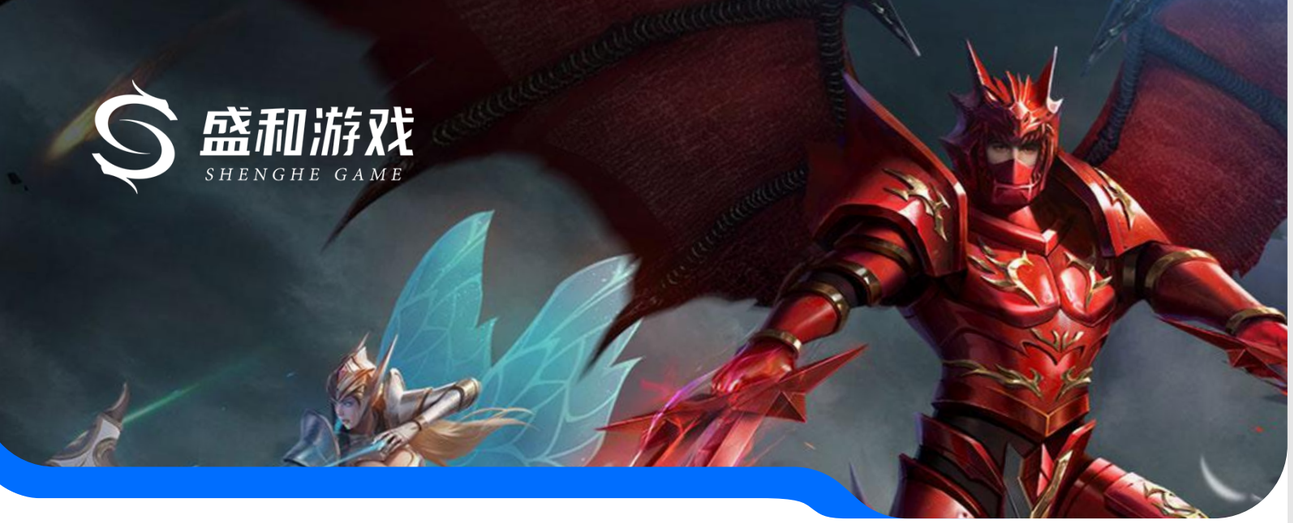
经验底座支撑
腾讯云 DBA 团队

13年
海量运维经验沉淀

十万级工单沉淀
Skill 与线上工单诊断同源迭代，持续进化

专家级知识图谱
DBA 判断路径 / 处置动作结构化归档

最佳实践模板库
覆盖典型故障、性能优化、变更审核场景



标杆案例

盛和

「传奇品类头部研发商的 AI 跃迁：代码到运营全链路提效，项目节奏与营收增长同步跑出加速度。」

业务场景

盛和游戏（浙江盛和网络科技有限公司）是A股上市公司恺英网络旗下核心研发子公司，国内传奇品类赛道头部研发商。代表作《蓝月传奇》《王者传奇》《原始传奇》等多款传奇产品长线运营表现稳健，旗下996传奇盒子是国内传奇品类领先的垂直用户平台，积累了规模可观的注册用户。2026年初，盛和接入腾讯云AI智能体矩阵，以CodeBuddy加速研发、ClawPro智能化运营，构建覆盖游戏链路的AI工作流。

业务挑战

▶ 高频版本迭代下的研发提效

传奇品类版本更新节奏快，版本窗口期通常仅2–3周。引擎升级涉及大量API迁移与文件适配，手工处理周期长，容易挤占本就紧张的窗口期；功能模块开发、代码审查与测试编写均以人工为主，Code Review覆盖率有待提升，核心模块测试覆盖率偏低，代码质量一致性难以保证，线上缺陷率与发版延误风险随之上升。

▶ 运营监控依赖人工巡检，数据分析人力投入大

996盒子每日需关注DAU、流水、交易成功率等数十个核心指标，人工巡检难以做到全天候覆盖，异常发现与响应存在延迟；运营日报、活动复盘等数据分析工作每天都需要固定人力投入，一次活动复盘耗时数小时，运营团队难以把精力集中在策略分析与增长决策上。

解决方案

- **CodeBuddy 贯穿研发全流程：**开发者以自然语言描述需求，CodeBuddy生成符合项目规范的代码；引擎升级场景下，CodeBuddy批量扫描代码库，自动定位并修复API迁移与导入错误；在代码提交环节自动执行Code Review，并为核心模块自动生成测试用例。代码生成、批量修复、自动审查、智能测试四项能力嵌入日常开发流程，覆盖从需求到提测的完整链路。
- **ClawPro 运营监控与数据分析智能化：**企业级智能体ClawPro（「龙虾」）可快速完成全员部署上线；接入996盒子核心运营数据，7×24小时自动巡检DAU、流水、交易成功率等核心指标，识别异常后第一时间推送IM告警（可对接企业微信、飞书、钉钉等主流办公协同工具）；自动生成运营日报、辅助活动复盘，并支持运营人员以自然语言实时查询运营数据。

架构图

盛和游戏×腾讯云 AI 智能体矩阵 业务架构

业务场景

研发场景	运营场景
代码生成・引擎升级・代码审查・测试生成・批量修复	数据巡检・异常告警・日报・活动复盘・实时查数

AI 智能体矩阵

CodeBuddy	ClawPro
代码生成・批量修复・智能审查・测试自动化	龙虾 7×24 巡检・IM 集成・弹性扩容・故障自愈

平台底座

CloudAgent 统一内核，云端沙箱，权限管控	企业 IM 集成（企业微信 / 飞书 / 钉钉），安全审计

统一交互入口：企业 IM 群聊（企业微信 / 飞书 / 钉钉等）+IDE / CLI / Web

客户价值

研发能力转化为交付加速度

在CodeBuddy加持下，功能模块开发耗时明显缩短，AI生成代码采纳率保持在较高水平；引擎升级等批量修复任务效率大幅提升；Code Review覆盖率大幅提升、基本实现全覆盖，核心模块测试覆盖率明显提升，线上缺陷率有所下降。项目在2–3周版本窗口期内稳定交付，显著减少因引擎升级、代码质量等环节导致的发版延误风险。

运营从"人盯盘"升级为"AI 盯盘"

996盒子核心运营指标实现7×24小时自动巡检与快速异常告警，异常事件自动识别与拦截效果良好；运营日报编写全自动化，活动复盘耗时从数小时大幅压缩，数据查询支持实时自然语言交互；智能体服务保持稳定可用。运营人员从机械重复的盯盘与报表工作中释放，聚焦策略分析与增长决策。

相关产品

CodeBuddy　ClawPro

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

淘米

「二十年的老IP们集体“上AI”，八成员工人手一个Agent，做素材、改版本不再排队等人，内容产能直接翻到近两倍。」

业务场景

淘米网络旗下拥有《摩尔庄园》《赛尔号》《小花仙》等多个国民级IP，并行长线运营近二十年。每个IP都有独立的世界观、美术风格与玩家社群。为破解“多IP并行、产能跟不上内容消耗”的核心矛盾，淘米引入腾讯云WorkBuddy与CodeBuddy，让Agent进入游戏制作的研发、策划、美术、运营、市场全链路，把过去“串行等待”的接力跑改造为“并行协同”的同步交付模式。



业务挑战

▶ **研发被重复劳动拖累**

UI适配、接口联调、配置表解析、数据埋点等代码必须写，却不直接产生玩法价值，挤占核心研发精力；同时多IP代码仓库交叉并行，带来权限与安全管控隐患。

▶ **内容生产串行接力**

策划出方案、美术画素材、IP衍生内容迭代三件事一棒接一棒地交，版本更新常卡在“等素材”，直接拖累玩家留存曲线。

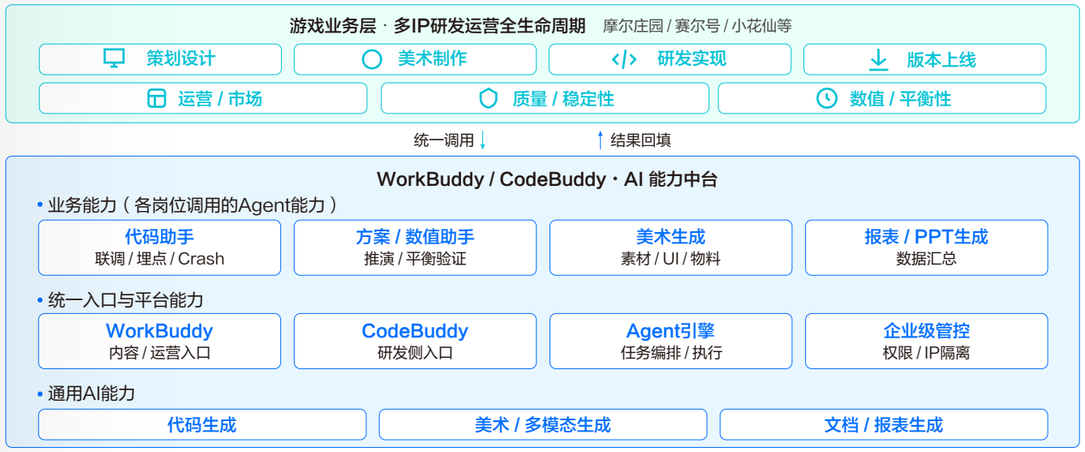
▶ **多IP抢占有限产能**

运营多个国民级IP时，有限的策划、美术与研发产能要在各IP之间分配，团队不得不权衡资源投向，难以让所有IP的节日活动与版本大更新齐头并进。

解决方案

- **研发流程全面提效：**CodeBuddy承接重复性编码，工程师转而聚焦核心玩法设计、性能优化与Crash根因定位；企业级安全管控让各IP代码仓库权限严格隔离，多条线并行不再是安全隐患。
- **工作流并行提速：**WorkBuddy渗透内容团队，策划用它快速搭建版本方案框架、美术用它加速素材产出与风格迭代，把接力跑改造成同步推进的并行协同。
- **全岗位协同同步：**策划做数值推演与平衡性验证、美术批量产出UI与宣传物料、运营市场一键生成数据报表与汇报PPT，让所有IP、所有环节同步启动、同步交付。

架构图



客户价值

版本迭代不再受限于“谁最慢”

所有环节同步启动、同步交付，多IP之间也不必再做资源取舍，能同时推进。

研发与创意精力回归核心

把重复劳动交给Agent，工程师、策划、美术得以把时间投入真正产生玩法与体验价值的工作。

沉淀二十年IP价值

把长期积累的IP资产转化为可持续、可规模化的内容产出能力。

业务效果

01 内容生产效率提升近80%

美术素材与IP衍生内容产出效率显著提升，同时版本方案产出周期大幅缩短。

02 WorkBuddy 覆盖约80%员工

深度融入游戏制作全流程，是从上到下的日常工作方式，而非创新小组的孤立实验。

03 多IP并行能力增强

团队可同时高质量推进多个IP的节日活动与版本大更新，不再因产能瓶颈相互挤占。

相关产品

CVM　CDB　WorkBuddy　CodeBuddy



标杆案例

游族

「出海老兵让 AI 进了每个工位，资讯、买量素材、数据复盘一条龙打通，海外收入和利润率双双往上走。」

业务场景

游族网络是国内最早一批出海的游戏公司，海外营收占比已连续11年超越国内。全球多区域发行让素材资产、工程开发、本地化与素材迭代等链路成倍拉长，成本与管理复杂度随之放大——“效率提不上来，出海优势就难以建立”。游族的选择是让Agent进入每一个岗位：算力公司买单、全员免费无限额使用、接入47个模型，并以腾讯AI办公助手WorkBuddy作为图形化企业级入口，将数据中台、资讯系统、广告素材链路封装成Skill供业务一键调用，完成资讯、素材、数据复盘三大业务场景的全覆盖。

业务挑战

▶ **全员铺开难管控**

个人用Agent看的是功能，企业用还要看权限管理、行为审计、数据隔离；且大量非技术岗员工也要能用，使用门槛不能太高。

▶ **资讯素材链路慢**

行业资讯靠人工盯信源、摘录写周报，易遗漏耗时；广告素材从创意→脚本→分镜→建模→后期线性串行，任一节点慢则全链路等待，而素材正是全球化获客入口。

▶ **数据库运维压力大**

实例规模与架构复杂度持续攀升，日常巡检、慢SQL排查等重复劳动繁重，故障定位与容量规划高度依赖资深DBA。

▶ **数据慢经验难复用**

业务看自定义指标得等报表模板、找数分排期，决策窗口被拖慢；定制化能力散落在中台、业务团队与个人经验里，跨部门才能实现。

解决方案

- **企业级入口托底：**WorkBuddy作为企业级入口，技术岗自选CLI/IDE、非技术岗用界面发起任务，统一管权限、安全与管理（门槛低、跑得稳、管得住）；算力公司买单、全员无限额、接入47个模型把Agent“养起来”。
- **双链路AI提速：**资讯场景Agent自动抓取汇总、大模型做情感判断后结构化输出并实时推送全员；素材场景AI覆盖从创意到分镜的文本环节，叠加图片/视频/BGM/配音能力压缩“创意到初稿”周期，最终投放仍由团队判断。
- **能力沉淀成工具：**自然语言出报表，并把数据中台能力打包成Skill供业务在自有Agent内自助查数归因；再用WorkBuddy“专家团”组合多方向专家协同，把隐性经验写进工作流。
- **标准化Skill接入DatabaseClaw：**依托SkillHub标准化Agent Skill生态加载DatabaseClaw官方Skill包，用户在WorkBuddy内即可无缝调用故障诊断、性能分析、安全巡检、容量评估等数据库智能体能力，零开发、即插即用。

架构图

出海行业业务层・全球多区域发行 本地化 / 多区域投放 / 跨时区运营

广告素材 / 买量 | 行业资讯/舆情 | 工程研发 | 运营数据复盘/经营分析 | 本地化&多区域发行运营

统一调用 ↓ ↑ 结果回填

WorkBuddy・AI 能力中台

- 业务能力（各业务调用的Agent）
  - 资讯情报Agent：抓取/判断/推送
  - 广告素材Agent：创意→分镜+多模态
  - 数据复盘 / ChatBI：查数/定位/归因
- 业务能力沉淀（Skill / 专家团）
  - 数据中台Skill
  - 资讯 / 素材链路Skill
  - 专家团（多专家协同）
- 统一入口・平台能力・管控
  - WorkBuddy入口：图形化 / CLI / IDE
  - Agent引擎：上下文 / 记忆 / 编排
  - 企业级管控：权限 / 审计 / 隔离

客户价值

技术能力变成业务可直接调用的工具

定制化需求不必再跨部门提需求、等排期、等开发，业务在自己的Agent里即可完成。

隐性经验沉淀进工作流

行业情报的判断规则、素材投放的区域品类经验、数据复盘的查询归因能力被封装复用，Agent越用越好。

放大全球化长链路的系统性价值

在多语种本地化、多区域投放、跨时区运营的复杂场景下，用Agent把组织能力转化为可持续的出海竞争优势。

专家级DBA能力普惠

腾讯云十万级DBA工单经验沉淀的专家能力开箱即用，对话即可完成数据库诊断与巡检，降低对资深人员依赖，普通员工也具备专家级分析能力。

业务效果

01 Agent渗透率超75%

覆盖资讯、广告素材、运营数据复盘三大高频场景，成为全员日常工作方式。

02 海外收入与利润率双增

2025年海外收入同比增长8.06%，财报将部分增长归因于AI与广告业务深度融合，推动主营业务毛利率提升2.46%。

03 能力沉淀持续推进

已完成4个系统的Agent改造、另有七八个进行中并计划年内全部完成；专家团已能将创意脚本约20分钟产出完整视频成片。

04 数据库运维智能化

DatabaseClaw沉淀故障诊断、健康巡检、容量评估等 30+专业场景，在 WorkBuddy对话内完成相关任务，从被动响应转向主动巡检预警，问题定位压缩至分钟级，DBA得以从重复任务中抽身，聚焦架构优化。

相关产品

CVM | EO | CDB | WorkBuddy | DatabaseClaw

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

游卡

业务场景

游卡网络是中国卡牌游戏赛道的领先者，旗下涵盖《三国杀Online》《三国杀移动版》等多款精品游戏，员工超千人。业务线众多、技术栈多元、跨多端多区多系统的故障排查的痛点，正因AI的到来迎来关键变化。

业务挑战

游卡在运维管理上长期面临「事务繁、链路长、协作散」的痛点，传统人力与工具难以匹配业务扩张节奏。

▶ **故障链路长且传导快**

多款游戏并行运营，每款至少包含四层逻辑，一层异常可在分钟内传导至全局，传统线性推进耗时长且泛化性差。

▶ **依赖人工经验，经验难传承**

资深工程师排障经验高度依赖个人记忆，难以跨产品复制与规模化。

▶ **跨层关联难捕捉**

单点故障可能涉及多层，跨层因果关联需人工推演。

▶ **告警碎片化与定界困难**

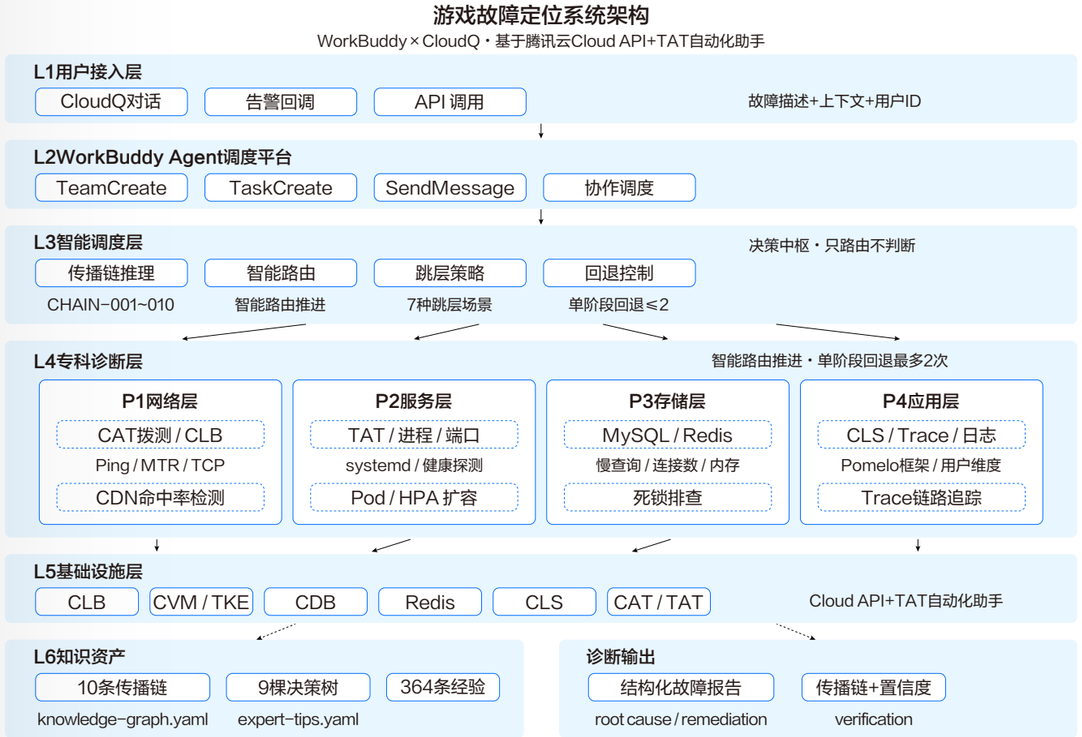
告警量级成倍增长，大量误报夹杂其中，跨团队定界成本高。

解决方案

以「故障诊断专家团」为核心的企业级AI辅助诊断方案，将AI能力嵌入四层诊断流程，包含五个核心能力模块：

- **标准化诊断流程：**按P1→P4四层架构组织诊断师，基于故障特征和传播链动态选路，支持单用户诊断等快捷路径。
- **故障传播链推理：**内置数十条传播链，涵盖缓存雪崩/网络级联等全场景，实现跨层因果推理。
- **专家经验库佐证：**覆盖33大类专家经验，采用「症状→原因→验证→修复」四段式结构，实时佐证诊断结论。
- **智能路由与跳层加速：**故障特征自动匹配诊断路径，传播链驱动跳层策略，直接锁定根因。
- **团队协作闭环：**主理人+四位诊断师组成协作链路，信息流经主理人中转，完成后输出结构化报告。

架构图



- **用户接入层**，通过CloudQ对话、告警回调、API三种入口接入故障描述与上下文。
- **WorkBuddy Agent调度平台**，负责团队创建、任务分配、消息中转与多Agent协作调度。
- **智能调度层**，基于传播链推理与智能路由动态选路，支持跳层策略和受控回退，是决策中枢。
- **专科诊断层**，P1–P4四位专科诊断师分别覆盖网络、服务、存储、应用，按需调度。
- **基础设施层**，依托腾讯云Cloud API与TAT自动化助手，调用CLB / CVM / CDB / Redis / CLS等云服务。
- **知识资产与诊断输出**，产品持续沉淀传播链、决策树与专家经验；优化输出结构化报告、根因与修复建议。

客户价值

定位速度显著提升

MTTR从小时级缩短至5~15分钟，传播链+智能路由避免泛化搜索。

知识资产沉淀与积累

排障经验编码进Agent逻辑，新人培养周期显著压缩，人员变动不再导致经验断层。

诊断一致性增强

标准化SOP输出量化判定，专家经验确保同一症状不会得出相悖结论。

运维成本降低约40%

从被动响应转向主动研判，关键节点前置风险巡检，消除团队间推诿成本。

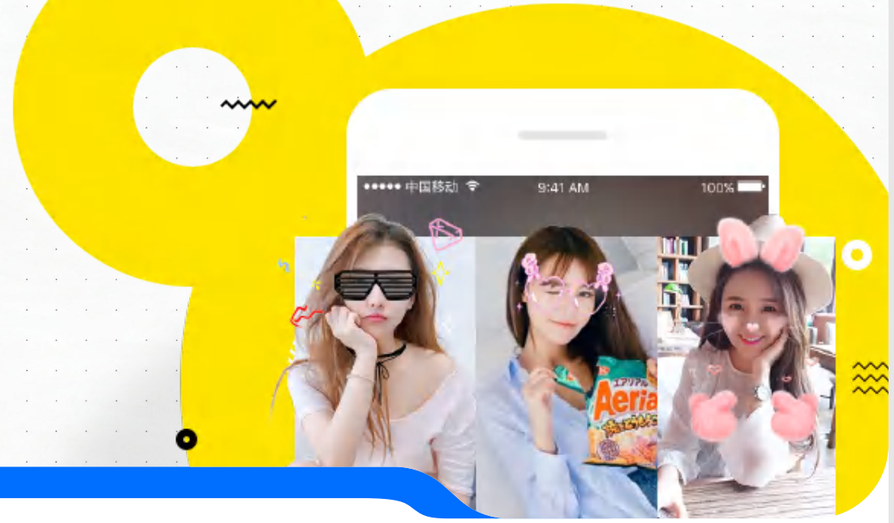
跨层级联不再遗漏

全场景传播链覆盖4阶段症状/影响/缓解措施，自动预测下游影响。

相关产品

WorkBuddy　CloudQ

花椒直播



标杆案例

花椒

业务场景

花椒直播是国内领先的移动社交与娱乐直播平台，集视频聊天、语音交友、才艺展示于一体，为年轻人打造沉浸式社交体验和潮流文化表达窗口。过去两年，花椒持续推进AI原生化转型，从面向用户的AI互动能力，逐步拓展到面向企业内部的Agent应用与业务Skill建设，让AI进入更多部门和真实工作流程。

业务挑战

▶ **要改的东西，比想象的多得多**

企业Agent建设并不是在现有RAG之上增加一个对话入口，而是需要同步升级知识访问、系统接口、产品形态、业务流程和组织协作方式，业务需求被理解、拆解和交付的方式也会随之发生变化。

▶ **我们把AI Coding推到了每个技术组，分组看效果明显，但整个研发流程未提效**

前端、服务端、Android、iOS、PC和测试团队都在使用不同的AI编程工具。各团队分别形成了自己的实践方法，也取得了局部提效，但由于缺少统一流程、上下文和质量标准，局部效率尚未完全转化为端到端的研发交付效率。

解决方案

构建公司级Agent平台——技术团队统一沉淀企业Agent所需的共性工程能力，并对身份权限、Agent、Skill、Tool、模型和运行过程进行统一治理。业务团队使用业务语言描述目标、规则和SOP，平台将其转化为可治理、可执行、可持续迭代的Agent能力。

- **业务系统的Agent化接入：**将原本依赖人工页面操作的业务能力，以及分散的BI查询和内部服务，封装为面向Agent、受身份权限控制的API、CLI和Tool。
- **业务SOP的Skill化沉淀：**将各部门和岗位的业务知识、操作方法、判断规则与异常处理流程，沉淀为可被Agent调用和持续迭代的Skill。
- **从真实问题到任务闭环：**Skill并非对API的简单封装，而是将业务目标、SOP、系统能力、判断规则和异常处理组合成可执行、可验证的任务链路。

架构图

企业Agent平台分层架构

第1层 用户与业务系统
平台管理员 | 业务团队 | 外部业务系统
典型场景：Agent 运营、业务问答、自动化任务、系统集成、知识与工具复用

↓

第2层 体验入口

管理控制台	对话工作区	API集成
资源管理、发布审核、权限配置、运行观测	多Agent入口、文件输入、流式对话、工具审批	身份接入、会话调用、事件订阅、结果回调

↓

第3层 Agent平台服务
身份与访问 | Agent 生命周期 | 技能与工具治理 | 模型管理 | 审计运营
支撑概念：角色授权、版本发布、依赖校验、能力目录、策略配置、操作审计

↓

第4层 AI运行时
会话编排 | 上下文管理 | 模型调用 | 工具协同 | 执行环境调度
支撑概念：流式响应、中断恢复、工具审批、任务执行、用量计量、错误可追踪

↓

第5层 数据与资产
平台配置 | 会话记录 | Agent 版本 | 技能 / 工具包 | 运行证据 | 用量指标
支撑概念：可追溯历史、可复用资产、发布快照、审计证据、统计分析

↓

外部生态 模型供应商 | 企业身份 | 对象存储 | 执行沙盒 | 通知渠道 | 业务系统
集成方式：统一认证、可信回调、文件资产、远程执行、消息送达、API 对接

平台承诺

统一入口
管理、对话、API三类入口，面向不同角色保持一致接入

权限治理
围绕用户、角色、Agent、技能、工具建立可管理的访问边界

版本可控
Agent、技能、工具可以发布版本，承载稳定运行与回溯

运行可追踪
会话、工具、模型调用和任务结果形成可观测运行证据

生态可扩展
模型、身份、存储、执行环境、业务系统按清晰边界接入

客户价值

01
以往需要打开后台、查找入口并跨系统完成的操作，逐步转变为通过Agent发起任务。在身份权限、审批策略和审计机制约束下，Agent可协助完成主播提醒、数据提取、活动配置和用户运营等高频任务，部分场景已形成从指令理解、工具执行到结果反馈的闭环。

02
运营人员即使不在工位，也能实时调度系统资源。效率的提升，已经超出了传统意义上的百分比衡量，而是进入了新的量级区间。

相关产品

WorkBuddy Managed Agents

采用腾讯云WorkBuddy Managed Agents作为企业Agent的云端托管与运行底座，并结合自身业务系统、身份权限、数据资产和业务SOP，持续建设面向多部门的Agent应用与Skill体系。托管能力降低了底层运行环境的建设与运维成本，使团队能够更聚焦于业务场景、Skill沉淀和应用效果验证。

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

搜狐畅游

业务场景

畅游有限公司（Changyou.com Limited，纳斯达克代码：CYOU，简称“搜狐畅游”），总部位于北京市石景山区，是中国领先的在线游戏开发与运营商。其前身为2002年成立的搜狐公司游戏事业部，2007年分拆独立运营，2009年4月登陆纳斯达克全球精选市场，2020年4月由搜狐完成私有化、现为搜狐全资子公司。公司拥有二十余年游戏研发经验，业务覆盖端游/手游研发、运营、发行与海外出海，并运营综合游戏媒体平台17173.com。其自主研发的旗舰产品《新天龙八部》端游是国内最经典的MMORPG之一，运营超十年、注册用户近4亿，同时持续拓展《天龙八部》系列手游及《西游：笔绘西行》等新品。

业务挑战

游戏运维

运维工作维护，脚本运行，巡检等机械化操作占用大量时间，夜间节假日告警、招人留人难、人肉值守本身成为新故障源，资深SRE离职意味着业务线能力断档，新人独立值班周期长，能力的不同，导致处理故障、问题时间恢复慢， 对业务损失大。

内部效率提升

公司内部全面拥抱AI，财务、采购、IT、HR、运营等部门各自寻找模型、独立接入API，模型接入的边际成本居高不下，与此同时，模型调用缺乏统一的评估机制，选型往往取决于“先到先得”而非场景适配度，模型迭代升级时也难以平滑切换，更棘手的是，调用量、成功率、Token消耗等关键数据对运营方完全不可见，不仅难以发现异常、优化调度，模型调用费用也在无人察觉间累积成高额账单，AI投入的业务价值难以量化、难以汇报。

云上安全

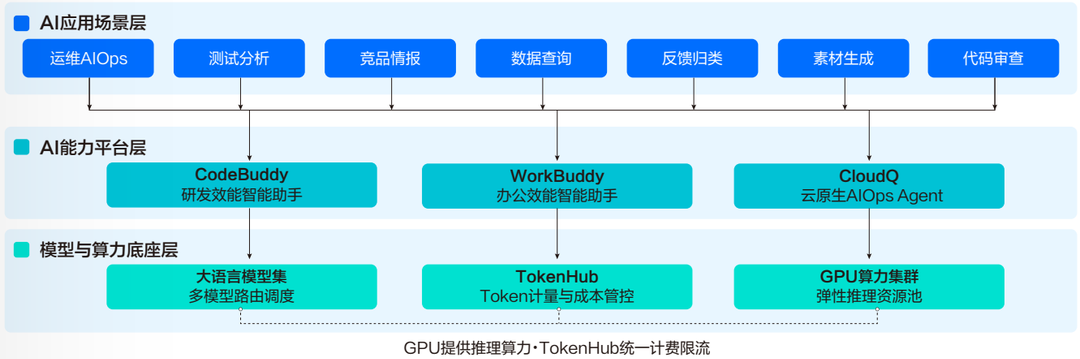
漏洞爆发后要在主机安全与云防火墙两个控制台手工关联资产与流量，各安全产品查询语法不统一、依赖资深安全工程师经验，跨云攻击链路断裂，暴露面盘点需数小时、跨云攻击链路无法贯通、安全人力成为瓶颈。

解决方案

- **WorkBuddy：**通过使用 WorkBuddy，脚本 / 重启 / 巡检沉淀为 Agent Skills，重复操作下降70%，资深 SRE 转向架构治理与风险预防，故障处理、巡检经验沉淀到 Agent。
- **CloudQ：**CloudQ 提供统一调用主机安全与云防火墙，高效完成漏洞影响资产检测以及 IP 流量分析，同时在多云架构下，支持做跨云联合流量分析，打通多云安全数据孤岛。

- **TokenHub：**客户内部平台接入 TokenHub 大模型服务平台，企业获得统一的 AI 能力底座：对外只暴露一套标准 API，财务、采购、IT、HR、运营等各部门一次接入即可按需调用多家模型，调用过程中，TokenHub 提供灵活计费与成本归集，按部门、场景、项目维度清晰切分 Token 消耗，让 AI 投入从“糊涂账”转为可复盘的经营数据。

架构图



客户价值

通过游戏运维AIOps与WorkBuddy联动，把资深SRE的故障处理经验沉淀为可复用的Agent Skills，新人独立值班周期从“月”级压缩到“周”级，组织能力不再随人员流动而断档。

多部门多模型的AI调用按部门、场景、项目清晰归集，模型选型从“先到先得”转向按场景智能路由、可平滑切换，AI投入价值可量化、可对账、可向管理层汇报，从成本中心走向价值中心。

CloudQ统一编排主机安全与云防火墙，高效完成漏洞影响资产检测，IP流量分析，并支持做跨云联合流量分析，打通多云安全数据孤岛，降低安全门槛。

业务效果

01

游戏运维AIOps通过AI Agent将P0级告警的故障恢复时长从传统 15–30分钟压缩至5分钟，实现响应提速83.3%、重复运维操作下降 70%，操作成功率达98%以上，可支撑7×24小时无人值守与脚本、重启、巡检等操作自动化。

02

TokenHub 以统一标准 API 支撑企业各部门快速接入、灵活迭代多家大模型，调用用量与异常报错均可快速定位追溯，消费明细清晰透明，让各部门 AI 投入花得清晰。

03

WorkBuddy 以对话式低门槛操作让非编程员工也能快速上手使用，同时可将资深员工的工作能力沉淀为可复用的Agent，供新员工直接复用，既实现了团队经验的高效传承，也能借助AI全面提升办公效率。

相关产品

TokenHub　CodeBuddy　WorkBuddy　CloudQ



标杆案例

紫龙游戏

业务场景

紫龙游戏是一家集研发与发行于一体的移动游戏公司，总部位于中国北京，并在北京、上海设有研发中心，在中国台北、日本东京、韩国首尔设有分公司，团队成员遍布全球。其产品覆盖中国大陆及中国港澳台地区、日本、韩国、东南亚以及北美、欧洲、澳洲等全球30余个国家和地区。在品类上，紫龙游戏以战棋（SRPG）赛道见长，代表作《梦幻模拟战》长期是其核心营收产品；此外还拥有《天地劫：幽城再临》《龙魂旅人》等战棋、卡牌、MMORPG多品类产品线，并持续加码海外市场。
在游戏行业，数据是核心生产资料。一款产品从买量、上线、版本迭代到长线运营，每一个决策背后都需要数据支撑。在AI发展的大趋势下，DATA+AI已然成为每个游戏公司都需要重点考虑的必选项。

业务挑战

游戏公司的数据查询、分析需求呈现“高频、临时、多变、跨域”的特征。而更深层次的挑战是如何利用好这座数据金矿，让数据真正的开始驱动公司发展。下面按决策层、运营层、客服层三个角色，分别拆解其真实痛点。

➤ 决策层（CEO / CTO / 制作人 / 发行负责人）

决策层关注的是全局经营视角：流水好不好、用户健不健康、钱花得值不值。他们的需求特点是“要得急、看得宏观、还要能随时下钻”。

典型场景

- “这个月流水跟上月比怎么样？帮我拆到渠道和区服看看。”
- “新版本上线后，大R的留存和付费有没有变化？”
- “买量花了这么多钱，D7、D30回收到几成了？哪个渠道亏了？”

核心痛点

痛点	具体描述	业务影响
看板僵化	固定Dashboard无法回答临时、组合型问题	决策依赖“拍脑袋”或等取数
下钻断层	看到异常指标却无法即时下钻定位原因	发现问题≠定位问题，贻误战机
跨源割裂	买量数据、游戏内数据、财务数据分散多系统	算一个真实ROI要跨多个平台手工拼
响应滞后	临时需求排期1－3天，错过决策窗口	版本调优 / 投放止损不及时

➤ 运营层（活动运营 / 商业化 / 用户运营）

运营层是数据消费最频繁、变化最剧烈的一层。他们几乎每天都在验证假设、复盘活动、调优策略，需求几乎“一天一变”。

典型场景

- “上周限时活动的参与率和流水增量，跟上上期那个活动对比一下。”
- “第30关通过率只有12%，卡在这里的玩家后面流失了多少？”
- “金币的日产出是不是超过消耗了？经济系统有没有通胀？”

核心痛点

痛点	具体描述	业务影响
需求高频多变	每个活动/版本都要定制分析，报表永远滞后	运营被迫“凭感觉”，A/B验证难落地
漏斗自定义难	转化漏斗、留存分层等分析口径频繁调整	每次改口径都要重新提需求排期
经济系统黑盒	货币产出消耗、道具流通缺乏实时监控	通胀/通缩发现滞后，破坏游戏体验
多维交叉分析	要按渠道×区服×版本×付费档同时切片	手工透视表易错、耗时、不可复用

➤ 客服层（客服 / 用户支持 / GM）

客服层是离玩家最近的一层，需要在与玩家沟通的“当下”即时获取信息，并对突发舆情/异常做出快速反应。他们的需求特点是“要得快、要得准、还要能定位到单个玩家”。

典型场景

- “这个玩家充了多少钱、是什么VIP、之前投诉过几次？”
- “今天充值异常的工单是不是比昨天多了很多？集中在哪个渠道？”
- “应用商店评分最近是不是掉了？差评主要集中在什么问题？”

核心痛点

痛点	具体描述	业务影响
查询割裂	玩家信息散落在充值、行为、工单等多个系统	服务一个玩家要切换多个后台
响应延迟	复杂玩家画像需提工单查询，无法即时响应	大R服务体验差，投诉升级
预警缺失	异常工单、突发舆情缺乏自动监控与告警	事故/舆情发现全靠人工，容易失控
缺少沉淀	问题分类、话术、处理经验难以数据化复用	服务质量依赖个人经验，难标准化

解决方案

腾讯AI办公助手WorkBuddy作为业界领先的桌面版智能办公助手，具备本地文件访问、联网检索、代码执行、多数据源连接与自然语言交互等能力。将 WorkBuddy与企业既有的大数据基础设施（数据仓库 / 数据湖 / 实时计算 / 指标平台）结合，能够以极低的使用门槛，把“数据分析”这件事交还给每一个业务角色。
借助模型内化的分析方法论与业务常识，它能够像一名分析师那样，围绕一个模糊的目标，自主决定拆什么维度、下钻到哪一层、提出什么假设并交叉验证，最终给出带根因和建议的结论。

更多免费行业报告 并购家 www.ipoipo.net

架构图

WorkBuddy × 大数据基础设施 — 整体架构

从「自然语言取数」跃迁到「自主分析」：说出目标，LLM 带着分析方法论下钻、归因、给建议

决策层
Executive：全局经营视角

运营层
Operation：活动 / 商业化分析

客服层
Support：玩家服务与舆情

↓

交互层 INTERACTION
腾讯AI办公助手WorkBuddy—提目标，而非写查询
自然语言提目标 | 下钻结论 / 图表 | 报告・名单一键产出

↓ 自然语言提问

智能层 INTELLIGENCE
LLM分析Agent + 方法论库 + 语义层
意图理解 + 任务拆解 | 自主下钻・归因推理 | 口径对齐 + 行动建议

↓ 多轮取数・边取边推理

服务层 SERVICE
指标平台・分析方法论库・权限中心
统一指标口径 | 下钻 / 归因剧本 | 行 / 列级数据权限

↑ 根因结论 + 图表 + 行动建议回传

数据层 DATA
数仓 / 数据湖 / 实时计算 / OLAP引擎
海量明细存储 | 高性能聚合计算 | 实时 + 离线

贯穿能力 CROSS-CUTTING
安全合规
数据权限
指标治理
监控告警
审计日志

核心理念：业务人员只需提出目标，LLM 带着分析方法论在既有底座里自主下钻、归因、给建议
从取数到分析・复用既有底座不改架构・方法论库保证下钻不跑偏・语义层锁口径・权限下沉

各层职责如下：

层级	组成	职责
交互层	腾讯AI办公助手WorkBuddy	自然语言问答、报表生成、图表可视化、文档导出
智能层	LLM+语义层 / 指标定义	意图理解、NL2SQL、指标口径对齐、结果解读
服务层	指标平台 / API 网关 / 权限	统一指标、数据权限管控、查询路由与缓存
数据层	数仓 / 数据湖 / 实时计算 / OLAP	海量明细与聚合数据的存储与高性能计算

关键设计原则：

- **不改底座：**复用企业既有数仓（如 Hive/ClickHouse/Doris/StarRocks）与指标体系，WorkBuddy 通过连接器（MCP）接入，无需推倒重建。
- **语义层兜底：**通过预定义指标口径与数据字典约束 LLM，保证“同一个指标算出来的结果一致、可信”。
- **权限下沉：**在服务层统一做行/列级数据权限，不同角色看到的数据范围受控。

客户价值

WorkBuddy + 大数据基础设施的组合，为紫龙游戏带来了业务效率、成本、决策能力的大幅提升，体现了DATA+AI的业务实践。

决策层
- 从“看固定报表”升级为“与数据对话”，全局经营视角随时可得、可下钻。
- 跨源真实 ROI 一键可算，投放与版本决策更快、更准。

运营层
- 摆脱排队取数，分析需求自助闭环，A/B 验证与活动复盘效率大幅提升。
- 经济系统与关键指标实时可监控，运营从“事后复盘”走向“事中干预”。

客服层
- 玩家信息一屏聚合，服务响应更快、更专业，尤其提升大R体验。
- 舆情与异常自动预警，从被动救火转为主动防控。

让数据能力平权——把原本只有数据团队才能完成的分析，交还给每一个业务角色，让“人人都是分析师”在游戏公司真正落地。

相关产品

WorkBuddy | 混元大模型



标杆案例

马帮

跨境电商招聘团队：从「自动找人」到「把人看准」的全流程实践 | 上海马帮科技 · 腾讯云客户案例

业务场景

某头部跨境电商招聘团队引入【马帮】的「猎集AI科学选人系统」，AI覆盖招聘全链路：简历解析、人才画像、面试题生成、面试录音分析与总结等核心环节，用一张持续生长的认知画像把每个候选人看准。该团队以运营、客服、开发实习等岗位的高频、批量招聘为主，旺季扩招时简历量大、面试频次高，HR大量时间被压在筛选、出题与面试总结等重复性工作上。接入猎集后：HR只需上传简历即可自动解析入库、生成候选人画像；面试前可一键生成针对性面试题，面试录音自动转写并解析成猎头式推荐书，直接转给老板决策。

业务挑战

市面上的HR工具在「记录」，而他们真正需要的是「判断」——四个坑，一个都绕不开：

▶ 手动招聘，前端吞掉大半时间

市面上的 HR 工具都在帮人事「记录」，没人帮人事「判断」招人的真正难点从来不是把流程走完，而是想清楚「这个人到底值不值得招」。猎集要做的，就是把 AI 从「记账」环节推进到「拍板」环节：帮你把招聘前端的杂活自动化，把跨轮次的评估信息串起来，让每一次用人决策都有依据、能交代。

▶ 已用系统只会「记账」不会「参谋」

市面上的系统普遍解决考勤薪酬绩效，招聘模块偏记录；报表一堆，却没有一条告诉你「这个人到底值不值得招」。

▶ 评估靠印象、跨轮次信息断层

初筛靠经验、面试靠感觉，多轮结论互相孤立；AI又常把「没有负面证据」当正向信号，看不准也说不清。

▶ 错招成本高、专业价值却看不见

一次错招≈6个月工资+团队效率下降+业务受损+重新招聘；HR的判断难沉淀成让老板信服的结论。

解决方案

系统功能全览 · 招聘闭环与数据资产——HR 不只需要记录工具，更要一套能「替你做判断」的完整闭环。

➤ 五步招聘闭环

01 定义岗位 · 建岗位画像
把「招什么样的人」讲清楚——画像即判断标准，后面对比都有依据。

02 简历汇入 · 自动解析入库
简历统一解析、结构化归档，信息不再散在各处。

03 匹配评估 · 四维匹配分析
对照岗位画像做多维打分，先过滤、再聚焦——把评估从「靠印象」变成「有维度」。

04 面试取证 · 定向问题+录音
按候选人短板出题，面试有记录、结论有证据。

05 决策推荐 · 猎头推荐书
综合各轮证据输出推荐结论，给老板汇报「为什么是他」说得清。

➤ 每一轮招聘都在沉淀数据

数据资产 01	数据资产 02	数据资产 03
人才库沉淀 面过的人留下认知画像，下次用得上。	**招聘数据漏斗** 各环节流转可见，问题出在哪一目了然。	**报表分析** 结论可复用、可复盘，判断能力随数据越攒越准。

闭环不是终点——每一次招聘都在为下一次判断积累依据，猎隼让 HR 的招聘从「走流程」变成「攒判断」。

架构图

AI原生Agent架构
非固定脚本、非黑盒。具备意图识别、自主执行与异常自愈能力：一句话需求经结构化为任务卡后，在后台7×24静默执行，全过程日志可视、可追溯。

大模型 + 语音识别双引擎
基于腾讯云大模型API（混元 / DeepSeek）完成简历结构化抽取、60+维度评估与猎头式推荐书生成；面试录音经ASR转写后进入五步分析工作流。

单一数据源的认知画像
简历/面试/录音写入同一数据源（Single Source of Truth），冰山六区域布局，信息分「已确认 / 初步认知 / 存疑」三态；confirmed 状态不可无故降级，防止AI静默篡改已确认信息。

独有复核自纠层
五步工作流第四步强制检查光环效应、平台归因与追问缺失，让AI不把「没有负面证据」当正向信号——把「面试印象」变成「可溯源的证据」。

更多免费行业报告 并购家 www.ipoipo.net

猎隼AI科学选人 · 技术架构

AI原生招聘智能体：从「自动找人」到「把人看准」的端到端链路

用户接入层

跨平台客户端 · KITTPLUS AI工作平台链接

KITTPLUS桌面客户端	KITTPLUS录音分析	邮箱简历中转	老板 / 决策端
过程分析 · 结果查看	市场洞察 · 行业情报信息	简历自动回流解析	猎头式推荐书一键转交

应用能力层 · 六大模块

一套系统贯通全链路 · 真AI原生，给判断而非记账

人才画像	岗位画像	冰山认知画像	定向面试题	录音五步分析	多候选人PK
自动生成 · 动态迭代	3分钟创建动态生长	六区域 · 显隐性素质	对准疑问 · 精准追问	独有复核自纠层	must-have 横向对比

AI智能体 · 大模型能力层

编排 · 记忆 · 自愈 —— 有决策力的真Agent，不是固定脚本、不是黑盒

意图识别 · 任务卡	Agent自主执行	五步分析工作流	画像记忆引擎
一句话需求→结构化任务	初筛/触达/要简历 · 异常自愈	维度→整理→分析→复核→更新	跨轮次更新 · 三态防篡改

大模型API · 腾讯云混元 / DeepSeek + 语音识别ASR（面试录音转写）

数据层 · 候选人认知画像（Single Source of Truth）

简历 / 面试 / 录音写入同一数据源，随每次接触持续生长、信息不断层

招聘 / 实时数据	认知画像数据源	规则与评估知识
简历库 · 简历解析 · 人才画像	冰山六区域布局 · 已确认 / 初步 / 存疑 三态	岗位must-have · 评分锚点 · 红旗信号 · OBER 题库

腾讯云基础设施底座：大模型服务 | 弹性计算 | 对象存储 | 语音识别 | 网络安全 · 合规

客户价值

45秒 / 份出报告

简历到手即用60+维度自动评估、生成画像，初筛从「靠经验」变「有证据」——判断更快，理由更清楚。

评估有据、决策可交代

四维匹配 + 定向面试题 + 面试记录，结论不是「感觉」，是每一条都拿得出证据的判断书。

沟通成本大幅下降

面试录音一键解析成推荐书转老板——深度用户原话：「沟通成本减少很多」。

专业价值被看见

把 HR 的专业判断显性化、沉淀成可复用的报告，让老板看得见「为什么是他」。

真实使用验证

产品上线即被13家企业每天深度使用，深度用户每天处理约200份简历，并主动愿意向同行推荐。一次错招=6个月工资 + 团队效率下降 + 业务受损 + 重新招聘；而避免这次错招，可能只是一杯咖啡的钱。

品类定位

招聘平台帮你「找人」、HR 系统帮你「管人」，猎隼帮你「看准人」——简历已经在你手里，用科学方法把每个人看准。真 AI 原生，不是管理平台。

业务效果

3分钟 建好一个岗位画像

45秒 生成候选人分析报告

60+ 候选人评估维度

1张 贯穿全程的候选人画像

相关产品

混元 | DeepSeek | ASR

03

技术创新

从模型训练到智能体应用的技术底座

方案

3.1 Agent沙箱提升大模型后训练效率 / P49

3.2 全新Agent原生基础设施助力Agent应用快速迭代 / P51

案例

标杆案例——阶跃 / P53

标杆案例——Kimi / P55

3.1 Agent沙箱提升大模型后训练效率

1 业务场景

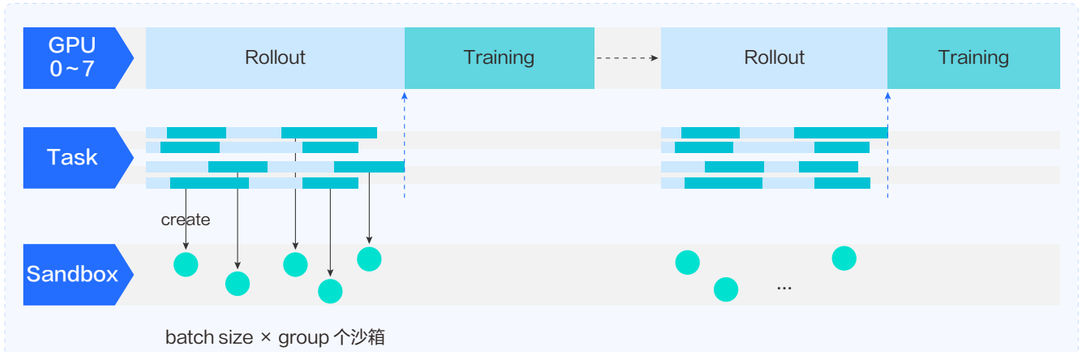
在通用大模型迈向AGI的关键阶段，Agentic RL成为了突破模型逻辑推理与自主决策瓶颈的核心路径。过去的强化学习手段如RLVR或RLHF，通过可验证结果或者偏好来优化模型的回答质量，让模型能搜索、做题或者写作，基本都是围绕单轮输出展开，在多轮Agentic场景下效果并不好。模型能力进一步提升都需要通过Agentic RL来提升在开放世界解决问题的能力，一般会通过让大模型完成真实的多步骤任务，常见的可能有手机、终端操作，比如在手机上打开12306 买票、或修复GitHub上特定的问题，再用这种数据集进行强化学习时就需要我们提供对应的真实环境，让模型在环境里不断观察、行动、接收反馈，最终通过对执行轨迹打分，让模型学会执行。Agent沙箱可以为各种复杂的真实环境提供运行空间，已经逐渐成为Agentic RL的核心基础设施。

2 业务挑战

➤ 挑战一：如何突破底层资源调度极限，支撑大规模并发与毫秒级弹性

训练阶段启动沙箱具备大规模、脉冲式特征。在部分场景下，镜像个数从数千至数十万量级。因此沙箱必须克服大规模镜像分发瓶颈，实现数万至数百万沙箱实例的极致弹性，并在高并发高峰期保障极细粒度的成本可控。

在一次典型的 Agentic RL 训练中，Rollout 阶段每个 Step 都需要生成大量沙箱，比如 batch size 为 512、group 数量为 16，这意味着系统需要在瞬间同时拉起 8192 个沙箱环境，若沙箱创建时间较长就会出现 GPU 空跑，甚至训练直接失败。



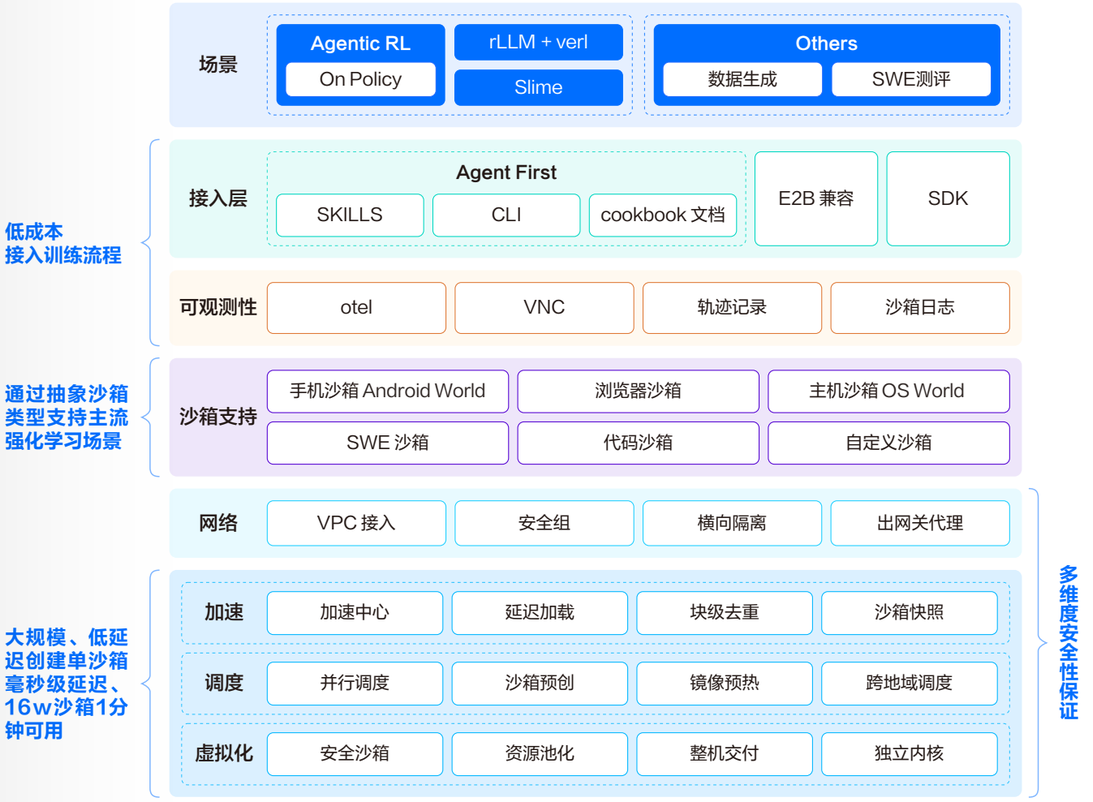
➤ 挑战二：如何构建全场景交互环境

智能体不仅要“说”更要“做”。沙箱必须演进为“全场景数字世界模拟器”，不仅要涵盖代码调测（Coding），还要全面支持浏览器操控（Browser Use）、移动端交互（Mobile Use），电脑端交互（OSWorld）、甚至是复杂的软件工程环境（SWE），让 Agent 无缝接入真实世界。

➤ 挑战三：如何提供深度实验支持与观测

高质量的“交互轨迹”和“执行环境 CheckPoint”是 Agentic RL 的核心。沙箱必须精准记录多步长交互的思考路径与环境反馈，支持复杂交互断点的精准状态回放，支持分叉实验等，并提供全链路可观测性，帮助算法团队高效提纯数据并加速迭代纠错。

3 解决方案



极致弹性保障突发任务不阻塞，毫秒级启动让GPU不空跑

通过资源解耦/池化，避免外部依赖和内核资源竞争干扰，降低环境初始化耗时；同时通过将 Ready 的沙箱制作成快照，通过快照恢复的方案可以跳过繁琐的初始化步骤，做到 60ms 交付服务可用的沙箱。同时针对Agentic RL镜像数量多的场景，通过海量镜像分发与按需加载的极致优化，可以做到分钟级启动16w不同镜像的沙箱实例。

全场景交互沙箱支持后训练全链路需求

围绕主流 Agentic RL 训练、评测需求，比如常见的有 PPT、表格操作，网页操作，手机 App 操作，写代码等。腾讯云沙箱支持代码、手机、电脑、浏览器四类沙箱，也支持常用于数据采集、训练、评测等场景的沙箱，如：SWE 系列、Android World、OS World 等。

镜像快照及快速复制能力支撑业务快速迭代

腾讯云沙箱可以通过快照记录沙箱内部的状态，在正常的 Stage 里制作快照，出现问题后，可以通过快照秒级回滚到之前的状态，也可以创建多个副本并行加速训练。除了任务崩溃恢复外，对于蒙特卡洛树搜索这类强化学习，也可以通过快照秒级回溯到上一个节点，或者从某个节点并行执行，加快搜索过程。

更多免费行业报告 并购家 www.ipoipo.net

3.2 全新Agent原生基础设施助力Agent应用快速迭代

1 业务场景

随着 Agent 从工具调用逐渐演进到真正的 Agentic Agent ——能自主规划、动态调用工具并与环境交互，对运行环境提出了更高标准。为了让这些数字员工安全、稳定地完成复杂任务，迫切需要一个数字世界的沙箱，为 Agent 提供专属环境。目前Agent应用主要覆盖如下场景：

- **Coding Agent**

  AI 生成的代码在云端沙箱中运行，与本地隔离，避免安全风险，具备弹性算力。支持 CodeBuddy Code / Claude Code 等沙箱环境。

- **Deep Research**

  浏览器沙箱提供安全隔离的浏览器交互环境，Agent 通过 Playwright 等工具完成深度信息调研。

- **数据处理与办公自动化**

  通用 Agent 通过代码生成完成数据清洗、分析、图表与 PPT 制作，所有代码在云端沙箱中运行。

- **GUI Agent**

  在浏览器/虚拟桌面/虚拟手机中模拟人机操作（Browser Use / Computer Use / Mobile Use），实现从“代码驱动”向“操作驱动”演进。

2 业务挑战

➤ 挑战一：Agent自主执行，行为不可完全预测

Agent完全由模型进行驱动自主执行，执行过程中可能会对执行环境产生不可逆破坏，例如删文件、改配置、误操作生产数据等。需要给到每个Agent一个完全隔离的运行环境，使得破坏不会外溢。

➤ 挑战二：常驻型Agent长期运行成本高

随着OpenClaw/Hermes等开源Harness框架逐步出现，Agent从原先的任务型执行更多的转变为常驻型执行，Agent成为一个有状态的服务应用。Agent的执行环境需要长期存在，Agent随时被唤醒继续工作，状态持续积累。Agent环境如果7×24运行，会导致资源成本较高，需要支持按需拉起按需暂停。

➤ 挑战三：Agent访问外部系统缺乏安全管控

Agent在执行过程中需要调用外部API、操作文件系统、访问第三方服务，越权访问可能会导致敏感数据泄露。访问行为缺乏可见性，操作无法留痕，异常难以定位，合规审计无从开展。

3 解决方案



虚拟化级安全硬隔离

基于 Cube 等安全沙箱技术，提供内核级强隔离的运行环境，有效防止因 Agent 执行任意代码而导致的数据泄露与越权风险。提供内置的监控、审计和日志能力，确保整个执行过程透明、可控、可追溯。

沙箱全生命周期管理

腾讯云沙箱支持完整的生命周期管理：启动、暂停、恢复、超时销毁。可以实现动态资源调度，沙箱可在运行与暂停状态间灵活切换，暂停时释放资源、按需快速恢复。不仅覆盖多样化业务场景，还能有效降低资源消耗，帮助用户实现更高效的使用体验。

Agent网关针对Agent出入流量统一管控

腾讯云Agent沙箱通过在操作系统底层自动拦截 Agent 所有网络请求，Agent 完全无感知、无需改代码，且无法被绕过。通过域名白名单、IP白名单、协议限制等规则对出入流量进行统一管控。同时通过零信任凭证代理，Agent 在任何情况下接触不到真实密钥，系统在网络层自动拦截并替换注入正确凭证。

StepFun



标杆案例

阶跃

业务场景

阶跃星辰2023年成立，聚焦AGI，构建Step系列多模态大模型、构建Agent应用，并推进 AI手机、汽车等多个场景端云落地。 预训练基座虽具备知识与语言潜力，但距离合规对话、指令遵循、工具调用和行业任务仍有差距。后训练项目通过大规模强化学习，增强复杂推理、工具调用能力，抑制幻觉与安全风险； 再结合Step系列在PDF、表格、图表、UI等场景的多模态理解能力， 支撑开放平台与企业级Agent稳定部署。

为了更好地满足大规模强化学习需求，阶跃团队采用腾讯云沙箱方案，支撑了百万级镜像存储和预热，数万并发沙箱的大规模极速启动，完善的TokenHub平台大模型能力，完整支持了从镜像构建，存储，预热，各种沙箱环境模拟，任务轨迹合成，多模型横向评测等全业务链路，大大提升了Step 模型的后训练效率，加速了模型迭代，助力业务增长。

业务挑战

▶ **百万级镜像存储和预热**

大模型强化学习场景，需要针对包括SWE，Terminal-Bench在内的各种软件工程和任务要求，构建大量的环境镜像，为大模型解决实际问题提供必要的环境模拟，每天会产生数万镜像，累计百万级镜像。对于海量镜像的传输带宽，存储空间，预热吞吐提出较大挑战。

▶ **大规模并发极致性能**

基于海量镜像，通常会以数万并发大规模启动沙箱，模拟任务环境，让大模型通过自身推理和工具调用能力解决特定问题。在数万并发下，对传统的K8S调度和性能提出严重挑战，亟需新的技术方案，解决性能瓶颈，加速后训练进程。

▶ **完善的大模型矩阵**

在多模型能力在不同榜单的横向评测环节，会需要国内外多种先进模型，进行同等任务下的横向对比和改进。

解决方案

- **镜像存储和加速中心**

腾讯云TCR镜像依靠后端海量COS存储空间，稳定支持客户每天3w级镜像上传和总量百万级镜像存储，通过几十Gbps的内网传输，加速预热到离沙箱集群最近的镜像加速中心，高效支撑了大规模镜像存储和加速的要求。

- **大规模极速启动**

腾讯云AGS沙箱，通过极简的自研调度层，避免了传统 K8S 多 Controller 的低效调度，通过资源预创，设备虚拟池化，镜像去重和懒加载等多种创新技术，在数万并发下，帮助极速大规模启动沙箱，提升 30%后训练效率，避免昂贵的GPU集群空转。

- **完善的大模型平台**

腾讯云TokenHub平台支持多种先进模型的选择，在多任务类型的横向评测中，支持了客户多种灵活选择和组合，帮助加速了模型评测和改进。

架构图



客户价值

大规模并发秒级启动沙箱，帮助客户同时启动数万个沙箱，敏捷完成特定任务后释放，加速了后训练进程，提高了 GPU集群利用率。

大规模镜像存储和加速方案，支撑了客户大规模的构建，提升了云上构建的灵活性，支撑了镜像构建管线的稳定运行。

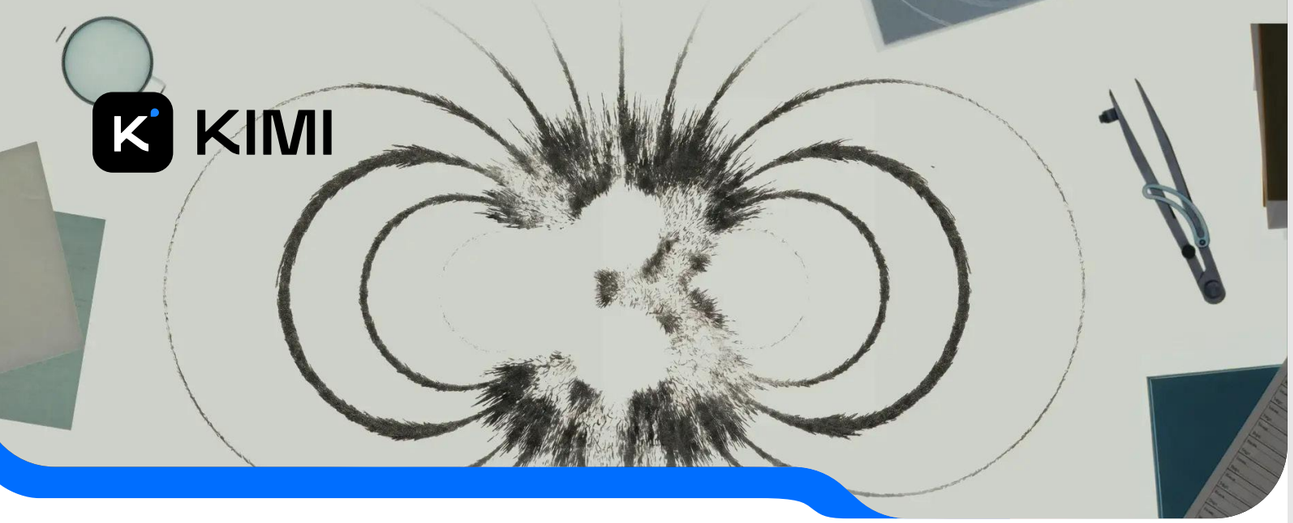
业务效果

大规模并发秒级启动沙箱，帮助客户后训练效率提升 30%，加速模型迭代的同时，大大提升了GPU集群的利用率。

相关产品

Agent Sandbox | TCR | COS | TokenHub | CCN

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

Kimi

业务场景

月之暗面（Moonshot AI）2023年初成立，核心团队涵盖Transformer-XL、RoPE、MuonClip等技术积累，主打Kimi系列大模型，其Kimi Agent项目承接Kimi长文本优势，从K2、“OK Computer”到K3持续迭代，基于强化学习做任务规划、工具调用与自主执行，内置建站、文档、表格、PPT、深度研究等多种工具 。进一步以Agent Swarm调度最多300个子代理、超4000次工具调用，面向研报、代码、办公与批量信息处理，把大模型从问答推进到可交付的自动化工作流部署。

为了更好满足Kimi Agent场景中，大规模并发启动多个Agent，Agent集群的需求，Kimi Agent团队采用腾讯云沙箱方案，使用敏捷高效的强隔离沙箱运行子Agent，按需调用。共同打磨了面向Agent的原生存储方案，既满足极致吞吐，又要给用户展示和协作Agent产物。同时基于出网关能力实现了严格和安全的权限管控，支撑了Kimi Agent业务的飞速增长。

业务挑战

▶ **强隔离敏捷启停沙箱**

Kimi Agent场景中，对于复杂任务，通常会批量启动多个子Agent去并行执行不同任务，此时需要彼此强隔离的系统，互不干扰，又要实现毫秒级别按需启动，执行完任务立即释放。对大规模敏捷启停提出较大挑战。

▶ **面向Agent的原生存储**

在Kimi Agent场景中，通常大模型驱动的多个Agent会生成数百个过程和结果文件，对存储吞吐要求较高。同时也需要对用户展示文件内容以及支持在线编辑协作。对存储系统提出前所未有的挑战。

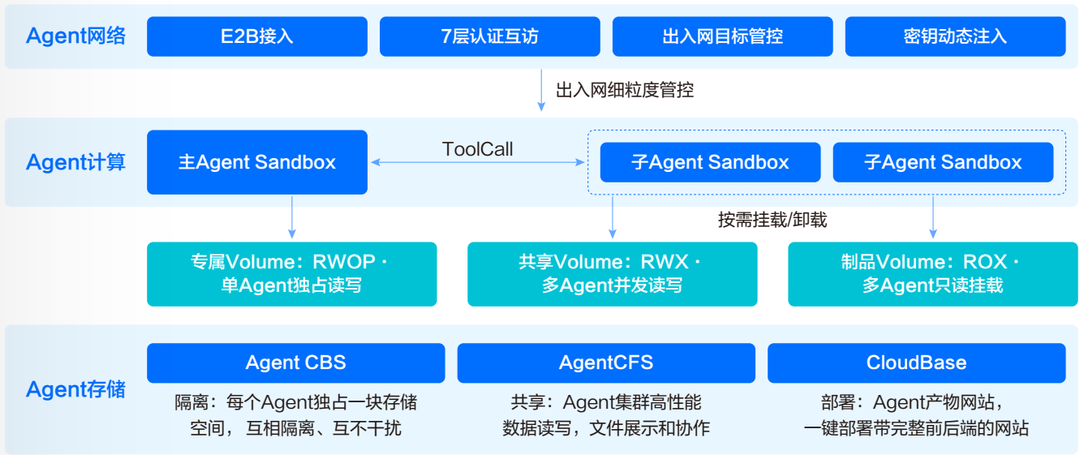
▶ **安全可控的Agent边界**

Agent中为了执行多种任务，通常具备较高的权限，在出入网的管控，工具调用中的权限认证，都是提出细粒度的管控需求，对传统安全管控能力提出新的挑战。

解决方案

- **敏捷强隔离沙箱：**腾讯云AGS沙箱，基于自研的MicroVM内核，实现了大规模毫秒级启动的同时，支持了OS级别的强隔离，特别适合作为Agent执行环境，按需启停，存算分离，敏捷支撑了Kimi Agent的大规模调用，也降低了使用成本。
- **性能强、可协作的存储：**腾讯云AgentCFS，面向Agent的原生存储场景，既支持了大量小文件的读写性能，又能通过标准S3协议展示和编辑文件内容，满足了Agent场景的人和大模型协作的需求。
- **灵活安全管控：**腾讯云沙箱出网络，实现网络流量的透明代理，严格把控权限和认证token，实现不入侵业务的同时，完成细粒度的目标管控和安全认证，保障了Agent平台的可靠安全运行。

架构图



客户价值

支撑客户数百万用户级别的Agent和Agent集群快速增长和稳定运行。

支撑Agent产物的极速读写和对用户展示及协作，实现了人和模型的协同工作。

细粒度的权限和目标管控，保障了Agent的大规模安全可靠运行。

业务效果

01 支持数百万用户大规模运行Kimi Agent，复杂任务启动沙箱耗时降低到 80ms级别，帮助Agent快速反应和高效执行，用完即销毁，敏捷的同时大大降低了使用成本。

02 细粒度的安全管控能力支撑Kimi Agent业务稳定可靠运行，筑牢安全底线。

相关产品

Agent Sandbox | TCR | Agent CBS | AgentCFS | COS

04

业务创新—SaaS Agent

AI Agent重塑行业SaaS体验

方案

4.1 SaaS Agent方案 / P59

案例

标杆案例——明略 / P61

标杆案例——熵简 / P63

标杆案例——慧算账 / P65

标杆案例——富通天下 / P67

标杆案例——蓝色光标 / P71

标杆案例——六度人和 / P73

标杆案例——光魔科技 / P75

标杆案例——天财商龙 / P77

更多免费行业报告 并购家 www.ipoipo.net

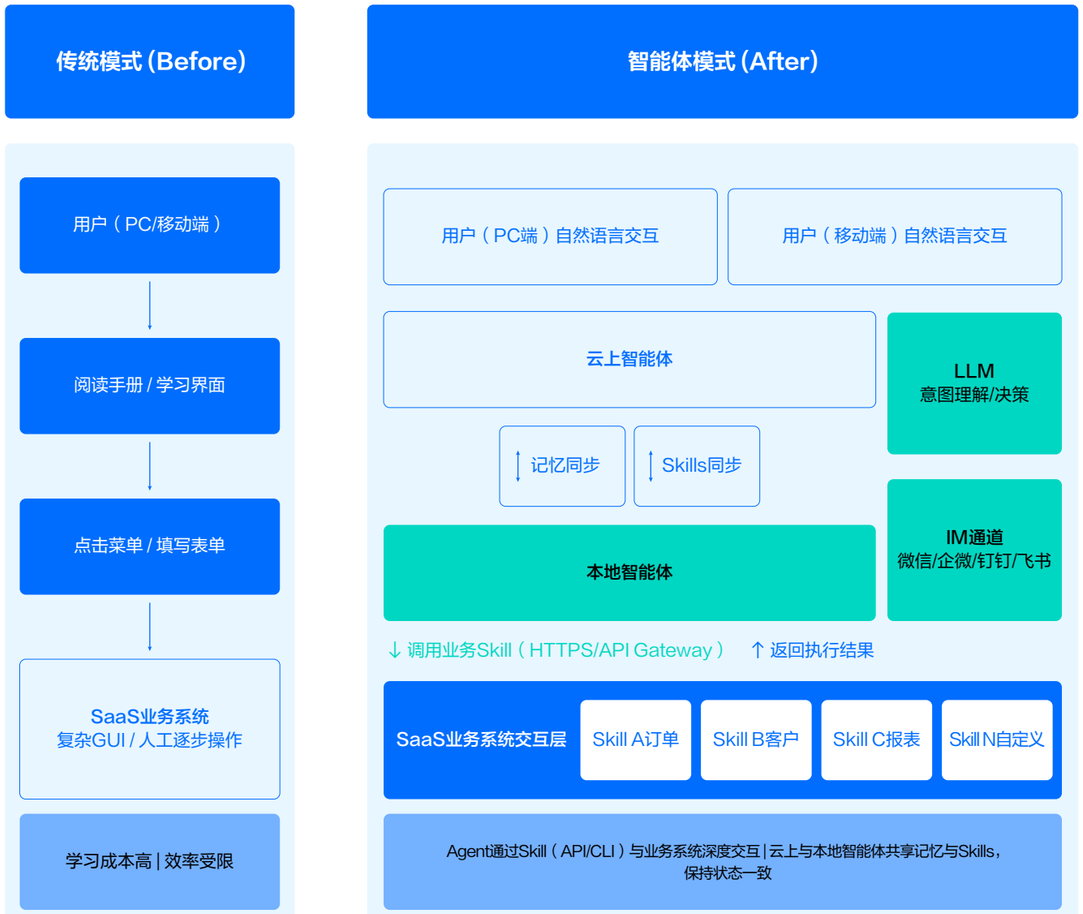
4.1 SaaS Agent方案

1 SaaS行业的Agent转型：从功能订阅到智能交付

转型路径：订阅式付费->任务Credits付费->按照交付结果付费

Copilot嵌入	Agent平台化	结果导向定价
现有产品嵌入AI助手	转向AI代理编排平台	按业务成果付费

转型方案：从GUI操作到智能体架构



2 腾讯云Agent底座助力SaaS应用架构重构

Agent Infra 云原生运行环境

- **快速唤醒：**100ms级冷启动响应，配合细粒度计费，在优化体验的同时降低成本；
- **极致弹性：**依托腾讯云百万核资源池，支持自定义规格与秒级弹性伸缩，从容应对业务峰值；
- **安全隔离：**基于 Cube 安全沙箱实现内核级强隔离，有效防止数据泄露与越权访问。

Agent Harness 工程配套

- **Agent 存储：**以Agent Bucket、Agent CFS 提供可共享、安全隔离、休眠态可达的数据底座；
- **Agent 记忆：**支持短期与长期记忆（四层分层金字塔），并支撑多用户、多 Agent 协同的团队记忆；
- **Agent 编排：**支持多 Agent 编排与子 Agent 调度。

全链路安全和可观测能力

- **Agent 安全：**提供凭据托管、Prompt 防注入、Skill 扫描与行为审计等防护能力；
- **Agent 可观测：**提供全链路 Trace、用量与成本监控、指标监控与智能告警。



明略科技®
MININGLAMP TECHNOLOGY

标杆案例

明略

业务场景

DeepMiner AI 数据分析的智能体，通过自研的 Mano（执行）与 Cito（推理）双模型，配合 Foundation Agent 中枢调度，实现复杂任务分解与协同，主要面向如下业务场景：

1. 企业内部决策场景

营销决策往往需要整合分析分散在电商平台、广告平台、社交媒体以及内部系统的多方数据，使用 Agent 可以一键接入多源数据，并自动完成数据清洗、结构化整理和指标对齐，最终形成完整的分析报告，辅助企业营销决策。

2. 广告营销全链路优化场景

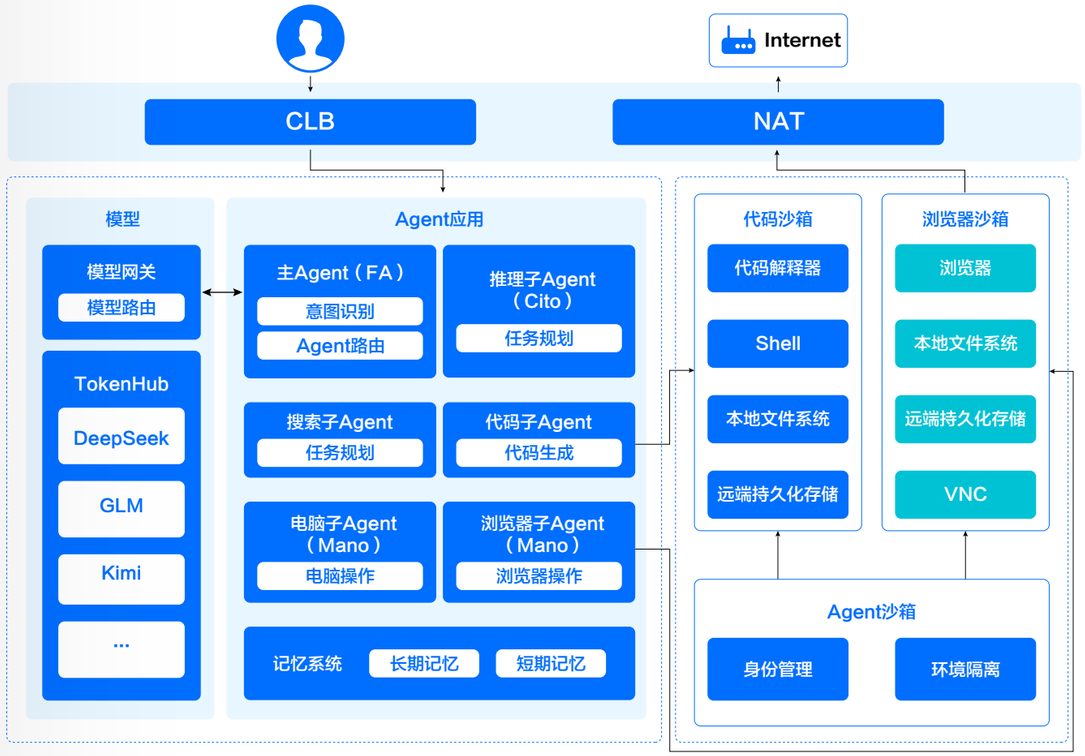
情感与趋势分析：通过对社交媒体和广告评论的情感分析，了解受众真实反馈，辅助营销策略调整。

效果预测：进行流量预测、销售预测和广告归因分析，优化投放成本和 ROI。

业务挑战

- Agent 应用面临不同于传统应用的安全与合规风险，缺少可控的“安全执行边界”、身份认证、权限管理，导致遭受 Prompt 注入攻击、越权执行等攻击。
- Agent 应用对性能与并发要求极高，缺少一个轻量级、快速启动、可水平扩展的执行环境来承载 Agent 任务，传统技术无法兼顾弹性、成本与性能。
- Agent 短期记忆依赖模型的上下文长度，存在上下文挤出（失忆）、上下文过长导致指令细节忽略以及成本与时延的多重挑战；长期记忆则面临信息压缩有损、情景记忆缺乏、检索精度低等问题。

解决方案



- **Agent 沙箱方案**

1、作为 Agent 应用新一代任务执行环境，提供全链路安全、Agent 原生资源架构、极致弹性与生态兼容的能力，让 Agent 在真实世界中具备安全、可靠的执行环境。

2、提供浏览器、code 执行、手机、computer 等丰富的沙箱类型，满足应用不同的使用场景。

- **记忆系统方案**

1、 提供历史交互信息的存储和检索，帮助 Agent 在长周期运行中表现出更强的连贯性、上下文理解力和个性化服务能力。

2、核心能力包括对话存储（会话消息、状态信息存储）、上下文管理（情景识别、上下文压缩）、长期记忆（个性化、事件记忆）、智能检索（多路检索、Agentic 检索）。

- **大模型方案**

1、对接腾讯云 TokenHub，一键接入 DeepSeek、GLM、Kimi、Hy 等国内顶尖模型，专属 TPM 保障。

2、利用腾讯云全球骨干网实现海外模型国内调用的网络加速。

客户价值

腾讯云以 Agent 沙箱、记忆系统与 TokenHub 模型接入，为明略 DeepMiner 提供安全隔离的执行环境、长期可用的记忆能力与稳定的多模型供给，帮助其 Agent 从演示走向生产：执行更可控、扩容更弹性、多轮任务更连贯、成本更可预期。

相关产品

Agent Runtime　TokenHub

更多免费行业报告 并购家 www.ipoipo.net

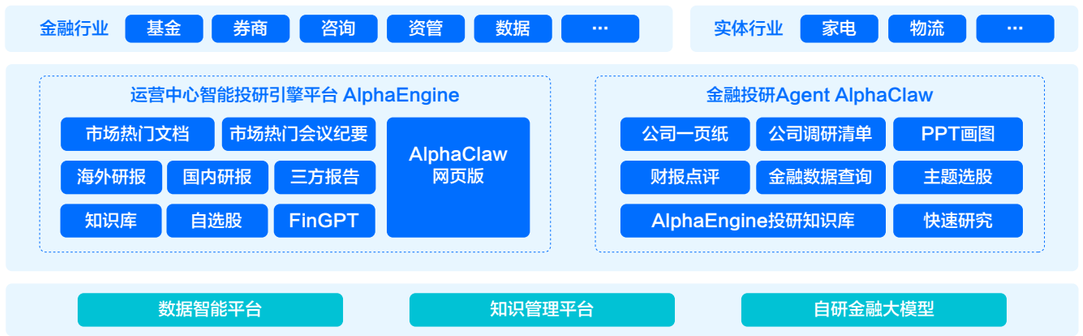


标杆案例

熵简

北京熵简科技有限公司（简称“熵简”）成立于2017年，是一家为中国金融资管行业提供数字化基础设施的专家型科技公司。熵简依托大数据、云计算、人工智能等先进技术，打造全资产、全流程的端到端智能资产管理平台，助力金融机构数字化转型。

业务场景



AlphaEngine

1、一站式商业洞察引擎，面向投研专业人士的标准化 AI投研引擎，被定位为“专业投资者获取阿尔法的发动机”，核心功能包括听会议、读观点、查数据。
2、内置AlphaClaw网页版，后端Agent运行在云端沙箱中，覆盖轻量级使用场景。

AlphaClaw

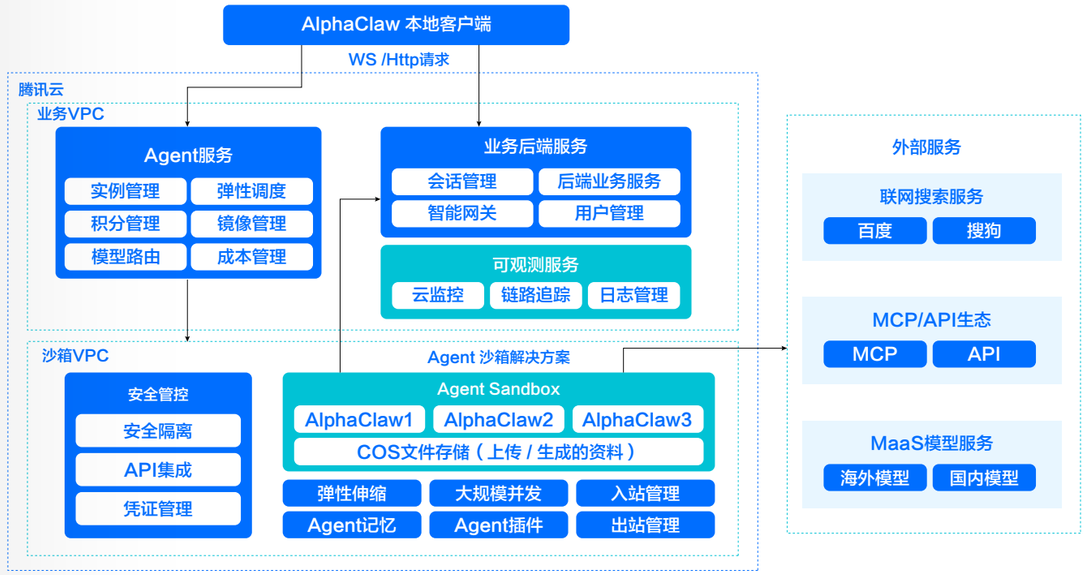
1、金融投研AI Agent，俗称“金融小龙虾”。它不是简单的“问答助手”，而是能自主跑完投研工作流的 AI 分析师。
2、内置公司一页纸、公司调研、财报点评、金融数据查询、主题选股、快速研究、AlphaEngine知识库查询等预制Skill。

业务挑战

- 云端Agent应用对启动速度与并发规模要求极高， 缺少一个轻量级、快速启动、可水平扩展的执行环境来运行常驻Agent应用，传统技术无法兼顾弹性、成本与性能。
- Agent应用面临不同于传统应用的安全与合规风险，缺少可控的“安全执行边界”、身份认证、权限管理，导致遭受Prompt 注入攻击、越权执行等攻击。

- 不同于传统常驻应用，云端Agent应用仅需要在用户有业务请求时运行，大部分时间无需常驻运行，缺少一个可以快速暂停、秒级拉起的执行环境，可以兼顾用户体验与资源成本。
- Agent应用使用的模型众多，原厂对接需要逐个接入，对接成本高且单一模型没有fallback机制，如遇故障无法切换。

解决方案



Agent云上组网安全	● **安全隔离：**结合现有业务统一设计云上组网，严格控制沙箱的网络边界，避免沙箱间以及沙箱与原有业务的安全问题； ● **安全策略：**通过安全组、CAM权限控制精细化管理沙箱与后台业务的访问策略。
Agent沙箱高效弹性	● **毫秒级弹性：**Agent沙箱提供虚拟机级环境隔离能力，同时具备毫秒级的启动、暂停恢复特性； ● **大规模并发支持：**Agent沙箱依托云原生弹性伸缩能力，支持万级沙箱的并发启动。
MaaS一站式模型接入	● **TokenHub合作：**一键接入 DeepSeek、GLM、Kimi、Hy 等国内顶尖模型，专属 TPM 保障； ● **模型加速：**利用腾讯云全球骨干网实现海外模型国内调用的网络加速。

客户价值

腾讯云Agent Sandbox为AlphaClaw提供毫秒级启动、万级并发的安全执行底座，让金融投研Agent从“能跑”走向“规模化可交付”，同时非常驻、强隔离、按需秒起等特性让AI投研Agent同时拿到弹性、安全与成本的三重最优解。

通过沙箱VPC与业务VPC独立组网、安全组与CAM精细化授权、凭证托管与出入站管控等整体安全方案设计，为 Agent划出可控的“安全执行边界”，有效应对Prompt注入、越权执行等新型风险，满足金融行业合规要求。

借助TokenHub一站式接入国内外主流模型并提供专属TPM保障与海外模型国内加速，配合模型路由与故障切换，解决多模型逐一对接成本高、单模型无 fallback的问题。

相关产品

Agent Runtime　TokenHub　CAM



标杆案例

慧算账

基于腾讯云 ClawPro（ClawPro 企业版）打造龙虾 AI 助手

业务场景

慧算账是国内领先的互联网财税服务平台，长期服务中小微企业及代账机构，业务覆盖发票、税务申报、社保工资、账簿报表、财税 RPA、政策咨询等高频场景。
2026年，慧算账将腾讯云 ClawPro（ClawPro 企业版）嵌入企业微信，打造面向会计一线作业的“龙虾AI助手”，通过若干财税Skill与MCP Server，将查社保、开发票、调账、跑财税RPA流程等任务接入AI Agent执行链路，实现从企微侧边栏到后端财税系统的一站式办理。

业务挑战

▶ **会计人效瓶颈**

代账行业每个会计平均服务约300家企业，日常工作中大量时间消耗在发票查询导出、社保数据核对、税务申报表下载、工资表调整等重复性操作上。人工操作效率低、易出错，难以支撑更大规模的客户增长。

▶ **销售能力参差**

公司积累的销冠话术、续费策略与转介经验沉淀在资深顾问脑中，初级销售人员缺乏系统化学习和训练手段，团队整体销售水平难以快速拉齐，影响客户转化与续费。

▶ **Agent规模化治理**

随着Skill/MCP数量持续增长，Skill边界重叠、准确率缺少量化口径等问题逐步显现，需要从“问题驱动修补”升级为“体系化治理”。

解决方案

慧算账基于腾讯云 ClawPro 平台，从以下三个维度构建财税 AI Agent 能力体系：

- **龙虾 AI 助手 —— 财税任务自动化执行**

将会计日常高频操作——查社保、开发票、调账、导出申报表、跑财税RPA流程——封装为标准化的Skill，通过企微侧边栏一键触发。Agent自动完成Skill选择、MCP Server调用、数据获取和结果整合，会计只需审核确认即可交付客户。

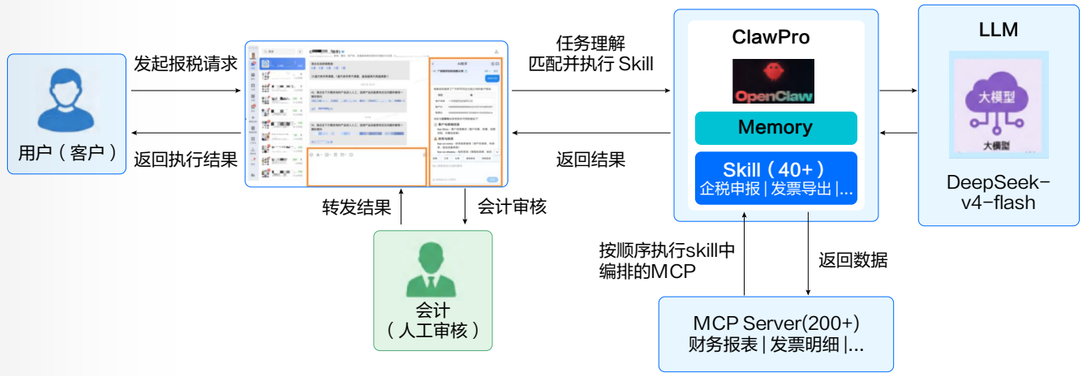
- **销冠经验Skill化 —— 销售能力标准化复制**

将公司Top销售的话术、客户续费策略、转介技巧等核心经验沉淀为Skill，嵌入企微AI助手。销售人员随时随地调用，无需依赖人工培训和经验传承，帮助初级销售快速掌握资深顾问的获客与成交方法。

- **Skill与MCP设计**

通过Skill与MCP的拆分与设计原则，构建高内聚、低耦合的Agent能力基座，提升系统可扩展性与维护效率。

架构图



业务流程以用户发起报税请求为起始节点，经企微对话提交至会计，再由会计转发至企微AI助手，最终完成自动执行与人工审核的闭环。

① **用户发起** → 会计在企微中与客户对话，提交报税需求

② **AI助手** → ClawPro 调用LLM进行任务理解与 Skill 匹配，执行对应 Skill

③ **Skill 执行** → Skill 编排一组 MCP Server，对接内部或外部系统完成业务操作

④ **结果返回** → 执行结果经 ClawPro 整合，返回企微 AI 助手界面

⑤ **人工审核** → 会计确认数据准确性与合规性后，将结果转发回用户

客户价值

自动化财税处理

将原本由人工完成的发票查询、申报表导出、社保数据查询等重复性工作交由 AI Agent 自动执行，会计可更聚焦于核心业务处理，而非机械操作。

销冠经验标准化复制

沉淀 Top 销售话术、续费策略和转介技巧为 Skill，初级销售人员可随时调用，系统性拉齐团队能力水平。

业务效果

ClawPro嵌入企业微信后，帮助“慧算账”实现了会计查社保、开发票、调账、跑财税RPA流程一站式完成。

400–500家
每位会计服务企业数
↑（原约300家）

↑50%
整体作业提效

相关产品

ClawPro　　TokenHub

更多免费行业报告 并购家 www.ipoipo.net

Joinf Claw

外贸企业专属AI专家团队

解决重复劳动 | 高效挖掘客户 | 提升营销效果 | 大幅业绩增长



标杆案例

富通天下

外贸整体成交转化率增长 100%+ —— 让AI智能体走进6万家外贸企业的真实工作流

业务场景

富通天下是深耕中国外贸领域24年的外贸管理与AI私域营销服务商，服务超过6万家外贸企业。2026年4月，富通天下将腾讯云ClawPro（OpenClaw企业版）深度集成到其AI私域营销管理云平台中，打造JoinfClaw智能体矩阵，围绕新客挖掘、老客盘活、独立站优化、社媒运营、数据分析等外贸核心功能，让「龙虾」AI智能体真正跑进外贸企业的真实业务场景中。上线后，整体成交转化率增长超过100%，新客转化率从3%~10% 区间提升到8%~20%，独立站实现7×24多语种在线接待。

业务挑战

▶ **响应滞后错失商机**

一条海外买家询盘往往同时发给多个供应商，谁先响应谁抢单；而询盘常在凌晨进入、业务员次日才处理，商机在时差里流失。

▶ **传统获客效率极低**

传统获客需要调研市场、搜索背调客户、撰写开发信、通过各种渠道触达客户，不仅速度慢、联系人不准，效果也极低。

▶ **Token安全与多租户**

富通天下需把AI能力开放给下游6万家企业客户，但底层龙虾真实Token与实例地址绝不能暴露给下游，且需按租户隔离配额与数据。

解决方案

基于腾讯云ClawPro构建JoinfClaw智能体矩阵，**一句指令就能完整交付结果。**









腾讯云 ClawPro 底座

统一承载多智能体的模型路由、私有Skill库、多租户配额、凭据托管与安全审计，让Agent真正跑进6万家外贸企业的真实业务环境

- **JoinfClaw智能体矩阵：**外贸企业专属的AI专家团队，一句指令，就能完整交付结果。解决重复劳动，高效挖掘客户，提升营销效果，促进业绩增长。
- **WebSocket安全网关：**由腾讯云CLB+AI Agent安全网关+ClawPro WS连接插件构成中间层。客户后台用网关公钥加密签发短期chatToken，浏览器仅持chatToken；网关侧解密后按实例ID动态查询真实ws-url并代理连接，真实Token与实例地址永不出网关。
- **多租户与统一管控：**依托ClawPro三级RBAC、凭据网关托管、租户隔离与全链路审计，配合动态实例发现，支撑面向下游客户的规模化、安全、可运营的AI能力输出。

WebSocket 安全网关

01 场景　富通天下把AI能力卖给下游客户，但龙虾Token不能暴露

Token 暴露风险
龙虾能力需真实Token+实例地址才能连；一旦下发到客户浏览器/APP，等于把密钥交出去

直连不可控
下游客户若直连OpenClaw实例，无法统一鉴权、限流、审计，实例地址也随之泄露

多租户规模化
富通天下有大量下游客户，逐实例配置 Token不可持续，需要统一、可扩展的安全入口

02 方案　CLB+AI Agent 安全网关，用chatToken代理WebSocket，真实Token永不出网关

签发	接入	解密鉴权	动态代理	回流
后台发chatToken	**CLB统一入口**	**网关插件校验**	**查ws-url转发**	**流式返回**
客户后台用网关公钥加密	浏览器仅chatToken连Agent，透传Upgrade	私钥解密与校验，取Token与instanceId	按instanceId查真实地址，动态代理对应实例	经网关回流并留审计

03 收益　安全、可控、可规模化

零暴露
密钥安全
真实Token与实例地址永不下发客户端，从根源杜绝泄露

统一入口
集中管控
鉴权 / 限流 / 审计集中在网关，合规可追溯

动态发现
规模化
不逐实例配置，凭instanceId 动态代理，平滑扩展

低改造
快速落地
业务系统仅需接一层chatToken，POC 已验证可上线

架构图

四层分层架构，安全网关居中：用户与接入层 → 腾讯云安全网关层 → ClawPro 智能体能力层 → 腾讯云基础设施层。

客户整体技术架构 ARCHITECTURE

ACCESS 用户与接入	**企业用户 / 业务员** 在ERP-CRM场景发起AI请求	**浏览器 / 客户APP** 仅持短期chatToken发起WSS连接	**富通天下业务系统后台** 保管龙虾Token，签发chatToken
GATEWAY 安全网关层	**CLB公网入口** 透传WebSocket Upgrade	**AI Agent安全网关** 鉴权 / 解密 / 代理 / 审计	**ClawPro WS插件** Token转换・动态查ws-url
Agent 智能体能力层	**OpenClaw企业实例** 承载会话、知识、Skills执行	**ClawPro管控平台** 多租户 / 配额 / 凭据托管 / 审计	**模型・Skill・记忆** 多模型路由+私有技能库
IaaS 云基础设施	**CVM计算** 弹性算力	**VPC・安全组** 同VPC网络隔离	**CLS日志** 全链路可观测	**安全防护** 端口隐身・边界防护

安全居中
网关作为唯一公网入口，真实Token与实例地址不出网关

分层解耦
接入 / 网关 / 能力 / 基础设施四层清晰，各层独立演进

多租户隔离
按租户隔离配额与数据，动态实例发现支撑规模化

云原生弹性
同VPC部署，网关无状态横向扩展，全链路可观测

客户价值

业务增长
整体成交转化率增长100%+；新客转化率从3%~10%抬升到8%~20%；跨时区询盘「问完不再流失」。

效率跃升
业务员从筛线索转向重点跟进高价值线索；独立站运营从「人操作工具」升级为「人操作AI自主完成」。

安全可控
龙虾Token零暴露，WS网关统一承担鉴权、限流、审计；多租户隔离满足企业级合规，底座跟随开源社区演进、升级零感知。

业务效果

整体成交转化率增长 **100%+**

服务外贸企业 **60,000+** 家

独立站 **7×24** 多语种在线接待

新客转化率跃升 **3%~10% → 8%~20%**

相关产品

ClawPro　AI Agent安全网关　CLB　CVM　VPC　CLS

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

蓝色光标

业务场景

蓝色光标成立于1996年，2010年在深圳证券交易所创业板上市，是全球TOP10营销传播集团。公司业务总部位于上海，行政总部位于北京，在中国各大区域设有分支机构，并在北美、欧洲以及亚太其他国家和地区拥有国际业务网络。主营业务包括出海营销服务、智能整合营销、数字广告以及元宇宙业务，服务内容涵盖营销传播整个产业链。2023年蓝色光标发布AI²战略，推出AI多模态营销引擎Blue AI，打造面向营销全链路的AI战略级产品矩阵，覆盖创意、内容、广告投放、社交及达人营销等核心场景，推动AI从单点提效走向决策、生产与执行全链路重构，加速向AI Native营销科技企业进化。

业务挑战

将AI智能体规模化落地到多项目、多团队、多客户的复杂营销场景时，客户遇到三类企业级瓶颈：

▶ 协作与隔离

多项目、多团队并行时，智能体的上下文与长期记忆易相互串扰，缺乏实例级、数据级的强隔离机制，难以支撑海量实例稳定运行。

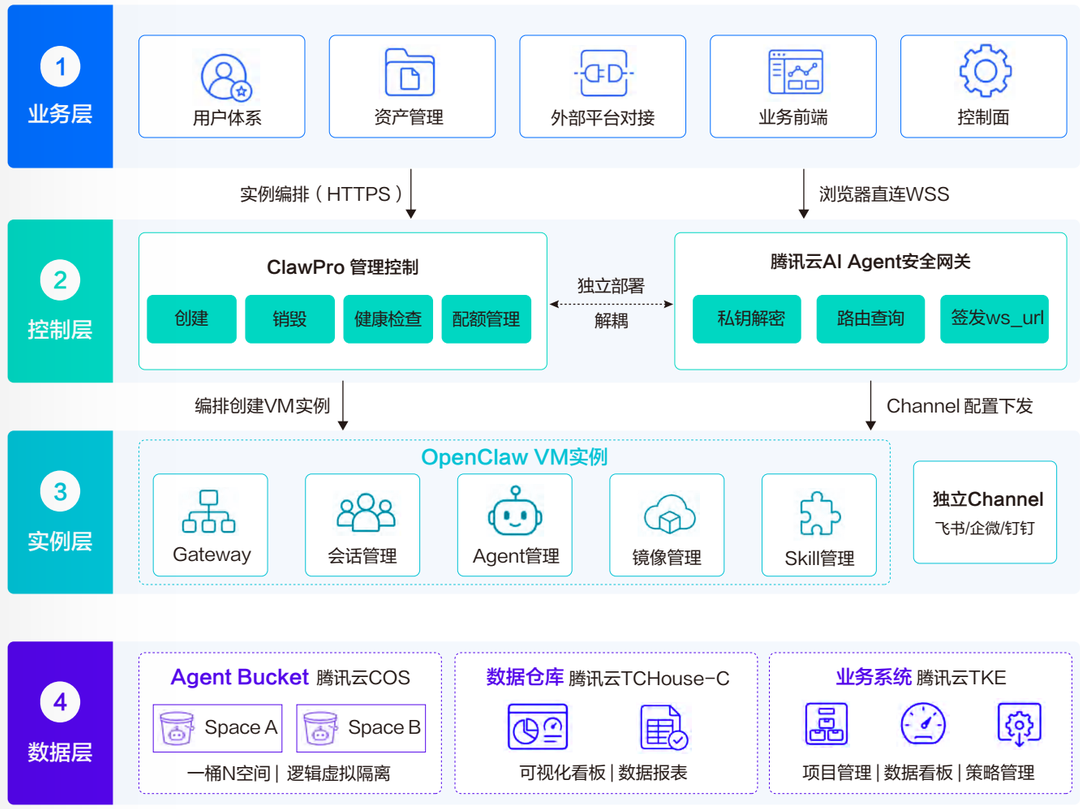
▶ 权限与安全

需在不同项目、客户之间实现安全的权限隔离、防止数据泄露；浏览器直连智能体还存在暴露实例真实地址的隐患，需要可控的安全边界。

▶ 文件操作不稳定

广告素材需经IM渠道转发给AI，流程不稳定、易丢失、难沉淀，客户希望为每个智能体提供固定路径、可长期存储、可安全分享的素材空间。

架构图



解决方案及客户价值

ClawPro实例编排

基于腾讯云ClawPro为每个用户独占分配一个智能体实例，生命周期受控，实例内再按会话做上下文隔离，多会话并行互不干扰，从根本上消除多项目并行时的上下文串扰。

AI Agent安全网关

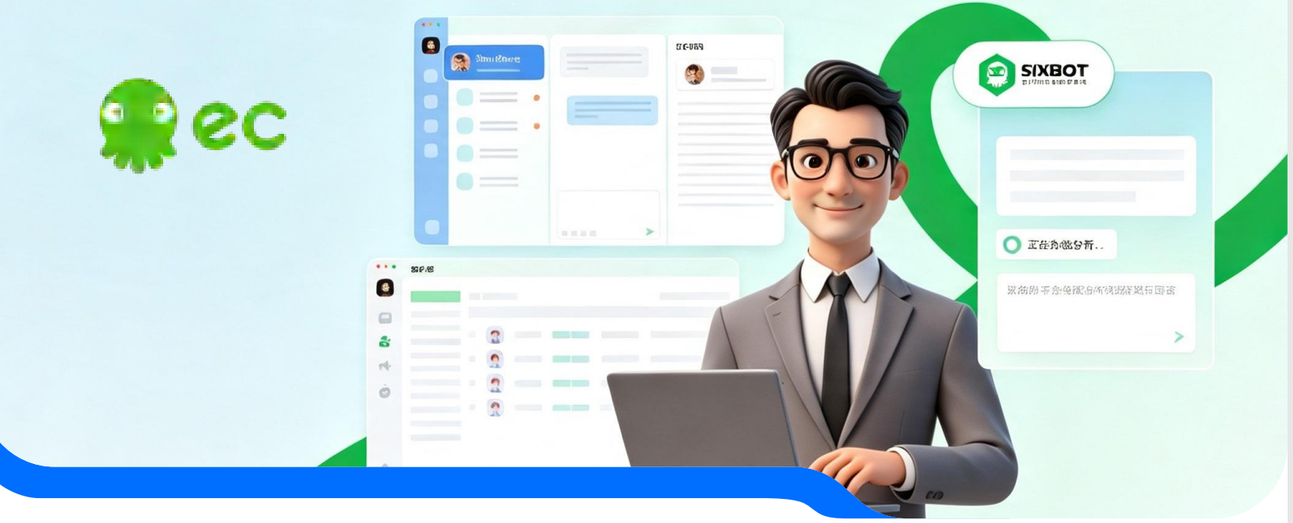
平台经腾讯云AI Agent安全网关与智能体通信，网关负责私钥解密、路由查询、签发短时效ws_url，浏览器全程拿不到实例真实地址，并为外部客户单独开辟接入通道，实现内外部分权分域，杜绝跨项目、跨客户的数据泄露。

Agent Bucket（COS 增强型桶）

采用对象存储COS的智能体桶，平台级一次性建桶，桶内以“空间Space”为逻辑隔离单元，每个智能体对应一个独立Space，单桶可承载亿级空间。

相关产品

ClawPro　TokenHub　Agent Bucket　AI Agent安全网关



标杆案例

六度人和

业务场景

六度人和是腾讯、用友、中信等多家知名机构投资的国家级高新技术企业，中国社交化客户关系管理的开创者与引领者。主营产品EC是一款AI驱动的CRM系统，每天支撑百万级销售谈客户、冲业绩，已助力企服、会展、教育、招商、制造、贸易、金融等多个行业，超过3万家客户实现增长。

业务挑战

▶ 获客效率瓶颈

人工手动找商机速度慢、覆盖窄，线索精准度低，获客成本高且难以规模化。

▶ 转化能力不均

销冠经验依赖个人悟性难复制，普通销售缺乏实时指导，客户跟进停留在表格管理而非精细化运营，成交率受限于个人能力天花板。

▶ 管理决策滞后

管理者深陷汇报审核、报表整理等杂务，团队人效看不清，问题发现靠事后复盘而非实时洞察，难以从“管人”转向“管增长”。

解决方案

• 业务增长

标准化销售获客流程规范，帮助销售更精准触达线索、提升转化。

• Agent云上组网安全

安全隔离：结合现有业务统一设计云上组网，严格控制沙箱网络边界，避免沙箱间及沙箱与原有业务的安全问题。

安全策略：通过安全组、CAM 权限、域名/IP 白名单、文件安全检测与行为意图分析，精细化管理访问策略。

• Agent 沙箱高效弹性

毫秒级弹性：Agent 沙箱提供虚拟机级环境隔离能力，同时具备毫秒级的启动、暂停恢复特性。

架构图



客户价值

全链路AI销售闭环

从获客（商机助手）→触达（话务专员）→转化（销售顾问）→管理（业务助理），AI覆盖销售全流程，不是单点提效而是系统性重构。

销冠能力产品化

将顶级销售的沟通模式、客户判断、跟进策略提炼为 AI 能力，让每个销售都有“销冠副驾”，实现经验从个人资产到组织能力的转化。

从经验驱动到数据驱动

客户画像360°自动生成、意向等级精准识别、团队人效实时可见、问题主动发现——用数据替代直觉，让获客、成交、管理三个环节都“有据可依”。

业务效果

30%
销售转化率提升

40%
成交周期缩短

相关产品

Agent Engine　Agent沙箱　TokenHub

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

光魔科技

基于腾讯云ClawPro托管 Hermes，打造Octask数字员工内容创作平台

业务场景

光魔科技（深圳光魔科技）是一家专注于AIGC内容创作的科技企业，通过自研生成模型与内容生产能力，为创作者提供连续图像、视频及多模态内容生成服务。

面向更广泛的内容创作者和C端用户，光魔科技推出Octask AI内容工作室：将AI能力具象化为可协作的“数字员工”，用户以Boss身份组建专属AI团队，在同一平台完成选题研究、文案撰写、图片生成、视频策划与内容交付。

光魔科技基于腾讯云ClawPro提供的实例交付能力与Hermes运行环境，完成Octask Agent产品的工程化建设；在基础环境之上，自主开发账户、积分、任务状态、场景工作流及多模型内容生产能力。

业务挑战

从单点生成工具迈向面向用户的 Agent 产品，光魔科技需要同时解决内容生产、运行环境和规模化运营三类问题：

▶ 内容创作链路长、工具分散

完整内容生产往往需要经历需求理解、资料收集、脚本编辑、分镜生成、图片与视频制作等多个环节。用户需要在多个工具之间切换，流程难以标准化，也难以沉淀个人方法论。

▶ Agent产品化工程门槛高

面向C端用户提供Agent服务，不仅需要模型调用，还涉及运行环境交付、用户隔离、Skill管理、长任务执行等完整产品能力。

▶ 复杂任务对稳定性与扩展性要求高

短任务可以在秒级完成，而视频等复杂任务可能持续数十分钟甚至更长。随着测试用户和Agent数量增加，平台需要兼顾常驻运行的稳定性与后续资源弹性。

解决方案

光魔科技基于腾讯云ClawPro与Hermes Agent环境，从以下三个维度构建Octask的产品能力：

▶ 标准化Agent运行环境

通过ClawPro完成云上实例交付和Hermes基础环境部署，为用户提供相对独立的常驻运行空间，支撑Agent持续执行长任务。光魔科技团队可将研发重点聚焦于业务体验与内容能力。

▶ 数字员工驱动的一站式内容工作流

Octask将不同内容场景封装为“数字员工”。一个场景Agent可从需求理解开始，连续完成脚本、分镜、图片、视频等步骤，并在工作台中呈现任务状态、中间产物与最终结果。

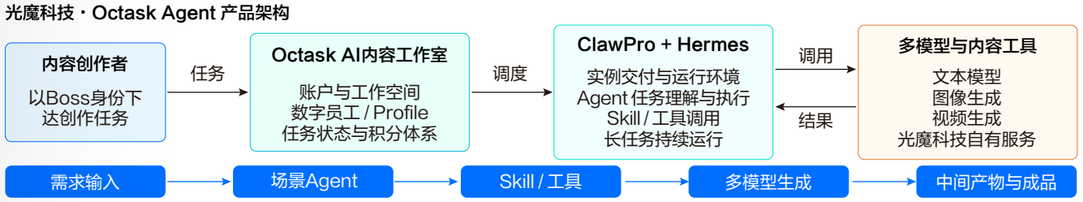
▶ 多模型与光魔科技能力协同

Hermes Agent通过Skill和工具调用衔接文本、图像、视频等多类模型服务，并接入光魔科技自有后台、账户与积分体系，形成从用户指令到内容交付的产品闭环。

架构图

整体架构以内容创作者发起任务为起点，经Octask工作台完成数字员工选择与任务调度，再由ClawPro上的Hermes Agent执行Skill、工具和多模型调用，持续返回中间产物及最终内容。

光魔科技 · Octask Agent 产品架构



① **用户发起：**内容创作者以Boss身份在Octask 下达创作任务。

② **Octask 调度：**平台完成账户、工作空间、数字员工、任务状态与积分管理。

③ **Agent 执行：**Hermes Agent理解任务，选择对应Profile、Skill与工具并持续运行。

④ **多模型生成：**调用文本、图像、视频模型及光魔科技自有内容服务。

⑤ **结果交付：**将脚本、分镜、图片、视频等中间产物与最终成品返回工作台。

客户价值

降低Agent基础设施建设门槛

ClawPro 提供实例与 Hermes 环境的标准化交付，减少从零搭建 Agent 运行底座的工作量，让产品团队能够更快开展场景验证和用户测试。

将创作方法沉淀为可复用数字员工

光魔科技把内容生产方法、流程与工具封装进场景Agent，使复杂创作任务由一个数字员工端到端执行，降低普通用户使用多模型工具的门槛。

为规模化运营预留资源弹性

按用户隔离的运行环境可保障测试阶段的稳定执行：随着用户规模扩大，可进一步结合资源预警与弹性方案，支撑从测试走向商业化运营。

业务效果

在测试阶段，Octask 已完成数字员工、内容工作流及 Hermes Agent 环境的产品联调，并形成面向真实用户的内容创作体验。

相关产品

ClawPro



标杆案例

天财商龙

业务场景

天财商龙深耕餐饮行业数字化二十余年，服务全国十余万餐饮门店，覆盖正餐、快餐、火锅、烧烤等全业态，提供餐饮、会员、供应链、智慧厨房一体化系统。

LongClaw是面向餐饮行业的AI数字员工，让每一家门店都能用自然语言驱动经营分析与业务执行。

业务挑战

▶ 数据查询门槛高

经营数据分散在各系统，店长、运营想查数要找人、等报表。

▶ 分析依赖人工

巡店、复盘、营销分析靠经验，难以规模化复制。

▶ 执行标准不一

集团策略到门店层层衰减，千店难同面。

▶ 安全合规要求严

经营数据是核心资产，AI落地必须先过安全关。

解决方案

• ClawPro 实例编排 —— 独立可控的智能体运行环境

基于 ClawPro 为 LongClaw 的每个租户独占分配智能体实例，生命周期由管控端统一编排（创建、健康检查、销毁）；实例内按会话做上下文隔离，多任务并行互不串扰；长任务可持续执行、异常可自动恢复，支撑经营分析、智能补货等复杂任务稳定运行。

• AI Agent 安全网关 —— 出入流量统一管控、凭证零暴露

智能体经腾讯云 AI Agent 安全网关接入，由网关在通信链路中统一承担鉴权、解密、路由与审计：客户后台签发短时效访问凭证，网关侧解密后按实例 ID 动态查询真实连接地址并代理转发，实例真实地址与密钥全程不出网关；同时叠加 ClawPro 三级 RBAC、VPC 与安全组隔离、对话 I/O 全链路审计及 Skill 供应链扫描，高危指令实时拦截。

• 弹性调度 + 精细成本管控 —— 按需伸缩、成本可预测

ClawPro 管控端统一负责实例的创建、销毁与配额管理，基于腾讯云计算与容器调度能力随业务负载快速扩缩容，支撑午晚市高峰、月末结账等集中请求的并发处理；配合 Token / Agent 三级配额、用量预警与按组按人核算，让常驻型 Agent 的资源投入与业务量匹配。

• 企业专属技能库 + 三级权限分组 —— 一套底座、多级复制

ClawPro 兼容开源、自研与自定义 Agent，LongClaw 可直接接入并统一纳管；依托企业专属 Skill 库与 MCP 私有库实现能力批量下发，配合部门 / 项目 / 个人三级权限分组与组织架构映射，支撑总部—品牌—门店的分级落地。

架构图

LongClaw 企业级AI数字员工能力全景

让每一位员工都拥有懂餐饮 、能分析 、会执行的 AI 专家团队

LAYER 01 多端入口 随时随地 开口即用

- 第三方入口：企微 / 钉钉 / 飞书
- 商龙入口 — 龙管家 · 手机端：店长与一线员工移动使用
- 商龙入口 — 商龙云2.0 · 网页端：LongClaw统一工作台

LAYER 02 产品能力 可直接使用的能力合集

LongClaw 工作台： 统一完成任务发起 、专家调用 、技能配置 、自动化与结果查看。

- 对话任务工作台：选专家 / 选模型 / 选技能；上传附件 / 实时查看结果
- Claw专家中心：办公秘书 · 经营分析师；私域营销专家 · 企业自建专家
- 技能市场：商龙官方 · 第三方 · 企业自建；安装启停 · 安全审计
- 自动化任务：定时执行 · 多端推送；启停管理 · 执行历史
- 主流模型接入：Hy3 · DeepSeek · GLM；Kimi · MiniMax · Qwen
- 企业知识平台：知识文档 · 集团/品牌/门店分级；权限可控

LAYER 03 腾讯龙虾 ClawPro 可信智能体运行底座

- OpenClaw：开放生态与通用任务承载
- Hermes：复杂任务理解与工具协作
- LightClaw：轻量 、安全 、面向企业场景运行

LAYER 04 商龙业务连接 CLI+MCP

连接商龙现有业务系统，让 AI 从 “会回答” 走向 “能查数 、能分析 、能执行 ”。

- 餐饮经营：门店 / 订单 / 菜品 / 收银
- 会员营销：画像 / 权益 / 复购 / 预警
- 供应链：采购 / 库存 / 成本 / 损耗
- 智慧厨房：出品 / 厨显 / 后厨效率
- 经营大数据：指标 / 看板 / 趋势 / 预估

LAYER 05 商龙AI中台 沉淀餐饮 行业智能

- 商龙行业知识沉淀：餐饮经验与最佳实践
- 算法中台：预测 、推荐与经营算法
- 模型中台：多模型统一选择与治理
- 智能体中台：专家创建 、配置与运营
- 技能中台：技能建设 、发布与审核

LAYER 06 云基础设施 稳定、安全、可扩展

- 腾讯云计算
- 存储与网络
- 安全防护
- 监控与审计
- 弹性与容灾

企业级安全贯穿全链路

身份与权限可控 · 技能接入可审计 · 关键操作可确认 · 任务过程可追溯

客户价值

稳定运行

智能体7×24在线、长任务可续跑，客户无需自建运维体系。

安全合规

Agent 全链路行为可管可控可追溯，真实密钥零暴露，满足连锁企业审计要求。

弹性降本

算力随经营高峰按需伸缩，常驻 Agent 成本可控，投入与业务量匹配。

快速复制

一套底座支撑多品牌多门店分级落地，新场景以 Skill 批量下发，上线周期显著缩短。

相关产品

ClawPro

更多免费行业报告 并购家 www.ipoipo.net

05

业务创新—内容生产

AI驱动规模化内容生产与创意升级

方案

5.1 营销图片生产方案 / P81

5.2 营销视频生产方案 / P82

5.3 营销素材管理方案 / P84

5.4 游戏美术智能体创作方案 / P85

5.5 AI Agent加速游戏素材全链路提效 / P87

案例

标杆案例——筷子科技 / P89

标杆案例——LiblibAI / P91

标杆案例——必凡娱乐 / P93

标杆案例——玩心不止 / P95

标杆案例——紫讯 / P97

标杆案例——小影科技 / P99

标杆案例——懂球帝 / P101

5.1 营销图片生产方案

业务场景

在信息爆炸的时代，消费者的注意力转瞬即逝。文字需要逐句理解，而一幅精心设计的营销画面，能在0.1秒内直击人心，成为决定营销成败的“第一触点”。而诸多营销环节对于营销画面的要求各不相同：

- **品牌营销活动：**在品牌新品发布与节日促销等高强度营销活动中，为确保传播效果，需要快速产出海量且多维度、高质量的视觉素材，并确保其风格与品牌调性高度一致。
- **社交媒体营销：**为维持社交媒体账号的活跃度与用户粘性，运营团队必须持续不断地创作并发布具有强烈视觉冲击力、能够瞬间吸引用户眼球的图片与视频内容，从而有效维持用户的长期关注与互动参与。
- **广告投放素材：**在程序化广告投放中，为精准定位目标受众并实现最优转化，必须通过海量多元的创意素材进行持续测试与数据对比，从而不断优化投放策略并提升整体营销效果。
- **电商平台商品展示：**电商平台需要大量高质量商品图片来吸引消费者，包括主图、详情页图片、场景图等；特别跨境电商平台面临更复杂的场景，同一商品需要针对不同地区生成符合当地审美的图片。

业务挑战

成本高昂

传统的商拍模式链条长、硬性成本高，从场地器材到模特、妆造与后期，每个环节都产生不菲开销。多方投入累积之下，导致即便是一次常规的拍摄，整体费用也动辄数以万计，令营销预算水涨船高。

效率有限

创意工作长期依赖少数高素质人才，而海量信息的内容梳理、甄别与重构更是重人力的密集型劳动。这种传统模式导致响应迟缓、产能受限，现有工作效率已完全无法匹配业务对高频、海量内容生产的刚性需求。

合规隐患

当前，海量素材的运用使得传统人工审核方式出现漏洞，导致未授权素材被误用的合规风险激增。这不仅会带来法律诉讼与财务损失，其引发的负面舆情更将直接冲击品牌价值，凸显出现行管理模式的巨大挑战。

解决方案

营销图片素材生成方案：使用者仅需文字撰写图片要求，即可秒级生成无版权风险的营销图片。技术方案：以多模态图像生成模型为基础，结合爆款图像素材的细节描绘 Prompt，实现目标图像的批量生成。

Prompt（提示词）：

图片呈现出简洁而专业的风格，主体是一位穿着米白色缎面衬衫的男模特，坐在纯色（浅灰或白色）的背景前，姿态优雅自然。

» 人物与动作

人物的头发梳成低马尾，显得干练整洁；身穿的缎面衬衫质感柔和、光泽细腻，领口微微敞开，整体造型简约大气，传递出一种精致且不失亲和的气质；人物的右手抬起，手掌轻握着一瓶护肤品，左手则自然地放在身前的白色平面上（可能是桌面），动作舒展流畅，给人一种从容自信的感觉。

» 护肤品细节

被展示的护肤品是一瓶精华类产品：瓶身为金色长方体形态，搭配棕色金属质感的按压泵头，设计精致且有质感；瓶身上印有黑色文字，能看到“玻尿酸保湿精华”以及“30ml”的清晰字样，清晰传递出产品的核心信息（成分、功效、容量）。

» 整体氛围与背景

背景采用了极简的浅色调（接近纯白或浅灰），没有多余装饰，让视觉焦点完全集中在人物与手中的护肤品上；画面光线均匀柔和，既突出了人物与产品的轮廓，又营造出一种高端、清新的广告感，很符合美妆护肤类宣传物料的调性。



（生成图示意）

5.2 营销视频生产方案

业务场景

在当下注意力稀缺的数字环境中，营销视频已不再是可选项，而是决定品牌能否突围的战略核心。它以其强大的叙事能力和沉浸感，直接贯穿于用户从认知到转化的全链路。

在触达层面，动态视频在信息流中占据绝对视觉优势，能在3秒内抓住用户眼球，传递品牌信息。在沟通层面，视频集画面、声音、情节于一体，能生动展示产品功能、讲述品牌故事、营造情感共鸣，高效构建用户信任与品牌认同。在转化层面，从直播带货的实时互动到短视频的“种草”引导，视频能直接缩短用户的决策路径，创造“即看即买”的消费体验，成为驱动增长的直接引擎。

而诸多营销环节对于营销视频的要求各不相同：

- **产品发布视频：**为达成最佳传播效果，需制作融合3D动画、实拍、特效的高质量视频，以创造沉浸式体验，直观传递产品功能、使用场景与品牌核心价值。
- **社交媒体营销：**为全面覆盖主流社媒渠道，需针对性制作两类视频内容：(1)适用于抖音、视频号、快手的15-60秒竖版短视频。(2)适配B站、YouTube平台的1-3分钟横版深度视频，以满足不同场景下的用户观看习惯与平台内容偏好。
- **电商营销视频：**为全面提升产品营销效果，需要系统化制作三大类核心视频内容：(1)360°全方位产品展示视频，以动态形式清晰呈现产品设计与细节。(2)直观的功能演示视频，精准展现产品核心卖点与操作流程。(3)真实的使用场景视频，通过情景化叙事让用户深度感知产品价值与生活方式关联。
- **广告投放素材：**为确保广告效果，6-15秒的前贴片广告面临着极为严苛的注意力竞争，其开场3秒必须通过强烈的视觉冲击、情感共鸣或悬念设置等高效手段，瞬间抓住用户注意力，有效降低跳过率，为后续品牌信息的传递创造宝贵机会。

业务挑战

成本高昂

传统视频素材制作流程繁琐，常面临优质模特与场地预约周期长、海量原始素材筛选耗时费力等问题。整个过程不仅耗尽人力，更导致预算持续吃紧，素材制作的时间与经济成本急需控制。

创意紧缺

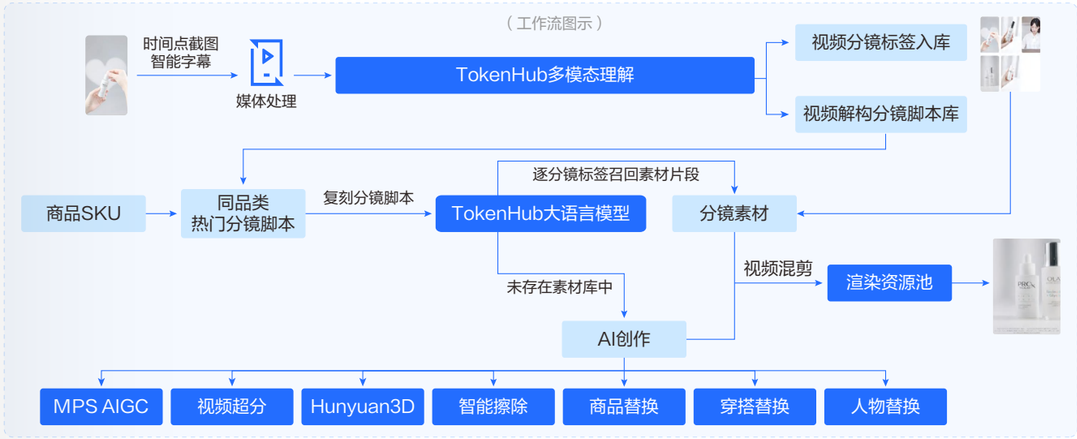
爆款视频的诞生离不开优秀的创意人才，但团队核心创作力的不足，使得高质量内容生产充满偶然性，无法稳定、规模化地产出爆款，严重制约了营销增长。

版权风险

海量素材可能涉及模特肖像、艺术品等第三方元素带来的版权风险，遗漏识别可能为平台或品牌带来负面影响和法律纠纷。

解决方案

视频素材解构方案：基于腾讯云媒体处理服务进行视频预处理，结合多模态图片理解模型及推理模型进行视频素材的结构化解析，实现爆款视频的精准分析，帮助创意团队理解优质内容，学习所长，提高内容制作能力。



更多免费行业报告 并购家 www.ipoipo.net

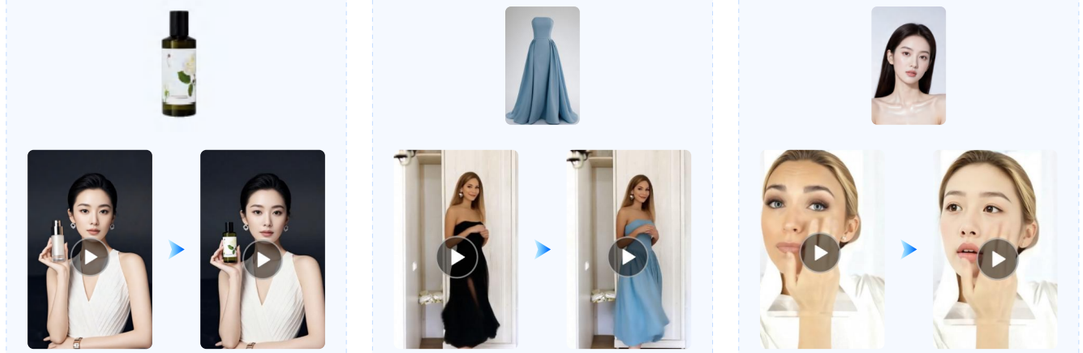
视频智能制作解决方案：(1)基于标签检索分镜脚本素材库实现视频片段召回进行素材复用。(2)基于多模态生成模型进行AI原生素材片段生成。(3)基于视频编辑能力进行素材二次创作。最后通过云上渲染资源池的多元算力进行最终混剪视频生成。

效果示意

- **效果一：**提供图片、视频、配音、音乐生成式AI能力，结合媒体处理能力覆盖内容制作全流程



- **效果二：**一次拍摄即可多次复用，极大减少反复实拍带来的时间和经济成本压力



- **效果三：**AI快速生成动态3D模型，降低技术门槛，高效赋能内容创作



5.3 营销素材管理方案

业务场景

在数字化营销时代，营销素材已成为营销类SaaS企业长期沉淀的核心战略资产。它不仅承载着品牌的内容价值，更构成了企业服务差异化竞争力的基础。

随着业务发展，企业通过多源渠道持续采集每日热点素材，同时平台用户也在不断创作，沉淀新的素材内容，因此素材库以指数级增长，亟需对海量的图片、视频等多素材进行规模化存储与系统化管理。

素材存储系统需要具备高可靠性、低成本与弹性扩展能力，以应对持续增长的非结构化数据存储；同时需要科学的管理体系能够实现素材的智能分类与高效复用，从而最大化素材资产的复用价值。

业务挑战

运营及成本压力

随着业务规模的不断扩大，企业素材库已快速积累至PB级别，庞大的存储需求直接带来了巨大的成本压力。这种压力不仅体现在硬件投入和云存储费用上，更涉及备份、容灾等一系列衍生成本，给企业的运营效益带来了持续性的负担。

资源浪费

由于缺乏体系化管理，这些珍贵的素材资产往往散落在不同的存储系统中，形成了诸多“数据孤岛”。大量高质量的素材因难以被快速检索和有效利用，实际上陷入了“价值黑洞”的困境--它们既占用了宝贵的存储资源，又无法为营销创作提供应有的价值支撑。

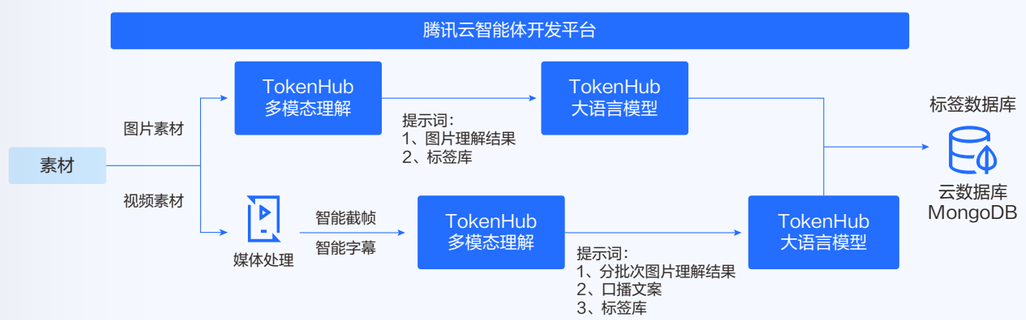
解决方案

素材采集存储方案：通过云函数及云手机等构建数据采集集群，素材文件存储在低成本的对象存储中，文件元数据通过腾讯云数据库MongoDB进行管理。



（工作流图示）

素材标签分类方案：基于腾讯云智能体开发平台构建自动打标签工作流，图片、视频素材落地对象存储触发打标工作流，结合多模态大模型以及推理模型自动打标分类，入库标签数据库；基于云数据库MongoDB构建的KV标签库满足标签动态新增调整的场景。



（工作流图示）

5.4 游戏美术智能体创作方案

业务场景

美术资产是游戏研发投入占比最高、迭代最频繁的环节，贯穿世界观设定、角色原画、UI 到序列帧与 3D 资产的全生命周期。AIGC 进入游戏美术以来仍停留在“单点生成”阶段：能力散落在几十个工具与插件之间，创意在跨软件搬运中反复损耗，美术团队亟需从重复劳作中解放，专注于风格创新与艺术表达。

业务挑战

▶ 单点生成缺乏工业级控制力

主流AI出图本质仍是“抽卡”，难以稳定命中项目的风格设定、视角与材质要求，产出是“图片”而非可直接交付的“数字资产”；角色三视图、序列帧、动态立绘等标准化内容仍依赖手工绘制与拼接，重复劳作持续挤压核心创意时间。

▶ 上手门槛高、经验难沉淀

复杂的节点工作流与参数调优只有少数专家能驾驭，策划与非美术岗位难以参与创作；主美的审美与方法论停留在个人手中，人员流动即意味着团队能力清零，无法形成可复用的企业创意资产。

▶ 孤岛式工具导致创作管线断裂

草图、生图、渲染、剪辑、建模分散在多个软件，从原画到多视图、序列帧的每一次交接都伴随格式转换与返工；活动页、官网等UI与运营物料从设计稿到开发实现仍靠人工标注与切图，AI在前端带来的效率增益在跨端搬运中被消耗。

解决方案

- **Agent 驱动、能记住风格的 AI 原生创意工作室：** Miora 以一块无限画布承载图像、视频、3D、UI 的全部创作，Creative Agent 读懂画布上下文，调度分镜、插画、UI/UX、视频、3D 等 Specialist 协同作业；Agent Memory 将项目的风格规范、禁忌与参考图沉淀为项目记忆，每次生成自动套用，风格一致、无限变体。局部重绘、六视图一键生成、动作姿势生成让每一步可介入、可回退；产出直接导出为透明底 PNG、SVG、GLB/OBJ 与 HTML 原型，交付可进入研发管线的数字资产。
- **Skills 开放生态与游戏专属能力沉淀：** 以自然语言对话替代节点编排，把上手门槛从“懂工具”降到“会表达”，策划、运营也能直接参与创作；主美的方法论通过上传 SKILL.md 或对话创建固化为可复用、可分享的 Skill，团队继承、市场发布。2D 角色序列帧 Skill 打通“参考图 → 多方向多视角生图 → 图生视频 → 云端抽帧 → 智能抠图”五步流水线，覆盖 ARPG 与像素俯视角等品类，单角色全套动作从 1–2 天压缩至约 10 分钟（纯生成时间）；角色多视图、分镜视频、资产量产、资产评审等 Skill 按需装载。
- **设计到开发的全链路落地：** 在 Miora 交付创意资产的基础上，由腾讯设计 Ardot 承接界面与运营物料的工程化落地：文生 UI 描述即生成布局与组件，上传图片智能分层秒变可编辑设计稿；组件、颜色、文本样式经设计系统全局同步、一键复用，Grid 布局适配多尺寸屏幕；开发直接查看标注并复制 CSS / iOS / Android 代码，游戏活动页、官网、角色介绍页与运营素材从设计稿到开发全链路闭环。企业级权限中心与操作日志追溯保障设计资产安全。

架构图



相关产品

Miora 无限画布 + Creative Agent + Skills 开放平台	**Ardot** 文生 UI、设计转代码一键交付	**游戏行业专属Skill** 2D角色序列帧生产 、游戏 UI 拆解 Skill 等
MPS 智能抠图、去水印、超分等 AI 能力	**TokenHub** 生文 / 生图 / 生视频等 AIGC 大模型	**3D模型** Miora 3D、混元 3D 等

更多免费行业报告 并购家 www.ipoipo.net

5.5 AI Agent加速游戏素材全链路提效

业务场景

竞争激烈的游戏市场中，高效、大规模、高频次的广告投放是获取用户的核心手段。随着AI内容创作能力普及，素材制作门槛大幅降低，海量素材版本快速涌入投放渠道，游戏买量已从“数量战”全面转向“质量战”——行业瓶颈正从“生产能力”转向“营销效果”，头部厂商需在海量素材中快速甄别爆款、沉淀创意资产并持续迭代。面对创意过剩而玩家注意力稀缺的新格局，如何借助AI对素材进行智能理解、打标、归因与二创，实现“分析—生成—投放—复盘”的规模化提效闭环，已成为投放团队保障ROI的关键。

业务挑战

▶ 海量素材“管不过来”，创意归因成黑箱

游戏买量每天产出、接触的素材量极大，但这些素材的处理、理解、整理、归类高度依赖人工，消耗投放团队大量精力；更关键的是，海量素材缺乏系统化的管理与分析，成功的创意元素难以被沉淀、复用，很难有清晰的素材归因洞察来反哺优化。

▶ 制作成本高、周期长，难匹配快速迭代节奏，且灵感易枯竭

游戏行业是“快速迭代 + 海量测试”的投放节奏，而传统人工制作素材在成本和周期上都很难跟上，同时创意团队长期面临灵感枯竭，素材生命周期越来越短，只能靠不断的高成本试错去撞新的爆款方向，缺乏数据驱动的高效裂变与二创手段，产能与创意质量两头都被卡住，制约规模化增长。

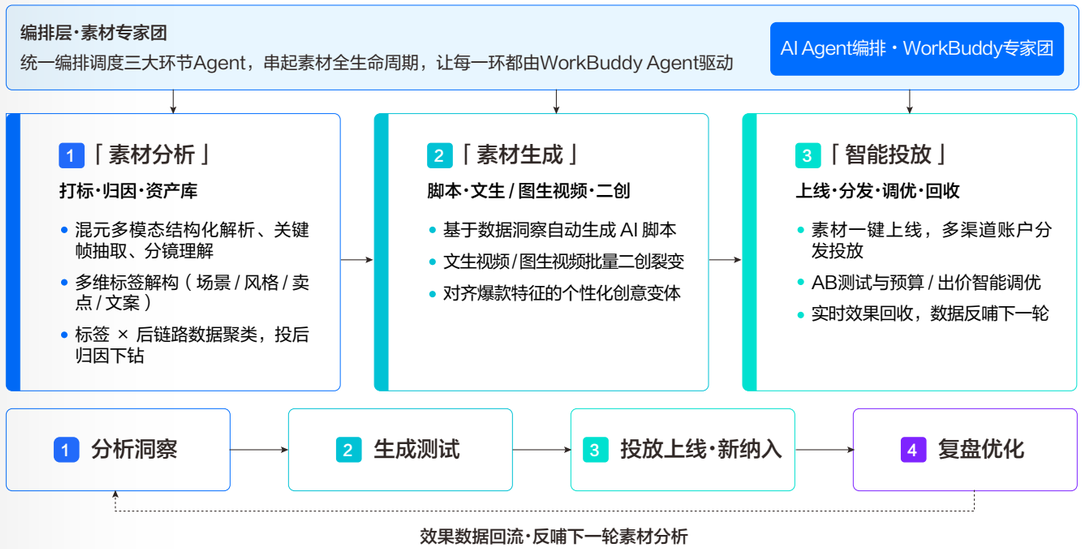
▶ 多渠道实时竞价靠人盯不过来，规模与精准双双受制于人力

游戏买量涉及多渠道并行，且竞价、出价、预算需实时调整，靠人工在后台逐条建计划、圈人群、盯盘调价，既反应慢半拍、又极易出错，投放规模的天花板被人力死死卡住——“想加量就得加人”，人力近乎线性增长，难以规模化扩张；同时创意元素与投放效果之间缺乏数据打通，投放决策高度依赖投手个人经验，做不到7×24小时不间断的精准优化。

解决方案

- **广告素材理解：**依托腾讯游戏云对游戏素材场景的理解，构建一套：自动完成智能转码、音视频对齐、智能关键帧抽取与分镜理解，并从场景、风格、核心卖点、玩法、BGM 等多维度对视频、图片、文案、音频统一标签化解构，将非结构化素材转化为可复用、可检索的结构化创意资产库。
- **数据化归因洞察：**将结构化创意标签与广告后链路性能数据（CTR、CVR、ROAS）深度关联，以媒体投放数据为权重做标签聚类与共性分析，并支持下钻到分镜层级持续打标、挖掘同类爆款素材的共性规律。这让创意元素（角色、场景、文案、风格）与投放效果之间原本的“黑箱”被打开，归因从“哪条素材跑量”深化到“是哪个元素在起效”的颗粒度，形成可指导下一轮创作的精准投后归因。
- **智能化创意生成：**视频分镜理解自动生成 AI 脚本，并调用文生视频、图生视频等生成式AI能力，围绕已验证的爆款特征快速批量生成个性化创意变体；同时支持在人工确认初稿后做局部修改（只改情节/台词/场景、不动整体节奏），兼顾“量”与“质”。产出交由素材专家团统一编排，形成“分析洞察 → 生成测试 → 投放复盘 → 效果优化”的高效迭代闭环，从源头系统性拉高广告投放整体ROI，也把创意团队从“灵感枯竭式试错”中解放出来。
- **AI Agent贯通全链路提效：**AI Agent从“造素材”延伸到“投素材”，打通生产到投放的全部环节，通过投放智能体替代繁琐的后台点击，一句话即可批量创建、执行投放计划，让投放规模不再受“加量就得加人”的人力天花板制约，并且根据素材归因的洞察和投放策略规则匹配，不断优化素材与渠道的匹配、出价和预算。

架构图



相关产品

WorkBuddy　ClawPro　TokenHub　ASR　ES

COS　MPS



标杆案例

筷子科技

业务场景

筷子科技（kuaizi.ai）是服务于内容商业生态的智能创意技术提供商，基于内容元素解构方法论、AI、云计算、创意内容大数据等核心技术，通过创意智能生产、运营优化、标签洞察、协作管理的一站式SaaS解决方案，链接全球数字化内容商业生态全链路，加快上千万品牌商家、互联网企业及内容服务商实现商业增长。

业务挑战

- PB级海量素材存储与管理的成本以及性能挑战。
- 在视频处理环节，高并发的视频渲染与处理任务对算力提出了极高要求。特别是在营销活动高峰期，同时处理数百个4K视频渲染任务时，传统计算资源往往出现性能瓶颈，导致项目交付延迟，直接影响营销效果。
- 在内容创作维度，如何通过AI技术实现规模化、批量化的素材创作成为关键课题。这要求企业不仅要掌握多模态生成技术，还要建立标准化的创作流程，在保持品牌调性统一的前提下，快速生成符合不同渠道需求的个性化素材。





生机（shengji.co）



AI影棚（airset.ai）

解决方案

- 高性能文件存储CFS+对象存储满足不同性能要求的素材文件读写需求。
- 通过弹性容器集群TKE动态弹性纳管CPU、GPU异构资源池，满足高峰期的视频处理任务算力需求。
- 多模态大模型辅助素材的批量创作。

客户价值

冷热分层存储架构有效控制素材存储成本，整体存储成本降低超 30% 。

qGPU方案显著提升了算力资源利用率与灵活度，常驻算力减少至 50% ，仍可保障视频处理任务稳定运行。

AI素材创作显著提升营销素材制作效率。



更多免费行业报告 并购家 www.ipoipo.net

LibLibAI



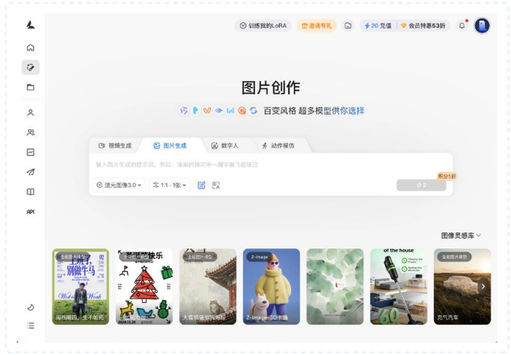
标杆案例

Liblib

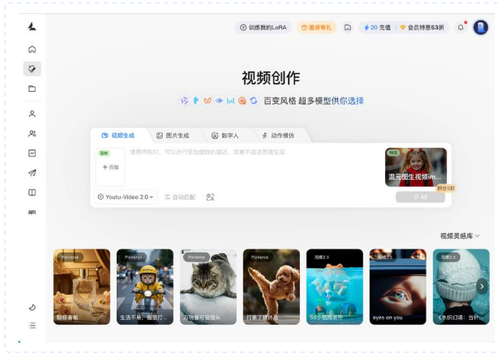
LiblibAI是中国领先的多模态 AI 创作与模型分享平台，致力于专业场景下的AI内容创作与分享，为广大创意工作者提供专业、可靠、高效的AI时代创作工具。LiblibAI创作者已突破三千万，日均保持数百万级创作交互，累计生成图片逾10亿张。

业务场景

图片生成器： Liblib 图片生成器是 Liblib AI 推出的 AI 绘画工具，基于多模态大模型，支持文生图、图生图、风格迁移等功能，可快速生成高精度个性化视觉作品。操作简单、模板丰富，覆盖插画、设计、艺术创作等场景，助力设计师、爱好者高效实现创意构想，让专业级绘图触手可及。



视频生成器： Liblib 视频生成器是 Liblib AI 推出的 AI 视频创作工具，基于多模态模型，支持文生视频、图生视频等功能，可一键生成动态短片、创意动画等内容。操作便捷、效果流畅，覆盖宣传、教学、娱乐等场景，助力创作者高效实现视频创意，让专业级动态影像制作更简单。



业务挑战

- 业务快速增长，自研和客户自训练模型众多，推理场景多样化以及资源需求量大，研发和运维人力不足。训练和推理能力决定产品质量及用户体验，需持续优化推理时间及降低训练推理成本。
- 内容创作平台需要持续更新及接入行业领先模型，保证创作平台竞争力。

解决方案

模型训练和推理	AI素材管理
云服务器、弹性容器集群TKE、GPU服务器	对象存储 COS、文件存储 CFS
AI素材生成	**AI素材处理**
Tencent HY、腾讯优图	媒体处理MPS

- **TKE容器化能力实现高效运维：** TKE提供一站式容器化平台能力，纳管多种GPU卡型进行管理，初始化脚本等能力自动化安装及升级驱动；qGPU能力可以做GPU卡切分，最大化利用显存提升资源利用率；批量设置自定义参数等实现高效运维。
- **高性能集群和存储助力训练推理：** H20高性能集群提供3.2T背板带宽，多机并行通信提升效率；CFS Turbo高性能存储提供高吞吐及高IOPS并行文件系统，助力多机训练。
- **多模态模型及媒体处理助力素材生成：** 混元3D、混元生图3.0和优图视频模型助力Liblib AI素材生成，通过媒体处理超分、降噪和效果增强能力，实现AI素材的效果优化。



AI素材处理效果

客户价值

训练推理平台一体化，减少数据传输时延，提升模型训练及迭代效率 **20%** 。

多模态模型及媒体处理助力高质量素材生成，加速创新速度。视频分辨率从720p超分至1080p，在效果优化的基础上降本 **50%+** 。



标杆案例

必凡娱乐

业务场景

深圳必凡娱乐科技（Beefun）是一家专注微信小游戏自研自发、以 Merge（合成融合）为核心赛道的快速成长型游戏公司。团队规模200余人，秉持“自研自发”战略，已成长为国内合成小游戏赛道的头部玩家。

其自研二合（Merge-2）标杆小游戏《改装大作战》于2023年推出，上线首月即登顶微信小游戏人气榜、跻身畅销榜 TOP 10，最高日流水突破百万；上线两年多来长期稳居畅销榜 TOP 100，累计用户过亿，累计流水突破8亿元。

在长线运营中，必凡把“用 AI 提效”作为内容生产与团队协作的核心杠杆——聚焦如何以腾讯云 AI 产品矩阵（TokenHub、MPS 媒体处理、腾讯AI办公助手 WorkBuddy），系统性解决美术、广告创意与日常运营环节的海量内容生产与组织协作难题。

业务挑战

爆款微信小游戏的长线运营，对研发内容产出、买量营销与跨岗位协同提出了极高的敏捷度要求，传统工作流在以下两方面遭遇瓶颈：

▶ **重度买量与高频内容迭代对素材产能的极限拉扯**

小游戏买量生命周期极度依赖海量宣传素材的持续供给，同时版本需要高频更新玩法、活动与欧美卡通视觉剧情。传统人工绘制、剪辑周期长、成本高，素材产能极其容易成为版本和投放节奏的瓶颈。

▶ **跨专业/跨岗位的日常运营重复性劳动负担**

在游戏直播、创作者运营、知识沉淀等高频业务中，运营、策划、运维岗位沉淀了海量碎片化、重复性的事务（如多渠道中奖记录交叉核对、薪酬手工结算、文档整理等），不仅耗费大量人工，且极易出现人为漏发或算错，挤占了团队用于高价值创意与商业决策的时间。

解决方案

必凡娱乐联合腾讯云，以“模型底座-内容生产-应用入口”为核心链路，搭建了服务全团队的三层 AI 生产力底座：

TokenHub（模型调度层）：企业大模型统一底座

一站式接入并调度多家主流大模型，提供统一接口、按场景智能路由与统一计量能力。将不同模型的能力平滑转化为团队随手可调用的通用基础设施，显著降低大模型接入门槛与调用成本。

- 多模型统一接入：一套接口对接主流模型，免去研发重复对接成本。
- 场景智能路由：根据任务类型、效果表现与算力成本，自动路由至最匹配、性价比最高的大模型。
- 用量统一计量：各岗位调用明细集中可见，成本完全可见可控。

MPS 媒体处理（内容生产层）：美术与广告内容工业化生产

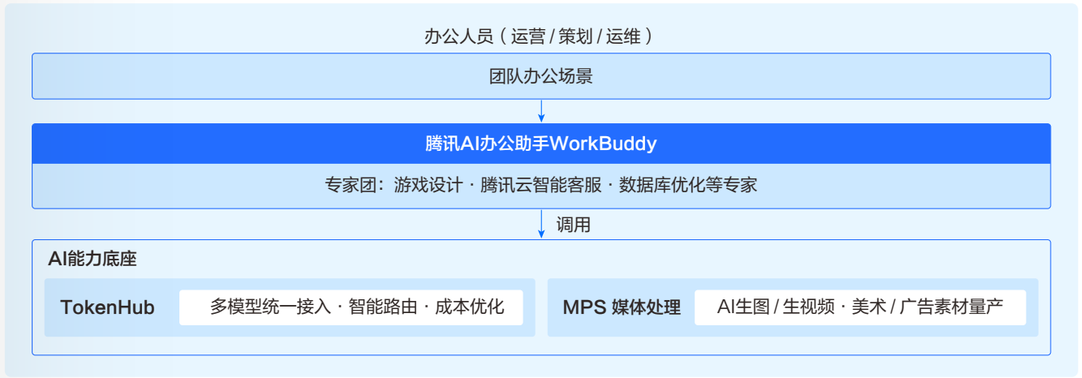
基于 AI 生图与生视频能力，规模化、弹性化量产游戏美术资产与广告创意，让版本迭代与买量投放节奏更快、单位成本更低。

- AI 生图：支持欧美卡通风格的角色立绘、场景原画、活动 banner 批量高质量出图。
- AI 生视频：玩法演示、剧情短片、广告素材快速自动化生成。

腾讯AI办公助手WorkBuddy（应用入口层）：全岗位 AI 协作入口

作为团队日常办公的统一 AI 入口，内置游戏设计、腾讯云智能客服、数据库优化等多领域专家，底层灵活调用 TokenHub 与 MPS 的 AI 底座能力，实现日常运营事务的自动化。

架构图



客户价值

通过“大模型底座 + 业务场景”的深度融合，必凡娱乐实现了从“单点 AI 尝试”到“企业级全场景 AI 增效”的价值跃升：

内容生产工业化升级：破局买量与迭代瓶颈

依托 MPS 媒体处理的 AI 生图与生视频能力，必凡打破了传统人工制作的产能瓶颈，实现了美术立绘与广告素材的规模化量产，广告买量与版本迭代整体节奏显著加快，单位素材成本大幅下降。

大模型能力统一管控：性价比与灵活性最大化

通过 TokenHub 这一集中式大模型调度底座，研发团队彻底告别了多模型单独对接的繁琐流程，不仅保证了技术底座的极高灵活性，更实现了按场景智能路由（高复杂度任务调大模型、轻量化任务调小模型），大模型调用成本得到系统性优化。

组织协同与日常运营敏捷响应：释放团队创造力

通过 WorkBuddy 的能力编排与日常应用，运营团队在创意、任务和知识三个维度实现了“人工劳动向 AI 自动化”的平滑过渡，日常繁琐业务的处理效率呈数量级提升，团队得以更专注地投入到核心玩法和商业变现的创新中。

相关产品

WorkBuddy　TokenHub　MPS

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

玩心不止

业务场景

玩心不止集团成立于2018年，是传奇 IP 授权与买量发行核心企业。公司聚焦传奇品类小游戏的研发与全球发行，代表作《烈焰觉醒》、《约战沙城》、《太古封魔录2》等多次登顶小游戏买量强度榜 TOP 10，全球注册玩家超1.4亿、海外用户超8000万。

作为典型的买量驱动型发行商，玩心不止每日需对海量竞品与历史跑量素材进行分析复盘，素材分析的深度与效率直接决定买量 ROI。通过使用腾讯游戏云素材分析方案，将素材打标、共性分析到创意二创的全流程从人工经验判断升级为多 Agent 协同的标准化 AI 流水线，单条素材分析效率提升10倍以上。

业务挑战

▶ **素材分析链路割裂，全流程难以贯通**

玩心不止的素材分析涉及竞品 / 历史素材获取、内容理解、多维打标、共性分析与创意二创等多个环节，但各环节工具分散、数据不互通——头部竞品素材常因平台安全机制难以一站式抓取接入，分析结果也难以沉淀复用，全流程始终无法形成贯通的闭环。

▶ **多模态分析门槛高，人工拆解效率低**

广告素材需同时从素材类型、开头钩子、核心卖点、画面与话术亮点、BGM 等多维度拆解，纯人工逐条分析单条耗时数十分钟，面对每日数百条的测试素材量，复盘严重滞后，难以快速沉淀爆款规律指导新素材生产。

▶ **自建智能体平台屡屡踩坑，难以稳定落地**

玩心不止曾尝试自建智能体平台（单模型问答→向量知识库→多智能体群聊），却接连遭遇模型能力不稳定、知识库易被污染、多智能体协同频繁出错且无法定位修复、依赖外部平台难以排查问题等瓶颈，AI 能力始终无法稳定支撑业务。

▶ **工程化与迭代门槛高，AI 能力难持续运营**

从模型接入、Agent 编排到持续调优，自建与自运维一套多 Agent 系统的工程化门槛极高，团队需投入大量精力在底层适配与故障排查上，难以聚焦素材分析业务本身，AI 能力的持续迭代与规模化复用受限。

解决方案

- **全流程业务建模，吃透素材分析每一环：** 腾讯云结合游戏行业买量素材分析的实战经验，将素材分析全流程标准化拆解为「素材获取—内容理解—多维打标—共性分析 / 下钻—创意二创」五大环节，明确每个环节的输入、产出与衔接关系，形成可复用、可度量、可迭代的业务流程闭环，让素材分析从依赖个人经验的黑盒变为人人可执行的标准化流程。
- **WorkBuddy 多 Agent 协同，承载全流程一站式落地：** 依托 WorkBuddy 智能体平台，以「项目工作区 + 多 Agent 协同」承载全流程落地：在统一项目工作区沉淀知识库、提示词与上下文，按业务环节配置素材采集、内容理解、打标分析、共性下钻、脚本二创等多个专家 Agent，由项目进行路由编排与群聊式协同调度，Agent 间通过标准协议传递中间产物，协同更稳定、问题可定位、迭代更高效，且免运维。
- **TokenHub 大模型底座，夯实多模态理解与打标：** 素材打标环节通过 TokenHub 统一模型网关接入多模态大模型，借助多模态视频理解能力完成关键帧抽取、音视频对齐与内容还原，并按素材类型、钩子、卖点、画面/话术亮点、BGM 等多维体系自动打标签，让海量素材实现标准化标签生成与数字资产化。
- **数据驱动闭环，沉淀可复用创意资产：** 以向量数据库 VDB 承载素材语义向量与分析结果，支撑语义检索、相似召回与 RAG 增强；以投放数据为权重对标签分布做共性分析，挖掘爆款根因，并支持下钻到分镜层级继续打标与召回，最终输出核心爆点与可复用创意点，指导 AI 分镜脚本二创与素材裂变，形成「分析—验证—迭代」的创意资产闭环。

架构图



客户价值

全流程理解，把素材分析做深做透

方案沉淀了腾讯云对游戏买量素材分析全流程的深度理解，将分散、依赖经验的分析动作标准化为可执行、可复用的五环节闭环流程，让团队不再受限于个人经验，从源头提升素材分析的深度与一致性。

数据驱动闭环，沉淀可复用创意资产

从素材获取到创意二创全流程打通，帮助玩心不止把「素材卖点分析」从主观判断升级为「卖点 + 数据证据 + 测试假设」的标准化复盘，建立起标准化分析体系与可复用创意资产，显著提升内容迭代的效率与确定性。

WorkBuddy 多 Agent 协同，落地轻、协同稳

通过WorkBuddy「项目+多Agent」模式，玩心不止无需自建和运维复杂的智能体平台，即可在统一工作区快速配置、编排多 Agent 协同完成全流程任务；相比早期自建方案，协同更稳定、问题可定位、迭代更敏捷，彻底解决了自建平台模型不稳、知识库污染、多 Agent 难定位三大关键难点。

相关产品

WorkBuddy | TokenHub | CVM | MySQL | VDB | COS



标杆案例

紫讯

业务场景

紫讯是中国领先的跨境电商科技公司，致力于结合行业知识与AI技术革新电商运营模式，紫讯旗下有紫鸟浏览器、LinkFox AI等核心产品。其中，LinkFox AI面向电商商家及品牌方，提供以“AI生图+智能营销”为核心的一站式商品视觉内容生产平台。平台聚焦电商商品图场景，覆盖商品主图生成、场景图合成、模特换装、营销海报等高频内容创作需求，帮助商家以更低成本、更高效率完成大规模商品视觉素材的生产与迭代。

业务挑战

随着LinkFox AI用户规模与日活的快速增长，平台对底层多模态大模型（尤其是图像生成模型）的调用量呈指数级上升，客户在业务发展中遇到了三大核心瓶颈：

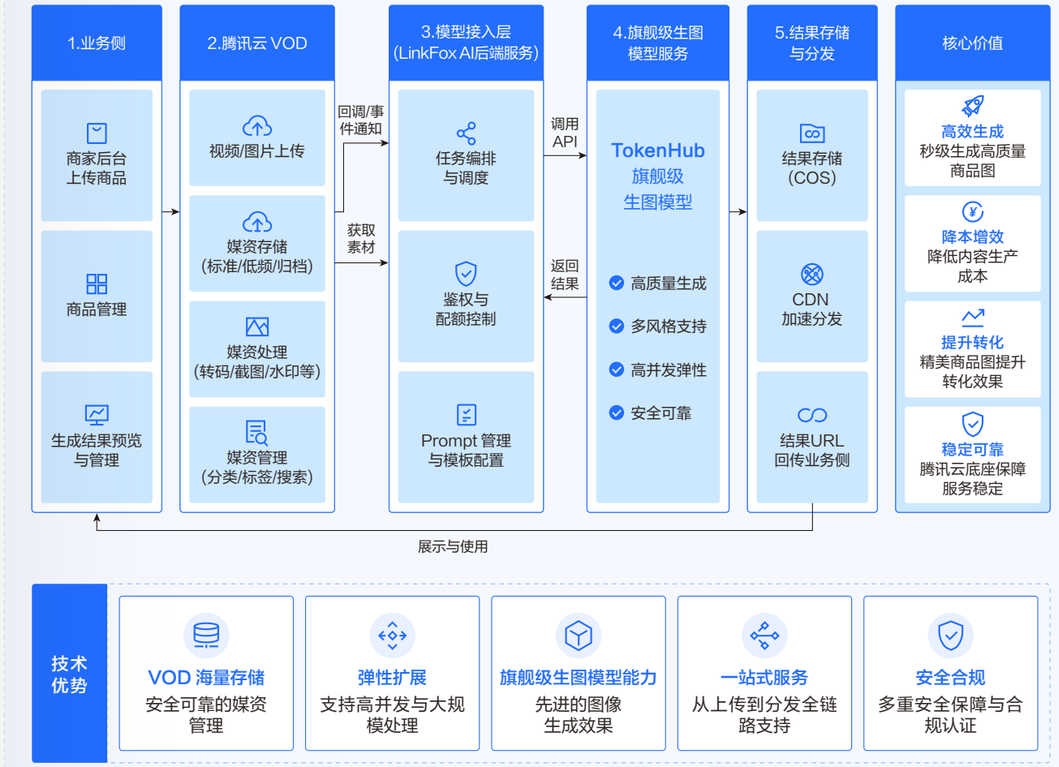
- **原厂模型TPM/并发限流严重：**直接对接原厂旗舰级生图模型时，TPM与RPM配额上限固定，业务高峰期频繁触发限流，导致用户生图任务排队、超时甚至失败，直接影响C端体验与付费转化。
- **多模型对接与切换成本高：**电商商品图场景对模型能力差异敏感，紫讯需要同时承接多家第三方模型，自行做模型路由、计费对账，研发与运维成本居高不下。
- **业务增长受制于配额：**客户为承接更多商家、扩大生图业务规模，需要稳定、可扩展的高并发模型调用能力，但单点原厂账号难以快速扩容，业务天花板被基础设施卡住。

解决方案

针对紫讯科技业务调整，腾讯云解决方案团队联手LinkFox AI业务方，共同设计出一套“多模型聚合 + 高并发”的整体方案：

- **多模型统一接入：**通过TokenHub提供的模型API服务，向客户侧统一提供多款旗舰级生图模型，客户侧只需对接一套标准化API，即可灵活调用多模型，免去自建多家厂商对接与凭证管理的负担。
- **配额池化与高并发承接：**基于腾讯云与上游模型厂商的合作配额，将紫讯多个UIN账号的TPM/RPM配额进行提升，并配合智能路由与排队削峰策略，稳定承接客户的高并发生图请求。
- **稳定性与失败兜底：**推动客户侧优化当前业务方案，在客户自建的AI网关中增加多模型间的故障切换能力。当某一模型出现限流或故障时，自动切换到同模型的其他供应商，保障客户业务连续性与最终用户体验。

架构图



客户价值

彻底解决并发瓶颈

客户底层生图能力由“原厂限流卡脖子”升级为“高并发可弹性承接”，业务高峰期不再出现限流报错，C端用户体验与付费转化同步提升。

大幅降低多模型对接成本

客户从对接多家原厂、自建调度路由，转为对接一套统一 API，研发与运维投入显著下降，资源更聚焦于上层产品与商家服务。

业务规模可持续扩张

底层模型供给能力打开后，LinkFox AI得以承接更多商家与更大体量的生图业务，业务增长不再受基础设施天花板制约。

相关产品

TokenHub　VOD

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

小影科技

业务场景

小影科技深耕出海 14 年，面向全球个人用户与企业客户，推出生成式 AI 驱动的视频内容创作产品及解决方案，助力其快速落地创作愿景、产出高品质视频内容，推动全球数字内容生产迈向 AI 原生时代。其中面向企业客户的解决方案，设有独立品牌“影伙引擎”，主要提供智慧视频创作解决方案、AI 短剧创作解决方案。AI 短剧创作解决方案服务九州文化等短剧出海企业，提供 30 余种语言的一站式翻译、配音等本地化服务。智慧视频创作解决方案聚焦荣耀等客户，就手机端侧方向提供 AI 创作能力，落地十倍生产力的视频本地化服务。

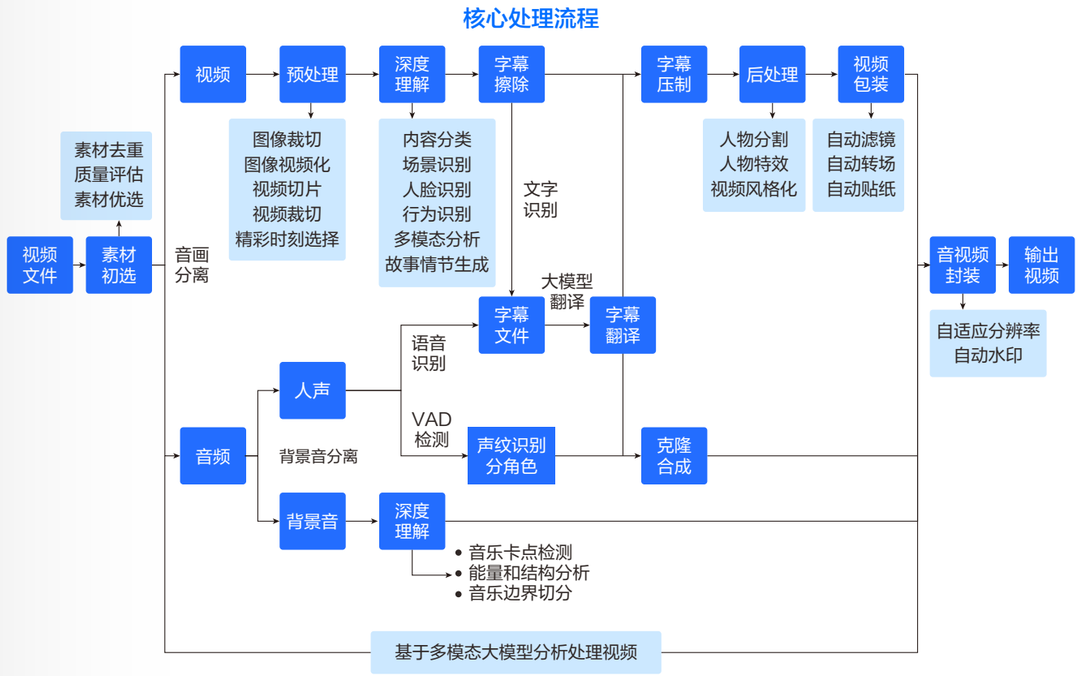
业务挑战

- **GPU资源性价比：** GPU（L20/A10/GC88），提高 GPU 显存使用率，降低 GPU 机器成本负载，提升性价比。
- **打破语言隔阂：** 覆盖 10 多种东南亚语种，服务 15 亿人口市场，自训练垂类语言翻译模型效果达到应用级标准。
- **内容出海译制成本：** 降低 90% 效果不变，低成本制作海量投放广告视频，提升客户信赖。
- **端侧技术支持：** 手机端显存性能较低，小模型效果不及大模型，无法满足需求。

解决方案

- 采用腾讯云全栈方案: TKE容器编排 + CVM云服务器 + GPU原生节点调度 + MPS 可灵API视频生成&接入MPS原生能力，构建统一的AI中台架构。
- 采用端侧小模型 + 云侧大模型组合方案: 手机端意图识别 + 传统音视频算法 + 云端模型效果处理，构建一体化内容生成架构。
- TKE容器服务、GPU云服务器、MPS媒体处理服务（模型、媒体处理接入）、COS对象存储。

架构图



客户价值

资源利用率提升

GPU性价比高，资源利用率提升，运维复杂度降低。

AI能力增强

自训练模型符合东南亚应用级标准，数据安全合规，降低海外API延迟。

本土化译制

低成本制作，深受客户信赖，提供本土化视频内容，与全球用户建立联系。

端侧技术合作

与手机厂商通力合作，借助端侧小模型 + 传统自研音视频算法 + 云侧大模型满足基本内容需求。

相关产品

GPU | MPS | TKE | COS



标杆案例

懂球帝

业务场景

懂球帝是国内最大的足球垂直社区平台，由北京多格科技有限公司运营，月活用户超1,800万，覆盖全球500余项赛事。平台集赛事资讯、实时数据、社区互动、赛事直播于一体，是国内足球迷的首选数字平台。目前，懂球帝国内和海外站点已经全面迁移至腾讯云，借助腾讯云稳定全面的云基础设施持续保障社区平稳运营。

2026年美加墨世界杯期间，懂球帝迎来爆发式增长。借助腾讯云AI能力，懂球帝全新上线了AI赛事解说、实时比赛战报生成、AI互动剧情设计三大创新功能，从传统足球资讯平台升级为AI驱动的智能互动体育社区。世界杯期间，平台单日活跃用户峰值突破1,500万，日均评论超35万条，全链路零故障平稳运行。

此外，懂球帝团队通过腾讯AI办公助手WorkBuddy等Agent产品，将热点运营效率提升至全新高度——产品与运营人员可直接将活动方案转化为代码并接入自动化发版流水线，世界杯期间快速上线了几十种互动玩法，吸引超200万用户参与。

业务挑战

内容生产瓶颈

世界杯赛程密集，单日多场比赛同时进行，传统人工编辑模式难以满足海量实时赛事报道需求。战报产出效率与覆盖度成为核心瓶颈——48支参赛球队、104场比赛，每场比赛都需要快速产出高质量内容。

AI成本管控

世界杯期间AI能力全面铺开——研发侧代码生成、C端赛事解说与剧情生产、内容安全审核多线并行，研发人均每日Token 调用量超1亿。如何在保障体验的前提下控制模型调用成本，是规模化落地的核心挑战。

互动体验升级

球迷不再满足于被动阅读资讯，对个性化、沉浸式的互动体验需求日益增长。如何利用AI能力打造差异化的互动玩法，提升用户粘性与社区活跃度，是懂球帝在世界杯期间的关键命题。

热点运营敏捷度

世界杯热点瞬息万变——冷门、逆转、争议判罚随时引爆话题。传统“需求–开发–测试–上线”流程周期长，无法支撑赛事期间的敏捷运营，亟需更高效的研发协作模式。

解决方案

- **AI赛事解说与实时战报：** 基于腾讯云的大模型底座能力，懂球帝构建了从比赛数据到内容生产的端到端AI链路：实时接入赛事数据流，由大模型自动生成多风格AI解说文本与赛事战报，覆盖进球、红黄牌、半场/全场等关键节点。用户可按偏好的解说风格（专业战术分析型、激情澎湃型、幽默调侃型等）自由切换，真正实现“千人千面”的观赛体验。战报生成覆盖全部48支参赛球队的全部比赛，产出效率相较人工提升数十倍。
- **AI 互动剧情设计：** 懂球帝在 APP 内推出“剧情创作”新功能，用户可借助腾讯云 AI 能力自行设计互动剧情——设定比赛场景、球队对阵、关键事件节点，AI自动生成沉浸式的互动叙事内容。该功能将内容创作权交还给用户，极大激发了社区的创作热情与互动活力。世界杯期间，剧情设计功能迅速成为社区爆款玩法，吸引了超 200 万用户参与互动创作，用户平均停留时长显著提升。
- **WorkBuddy 驱动敏捷运营：** 产品与运营团队通过腾讯AI办公助手WorkBuddy等Agent产品，直接将活动方案转化为代码，校验后接入自动化发版流水线。世界杯期间，懂球帝基于热点快速上线了创作互动剧情、赛事竞猜、主播人气评选等几十种互动玩法，从想法到上线的周期大幅缩短 80% 以上。Agent 化开发让非技术人员也能参与到功能交付中，释放了研发团队在核心系统上的精力，实现了世界杯期间业务与技术的双线高效运转。
- **TokenHub 统一管控模型成本：** 通过接入腾讯云 TokenHub，懂球帝可根据不同业务场景灵活匹配模型规格：强推理模型保障代码生成与复杂解说的质量，高性价比模型控制剧情生产与常规战报的单次成本。统一调度、统一计费、成本可视化，使懂球帝在高并发 AI 调用场景下实现了精细化成本管控。

架构图

应用层	AI 赛事解说	实时战报生成	AI 剧情设计	赛事竞猜	主播人气评选
AI 平台层	腾讯混元大模型	WorkBuddy Agent	TokenHub		
数据层	实时赛事数据流	球员数据库	内容知识库	用户行为数据	
基础设施	CVM 弹性计算	TDSQL-C	EMR 大数据	CDN 分发	直播 & 点播

客户价值

内容生产效率革命

AI战报与解说覆盖全部48支参赛球队的全部比赛，产出效率提升数十倍。内容团队从重复性的战报撰写中解放出来，聚焦深度分析与独家策划。

稳定压倒一切

世界杯期间全赛程全链路零故障，从容承载1,500万DAU峰值。赛事直播、社区互动、AI内容生产多条线同时平稳运行，在保障千万级用户极致观赛体验的同时，为同类应用构建高并发互动社区提供了可借鉴的范本。

模型成本精细化管控

通过TokenHub统一接入与智能调度，懂球帝在研发人均日调用量超1亿Token的规模下实现了模型性能和成本的平衡可控，保障AI体验与商业可持续性的平衡。

互动玩法敏捷交付

依靠WorkBuddy Agent能力，几十种互动玩法从创意到上线周期缩短80%以上。产品和运营人员可直接驱动新功能的开发和交付，解放研发人员生产力，真正实现“热点即功能、想法即上线”。

更多免费行业报告 并购家 www.ipoipo.net

业务效果

美加墨世界杯期间，日均活跃峰值达到1500万+，日均评论数达到35万+，用户平均使用时长100分钟+。

懂球帝的平台核心数据

懂球帝--所有热爱的发“声”地

懂球帝作为全国最大的足球社区，覆盖全国85%+足球球迷；用户粘性达到63分钟/人/天，是全网体育类App的TOP级别



	日常		美加墨世界杯期间	
	注册用户数量 9000万+	日活最高峰值 800万+	日活最高峰值 1500万+	↑ 700万+
	日均评论数 20万+	日均打开次数 16次	日均评论数 35万+	↑ 75%
	日均使用时长 63min	UGC与PGC占比 6:4	日均使用时长 100min+	↑ 59%

相关产品

CVM　CLB　TDSQL-C　EMR　COS　TokenHub　WorkBuddy

06

业务创新—AI社交陪伴

AI角色沉浸陪伴，情感化互动体验

方案

6.1 AI社交陪伴方案 / P105

案例

标杆案例——测测 / P107

标杆案例——即刻小宇宙 / P109

6.1 AI社交陪伴方案

业务场景

主体间的互动对话是AI陪伴产品的基础内核，其核心均聚焦于构建多样化的对话场景与深度强化特定情感倾向，这些对话模式之间并非彼此排斥，在实际的对话发生中，长段的对话过程往往会兼有不同的模式，而用户的幻想、创作、被陪伴、被疗愈等需求，也随着不同的聊天模式而被满足。

- **交互创作类：虚拟 IP、小说主角等**

  典型深度用户的对话模式属于交互创作，用户自身多为二次元动漫游戏、追星等亚文化爱好者或网文读者，通过与虚拟对象交流，演绎剧情，获得幻想和创作的快乐，但这类聊天的体验依赖于一定的技巧（主控设定引导）。

- **IP探索类：好奇真实名人、畅聊剧情等**

  泛用户群体大多是为了满足自己的好奇心而选择和真实名人和IP角色互动，如好奇AI角色和原作设定或者真实人物有什么不一样，不理解角色想法或者剧情而借由AI角色追问等。

- **虚拟咨询类：情感咨询、心理疗愈等**

  在该类对话中，用户往往会将AI角色当做一个“树洞”去倾诉和梳理生活中的烦恼困惑、付诸心事，并获得一个让自己感受更好的“答案”，部分用户也表示让自身感情得到宣泄和释放也是一种“答案”。

- **关系投射类：虚拟男友/女友、虚拟朋友等**

  在该类对话中，用户倾向于将某个角色定义为自己私人关系中的虚拟人物，比如男友、红颜知己、闺蜜，可能该类角色也存在于现实生活中，但达不到陪伴频率或者有一些想说却不方便说的话（如社会话题讨论-女权），可以通过AI角色扮演进行倾诉。

业务挑战

▶ 角色模型自部署

海量人设 / 对白数据需SFT+偏好对齐，拟人化与多轮稳定性双目标

TI 平台 · 训练套件　SFT + DPO 偏好对齐

▶ 高并发长对话推理

数十上百轮对话上下文，单用户日均 250+ 轮，成本与延迟双压

智算平台 · 推理优化　PD 分离 / FlexKV / Prefix Cache

▶ 多模型灵活调度

不同角色场景最优模型不同，需按场景切换 · 多云容灾 · 成本账本

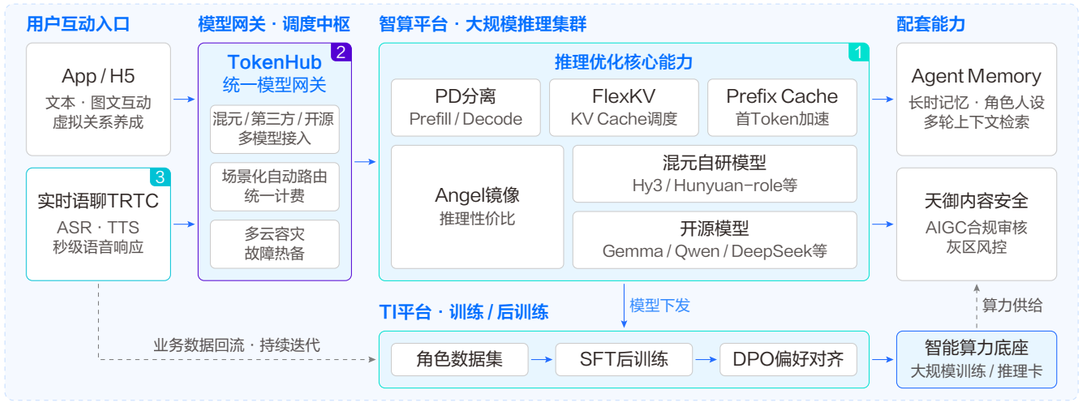
TokenHub · 多模型网关　混元自研 / 第三方闭源 / 开源

▶ 实时语音互动

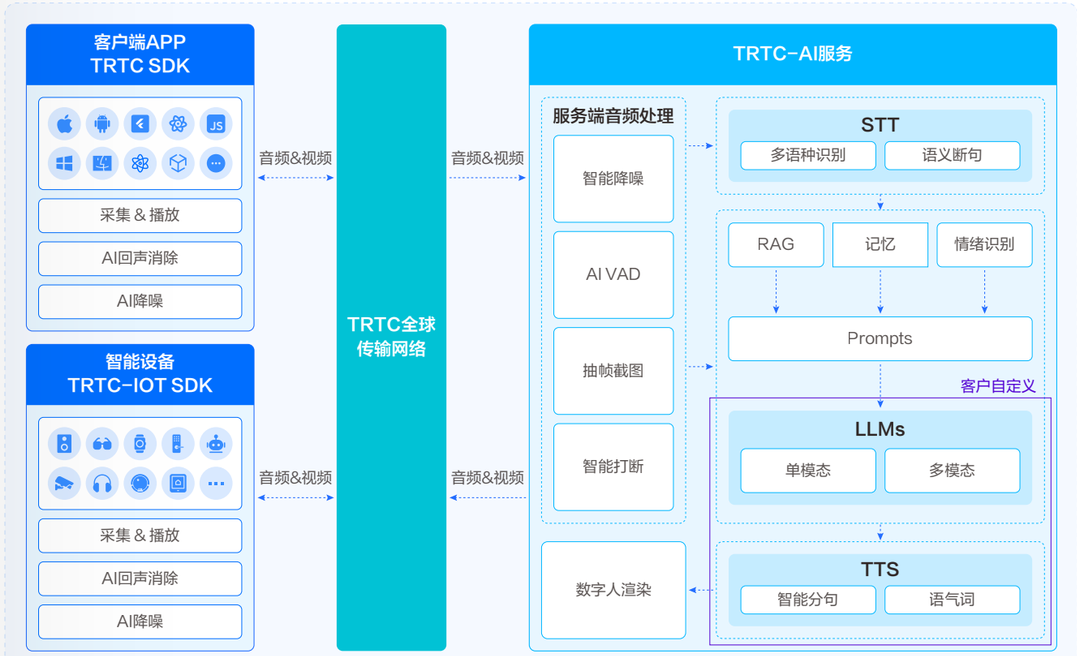
“听-想-说”端到端链路，拟人化音色与秒级响应缺一不可

TRTC + TTS + ASR　全链路秒级 · 角色音色

解决方案



架构图



- **腾讯侧：**客户端和服务端的一体化集成方案，提供智能打断、降噪能力。
- **客户侧：**自定义大模型和TTS，接入到腾讯AI服务，保持业务灵活性。

功能特点

实现超低延迟的AI对话

音视频端到端延迟低于300ms，AI对话延迟低于1000ms。

精准的ASR识别

支持多种语言，包括英语、西班牙语、日语、韩语、中文以及23种方言和130种国际语言。

第三方LLM和TTS无缝集成

配置LLM和TTS服务的账户凭证，无缝集成到服务后台。

高兼容性

支持多个平台，兼容超过20,000种设备模型。

技术优势

双讲打断

AI聊天过程，实时打断，真正实现与人聊天的真实体验

AI降噪

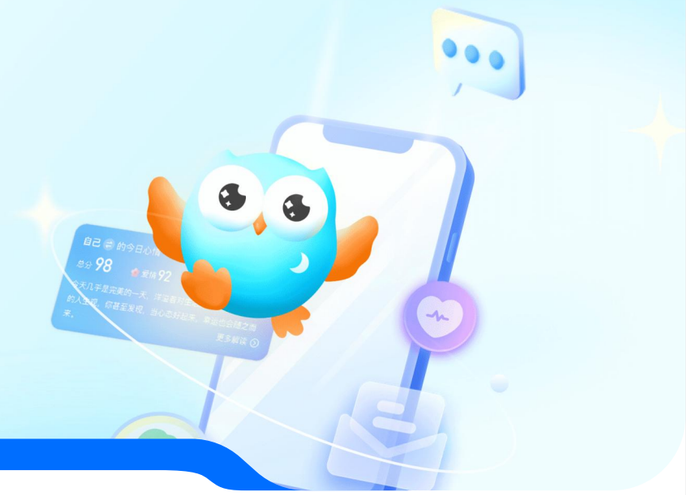
全新降噪引擎，输入端音频实时AI降噪，提升ASR的识别准确度

服务稳定

全球2800+节点，多服务容灾，高稳定超低延迟

更多免费行业报告 并购家 www.ipoipo.net

测测



标杆案例

测测

业务场景

测测通过人工智能和在线人工相结合为用户提供情感陪伴、情绪疏导、心理科普、心理测评、社区互动交流、3D心理沙盘、替代型心理服务工具等会员服务，由专业咨询师和AI虚拟咨询师提供专项咨询服务。随着用户规模与对话轮次持续增长，平台面临多垂类场景的专业度保障、高峰期并发响应、算力成本控制等多重挑战，亟需一套兼顾体验与成本的AI基础设施方案。

业务挑战

▶ 垂类场景多、单一模型效果天花板低

星盘、性格、泛心理等垂类场景各自有专业知识体系与话术风格，单一大参数模型难以兼顾所有垂类的专业度与人设一致性，效果受限且成本高昂。

▶ 异构算力调度与资源利用率

多垂类模型并发调度需错峰进行，既要保障模型推理稳定性，又要减少GPU资源空跑浪费，对自动化调度能力要求高。

▶ 体验与成本的刚性矛盾

AI陪伴场景下体验越好（响应越快、人设越拟真）成本越高，用户迁移成本低意味着留存对体验高度敏感，需在成本可控前提下保障秒级响应与对话质量。

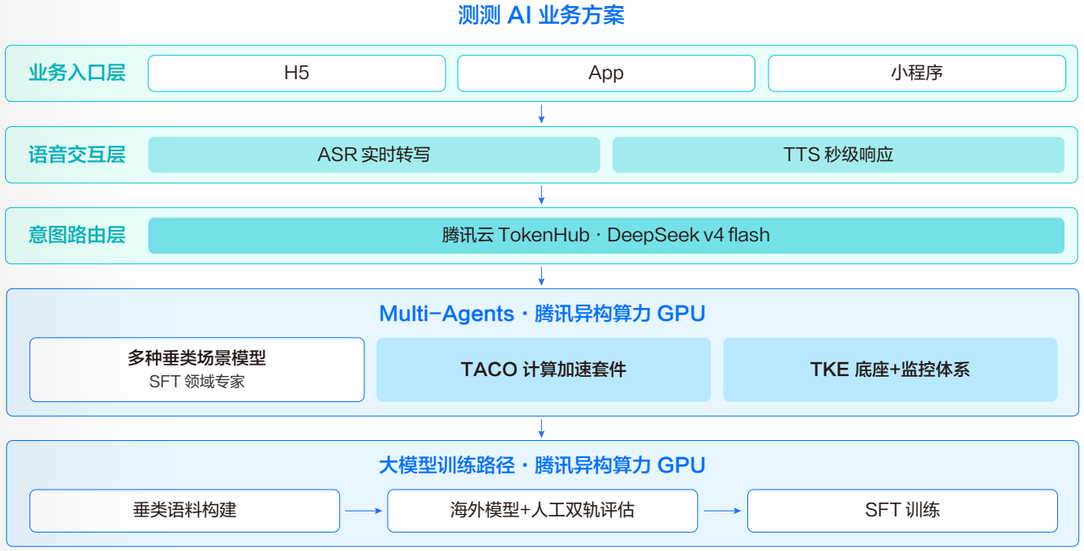
▶ 数据安全与行业合规

泛心理场景涉及用户情感隐私与心理数据，对数据不出域、模型自部署合规有强制要求，公有云大模型API直连方案无法满足。

解决方案

- **GPU弹性调度与动态扩缩容：**基于TKE完善的监控体系，包括GPU利用率、并发量、产线流量等趋势来预判模型需要的部署规模，再结合rolling-update和腾讯云提供的qGPU技术，最大化地实现GPU资源有效利用，完成模型部署规模的动态缩扩容调整。
- **自建Multi-Agents平台（画布+CLI）：**搭建决策模型+通用大模型+垂类大模型的Multi-Agents，并辅助有mcp、skill、长期记忆等模块，支持不同的业务侧对Agents的需求，并方便产研进行快速研发迭代。
- **训练与推理混部：**在产线业务低峰期部署开源模型，执行数据蒸馏以提升GPU利用率，减少空跑带来的资源浪费。
- **数据安全与合规部署：**垂类模型自部署在客户专属TKE环境内，垂类语料与用户对话数据全程不出域；结合腾讯云安全审核产品对AI生成内容进行实时合规检测，满足泛心理行业的数据安全与内容合规要求。

架构图



客户价值

模型鲁棒性
业务高峰期间（晚间情感倾诉高峰）模型响应成功率高于90%，TKE弹性调度与监控体系兜底，保障用户陪伴体验不中断。

路由降本
意图识别与对话生成解耦，每轮对话仅做分类不生成，路由层延迟与成本压到最低（具体降本比例见业务效果章节）。

垂类效果保障
各垂类场景拥有专属专家模型，专业度与人设一致性不受通用模型能力天花板限制，用户体验显著优于单一大模型方案。

算力错峰复用
在线推理与离线训练共享同一套GPU资源，业务空闲期跑训练任务，算力利用率最大化，避免资源空跑浪费。

成本优化
以蒸馏垂类小模型替代单一大参数MoE“全扛”模式，性能、效果、成本三者兼顾，训练与推理成本同步下降。

数据安全合规
垂类语料与用户对话数据全程不出域，AI生成内容经实时安全审核，满足泛心理行业的数据安全与内容合规要求。

相关产品

GPU异构算力
提供支撑陪伴模型大规模调用的底层算力资源，保障资源的同时提供多种算力选型，同时提供TACO计算加速套件，降低AI优化门槛的同时大幅提升训练和推理业务性能，节约算力成本。

TokenHub
星言 / 星座 / 塔罗等头部场景的AI角色化问答接入腾讯云TokenHub大模型调用平台，统一调度混元自研、第三方闭源与开源等多家大模型，按场景自动路由。

RTC / ASR / TTS
使用腾讯RTC / ASR / TTS产品给予AI角色“耳朵”与“嘴”，拟人化音色与秒级响应提供更生动的陪伴互动能力。

TKE（腾讯云容器服务）
承载垂类模型的容器化部署与弹性扩缩容，基于监控体系实现模型部署规模的动态调整。

安全审核 对AI生成内容进行实时安全审核，保障泛心理场景下的内容合规。

小宇宙

标杆案例

即刻小宇宙

「借助腾讯云 TokenHub 统一调度、向量数据库与全球加速能力，小宇宙把 AI 用在了刀刃上——让每一集播客都能被理解、被精准送达每一位听众。」

业务场景

小宇宙是即刻团队孵化、国内头部的中文播客平台，于2020年3月上线，覆盖播客收听、发现、社区互动等全链路体验，累计用户已超千万，是即刻产品矩阵中体量最大、微服务体系最规范的一套系统。作为一个以「声音」为核心的内容平台，小宇宙沉淀了海量优质播客内容与活跃的播客社群，并在内容分发、个性化推荐与用户互动上持续深耕。

面向「每天海量新增单集」的内容运营压力和「面向全球用户」的稳定接入诉求，小宇宙携手腾讯云，以TokenHub统一推理调度、向量数据库（VDB）、TKE自建训练平台、ASR/TTS多模态能力，以及GAAP全球加速为底座，构建了一套「业务系统 + 大数据 + 推荐 + AI」四位一体的云上架构，把过去因成本与人力只能凑合的环节，用算力补齐，让团队更专注于产品本身。

小宇宙业务功能全景图

内容消费	发现推荐	订阅关注	社交互动
播客节目	个性化推荐	订阅播客	评论 / 回复
单集Episode	搜索	更新提醒	点赞 / 转发
播放器（倍速 / 定时）	排行榜 / 热门	关注用户	时间轴评论
离线下载	分类 / 频道	收藏单集	

个人中心	创作服务	消息通知	会员 / 设置
我的订阅	播客创建	推送通知（更新 / 推荐）	会员权益
播放历史	单集发布 / 音频转码	系统消息	分享 / 分享卡片
播放进度多端同步	创作者后台	私信 / 互动提醒	设置 / 偏好
我的收藏 / 稍后听	内容数据分析		账号与安全

业务挑战

海量新增内容的理解与打标

每天新增大量播客单集，靠人工听审、归类、打标已不现实；引入完整数据系统成本过高、对当前体量超配，基础工具又无法满足需求。

多模型接入与成本控制

热词提取、打标分类等高频批量任务与深度推理任务并存，需要在「又快又省」与「够强够准」之间灵活切换，同时避免为每家模型单独对接接口、计费与限流。

跨语言内容的体验天花板

大量优质播客为英文内容，用户受限于语言能力只能消费「二手的、被压缩过的」信息，原汁原味的表达大量流失。

面向全球用户的稳定接入

源站位于上海，需为分布在日本、欧洲、北美、中东等地的海外用户提供稳定、低延迟且安全的接入与容灾能力。

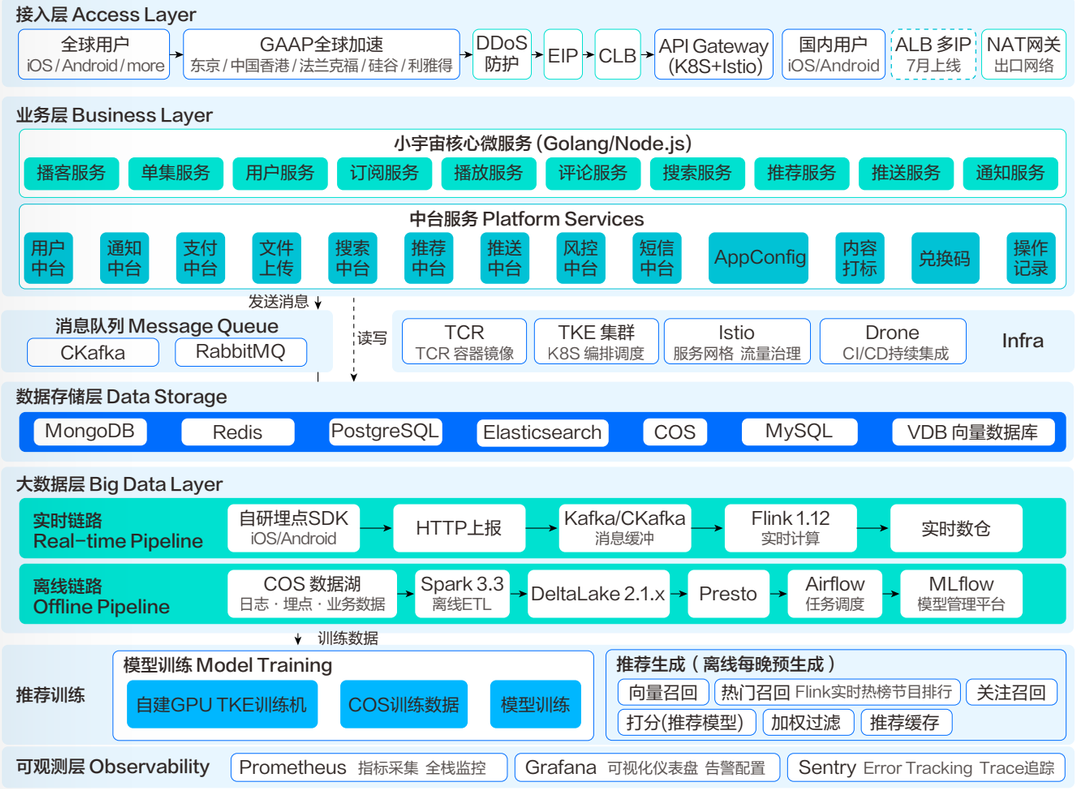
推荐系统的持续迭代

播客推荐是核心体验，其背后是需要持续训练迭代的推荐模型，对GPU算力、训练平台与数据链路提出完整要求。

解决方案

- **AI 能力矩阵（性能优势）：** 通过 TokenHub 统一接入 GLM-5.1/5.2、DeepSeek-V4 等多款主流大模型，重推理任务走强模型、高频批量任务走高性价比模型；DeepSeek-V4-Flash 内置缓存能力，对重复或高度相似请求直接命中已有结果，热词提取 / 打标分类等高频任务 >70% 请求无需重新推理，单次调用输入成本最低降至 0.02 元。
- **向量推荐（功能优势）：** 借助腾讯云向量数据库存储 数百万+ 多维度内容向量，将每集播客的内容摘要、主题风格、叙事节奏提取为特征向量，从「内容气质」维度做相似匹配与定时推送，相比单维度检索更贴合用户偏好。
- **多模态交互：** ASR 语音识别、TTS 语音合成同步接入，结合同声传译 + 音色复刻，实现「原声 + 中文」的跨语言消费体验。
- **推荐训练链路：** H20 / L20 / PRO5000 多型号 GPU 算力 + TKE自建训练平台，打通「埋点日志 → Kafka → Flink/Spark 实时处理 → Spark 批处理 → TKE自建训练 → 推理服务」的完整数据流；COS 提供训练数据分路写入与稳定挂载方案。
- **全球化接入（架构优势）：** GAAP 全球加速为海外用户提供稳定接入与访问日志能力，CLB 承接接入层流量分发与多实例容灾，WAF 提供接入层安全防护，共同保障全球用户的访问体验与业务安全。

架构图



更多免费行业报告 并购家 www.ipoipo.net

客户价值

把「最笨的活儿」盘活

对海量单集自动做内容理解与打标，运营同学从「读懂每一集讲了什么」中解放出来，转而做更需要判断力的事，同样人数能扛的事情多了一大截。

打破语言门槛、创造内容增量

翻译 + 音色复刻让原本「听不懂」的英文节目变得可消费，既保留创作者本人的语气，又让用户听懂内容，为内容、创作者、平台各环节都创造了增量。

成本与精力回投产品

Token统一调度 + 命中缓存降本显著压低了高频 AI 任务的单次成本，省下来的成本和精力被重新投回到产品打磨上。

自主可控的技术选型

多模型能力如何选、如何组合由客户自己主导，不被绑死在某一家。

稳定的全球化底座

把稳定性、算力、实时语音与全球接入交给云服务，团队得以更专注在产品创新本身。

业务效果

在腾讯云能力的支撑下，小宇宙在降本、推荐、全球化与提效等维度取得了可量化的成效：

缓存命中降本

高频打标 / 热词任务>70%请求命中缓存，单次调用输入成本最低降至0.02元

向量规模

数百万+多维度内容向量，日增约数千条，支撑精准个性化推荐

全球化接入

上海源站+5大海外加速节点，稳定承载全球用户访问

流程提效

AI分流后，原需1-2周的内部流程压缩至1-3天

内容破圈

马友友×lululemon同声传译单集累计播放30万+，打破英文内容触达天花板

相关产品

WorkBuddy

部署于企业自有环境的研发/运营提效Agent，确保全程可控

TokenHub

统一多模型接入、路由分发、缓存命中与成本核算的推理调度层

VDB

存储数百万+播客内容特征向量，支撑个性化推荐与相似内容匹配

TKE

承载推荐模型训练与上线，打通从埋点日志到推理服务的完整数据流

GPU（H20 / L20 / PRO5000）

为推荐模型训练与推理提供多型号GPU资源保障

ASR、TTS

语音识别、语音合成，支撑同音色翻译

GAAP

上海源站+东京 / 法兰克福 / 硅谷 / 利雅得及中国香港等全球节点接入加速

CLB

接入层流量分发与多实例容灾

WAF

接入层安全防护，保障全球用户访问安全

COS

训练数据集挂载与数据湖存储，稳定支撑TKE训练链路

07

业务创新—游戏内AI玩法

AI驱动NPC智能化与玩法创新

方案

7.1 游戏NPC方案 / P115

案例

标杆案例——游卡 / P119

更多免费行业报告 并购家 www.ipoipo.net

7.1 游戏NPC方案

业务场景

把游戏里数量最多、玩家接触最频繁、却最“没有生命”的那部分内容，变成真正的角色。

NPC 是玩家在游戏中接触频次最高的内容单元 —— 从新手引导的第一个村民，到主线剧情的关键角色，再到商店老板、任务发布者、副本队友。一款中大型游戏的 NPC 数量通常在数百到数千量级，覆盖玩家从下载到长线留存的完整生命周期。

本方案面向以下四类典型业务场景：

▸ 场景一：世界沉浸型 —— 开放世界 / MMORPG

玩家在庞大世界中长期驻留，NPC是世界“活着”的证据。城镇里的居民各有作息与故事，玩家在探索中随时可以停下来与任意角色深入交谈，世界不再是空布景。

▸ 场景二：情感陪伴型 —— 二次元角色养成 / 恋爱互动 / 虚拟偶像

角色本身就是核心卖点与付费对象。NPC需要记得玩家、认得玩家、随关系推进而改变态度，玩家为“这个角色和我之间的关系”付费，而非为一段固定剧本付费。

▸ 场景三：引导与转化型 —— SLG / 卡牌 / 休闲模拟

NPC 承担新手引导、玩法讲解、活动播报、付费转化承接。用会说话的角色替代弹窗与教学关卡，把“读说明书”变成“和人聊天”。

▸ 场景四：内容与社交补位型 —— 全品类通用

在版本空窗期、非黄金时段、低活跃服，AI NPC 作为可控的内容与陪伴供给，填补人力内容和真人社交的空缺，并支撑多语种出海。

业务挑战

▸ NPC 不像人，沉浸感在对话框里被击穿

传统 NPC 依赖预设台词树：同一句台词重复上百遍，问不了脚本之外的问题，玩家昨天刚救了他的村子，今天见面还是那句“欢迎来到新手村”。美术、世界观、配乐投入巨资构建的沉浸感，往往在玩家点开第一个对话框的三秒内碎掉。玩家社区高频出现的“NPC 是木头人”“台词尴尬只能跳过”，是最直接的差评来源。

▸ 内容产能撞上天花板，更新永远追不上消耗

剧情、台词、配音、本地化全靠人力。一个大版本动辄数十万字文本，叠加配音录制、多语种翻译、QA 验收，周期以月计；而玩家把内容消耗完只需要一个周末。结果是版本空窗期长、玩家断线、回流成本高。每多支持一个语种、每多加一条剧情分支，成本近乎线性增长。

▸ 陪伴与社交缺口：多人游戏的“单机化”

开服后期、深夜时段、低活跃服，玩家实际处于“一个人跑图”的状态。社交是长线留存的核心变量，但社交供给天然不稳定，而这块空缺是固定脚本 NPC 无法填补的。

▸ 同质化严重，宣发缺少差异化记忆点

玩法框架高度趋同，宣传物料只能拼画质、拼卡面、拼代言人。缺少一个讲得清、看得懂、玩家愿意自发传播的核心卖点。

▸ 长线缺乏二刷动力，LTV 见顶

剧情固定，通关即完结，玩家没有理由回来第二次，生命周期价值被牢牢锁在天花板下。

▸ 引入 AI 本身也带来新挑战（方案必须回答）

出戏风险	剧情脱轨	内容安全	成本与延迟
角色说出不符合世界观、时代背景或人设的内容	自由对话与主线任务、关键道具、进度校验发生冲突	玩家诱导下产出违规、敏感或有害内容	大规模在线调用带来的算力成本与响应速度压力

解决方案

核心定位：把 NPC 从“游戏资源”升级为“游戏角色”。

不是在对话框后面接一个通用聊天机器人，而是用大模型重构 NPC 的底层构造 —— 让它拥有稳定人格、记得住玩家、并且会自己做事。

▸ 人格内核：每个 NPC 都是“这一个”

为每个角色建立完整档案：身份背景、性格特质、价值观、与其他角色的关系网、语言习惯与口头禅、以及知识边界（村口铁匠不该知道千里之外的宫廷秘辛）。同一句提问，暴躁老兵与怯懦书生会给出完全不同的回答。人格通过配置化“角色卡”定义，策划可控、可调、可持续迭代。

腾讯云能力支撑：角色卡以结构化 Prompt 注入 TokenHub 大模型（DeepSeek / 混元系列），保障人设一致性与多轮稳定输出；角色资产沉淀于 Elasticsearch，可语义检索、跨项目复用。

▸ 记忆系统：关系是会长出来的

短期记忆	长期记忆	关系状态
单次对话上下文连贯，不会聊三句就失忆	跨会话记住玩家姓名、职业、做过的选择、许下的承诺。玩家上周帮他找回女儿，今天见面他会主动道谢	好感、信任、敌意随交互动态演化，并反向影响 NPC 态度、可发布的任务、可解锁的剧情与商店折扣

这是“NPC 像活人”最强的感知点。玩家买单的不是对话本身，而是被记住。

腾讯云能力支撑：长期记忆（交互事件、承诺与选择）向量化写入 Elasticsearch，语义召回让 NPC“记得你”；短期会话上下文与好感度等关系状态由 腾讯云 Redis 高速承载，支撑高并发实时读写。

▸ 自由对话：跳出台词树

玩家可以随意提问，NPC 在世界观与人格允许范围内自然应答；支持文本输入、语音输入、以及 AI 实时配音输出。多语种由同一套人格直接生成，无需为每个语种重复编写与录制。

腾讯云能力支撑：TokenHub 大模型流式生成、首 token 响应快：ASR 语音识别 + TTS 语音合成（多音色、边生成边播）构成语音交互闭环；世界观设定与攻略 FAQ 沉淀于 ES 向量知识库（BM25 + kNN + RRF 混合检索），RAG 增强让回答有据可依、不出戏。

▸ 自主行为：NPC 有自己的生活

NPC 具备作息与目标 —— 白天摆摊、傍晚回家、雨天躲进酒馆；具备情绪状态，会因世界事件改变言行；具备主动性，会在玩家上线时主动打招呼、提醒未完成的约定、甚至主动上门委托任务。世界从“等玩家触发”变成“自己在运转”。

▸ 安全与可控围栏

世界观约束层	剧情主干保护
把设定、时代背景、禁用词、知识边界写成硬约束，角色不出戏	主线关键节点仍由策划脚本掌控，AI 只在枝叶自由生长，不影响任务判定与进度校验
内容合规审核	**成本与性能控制**
输入输出双向过滤，敏感话题拦截、越狱诱导识别、违规日志留档	分级调用策略 —— 主线关键角色用高规格模型，路人 NPC 用轻量模型；高频通用应答走缓存；保证响应速度与单用户成本可控

腾讯云能力支撑：TMS 文本内容安全对输入输出双向审核，敏感拦截与违规留档开箱即用；TokenHub 统一接入混元 Hy3、DeepSeek-V4-Pro / Flash等多规格模型，一套接口按 model 参数灵活路由——主线角色走旗舰模型、路人 NPC 走轻量模型；配合模型 Cache 缓存复用上下文计算结果，成本与延迟可控。

▸ NPC 内容工厂：让产能不再是瓶颈

配备一套生产工具链：角色卡批量生成、台词批量生产、AI 配音、多语种一键铺开、自动化质量验收（人设一致性、出戏检测、合规抽检）。角色资产可跨项目复用，一次积累、多产品受益。

架构图

整体分五层：

用户层

为游戏客户端（手游 / PC / 主机）与官网社区。

接入层

由 CLB 负载均衡承载 HTTPS 加密接入与流量分发。

应用服务层

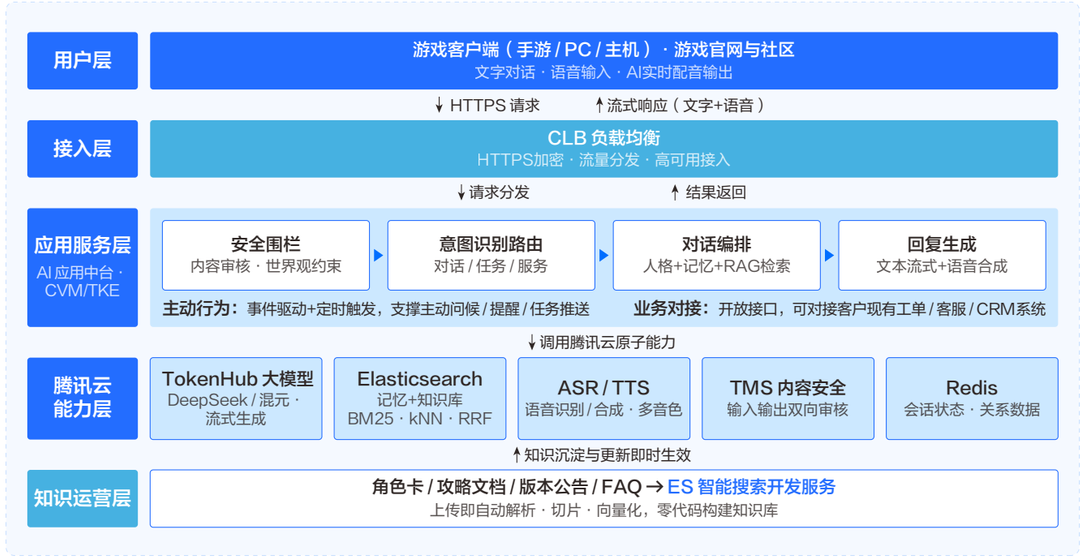
为 NPC 中台（部署于 CVM / TKE），内含安全围栏、意图识别路由、对话编排（人格 + 记忆 + RAG 检索）与回复生成，事件与定时任务驱动 NPC 自主行为，工单服务可对接客户现有客服工单系统。

腾讯云能力层

提供 TokenHub 大模型、Elasticsearch 向量检索、ASR / TTS 语音、TMS 内容安全与 Redis 状态存储。

知识运营层

通过 ES 智能搜索开发服务，实现角色卡与攻略文档上传即自动解析、切片、向量化，知识更新即时生效。



客户价值

玩家体验：从“看内容”到“有关系”

沉浸感提升
NPC 应答自然、认得玩家，世界可信度显著上升

内容近乎无限
每个玩家的对话体验唯一，不再“人人一套台词”

陪伴常在
任何时段都有角色可以互动，孤独感下降

核心经营指标

指标方向	预期效果	驱动来源
次日 / 7 日留存	提升	新手引导拟人化、开局即建立角色关系
长期留存与回流	提升	记忆与关系形成情感沉没成本，NPC 主动召回
单日在线时长	提升	自由对话与自主行为带来非任务型停留
付费转化与 ARPU	提升	角色相关皮肤、语音、剧情、亲密度道具形成新付费点
LTV	提升	内容不重复带来二刷动力，生命周期延长
内容生产成本	下降	台词、配音、本地化由 AI 批量产出
版本迭代周期	缩短	内容工厂让更新频率不再受人力线性约束

商业化与市场价值

新增付费点
角色亲密度、专属语音包、私人剧情线、AI 角色定制

降本增效
内容与本地化成本结构性下降，人力从“写台词”转向“做设计”

出海加速
多语种由同一人格自动生成，海外版本上线周期大幅压缩

宣发传播价值

可传播的核心卖点
一句话说得清 —— “这个游戏里的角色，会记得你。”

天然的 UGC 素材
玩家与 NPC 的高光对话截图、离谱对话、感人瞬间，是社区自发传播的绝佳内容，二创与短视频转化率高

口碑差异化
在同质化红海中建立“技术领先 + 情感体验领先”的双重认知

媒体与行业议题
AI 原生玩法本身具备行业话题度，有利于获取媒体报道与曝光资源

组织与资产沉淀

一次投入形成可复用的中台能力：角色资产库、人格配置规范、安全审核流程、生产工具链，可平移到后续新项目，形成长期技术壁垒。

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

游卡

「国民级卡牌游戏，牌桌上请来 AI 军师，武将推荐、局内问答、武将介绍一站打通，新手少走弯路，牌局体验再上一个台阶。」

业务场景

《三国杀》是杭州游卡网络技术有限公司（以下简称“游卡”）旗下的国民级游戏IP，矩阵产品长期稳居国内卡牌品类头部。它以三国历史为背景，深度融合武将技能、卡牌策略与身份博弈三大玩法内核，凭借丰富的武将搭配与身份对抗机制，沉淀出规模庞大、粘性极高的核心玩家群体。

《三国杀》小游戏是游卡今年在三国杀 IP 发展上的重点方向之一。依托“即点即玩、社交裂变”的天然优势，客户将其定位为 IP 用户增量的重要引擎，微信小游戏与抖音小游戏双端今年均已实现显著增长突破。而随着用户规模的快速攀升，客户对架构弹性、稳定性与成本效率提出更高要求，并希望在云服务与流量资源两端与我方展开更深入的一体化合作。

业务挑战

▶ **新玩家上手门槛高，增长与留存承压**

《三国杀》拥有众多武将，技能各异、流派繁多，叠加身份博弈与多模式玩法，新玩家很难快速掌握选将、用技与配将，学习曲线陡峭，直接抬高新用户留存与拉新成本。

▶ **AI 探索初期人力有限，开源自建处处受限**

团队希望聚焦游戏体验而非底层平台，但前期用开源框架自建工作流暴露众多短板，缺乏应用评测，权限与内容安全管控不足，简单功能也需大量额外工程，迭代慢，上线周期长等。

▶ **AI游戏助手的实时三重压力**

牌局内问答、武将推荐等场景属实时交互，对延迟、成本与回答质量三重敏感，通用模型在垂直游戏场景下难以同时满足。

解决方案

- **基于 ADP 可视化工作流构建 AI 能力底座：**以可视化拖拽方式编排“开始 → 大模型 → 回复”工作流，零代码即可完成 AI 功能的搭建、调试与发布。工作流中的核心大模型节点采用系统 / 用户双提示词架构——系统提示词每次基本固定，用于锁定模型角色与三国杀背景知识；用户提示词承载客户端每次实时上报的武将信息与玩家战局信息，二者结合后由工作流统一调度输出回答。
- **游戏场景化模型选型与隐式缓存调优：**模型采用 DeepSeek-V4-Flash，Hy3，在保证回答质量的同时兼顾低延迟与低成本，契合牌局内实时问答的诉求；并借助 ADP 对 DeepSeek以及Hy3隐式缓存的支持，让基本固定的系统提示词稳定命中缓存，进一步提升响应速度、压低调用成本。

- **三大AI功能落地，覆盖牌局内外：**基于上述能力，已上线“武将推荐”“局内问答”“武将介绍”三个功能并嵌入游戏——新玩家可即时了解武将技能与使用方式，并在不同模式下获得 AI 配将推荐，把武将技能转化为随手可得、通俗易懂的智能辅助。

架构图

① 玩家终端 | 《三国杀》手游 App（iOS / Android） — 武将推荐、局内问答、武将介绍，AI 回答嵌入牌局内外

↓

② 游戏业务服务 — 身份鉴权、协议转换，组装「武将信息+玩家战局信息」后调用 ADP

上报武将+战局信息 ↓ ↑ 返回 AI 回答

③ AI 能力底座 | 腾讯云智能体开发平台 ADP

在 ADP 内完成工作流编排与模型调用，无需自建 AI 服务

可视化工作流编排	双提示词架构	隐式缓存
零代码拖拽，调试发布一体	系统与用户提示词分层	系统提示词命中缓存

平台内置支撑能力

应用评测	权限管控	内容安全	数据观测
回答质量闭环	多角色访问控制	合规风险拦截	调用链路可追溯

↓

④ 模型层 | 混元 Hy3、DeepSeek-V4-Flash — 经 ADP 统一调度，按场景选型，兼顾回答质量、低延迟与低成本

客户价值

从“搭平台”回到“做体验”

ADP零代码编排替代Dify自建工作流，逐项补齐编排、评测、权限与观测短板，AI功能上线周期大幅缩短，有限人力回归游戏价值本身。

大幅降低上手门槛

AI推荐、介绍与局内问答把复杂武将技能"翻译"成即时可懂的辅助，新玩家无需查攻略即可快速理解与决策，改善新手体验、降低获新难度。

延迟、成本、质量三者兼得

游戏场景优化的混元模型与DeepSeek-V4-Flash叠加隐式缓存命中，在保证质量的同时持续压低延迟与成本，并依托ADP评测、观测、权限与安全能力支撑海量用户稳定迭代。

业务效果

- 从开源Dify自建迁移到ADP后，工作流编排、评测、权限与数据观测等短板被逐项补齐，单个AI功能的搭建与迭代周期从原来的数天压缩至1天以内，需求响应效率提升约70%，配套的工程维护投入大幅下降，团队精力得以回归业务体验本身。
- 武将推荐、局内问答、武将介绍三大功能上线后，牌局内问答的平均响应延迟控制在2秒左右、较早期方案下降约40%，叠加DeepSeek以及HY3隐式缓存命中，单次推理成本降低约30%，回答质量稳定；新玩家借助AI 辅助上手门槛明显降低，新手期次日留存获得双位数提升，为后续更多牌局内外AI玩法的规模化拓展奠定平台基础。

相关产品

ADP　CVM　TKE　CLB　内容安全　TokenHub

08

业务创新—运营 / 智能客服

AI驱动客户全生命周期智能化运营

方案

8.1 智能客服方案 / P122

8.2 智能质检方案 / P123

8.3 智能外呼方案 / P124

案例

标杆案例——天润融通 / P125

标杆案例——轻喜到家 / P127

标杆案例——智齿科技 / P129

8.1 智能客服方案

业务场景

客服Agent场景，指的是利用人工智能代理，来部分或完全替代传统人工客服，完成与客户对话交互的智能化场景。“Agent”不同于简单的“聊天机器人”，强调其具备自主感知、分析、决策和执行的能力。

电话、官网或者APP客服平台，通过接入AI客服降低人工客服运营成本，分为2个场景：文字对话客服和呼入式语音对话客服，可用于售前/售后咨询、订单查询、故障排查等场景。

业务挑战

▶ **需要强大的知识库处理能力**

当用户提出专业问题时，系统不仅能快速从海量知识库中精准定位相关信息，还能结合大语言模型的理解能力，对知识进行整合、提炼和重组，生成准确、完整且易于理解的回答，而非简单地罗列文档片段。

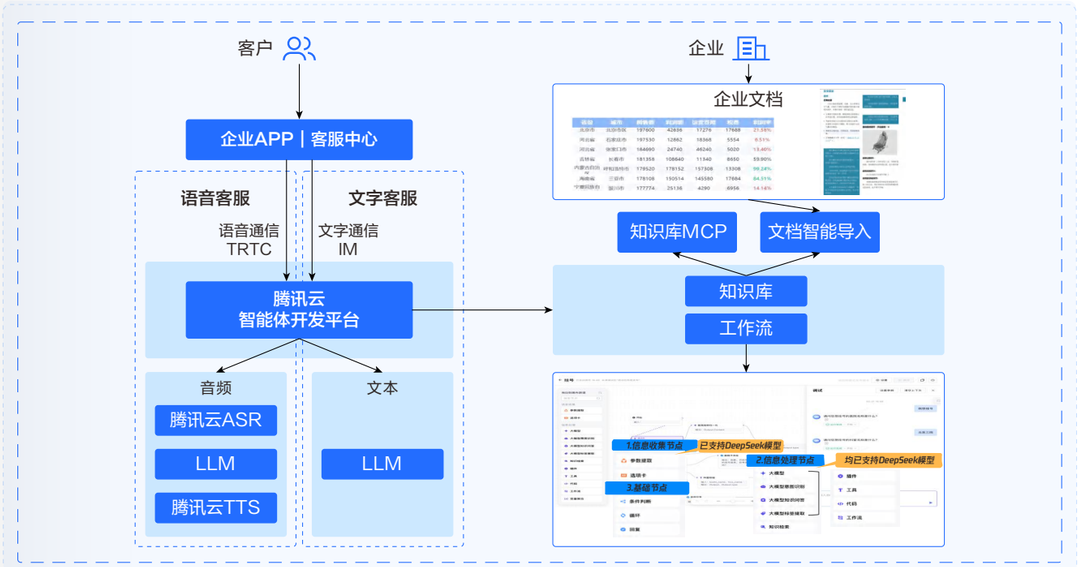
▶ **具备复杂业务流程处理能力**

不能仅限于问答，而要能够理解并执行多步骤的业务操作。

▶ **在多轮对话中识别用户意图，理解上下文，避免重复询问或逻辑混乱**

要求能够记住整个对话历史，准确理解指代信息，保持对话逻辑的连贯性。通过深度语义理解，系统可以避免重复询问已提供的信息，确保对话自然流畅，为用户带来接近真人交互的体验。

解决方案



（工作流图示）

- **知识库管理方案：**腾讯云智能体开发平台支持多种文档格式导入，通过大模型自动生产问答对，并实现文档对校验，提升大模型问答体验。
- **工作流管理：**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。
- **混元大模型方案：**混元大模型具备上下文理解和长文记忆能力，流畅完成各专业领域的多轮问答。

更多免费行业报告 并购家 www.ipoipo.net

8.2 智能质检方案

业务场景

智能质检指利用人工智能技术，对企业与客户之间的沟通内容（如电话、在线聊天、邮件等）进行自动化、全量分析，以评估服务质量、发现合规风险并提炼业务洞察的过程。

智能客服语音质检的核心场景包括：1）服务质量全量检测：实现自动评估客服态度、话术规范性及业务解答准确性；2）合规与敏感行为监控：支持实时识别违规用语、客户隐私泄露风险；3）客户情绪分析与危机预警：通过情感分析技术及时发现客户不满并触发干预，最终通过数据报表驱动服务策略优化和员工针对性培训。

业务挑战

质检效率与覆盖率不足

传统人工质检依赖抽样检查，耗时费力且主观性强，难以实现100%全量覆盖，导致问题遗漏和评估标准不统一。

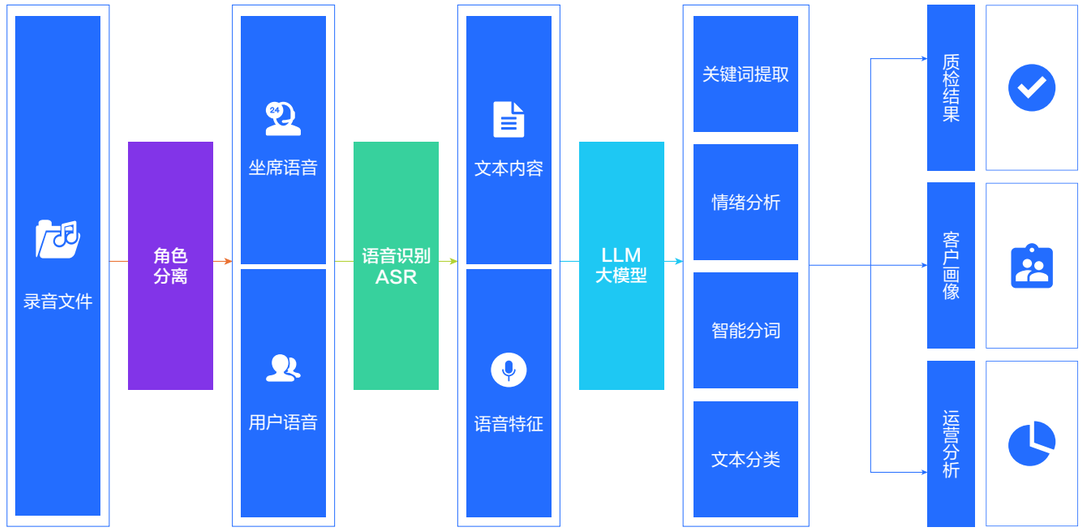
复杂场景识别准确率低

在多方言、语速快或背景噪音干扰的通话中，语音识别和语义理解的准确率易下降，影响关键词提取、情感判断的可靠性。

数据价值挖掘不充分

海量语音数据缺乏深度分析能力，难以提炼业务洞察，如客户需求趋势、服务瓶颈，导致决策支持不足。

解决方案



（工作流图示）

- **AI智能质检全流程方案：**通过腾讯云语音识别ASR将海量通话转写为文本，支持多种方言，同时在噪音环境通过降噪技术提升识别准确率；腾讯混元大模型的上下文理解能力精准判断服务意图与合规性，可以对客服内容进行批量深度分析，实现质检全场景覆盖，大幅提升质检效率。
- **质检数据分析方案：**利用腾讯云实时数仓TCHouse-D对质检内容进行大数据分析生成多维分析，深度挖掘数据价值，输出质检报告，反哺客服流程优化与个性化培训。

8.3 智能外呼方案

业务场景

智能外呼是指利用人工智能技术，由机器自动发起电话呼叫，与用户进行多轮、拟人化的语音交互，以完成特定商业目标的系统。核心在于，将传统需要大量人工坐席重复执行的外呼任务，转变为由AI驱动的、高效且可规模化的自动化流程。

智能外呼是企业“AI+营销”和“AI+服务”战略的关键落地环节。它能显著提升外呼效率，并基于用户数据深度分析实现个性化沟通，从而有效提升营销转化率与GMV，同时大幅降低企业运营成本，其拟人化的交互体验也优化了客户感知。

业务挑战

交互体验待提升

智能外呼系统在实际应用中面临端到端低延迟的严格要求，任何明显的延迟都会影响对话流畅度，降低用户体验。

噪声影响语音识别准确性

在复杂通话环境中，降噪优化成为技术难点，背景噪音容易干扰语音识别准确性。

缺乏应对自然打断能力

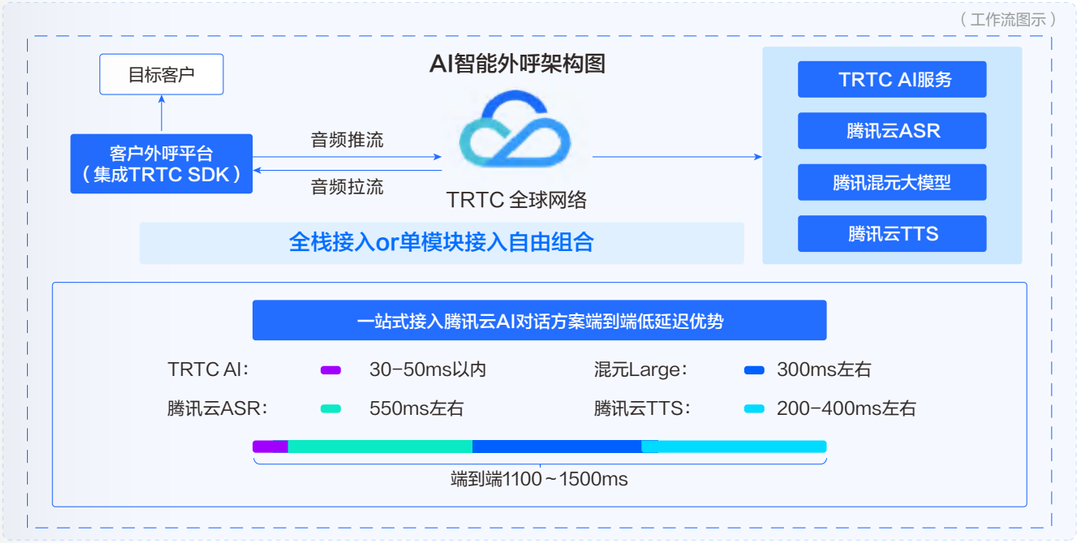
自然打断交互能力对AI系统提出高要求，需要系统能够准确识别用户意图，实现更接近真人对话的体验。

拟人效果仍需提升

灵活话术应答是外呼系统面临的核心挑战之一，传统外呼系统往往无法有效应对用户实时提问，导致转化率低；情感丰富拟人音色的生成也至关重要，机械化的语音容易引起用户抵触心理，影响沟通效果。

解决方案

- **智能语音识别：**定制8K电话模型，提升复杂场景下识别准确率与对话流畅度。
- **降噪与打断：**依托腾讯云TRTC，AI降噪结合智能打断，提升听觉清晰度与交互自然度。
- **混元大模型：**采用混元-Large低延迟模型，首Token延迟约300ms，实现实时理解与灵活话术生成。
- **智能语音合成：**基于腾讯云TTS超自然音色库，多音色多风格选择，打造拟人化情感语音体验。





天润融通

标杆案例

天润融通

业务场景

天润融通，AI驱动的一体化客户联络云平台，专注服务企业在“营、销、服”三大核心运营流程中的客户联络场景。

天润融通推出面向服务与营销场景的对话式智能体 ZENAVA，具备“能沟通、会执行、懂学习、可成长”的能力，实现从对话到执行，从学习到优化的智能闭环。它不仅是企业的AI员工，更是企业新一代智能伙伴，助力企业完成从“人力驱动”到“AI驱动”的生产力与组织形态实现。

客户通过接入天润融通客户联络平台，即可使用智能体在全渠道为客服及营销等业务场景提供一站式客户联络服务，该平台与企业自有业务系统无缝对接，所有能力统一运行在同一平台之上，流程高度协同、数据统一流转，显著提升运营效率，降低整体运营成本。

业务挑战

▶ **实现交互拟人化**

需要通过高度拟人化的交互体验降低用户的沟通门槛，具体包括能够理解语言的细微差别，包括口语化表达、隐含意图和地方性表述。在对话过程中，准确识别用户情绪变化，并以带有适当情感色彩的语气进行回应，支持流畅的多轮对话与自然的话题切换。

▶ **增强海量知识库管理能力**

需要处理海量知识包括产品文档、技术手册、常见问题、政策法规等，确保在任何专业领域都能给出准确、一致的答案。同时支持知识的自动化更新与版本控制，通过持续学习机制保持知识库的时效性与准确性。

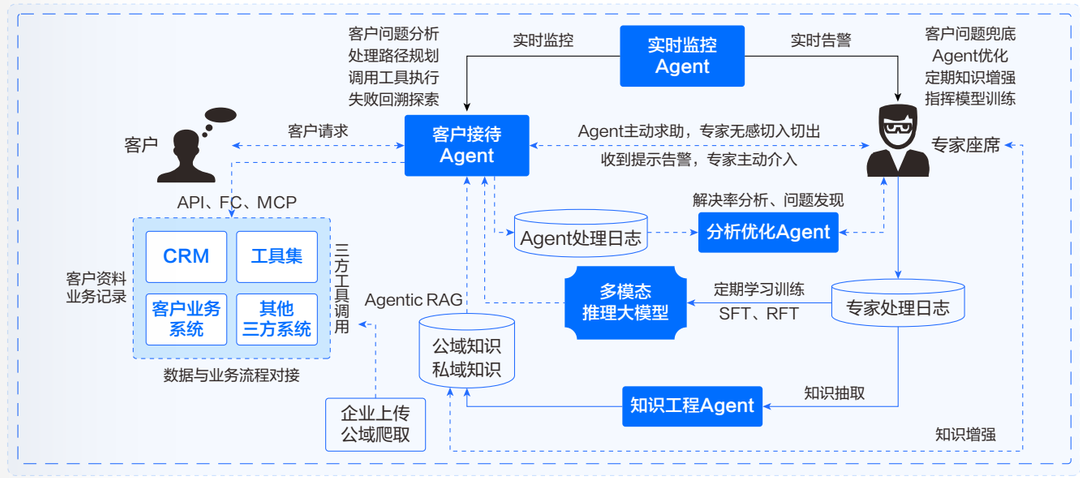
▶ **满足不同企业复杂业务场景需求**

需要适配企业各种业务场景，实现从单纯问答到业务办理的能力升级，真正成为企业业务流程中不可或缺的智能枢纽。

解决方案

- **智能客服拟人化方案：**天润融通基于腾讯云混元大模型能力，在自身客户联络平台中构建多模态智能客服，实现文本、图片、语音等多场景的自然理解与智能交互。
- **智能知识库管理方案：**依托腾讯云智能体开发平台能力，天润融通构建智能知识库体系，通过大模型自动生成问答与文档校验机制，持续优化问答质量与服务效率。
- **多Agent协同方案：**基于腾讯云智能体开发平台的Multi-Agent框架，天润融通搭建多Agent协同体系，实现任务拆解、工具调用与流程协作，提升服务灵活性与处理效率。

架构图





（技术架构图）

客户价值

在大模型能力加持下，客服回复准确率达到 **85%** 以上，响应时效在 **5s** 内。

多Agent能力实现的自助答复，拦截 **80%** 常见问题的咨询，节省 **50%** 人力。

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

轻喜到家

业务场景

深圳轻喜到家，到家服务平台（保姆 / 保洁 / 月嫂等）人员自营自养、服务标准化，采取「核心城市深耕 + 梯度扩张」策略，以深圳为样板城市逐步复制全国。

业务矩阵覆盖服务商城、轻喜优品、闪购秒杀、喜币商城等多元场景。

68城	1000万+	2万+	10亿+
覆盖城市	家庭用户	员工规模	年营收（元）

业务挑战

供需错配

客户需求与阿姨技能信息不对称，传统人工调度匹配成功率低、资源浪费严重。

效率瓶颈

派单、定价、人员管理高度依赖人工经验，标准不统一，难以规模化复制。

服务黑盒

服务交付过程缺乏有效监控与数据反馈，质量难保证，问题发现严重滞后。

解决方案

基于腾讯云 ADP 构建「小喜」智能体，全流程 AI 赋能

架构图



客户价值

智能客服降本增效

全渠道 7×24h 在线，人工客服压力显著降低

精准营销提转化

客户分层 + 动态标签，个性化推荐提升粘性

自动化质检控质量

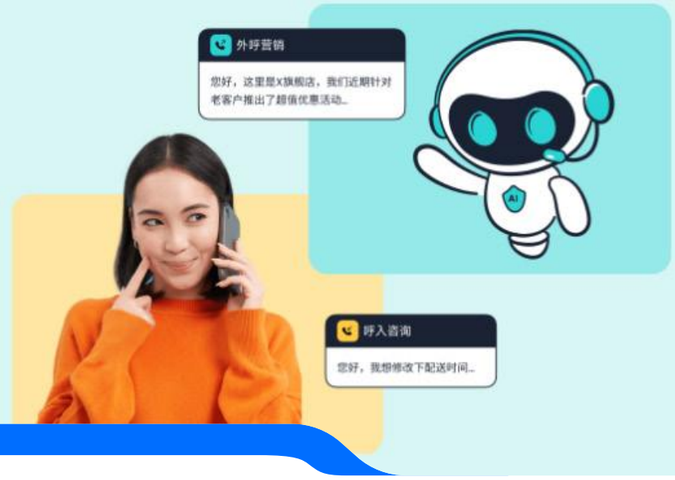
关键节点图片 AI 审核，服务质量可控可追溯

数据资产沉淀

碎片化经验转化为结构化数据，支撑规模化运营

相关产品

ADP	人脸核身	即时通讯 IM	VOD
LLM+RAG+Workflow 多框架，一站式构建企业级智能体	用户 / 服务人员注册核身，保障身份真实	管家、服务人员、用户消息互动与营销推送	服务录屏、服务商品视频的存储与分发



标杆案例

智齿科技

业务场景

智齿科技定位于一体化客户联络解决方案提供商，致力于帮助企业降低成本、增进效率，提升销售额及用户满意度。通过不断整合前沿的人工智能及大数据技术，智齿科技构建了包括呼叫中心、机器人「在线+呼入+呼出」、人工在线客服、工单、企微智客等在内的多个软件产品，并提供BPO外包服务。

业务挑战

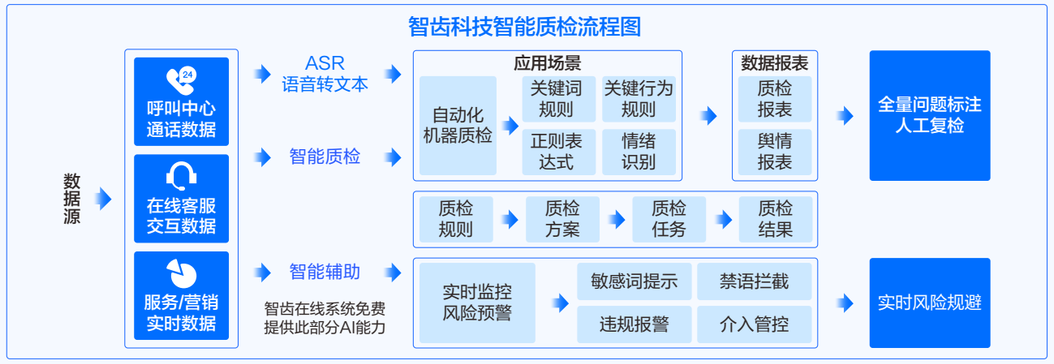
亟需面向质检复杂场景的语音识别能力

智齿科技日录音处理量峰值高达20000通，语音识别的时长将近700个小时，需要支持方言识别，噪音场景下的语音识别。

亟需构建质检数据分析平台

利用大数据分析生成多维度质检报告，如客户满意度趋势、服务短板统计，深度挖掘数据价值。

解决方案



（工作流图示）

- **质检语音识别方案：**腾讯云基于海量数据积累和业内领先算法，提供落地经验丰富、接入体验最优的语音识别ASR能力，其ASR在实验室环境下识别准确率高达98%，同时在噪音环境通过降噪技术提升识别准确率。
- **质检数据分析平台：**基于腾讯云TCHouse-D构建的质检数据分析平台，能够实现对海量语音、文本等质检数据的实时分析处理，帮助企业快速洞察服务质量、合规风险及业务趋势。

客户价值

助力客户实现高效的服务质量管理，提升客户服务品质，效率提升 **20%** 以上

提供全面的业务统计分析能力，风险发现的准确率提升 **15%** 以上，快速支撑客户经营决策。

更多免费行业报告 并购家 www.ipoipo.net

09

WorkBuddy生态共建

共创超级团队

方案

9.1 生态能力全景 / P132

案例

标杆案例——智慧芽 / P133

标杆案例——蜜度 / P137

标杆案例——MasterGo 莫高设计 / P143

标杆案例——天财商龙 / P147

9.1 生态能力全景

五大能力全面开放，伙伴把行业 Know-How 装进 Agent

Ecosystem Overview · 首期开放五大生态能力

No.1	10 万+	100+家	3000万+
桌面办公智能体	生态资产	共创伙伴	累计安装

上下文的三类限制：场景难整合 / 系统难串联 / 行业方法论难被 AI 用 → 开放五大生态能力，让伙伴的专业积累成为 Agent 可直接连接的能力

Buddy 应用
打包完整场景，开箱即用

让行业 Know-How 与专业数据，打包成开箱即用的应用

场景级

专家
把行业专家变成随叫随到的 AI

多年沉淀的经验与方法论，在 AI 专家中完美复现

角色级

Skill
给 Agent 装上任意新技能

一项专业能力、一个数据接口，被 Agent 随时调用

能力级

连接器
打通外部系统，Agent 直接读写数据

企业系统与外部数据，在 Agent的工作流中畅通无阻

系统级

硬件
给硬件注入 AI 灵魂

设备一键唤起 WorkBuddy，说话即可下令、状态随时掌握

设备级

一套底座，把伙伴的专业能力送到用户面前

接入链路：入驻 → 资质认证 → 创建 → 调试 → 审核 → 上线 → 运营

伙伴的专业积累（行业Know-How · 专业数据 · 系统与设备） → 五大生态能力接入（Buddy应用 / 专家 / Skill / 连接器 / 硬件） → WorkBuddy Agent 底座（模型 · 记忆 · 工具 · 权限，统一封装为标准 API） → 用户的日常入口（打开即用，行业能力跟着人进入工作）

已入驻软件应用伙伴（开放平台展示）

No.1桌面办公智能体 10万+生态资产 100+家共创伙伴 3000万+累计安装

LinkFox	美图	同程	企查查	金数据	WPS Office	轻流	利唐 i 人事
Moka	万方数据	智慧芽	钛动科技	云账房	鹏城信息	汇易联	通达信

接入规范：OAuth 2.1 · Open API · Scope 授权 —— 用户显式授权后按最小权限调用；腾讯把沉淀的 Agent Harness 能力封装为标准 API，底座由平台筑牢，行业场景交给伙伴。

智慧芽

标杆案例

智慧芽

智慧芽是一家服务全球科技创新的AI原生平台企业。其打造AI原生平台，为全球的研发工程、生物医药、材料科学、知识产权等业务场景提供AI原生的研发智能体、AI赋能的专业应用和科技创新数据支持，帮助企业等创新机构将概念想法转化为突破性的技术成果。

案例概要

WorkBuddy接入智慧芽连接器，用户可通过自然语言直接调用2.08亿+全球专利及2.16亿+科研文献数据。支持技术查新与立项调研、竞品专利动态监测等场景，将传统“逐篇检索”升级为自动化完成专利检索、趋势分析、风险识别及结构化报告生成的一站式创新研发工作流。

应用场景

▶ 场景一：技术查新与立项调研

研发立项前，用自然语言描述技术方向，WorkBuddy调用智慧芽全球专利和文献库，检索相关专利与论文，分析技术趋势、主要申请人及创新空间，快速生成结构化查新报告，大幅缩短立项决策周期。

▶ 场景二：竞品专利动态监测

输入竞争对手名称和技术领域，WorkBuddy通过智慧芽MCP按申请人过滤检索最新专利动态，结合IPC分类和时间维度分析竞品技术布局变化，识别潜在风险信号，输出带行动建议的竞争情报简报。

▶ 场景三：专利侵权风险排查

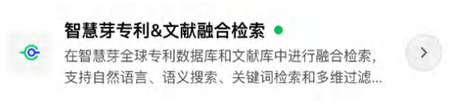
输入自身产品技术方案关键词，WorkBuddy调用语义搜索和关键词检索，定位高相似度专利，拉取权利要求和说明书全文进行比对分析，辅助研发团队在产品上市前快速识别潜在侵权风险点。

▶ 场景四：前沿文献速览与技术选型

针对特定技术方向，WorkBuddy同时检索全球专利与学术论文，按受理局、时间、IPC分类等多维度筛选，以Markdown格式呈现摘要和核心观点，帮助用户在技术选型时快速掌握领域前沿进展。

接入方式

点击左侧栏「专家 · 技能 · 连接器」→ 选择「连接器」→ 搜索“智慧芽专利&文献融合检索”→ 点击「连接」→ 手机号注册或微信快捷登录 → 获取并填写API Key → 按提示完成配置。



使用示例

➤ 技术查新与立项调研

以前的工作：反复查阅技术文档、下载专利相关资料、逐篇比对、画技术路线图……好几天才磨出一版材料，交上去还怕漏了关键对比文件。

现在我们只需要把项目说明书、技术方案和既有研究资料交给WorkBuddy：

提高锂电池能量密度的技术方案.md

请根据我的提高锂电池能量密度的技术方案，检索相关专利和科研文献，分析现有技术、主要申请人、技术发展趋势以及我们的潜在创新空间，并生成一份立项查新报告。

➤ WorkBuddy调用过程



更多免费行业报告 并购家 www.ipoipo.net

生成提高锂电池能量密度的技术方案

报告概要

检索规模：通过智慧芽检索全球 **74万+** 件相关专利 + **10,875** 篇学术论文，精读 95 件高相关专利。

主要发现：

维度	结论
主要玩家	LG Chem/LG ES（单晶正极）、Samsung SDI（固态电解质）、CATL（结构创新）、Toyota（全固态电池）、AGC（硫化物导率最高 >2 mS/cm）
技术趋势	单晶正极 → PDI精确控制 → Zr/Y共掺；硅碳负极 → 多孔碳限域 → Li_3P改性；固态 → 硫化物玻璃态 > 2 mS/cm
专利壁垒	LG Chem 的单晶 PDI 控制和 Zr/Y 掺杂需重点规避

5 个潜在创新空间：

1. 🔷 硅碳负极 + 固态电解质界面协同设计（空白地带）
2. 🔷 预锂化 + 干法电极工艺一体化（两个独立路径的融合）
3. 🔷 多孔碳分层三级功能化改性（外层导离子/中层硅/内层补锂）
4. 🔷 AI 驱动正极材料逆向设计
5. 🔷 电芯级多维度协同优化方法论

今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令

➤ 最终生成的报告

锂电池能量密度提升技术 — 立项查新报告

委托方：[内部立项]
查新目的：科研立项
检索日期：2026年8月6日
数据来源：智慧芽（Patsnap）全球专利数据库 + 学术文献库
检索范围：全球专利（US/EP/CN/JP/KR/WO）+ 核心学术论文

一、查新项目技术要点

本项目拟通过"高镍正极材料升级 + 硅基负极/预锂化 + 固态电解质 + 电芯结构创新"的多路径协同方案，实现锂离子电池能量密度从当前 250-300 Wh/kg 向 350-500 Wh/kg 的突破。核心技术要点如下：

1. **高镍单晶正极**：NCM9系单晶化 + Zr/Y 共掺 + 梯度浓度包覆
2. **高硅碳复合负极**：多孔碳限域硅沉积 + 纳米磷化锂界面改性 + 预锂化
3. **固态/半固态电解质**：硫化物玻璃态电解质 + 复合膜层结构设计
4. **干法电极工艺**：无溶剂辊压成膜 + 厚电极设计
5. **电芯结构集成**：CTP 3.0 + 超薄复合集流体

二、检索策略

检索维度	IPC分类号	关键词	时间范围	检索量
高镍正极	H01M4/525, H01M4/505	NCM811, single crystal cathode, concentration gradient	2021-2026	14,237 件
硅基负极	H01M4/38, H01M4/134	silicon anode, Si/C composite, SiOx, prelithiation	2021-2026	20,041 件
固态电解质	H01M10/0562, H01M10/0525	solid electrolyte, sulfide, LLZO, garnet	2021-2026	53,947 件
干法电极/CTP	-	dry electrode, solvent-free, CTP, cell to pack	2021-2026	641,878 件
学术论文	-	energy density, lithium metal, solid-state	2015-2026	10,875 篇

三、现有技术分析

3.1 高镍单晶正极

技术现状：

- 单晶高镍 NCM 是当前最活跃的专利申请方向，2021-2026年全球相关专利超过 1.4万件
- 核心技术聚焦于：单晶粒径均一性控制（PDI 1.0-2.0）、Zr/Y 双掺杂、表面 Al_2O_3/W 复合包覆
- 已有方案基本覆盖了从前驱体共沉淀、单晶粒子烧结到表面改性的全流程

主要申请人：

排名	申请人	国别	核心技术方向
1	**LG Chem / LG Energy Solution**	KR	单晶 NCM（≥50%Ni）、Zr/Y 掺杂、PDI 控制
2	**POSCO Future M**	KR	单晶金属氧化物颗粒型包覆层（≥35%覆盖率）
3	**Panasonic**	JP	Li-Mn-Co 无序岩盐结构（Fm-3m）、碳膜/碱金属修饰
4	**Contemporary Amperex (CATL)**	CN	钠电正极 Zn 掺杂层状氧化物、正极添加剂
5	**Guangdong Brunp (邦普)**	CN	三元材料包覆效力因子（杨氏模量×厚度比/方差）
6	**Toyota**	JP	CeO_2/ZrO_2 储释氧材料抑制容量衰减
7	**Sumitomo Metal Mining**	JP	高镍氢氧化物前驱体结晶控制（超饱和区体积分率）

3.2 硅基负极与预锂化

技术现状：

- 硅碳负极相关专利超过 2万件，是近 5 年增长最快的方向之一
- 技术热点：多孔碳限域硅 CVD 沉积（孔道内硅 + 外碳包覆）、硼/氮/硫掺杂多孔碳、Li_3P 原位改性
- 电解液适配方面：FEC + VC + 酯类电解质体系已成为 Si 负极标配

主要申请人：

WorkBuddy会通过智慧芽连接器，完成：

- 从技术方案中提取技术特征；
- 调用“专利&文献融合检索”帮你搜；
- 调用“全球核心专利库”查法律状态；
- 调用“专利图表分析”做趋势；
- 调用“创新与专利报告生成”输出结构化查新报告。

大幅缩短从资料检索到立项决策的执行周期。

midu 蜜度
垂直大模型企业



标杆案例

蜜度

案例概要

蜜度Skills：一站式赋能办公写作、内容校对全场景

蜜度提供公文写作、文本校对、时政内容检索、热点追踪、拼音校对与注音等AI技能，为政务、新闻、教育、金融、出版等各行业提供办公写作、数据洞察与编校审核的一体化智能方案。将AI+办公技能无缝融入WorkBuddy写作流程，随需调用，显著提升办公场景的创作效率与内容质量。



应用场景

▶ 场景一：日常办公写作

日常工作中，周报、总结、会议纪要、商务邮件等办公文档撰写需求频繁。在WorkBuddy对话中输入写作需求，调用蜜度公文写作技能，即可快速生成结构清晰、语言规范、可用性强的文档初稿，并支持通过多轮问答完善内容。

▶ 场景二：公文写作与文书起草

面向政务及企事业单位，适配严格的公文格式与语言规范。在WorkBuddy中调用蜜度公文写作技能，可快速起草通知、请示、报告等公文文稿，结构完整、用语专业，套用本单位格式即可发文。

▶ 场景三：文本校对与质量审核

在WorkBuddy中粘贴或上传待校对内容，调用蜜度文本校对技能，即可对错别字、标点、事实差错等35类问题一键检测。系统将输出带修正建议的勘误报告，支持逐条采纳或一键修正，全方位保障内容准确性。

▶ 更多场景

时政内容检索与政策研究	出版教学注音与拼音校对	热点追踪与数据洞察
会议纪要整理与待办事项提炼	新闻稿件与宣传文案生成	新媒体文案写作与质检
.....		

接入方式

1、在 WorkBuddy 技能市场中搜索「蜜度」，或浏览「推荐」分类。

2、找到「蜜度公文写作」「蜜度文本校对」等技能卡片。

3、点击「+」完成安装并使用，根据提示获取短信验证码激活账号。

4、后续可在对话中直接调用，或在技能面板中查看已安装的蜜度技能。



更多免费行业报告 并购家 www.ipoipo.net

➤ 调用蜜度公文写作技能示意



➤ 最终交付的校对结果示意



使用示例

以下展示一个任务示例：如何在WorkBuddy中一站式完成文稿快速撰写与校对。以防御台风“白海豚”应急通知为例，在一次对话中，串联使用蜜度公文写作与蜜度文本校对两个技能，同步完成写作与校对。

➤ 以前 vs 现在

一次对话完成写作与校对

环节	以前：人工方式	现在：WorkBuddy+蜜度技能
起草通知	台风临近，紧急查阅公文范文、对照格式规范，逐段撰写、反复斟酌用语，约1小时	一句话输入需求，蜜度公文写作技能基于时政资料库自动生成，分钟级完成
内容校对	逐字人工核对，错别字、标点、事实差错易漏检	蜜度文本校对技能自动检测35类错误，输出带修正建议的勘误报告
流程衔接	写作、校对分属不同环节，需切换工具处理	一次对话串联两个技能，“写作→校对”一站式完成

➤ 使用方式

一句话同时调用蜜度公文写作与文本校对

➤ 用户输入

「帮我写一份防御台风“白海豚”的应急通知，并对内容进行校对」

➤ WorkBuddy调用过程



➤ 最终交付的结果：



效果价值

效率

公文/报告写作时间从小时级缩短至分钟级，文本校对从人工逐字检查升级为技能快速校对。

质量

文本校对覆盖35类错误，帮助减少差错，提升发布内容质量；公文写作基于高质量时政资料库，保证内容准确性与规范性。

成本

减少重复性写作和校对人力投入，释放从业者精力至创意和策略层面。

体验

多轮对话交互，支持持续追问和多次优化，写作过程自然流畅。

蜜度 Skills 已入驻 WorkBuddy 技能市场，搜索「蜜度」即可体验。

更多免费行业报告 并购家 www.ipoipo.net



标杆案例

MasterGo 莫高设计

案例概要

MasterGo 莫高设计将最新升级的 MCP 能力上架 WorkBuddy。

通过 MCP 协议，MasterGo 将整块画布向大模型全面开放，AI 得以深度连接画布，并与 WorkBuddy 保持双向连接，使 AI 能够在画布上完成原生的读写操作，帮助设计师更高效地产出符合企业规范的设计稿，也让研发更准确地理解设计意图，实现高质量还原。

生成、读取、修改、同步，构成了 MasterGo 画布的四大核心能力。

- MasterGo MCP 能够直接生成完整的组件体系，同时创建多种风格的设计提案，大幅缩短从创意到落地的路径。同时，MCP 可自动读取既有设计稿的视觉语言并形成规范，在后续页面生成中严格保持视觉一致性。
- 凭借 MCP 对设计稿的深度解析，AI 最终交付的是还原度极高的前端项目。
- 选中即定位，无需再费力描述图层名称、位置和层级，让 AI 真正做到指哪改哪、精准执行。AI 能精准识别出设计稿改了哪些地方，并将修改内容更新到对应的前端代码中，不会破坏原本写好的工程代码与业务逻辑。
- 此外，MasterGo MCP 还支持生成图表，生成符合企业规范的前端代码，加强对团队库的识别和引用，支持使用团队预设中的字体。

使用方式

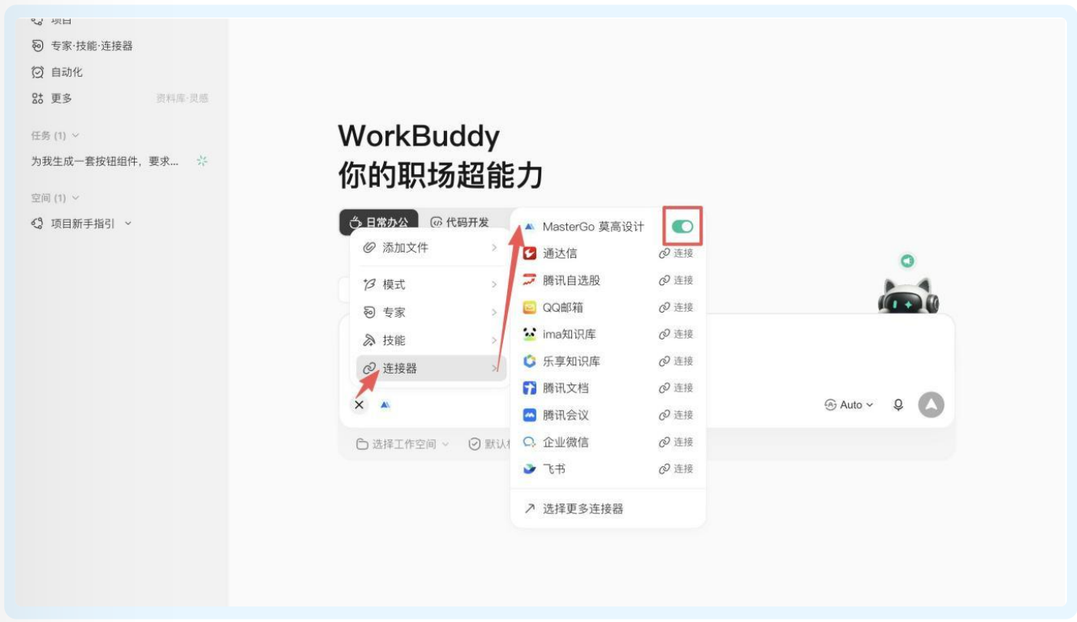
1.在WorkBuddy连接器中找到「MasterGo 莫高设计」并完成连接。



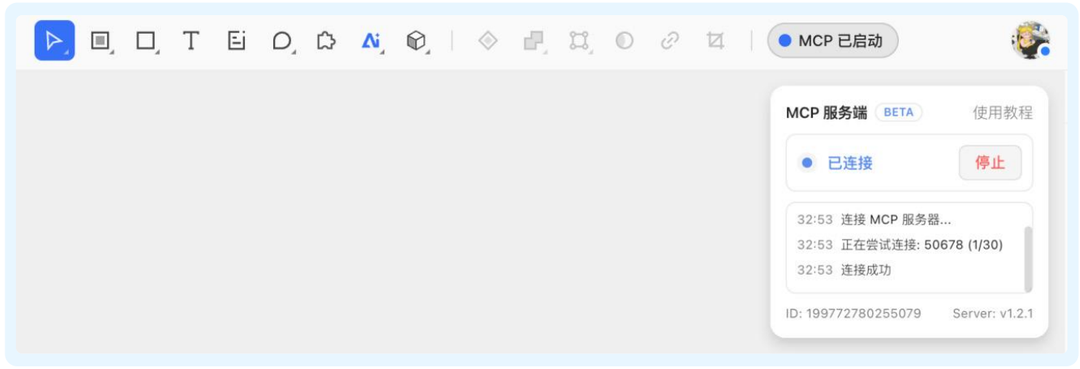
2.通过「去试试」按钮启动「MasterGo 莫高设计」连接器。



3.也可以在新建任务时添加「MasterGo 莫高设计」连接器。



4.打开 MasterGo 连接「MCP」



5.在 WorkBuddy 中输入指令「使用 MasterGo MCP，.....」就可以开始使用啦！



更多免费行业报告 并购家 www.ipoipo.net



标杆案例

天财商龙

天财商龙将最新升级的「天财商龙餐饮SaaS」连接器上架WorkBuddy。

案例概要

通过WorkBuddy连接器，商龙将二十余年餐饮数字化沉淀的运营方法和SaaS系统入口向大模型开放，企业通过授权可以安全连接餐饮、会员、供应链等核心真实业务数据。从此，门店经营者和总部管理者无需学习报表、无需找人取数，一句自然语言提问，即可获得经营分析结果。

数据沙箱隔离
每个商户数据完全隔离，互不可见

只读分析
只做读取与分析，不改动您的业务数据

权限内取数
按账号既有权限访问，不越权取数

留痕可审计
访问与分析全程留痕，可追溯

➤ 查询、汇总、对比、解读，构成了商龙经营洞察的四大核心能力。

- **一句提问直接触达营销ROI：**支持用一句话按时间、活动名称和活动创建人，查询发券量、用券量、收益等活动明细和核销情况。实现营销活动ROI自动分析、重点活动与异常定位，为活动复盘、营销策略优化和经营决策提供可靠的数据依据。
- **分析口径清晰、可追溯：**每一次回答都说明业务口径、时间范围与统计范围；当返回数据出现明显异常（如核销率明显低于正常范围）时，会在回答中提醒你先核查，而不是直接基于错误数据做分析。核心指标准确，趋势分析可靠。
- **源于二十余年餐饮数字化实践：**分析维度贴合餐饮经营实战，覆盖正餐、快餐、火锅、烧烤等全业态，把服务全国十余万餐饮门店的经验沉淀进每一次回答。
- **开放可扩展的数字员工体系：**营销活动分析之外，还可逐步扩展巡店巡检、门店经营预警、供应链协同等更多场景技能，用得越久，越懂你的企业。

使用方式

1.在WorkBuddy连接器中找到「天财商龙餐饮SaaS」。



2. 点击「去试试」，按指引完成授权，即可启用连接器。



3. 在WorkBuddy中输入具体指令即可开始使用，如「上个月的营销活动效果怎么样」。





➤ **你还可以这样问**

昨天各门店营业额、客流和优惠怎么样？

本周会员消费排行和复购留存有什么变化？

哪些菜卖得最好，毛利贡献最高？

各渠道营收占比环比有什么变化？

上月供应链原物料成本走势如何？

合作客户

测测
懂球帝
天富下通
octask
花椒直播
慧算账
小宇宙
StepFun
KIMI
kuaizi
BlueFocus
LibLibAI
ec
MABANG
莫高设计 MasterGo
midu 蜜度
明略科技 MININGLAMP TECHNOLOGY
熵简科技 Value Simplex
盛和游戏 SHENGHE GAME
畅游 changyou
天商财龙 连锁餐饮数字化 CATERING DIGITALIZATION 整体闭环解决方案
Taomee淘米
玩心不止 GAME FOREVER
游卡 YOKAVERSE
YOOZOO游族
智慧芽
智齿科技
ZIXUN紫讯

更多免费行业报告 并购家 www.ipoipo.net