

华为云 AI Token 服务客户案例集



华为技术有限公司

深圳龙岗区坂田华为基地

电话: +86 755 28780808

邮编: 518129

网址: www.huawei.com





目录 CATALOGUE



Hobby携手华为云Tokens服务
AI社交迎来“超级引擎”

01

赞奇科技 X 华为云Tokens服务
精准识别工业图纸，助力高效生产

03

数云 X 华为云Tokens服务
为消费者数字化运营装上“智能引擎”

06

方寸无忧 X 华为云Tokens服务
以“方寸智脑”重塑“AI+公文写作”数智新时代

09

合力亿捷 X 华为云Tokens服务
打造智能客服新范式

12

今日人才 X 华为云Tokens服务
以“数智员工”破解B2B销售困局

14

鲸海拾贝 X 华为云Tokens服务
给AI教育带来新范式

17

Hobby X 华为云Tokens服务 AI社交迎来“超级引擎”

■ 就现在，与AI交个朋友吧！

他就是年轻人最in的AI社交APP——上海微烽科技开发的Hobby，2025年新版上架就在苹果App Store社交类下载TOP10榜单停留近两个月，成为今年社交领域新应用的代表产品。

Z世代的社交已经不仅仅是聊天打卡，而是追求更进阶的沉浸式体验。作为一款为年轻人量身打造的AI社交APP，Hobby主打潮流派对式共创体验：用户既能拥有专属虚拟角色，又能和虚拟的AI朋友一起玩剧本杀、语聊甚至云养崽。这些“未来感满满”的场景，正在Hobby潮流社区里发生。

而要支撑这种高频、海量、个性化的互动，背后少不了一台“超级引擎”。Hobby选择华为云Tokens服务，通过DeepSeek大模型的加持，把年轻人的潮流派对玩出了新高度。

华为云MaaS平台就像是“Tokens发动机”，通过高吞吐低时延的算力，让大模型调用更快，从自然语言对话、剧情生成到内容审核，从虚拟角色塑造到社交体验优化，从高并发到内容合规，性能和体验一起飞升。



01 炸场的“AI社交”需要应对很多技术挑战

在Hobby“元宇宙版朋友圈”中，年轻人“搞事情”的方式千奇百怪，但要让派对“炸场”，背后其实有不少挑战：

- **虚拟角色性能压力大：**日活用户数十万，每个用户都需要一个专属虚拟角色，对大模型推理性能要求极高；
- **流量波峰波谷差异大：**业务波峰波谷流量，活动推广期流量猛增，需要算力弹性灵活扩展；
- **内容安全审核压力大：**虚拟社交氛围轻松自由，但内容合规必须严防，不然违规内容可能影响用户体验，严重时甚至会导致APP下架；
- **大模型响应速度要求高：**Hobby依靠AIGC实时驱动，剧情互动、实时对话、内容推荐等功能提供了多样化的社交体验，但对大语言模型的稳定性和响应速度提出了更高要求。

02 华为云Tokens服务“神助攻”快、稳、放心的三重保障

- 首先，大模型训推效率提升，训练推理速度提升10-30%，用户端侧时延大幅降低。与AI虚拟角色对话时，真正做到“秒回”；
- 其次，支持弹性扩容，高并发场景下性能稳健，就算派对“人满为患”，也能从容应对，平均时延优于业界水平；
- 最后，在内容审核方面，通过敏感词命中与语义拦截双重防护，确保社区氛围自由又合规。

在具体的核心任务场景，华为云MaaS平台相比业界水平做到了显著提升。在“用户兴趣爱好评估”场景中领先15%-23%，可毫秒间识别用户意图，实时做出推荐；“用户情感状态分析”场景中性能提升12%-13%，哪怕流量高峰也能迅速“读懂”用户情绪，回应又快又准；就连“生成开场白”这种高频任务，也保持了10%-12%的性能优势，确保成海量用户同时聊天，AI照样响应迅速、对答如流。

华为云MaaS平台让Tokens像水电一样触手可及，企业可以开箱即用，用户可以体验到更聪明的服务。未来，Hobby将在“潮流派对”的舞台上，给Z世代带来更多想象力与新玩法，而华为云MaaS平台也将把高吞吐低时延的Tokens源源不断地输送到万千场景，加速AGI时代的到来。



赞奇科技 X 华为云Tokens服务

精准识别工业图纸，助力高效生产

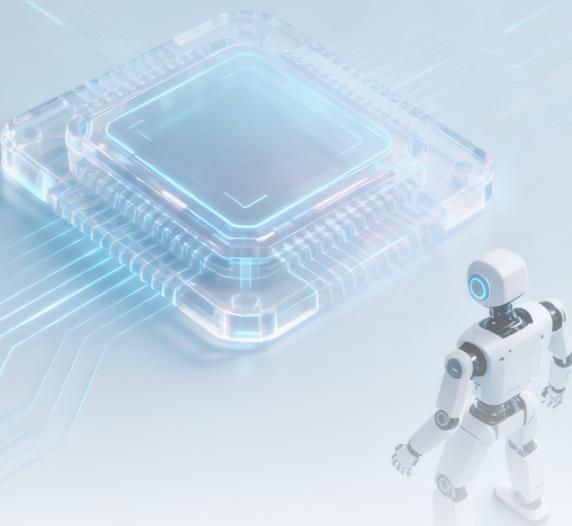
如今企业对AI的需求已从“单一功能应用”转向“全场景智能协同”

赞奇科技基于其在数字视觉行业多年的项目经验和行业积累，推出AIKNOW智能体编排服务，聚焦制造和互联网企业，以“图纸信息高效识别、业务流程紧密连接”的核心优势，成为AI赋能企业数字化转型的“关键助力”。

凭借“低代码部署+高定制化”的特性，赞奇AIKNOW智能体编排服务打破了传统AI应用“场景单一、适配性弱”的局限，让企业无需投入大量成本开发多套系统，即可通过AIKNOW快速激活不同业务环节的智能潜力。而支撑这一解决方案实现“多场景适配、高稳定运行”的核心，离不开强大的算力与模型支撑。

为此，赞奇科技选择与华为云Tokens服务深度合作，依托华为云MaaS平台的高吞吐、低时延特性，让企业在各业务场景下的AI应用更流畅、更高效，真正将AI的价值转化为企业的业务增长动力。

赞奇科技聚焦制造行业、服务不同企业的过程中，发现企业智能决策与图纸识别的“进阶之路”，核心需求主要集中在“图纸处理-标准查询-文档生成”三大环节，每个环节存在的痛点也逐渐显现，具体场景与需求如下：



01 日活研发任务量激增大模型面临严峻性能挑战

针对企业的非标产品，每张图纸均需通过大模型提取非标技术参数、异构零件、工艺注释等信息，且随着日活研发任务量激增，大模型面临严峻性能挑战：

- 推理速度滞后：**传统算力支撑下，当季度末交付高峰图纸处理量骤增3倍时，工程师需等待的时间大大延长才能获取结果，严重拖慢研发节奏；
- 识别准确率波动：**受图多种因素影响，大模型对模糊信息的语义理解能力不足，参数漏判、误判率高，需人工反复复核，额外增加30%的工作量。

02 高频查询场景下大模型响应与稳定性不足

研发过程通常需严格遵循大量的国家标准与技术规范，大模型需支撑“自然语言查询-标准定位-关联推荐”全流程。但高频查询场景下，大模型暴露明显短板：

- 响应时延高：**多位工程师同时发起查询时，系统平均响应时间超300ms，远超“毫秒级”的实际需求，导致设计决策卡顿；
- 并发稳定性差：**当查询请求量突破一定标准，大模型易出现“排队拥堵”，甚至偶尔触发服务中断，影响研发连续性。

03 弹性的高要求让传统服务难以满足

基于图纸数据生成数据报表、工艺指导文档时，大模型需完成“数据校验-标准匹配-模板填充”的全流程编排。但该场景对算力弹性的高要求，让传统服务难以满足：

- 算力弹性不足：**文档生成需求随研发进度波动，传统固定算力要么在波峰时“供不应求”，要么在波谷时“资源浪费”，算力成本居高不下。



针对制造研发场景中的大模型痛点，华为云Tokens服务以“快、稳、省”为核心，为大模型应用提供全链路支撑，成为企业研发数字化转型的“关键引擎”：

04 提升大模型训推效率实现“秒级响应”

华为云Tokens服务依托昇腾AI芯片的高性能算力，大幅优化大模型推理与响应速度，直击“慢”的痛点：

- 推理速度提升10-30%：**将图纸参数提取的大模型推理时延和文档生成的全流程处理时间大幅缩短，工程师上传图纸后“即时见结果”，彻底告别等待；
- 语义理解准确率升级：**通过Tokens服务的高吞吐算力对大模型进行持续优化，增强对模糊图纸信息的语义解析能力，参数识别准确率提升至95%以上，人工复核工作量减少80%，研发效率显著提升。

05 支持按需调用保障大模型“稳如磐石”

针对业务波峰波谷与高并发需求，华为云Tokens服务提支持算力按需调度与使用情况实时反馈：

- 高端算力即开即用：**华为云Tokens服务支持按需使用，避免资源浪费，整体算力成本降低30%以上；
- 使用情况实时更新：**华为云ModelArts平台提供实时业务负载情况报表，为用户提供用量峰谷值情信息，便于企业基于实际业务情况更加深入了解核心业务板块情况。

华为云MaaS平台让Tokens像“企业AI应用的智能动力源”，持续为AIKNOW输送稳定高效的算力，而AIKNOW则将这种动力转化为企业的“智能竞争力”。

未来，江苏赞奇科技将继续深化与华为云Tokens服务的合作，推动AIKNOW在更多行业场景落地，尤其是在图纸识别、工程数据处理等细分领域持续突破，加速AI在千行百业的深度渗透，共同构建“高效、安全、全场景”的企业AI应用新生态。

数云 X 华为云Tokens服务

为消费者数字化运营装上“智能引擎”

在数字化浪潮席卷零售行业的当下，无论是连锁快消、美妆品牌，还是鞋服商家，都绕不开一个核心课题：**如何盘活数据、优化营销，留住客户并提升业绩？**数云作为深耕全渠道消费者数字化运营的资深玩家，正是帮品牌及零售企业解决这一问题的最优解帮手。



从2011年起，数云已经服务了超10000家零售品牌与企业，从区域连锁到全球性零售集团，均借助其方案让全渠道消费者数字化运营更高效。近年来，数云持续升级产品生态，推出全渠道消费者AI智能运营解决方案，让营销更简单、智能且具有价值。

企业想实时掌握业务数据，数云麒麟“Data Agent”可轻松实现——无需懂复杂代码，随口问“最近一周哪个品类卖得最好”，就能快速得到答案；若近期销售额突然下滑，它还能第一时间分析原因，定位问题；品牌要开展促销活动，数云麒麟“AI营销团队”可基于业务需求和数据洞察，自动策划方案、配置活动规则、生成适配宣传素材，一个系统就能搞定从策划到营销落地的全流程。

此外，数云还有智能回复客户咨询的“私域AI运营专家”、助力出海品牌获取海外流量的“智能SEO”等多重功能，这些工具串联成“数据+营销+服务”完整智能运营解决方案，为企业数据决策与营销增长提供实打实的数智化支持。

要支撑这类高频、海量、个性化营销，离不开强劲的“超级引擎”。数云选择华为云Tokens服务筑牢底层支撑，凭借高吞吐、低时延的硬核算力，让大模型调用更高效，实现从数据洞察、营销生成到服务响应的全链路提速，为AI应用打下坚实基础。

01 数云“AI动机识别”典型场景无忧应对大模型核心挑战

在全渠道消费者运营中，品牌对“用户理解”的需求正从“行为层面”向“动机层面”深化——传统的收藏、加购、浏览时长等用户行为指标，虽能反映用户动作，却不足以解释消费者“为什么”做此决策。

比如两位购买抗老精华的用户，浏览、下单动作完全一致，背后却可能是“敏感肌试错焦虑”和“美业门店拓客需求”两种截然不同的动机。若无法精准识别这种差异，品牌的营销沟通就会陷入“一刀切”困境，难以触达用户真实需求。

这正是数云聚焦的典型场景之一：如何通过AI能力深度挖掘每位用户的真实动机，进而制定个性化沟通策略与运营路径？为此，数云构建了“AI动机识别”能力，依托大模型对用户语义、行为轨迹、会话记录、历史偏好等信息进行多维度解析和语境推理，目前已在多个美妆、快消品牌落地，实现从用户标签构建、营销内容生成，到策略匹配、用户响应的闭环智能运营，让品牌的每一次触达都更精准。

但这一场景对大模型应用提出了多项挑战：

- **语义理解需有深度：**不能只停留在关键词提取，更要精准推断出语义背后的情绪（如焦虑、期待）、潜在需求（如安全诉求、性价比追求）、使用场景（如自用、送人）等复杂意图；
- **调用频次需实时：**在私域会话、社群运营、企微触点等实时互动中，需秒级完成动机识别和话术生成，避免因延迟影响用户体验；
- **系统承载需扛高并发：**在促销节点、新品推广等高峰场景下，需同时支撑数百万级用户的内容推理与决策。

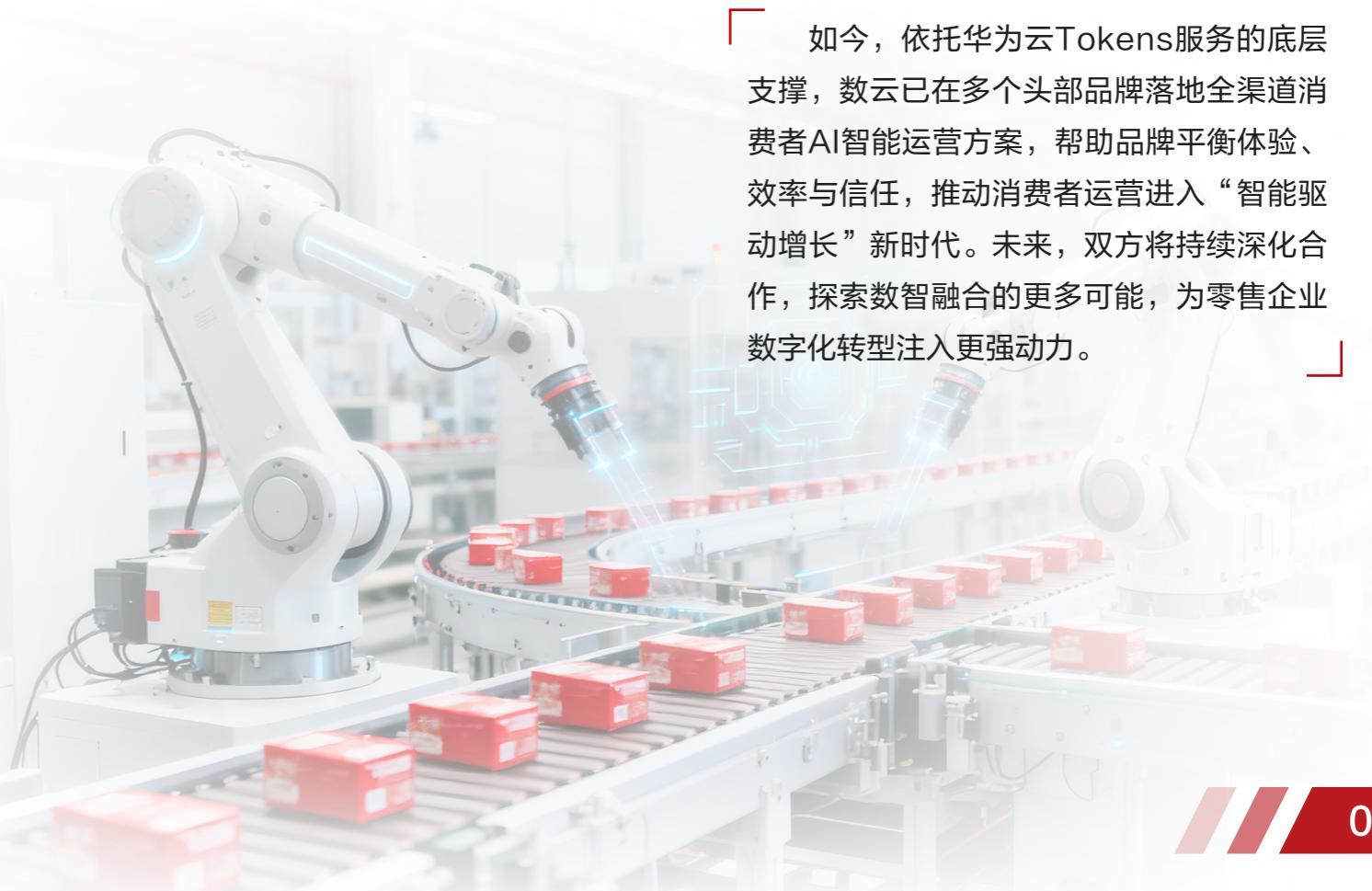


02 华为云Tokens服务为消费者智能运营保驾护航

数云依托华为云Tokens服务，搭建起高性能、高弹性、可控可扩展的AI推理架构，为该能力在复杂业务环境中的稳定运行提供强力支撑，具体价值体现在四方面：

- **推理提速，响应更即时**借助华为云Tokens服务优化的大模型调用链路，语义解析与动机识别速度平均提升30%，在导购对话、用户互动中实现毫秒级响应，真正做到“边聊边理解”，避免用户等待流失。
- **生成内容升级，沟通更贴心**依托华为云大模型的自然语言生成能力，系统可自动输出贴合用户身份、标签与场景的沟通内容，如“安心抗老计划”“项目推荐话术”等，让沟通更具策略性与人情味，提升用户感知价值与信任度。
- **弹性调度，高峰更稳定**华为云Tokens服务具备大规模算力调度与异构资源管理能力，在双11、新客招募、权益推广等高并发场景下，无需人工干预即可自动扩容，保障大模型稳定运营，避免因算力不足影响任务执行。
- **安全合规，运营更安心**华为云提供全链路内容审核与语义拦截机制，助力品牌在私域沟通、内容触达等场景中合规运营，降低违规风险，为企业级应用保驾护航。

如今，依托华为云Tokens服务的底层支撑，数云已在多个头部品牌落地全渠道消费者AI智能运营方案，帮助品牌平衡体验、效率与信任，推动消费者运营进入“智能驱动增长”新时代。未来，双方将持续深化合作，探索数智融合的更多可能，为零售企业数字化转型注入更强动力。



方寸无忧 X 华为云Tokens服务

以“方寸智脑”重塑 “AI+公文写作”数智新时代

它不仅是办公室笔杆子的智能助手，更是政企办公的“决策智囊”。北京方寸无忧科技发展有限公司推出的“方寸智脑”公文写作智能体，在公文材料写作水平上已经达到“人机难辨”程度，是国内首款通过“图灵测试”公文写作智能体。“方寸智脑”依托华为云Tokens服务的强大算力，正推动政企办公从“人工作业”走向“智能协同”新时代。

在政企办公领域，公文写作长期面临耗时费力、专业性强、容错率低等挑战。一台电脑、一堆材料的传统写作模式，早已无法适应AI时代高效、规范、智能的办公需求。

“方寸智脑”应运而生——它以国内知名公文写作专家写作经验为核心，以方寸9年积累的高质量海量数据为基础，基于通用大模型与方寸自研行业模型，结合华为云Tokens服务高性能算力，实现从信息检索到合规审校全流程智能化输出，真正做到“提笔即达意、落笔即规范”。

“方寸智脑”基于“专用模型底座+工具底座+文档底座”三大核心架构，不仅支持完全私有化部署，确保政务数据不出域、不泄露，还具备强大的跨领域知识融合与语义理解能力，可灵活适配各级党政机关、各行业的公文规范与表述习惯，是政企用户随需随用的“智能写作专家”。



01 复杂政企写作流程亟需技术加持的需求日益强烈



• 跨领域知识整合难

公文写作需融合政策法规、地方实际、行业术语等多维信息，要求系统具备动态知识图谱与跨域语义理解能力。例如，撰写一份关于制造业发展的计划，不仅需理解制造业政策、技术趋势、区域经济布局，还要熟悉本地企业情况、产业链短板及政策补贴导向。通用大模型往往因知识更新滞后、联网数据质量不高、领域适配弱而无法胜任，导致输出内容泛泛而谈，缺乏落地性。

• 语义合规性要求极高

政企公文不仅不能有错别字、语法和知识差错，更需准确把握政策文件的权威表述，避免各类政务性错误。通用大模型仅能完成基础字词、语法审校，且存在模型幻觉，无法及时跟进动态变化的政务领域各类表述。

• 写作流程复杂且依赖经验

从模糊需求到成稿，需经历搜索、分析、撰写、审校等多环节，传统工具无法实现全流程自动化。一名资深笔杆子通常需花费数天时间查阅资料、搭建框架、撰写内容并反复打磨，尤其在面对突发任务或跨领域题材时，压力极大。由于写作能力培养难度大，很多单位都面临笔杆子青黄不接的问题。

02 方寸智脑×华为云Tokens服务让智能写作“又快又准又稳”

“方寸智脑”依托华为云Tokens服务，接入方寸公文知识库与方寸自研行业模型，实现了对跨域信息的实时抓取、语义关联与逻辑整合，确保每份公文既符合宏观政策导向，又贴近本地实际需求。通过“一个大脑统筹，九位专家协同”的智能体架构，“方寸智脑”将写作任务自动拆解为信息搜索、主题提炼、结构生成、内容润色、合规审校等子流程，由专家智能体分工作业、流水线闭环。

以《台州市推进工业母机产业带建设工作三年行动计划（2025—2027）》为例，用户仅需输入标题，“方寸智脑”即可自动调度信息搜索、资料分析、政务审校等9个专家智能体协同工作。例如，信息搜索智能体可快速扫描百万级文献库与实时政策库，提取与“工业母机”“台州产业”相关的关键信息；资料分析智能体则对数据进行清洗、归类与关联分析，生成深度分析报告。

华为云Tokens服务基于CloudMatrix 384超节点和xDeepServe架构，提供单芯片最高2400TPS吞吐量与50ms超低时延，保障智能体间高效协同，20分钟即可完成传统模式下需一周的工作量，生成“信息密度高、贴近本地化、文风像人写”的公文。

依托华为云Tokens服务的高性能推理能力，“方寸智脑”通过自研审校模型与通用大模型结合，可对公文进行多轮语义扫描与风险模式匹配，精准识别表述偏差、政策冲突、敏感信息泄露等隐患，将人工审校漏检率降低至行业新低。

华为云Tokens服务提供的弹性算力保障了系统在高并发场景下的稳定性，即使面对用户高并发数同时发起复杂写作任务，仍能保持流畅响应与一致输出质量，真正实现了智能写作的“开箱即用”。

03 不是替代，是进化 人与AI的协同共赢

“方寸智脑”并非取代笔杆子，而是将其从重复性、高负荷的劳动中解放出来，转向更具创造性与决策性的工作。从“人机协同写得像”到“自主运行写得准”再到“辅助决策写得优”的跨越，“方寸智脑”重新定义了政企办公的智能范式。

在实际应用中，用户可将更多精力集中于政策研究、战略规划、创新设计等高层次工作，而“方寸智脑”则承担起资料整合、初稿生成、格式审校、风险筛查等基础性、重复性任务，形成人机协同、各有侧重的新办公模式。这种分工不仅提升了组织整体效率，更让专业人才得以发挥更大价值，实现“以人为本、AI赋能”的智能化转型理念。

随着华为云Tokens服务的持续升级，“方寸智脑”将在法律、税务、金融等更多垂直场景中落地应用，成为政企智能化转型中不可或缺的“智慧伙伴”。方寸无忧已计划基于华为云Tokens服务，进一步开发法律文书、税务报告、金融风控等专项智能体产品，深入赋能各行业办公场景。华为云也将通过持续优化Tokens服务的性能与成本，推动AI技术在各领域的规模化、普惠化落地。

方寸无忧与华为云的这一合作，不仅是一款产品的成功，更是AI技术与行业知识深度融合的典范，真正践行了方寸无忧“产品改变世界，AI助力梦想”的愿景。

合力亿捷 X 华为云Tokens服务

打造智能客服新范式

他不是传统客服机器人，而是具备“理解力、执行力、协作力”的客服AI员工。他可以 7x24 全天候在线，自动承接重复咨询、识别复杂请求、判断执行路径、联动工单与知识库完成服务闭环。合力亿捷打造的客服AI员工，已在多个真实业务场景中正式“上岗”。

作为一家专注智能客服二十余年的技术服务商，合力亿捷通过AI与大模型能力，帮助企业构建真正可用的客



服AI员工。从语音通话到在线对话，从自动应答到任务执行，合力亿捷正在帮助客户构建一支真正“可用”的AI客服团队。

为了让AI客服真正“用得起、用得稳、跑得快”，合力亿捷选择与华为云Tokens服务深度协同，借助其高吞吐、低时延的Tokens服务能力，打造了智能客服Agent的“超级引擎”，让AI真正融入服务现场，成为企业可信赖的服务合伙人。

01 打造“能干事”的客服AI员工 客服Agent要应对哪些挑战？

以合力亿捷服务的某头部社交平台为例，该平台拥有过亿注册用户、海量服务请求和复杂多样的客服场景，这些问题使用户对AI客服的响应速度、理解能力和执行能力提出了更高要求。在电话场景中，客服AI员工主要面临以下几项关键挑战：

- 毫秒级语音响应，保障通话连续性与用户体验；
- 抗噪精准识别，适配不同语速、口音、环境的语音输入；
- 上下文理解与意图跟踪，支撑复杂问题的多轮追问与跳转式表达；
- 语义驱动任务执行，真正实现从理解到处理的业务闭环。

这些挑战对模型推理性能、流程协同效率与系统稳定性均提出了更高要求，推动AI客服从“会说”走向“能干”，并对大模型的推理效率、服务流程设计、语音底层技术与上下游系统能力都提出了极高要求。

02 华为云Tokens服务 赋能真人级AI员工

为了让客服AI员工真正可落地、可扩展，合力亿捷基于自研平台接入了华为云Tokens服务，以高吞吐、低时延的Tokens服务为引擎，打造了“跑得快、接得住、干得了”的智能客服中台。

- 高效推理，保障多轮交互自然顺畅：**基于华为云Tokens服务的大模型推理能力，客服Agent可在复杂意图识别、多轮语义理解中保持高速响应，对话时延降低30%以上，用户体验显著优化；
- 弹性扩展，应对流量波动与高峰压力：**依托CCE容器引擎与专属云部署能力，客服服务可灵活弹性扩容，支持百万级并发请求承压，保障AI员工在活动高峰期稳定运行；
- 数据驱动，沉淀服务运营资产：**基于华为云大模型AI质检与统一看板能力，企业可全程监测AI服务表现、识别风险话术、抽取关键问题，有效推动服务质量与流程持续优化。

该项目已累计带动调用大模型 Tokens 超40亿+，显著提升了服务效率与客户体验。Agent对话满意度稳步提升，转人工比例下降超50%。

华为云Tokens服务以稳定的算力调度、高效的模型调用能力和企业级服务保障，为AI应用提供了持续演进的基础设施。

合力亿捷基于这一能力底座，打造了具备业务执行力的客服AI员工，跑通了从语义识别到服务闭环的全流程。

未来，合力亿捷将继续携手华为云，共同推动AI Agent的工程化落地，让智能服务真正“用得起、用得稳、跑得快”。



今日人才 X 华为云Tokens服务

以“数智员工”破解B2B销售困局

它就是B2B企业销售团队的“超级战友”——今日人才开发的“销售大师”数智员工，自推出以来，已助力超300家中小企业打通销售全流程，成为众多企业突破销售瓶颈、实现业绩增长的核心利器。

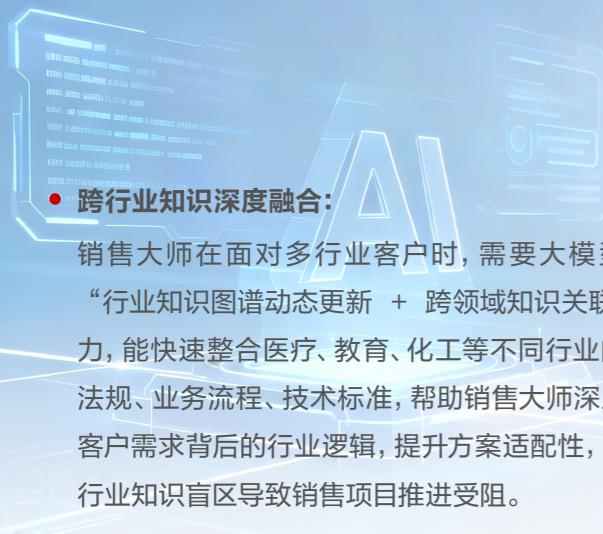
在B2B销售领域，企业早已告别“靠经验闯市场”的时代，精准获客、高效转化、风险可控成为核心诉求。作为一款为B2B企业量身打造的数智销售助手，“销售大师”数智员工依托全球LTC销售管理体系，结合227万名销售冠军的技能沉淀，融合领先AI智能体技术，从“线索获取、商机判断、需求分析、方案汇报、商务谈判、合同签订、交付回款”全流程陪跑，用“销售七阶三十六步法”为企业销售赋能，让每一步成交都有章可循、有据可依。

而要支撑这种全流程、高精准的销售赋能，背后离不开一套“专业且智能”的支撑体系。“销售大师”数智员工以海量销售实战数据为基础，结合AI实时分析与决策能力，从商机判断的精准度到合同条款的风险防控，从解决方案的个性化输出到交付回款的高效推进，全方位解决B2B企业销售难题，让销售团队告别“上坡爬滚打，无助又无力”的困境。

01 复杂销售业务场景亟需解决的技术挑战

销售大师数智员工在处理复杂的销售业务场景中，需要应对很多的技术挑战。





• 跨行业知识深度融合：

销售大师在面对多行业客户时，需要大模型具备“行业知识图谱动态更新 + 跨领域知识关联”能力，能快速整合医疗、教育、化工等不同行业的政策法规、业务流程、技术标准，帮助销售大师深入理解客户需求背后的行业逻辑，提升方案适配性，避免因行业知识盲区导致销售项目推进受阻。

• 线索深度分析与价值挖掘：

销售大师对线索的分析不能仅仅停留在表层信息的提取，需要通过大模型的“多维度数据关联分析+潜在需求预测”能力，整合企业工商信息、行业趋势、业务动态等数据，精准识别线索中的隐藏需求（如企业拓展计划、政策补贴机会），建立科学的商机价值评估模型，将高潜力商机识别率从12%提升至30%以上，提高线索资源利用率。



• 个性化商务谈判策略生成：

在复杂的商务谈判中，销售大师不仅仅提供通用的谈判策略，更需要通过大模型具备“客户画像精准刻画 + 行业谈判规则适配”能力，基于客户企业性质（国企、民企、外企）、行业竞争格局、历史合作数据，生成定制化谈判策略，包括价格机制设计、条款优先级排序、风险应对方案等，帮助销售员在谈判中占据主动，缩短合同签订周期，降低经营风险。

I 助力跨领域知识整合与需求深度拆解

在跨行业伙伴联合方案定制场景中，销售大师数智员工常常面临客户复杂且碎片化的需 求，以及跨行业知识储备不足的难题。华为云 Tokens 服务接入主流大模型及 Agent 平台，为销售大师提供了丰富的知识图谱构建与需求结构化拆解能力。

- 当软件企业与硬件厂商为制造业客户打造“智能生产管理一体化方案”时，销售大师利用 Tokens 服务，自动补全汽车制造业“精益生产”标准、电子产品行业“ESD 防护”要求等跨领域专业知识，并将客户提出的“降低设备故障率”“缩短生产周期”等模糊需求，拆解为“设备数据采集频率要求”“系统与现有 ERP 对接标准”等具体、可落地的方案要素。
- 在首次方案汇报时，使满足客户核心需求的比例从原本的约 42% (5/12) 大幅提升，有力推动合作顺利开展，提升方案的精准度与适配性，让销售大师在复杂方案定制中如虎添翼。

未来，随着华为云 Tokens 服务持续迭代升级，其与销售大师数智员工的结合将更加紧密与深入，为 B2B 企业销售赋能的效果也将愈发显著，助力更多企业在激烈的市场竞争中拔得头筹，实现业绩的飞跃式增长。

02 华为云Tokens服务“神助攻”

I 提供强大且稳定的算力支持，加速商机匹配对接

在商机判断环节，销售大师数智员工需快速处理海量潜在客户数据，从诸多线索中精准识别高价值商机。华为云 Tokens 服务基于 CloudMatrix384 超节点，通过 xDeepServe 架构创新，实现单芯片最高 2400TPS、50msTPOT 的超高吞吐与低时延性能，远超业界水平。这使得销售大师能在瞬间完成对线索中预算金额、决策人、项目周期等关键信息的提取与分析。例如，面对每月集成商推送的 200 + 条线索，销售大师借助 Tokens 服务，可在 1 个工作日内完成筛选，快速标注出如“预算 50 万 +、3 个月内启动项目、需求匹配核心产品”的高潜力商机，相较于传统人工筛选的 10 个工作日，效率大幅提升，确保销售团队能第一时间跟进最具价值的商机，避免优质线索遗漏，让销售流程的开端便高效且精准。



鲸海拾贝 X 华为云Tokens服务 给AI教育带来新范式

大模型时代，教育创新AI应用“新晋独角兽”——鲸海拾贝与华为云达成战略合作。

作为一家专注于教育领域的AI科技创新企业，鲸海拾贝的使命是构建「千人千面，AI驱动」的教育新生态。解构并催生“泛化大学”——即一种随处可以学习的“作坊式”教育形态，真正做到孔子的“有教无类，因材施教”。

从一个致力于公益分享“硬技能”的科技账号——“GenJi是真想教会你”起步到成立公司，凭借卓越的内容影响力，其泛知识内容累计触达超1.1亿人次，并获得了千万级粉丝的信任和认可。也因在教育垂直领域的深耕与创新，公司于2025年5月作为唯一的“AI+教育”企业荣登《2025福布斯中国人工智能科技企业TOP50》“新锐企业”榜单，为其推动产业智能化变革奠定了坚实基础。

鲸海拾贝依托华为云强大的技术底座，搭建击穿“四层引擎”的动态增长模型：涵盖面向个体的内容与教育（To C）、面向企业的复合智能体解决方案（To B）、面向公共标准模型的共建（To G）及技术与内容生态。双方在个性化教育与AI内容营销两大核心赛道展开深度技术共创与业务协同，旨在构建产业智能的新标杆，共同探索“AI赋能于人”的未来蓝图。

01 To C-让规模化“因材施教”解决当下“千人一面”的教育困境

长期以来，传统教育和在线学习都面临着一个核心痛点：同质化且低效。

无论是标准化的课程，还是固定的学习路径，都无法满足每个学习者独特的知识背景、认知节奏和兴趣点。“千人一面”的模式导致学习效率低下、兴趣衰减，难以实现真正的“因材施教”。



鲸海拾贝作为教育领域的前沿探索者，其自主研发的Omni-Learning平台。作为国内首个将“教—学—练—测—评—证”全链路打通的“个性化强化学习”多智能体协同架构，其目标是为每位学习者构建一个动态、自适应的“全知”学习环境，从过去的“内容为王”转向“反馈为王”，让组织和个人都进入一种“反馈驱动、终身进化”的学习模式。



*围绕“AI教育普惠”的目标，构建起完整的教育产品矩阵与知识供给模型

在这个教育体系中，AI不再只是辅助工具，而是重塑教育范式的核心动力，它重新定义了“教”与“学”的全程互动方式。

然而，实现这一愿景面临四大核心技术挑战：

- 动态个性化：**为海量用户实时生成并调整学习路径，需要澎湃的算力进行高并发的知识图谱计算和路径规划。
- 实时交互反馈：**在编程、写作等实际场景中，AI导师需毫秒级响应，并能从海量案例库中即时检索（RAG）以提供精准示范，这对模型推理速度和向量检索能力是巨大考验。
- 复杂任务协同：**模拟真实企业项目，需调度多个AI智能体协同工作，这对任务编排、状态管理和资源调度提出了极高要求。
- 深度能力评估：**自动分析学员全过程表现，并生成多维度的能力评估报告，需要强大的多模态分析与可靠的评估逻辑。

面对上述挑战，华为云Tokens服务提供的不是零散的工具，而是一个从算力、模型到开发运维的完整解决方案，精准破解了Omni-Learning落地的核心难题。

- 利用昇腾澎湃算力与高性能数据处理服务，结合鲸海拾贝的教育知识图谱进行精调，在“个性化学习路径生成”与“实时反馈”等核心场景，业务处理效率领先业界 15%-23%。
- 华为云原生服务的弹性扩容能力（如CCE云容器引擎），支撑了复杂多智能体系统的稳定运行与高效协同，从容应对超百万用户级的并发实操需求，保障核心服务稳定不中断。
- 基于华为云Tokens服务进行模型训练与推理，效率提升10%-30%，为AI导师的实时交互提供算力保障，带来拟人化的伴学体验。
- 利用华为云Tokens服务的模型微调与评估能力，训练出能够理解学生解题思路、分析知识薄弱点的深度评估智能体，提供立体化、过程性的学情诊断与反馈。

双方的合作，是AI时代“梦想”与“技术”结合的典范。它清晰地展示了，一个颠覆性的教育理念，如何通过与强大的云平台深度耦合，将复杂的AI能力转化为普惠易用的产品，真正构建起一个以学习者为中心的教育生态体系，真正推动“因材施教”走向规模化落地。

02 TOB-用AI重塑“品效合一”新范式 声量与转化双增长

在存量竞争时代，企业内容营销普遍面临“品牌价值（品）”与“商业转化（效）”难以兼顾的困境。内容生产成本高昂、创意迭代缓慢、投放效果难以归因等问题，严重制约了增长潜力。其核心问题在于，品牌知名度仅是第一步，真正的销量转化依赖于系统化的品牌策划与以优质内容为载体的用户价值传递。

为应对这一挑战，鲸海拾贝结合全网千万科技粉丝的自媒体内容创作经历，打造了内容人工智能的多智能体平台（Content+AI Multi-Agent），构建全新的人机协作内容生产模式，为品牌提供“内容交付 + 运营策略 + 产品工具”的整合方案，让企业的内容营销从“经验驱动”转变为“智能驱动”，真正做到“模型即服务，AI即增长”。

鲸海拾贝的核心创新动作主要体现在以下几个方面：

- 人机协作新模式：**基于品牌需求高效拆解，灵活配置“人+AI智能体”工作流。
- 整合化解决方案：**打通从内容策略、素材生产到效果转化的全链路。
- 规模化内容生产：**支持品牌批量化、规模化的内容素材生产需求，保证质效。
- 数据转化闭环：**通过实时监测与反馈形成持续优化的营销系统。

鲸海拾贝规模智能体集群的稳定运行与高效迭代，也离不开华为云提供的坚实技术底座。

- 华为云提供了一系列预训练的行业大模型，能够快速、低成本地打造具有深度行业知识的专属模型，大幅降低了领域迁移的初始成本。
- 企业级内容营销需要多个智能体在复杂的业务逻辑下协同作业。华为云提供的稳定基础设施和成熟的AI工程化工具，支持鲸海拾贝高效构建、部署和管理这一复杂的多智能体协作网络，保障了平台的可靠性与可扩展性。
- 华为云利用ModelArts的模型评估和持续学习能力，将人工审核的修正结果、内容的效果数据作为反馈信号，反向对生成模型进行迭代优化，确保内容风格与品牌调性始终保持一致。
- 华为云Tokens服务支持鲸海拾贝利用自有高质量数据，对盘古大模型进行精调，孵化出具备特定行业知识与品牌风格的垂直领域专用模型，确保了AI生成的内容不仅在技术上可用，更在商业上有效，符合品牌调性。

凭借Content+AI Multi-Agent，鲸海拾贝为华为等诸多头部品牌打造了规模超万、具备独特“人设”和专业能力的AI智能体集群。实践证明，这些智能体基于LLM微调、Prompt工程、知识图谱与向量数据库技术，不仅实现了跨场景的知识迁移与协同工作，还创造了多次现象级的营销事件，成功推动品牌产品DAU、MAU和CVR的指数级增长实现了“品效合一”。



华为云与鲸海拾贝的合作，不仅是技术上的强强联合，更是愿景上的高度契合。我们都相信，AI的最终价值是服务于人，提升每个人的潜能。华为云Tokens服务提供的澎湃算力、先进AI工具链和稳固的云原生底座，正是企业跨越技术鸿沟、实现教育普惠这一宏大愿景的基石。

华为云所扮演的，不仅仅是算力提供商，更是技术合伙人与赋能者，为鲸海拾贝这样的创新先锋铺平了通往未来的道路，共创产业智能化的未来，推动“人工智能+”战略在各领域的深度落地与长远发展。

