

亚马逊云科技 FOR SOFTWARE  
AND TECHNOLOGY

# Agentic AI ISV & Startup 客户成功实践

亚马逊云科技是构建 Agentic AI 的不二选择

# PREFACE

## 序言

Agentic AI 技术在软件行业迎来爆发式增长，软件行业正处于从被动的“数字化工具”向拥有自主认知与行动能力的“智能化伙伴”跨越的关键节点。据 Gartner 预测，到 2028 年，企业软件应用中将有 33% 集成 Agentic AI 功能，且有 15% 的日常工作决策将由 Agentic AI 系统自主完成。

面对 Agentic AI 如火如荼的发展态势，一些新兴的 Agentic AI 原生企业应运而生，他们依托 Agentic AI 技术，全面革新了传统软件的工作流以及客户交互模式，不仅让软件实现从“被动响应指令”到“主动提议，对话合作”的根本性跃迁，还让软件从“工具”升级为了“工作伙伴”。

在竞争全球化、体验个性化、业务复杂化的今天，传统软件企业单凭传统的功能迭代已难以保持强有力的市场竞争优势，Agentic AI 带来的技术革新，正驱动着软件行业的颠覆性变革，因此越来越多传统软件企业选择在已有的软件应用中增加 Agentic AI 模块，接入 Agentic AI 服务，以推动传统软件应用的 Agentic AI 转型升级。然而，将 Agentic AI 技术融入软件应用的过程并非一路坦途，传统软件企业需要在技术选型、产品设计、软件架构、基础设施等方面完成系统性升级。

为了让不同类型的企业全面掌握 Agentic AI 升级软件应用的清晰路线图，助力传统软件企业利用 Agentic AI 转型，并帮助 Agentic AI 原生企业在 Agentic AI 的新纪元中精准定位发展方向，本白皮书聚焦近 20 家不同类型软件企业的创新过程，深入研究了他们如何依托亚马逊科技提供的 Amazon Bedrock、Amazon AgentCore、Kiro 等丰富、成熟的技术服务，将 Agentic AI 深度融入软件应用内核，实现软件产品的创新与升级。同时利用亚马逊科技 APN，Marketplace 等成熟的网络、渠道和资源帮助这些企业进行全球化业务拓展，将先进的 Agentic AI 软件产品带给全球用户。

# CONTENTS

## 目录

<b>01</b>	<b>初创企业</b>	<b>01</b>
	<b>借力 Agentic AI，打造差异化创新产品</b>	
	知微行易	02
	MaxQuant	05
	Bika.ai	10
	Sandwich lab	14
	AIGCode	18
	Notta	22
	包小盒	27
<b>02</b>	<b>传统软件企业</b>	<b>31</b>
	<b>融入 Agentic AI，开启智能化发展新征程</b>	
	Intsig 合合信息	32
	智齿科技	35
	销售易	38
	特赞科技	41
	WallTech	44
	金蝶	48
	乐鑫	52
	飞轮科技	55

# 01 初创企业

借力 Agentic AI  
打造差异化创新产品



## ERP 不再手工操作， 知微行易借 Agentic AI 驱动企业智能运营

知微行易（上海）智能科技有限公司（以下简称“知微行易”）作为专注于服务中国 500 强企业的智能企业运营专家，依托自研的 AI 原生系统“笃行（Duxing）”，为大型企业提供包括 DxERP（经营管理）和 DxMES（生产管理）在内的智能企业管理软件及交付服务。

知微行易一直以将 AI 原生能力深度植入 ERP 等企业核心系统，使其具备自学习、自优化的能力，并成为企业运营的“AI 引擎”为核心使命，旨在为客户提供“从指令到执行”的全周期智能运营体验。



交付效率提升  
50%



运维成本降低  
70%



项目周期从数月  
缩短至数天

目前，知微行易采用的亚马逊科技主要服务包括：**Amazon Bedrock（仅海外业务使用）、Amazon EKS、Amazon RDS、亚马逊科技云安全与合规等服务。**作为亚马逊科技技术合作伙伴，知微行易希望借助亚马逊科技在全球的影响力，将其在制造业智能化转型领域的成功经验和解决方案输出到更广阔的领域。同时，知微行易也期望能够与亚马逊科技在智能制造领域展开深度合作，共同开发面向制造业的 AI 解决方案，为更多企业客户创造价值。



\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 突破外挂式 AI 功能困境让 Agentic AI 深植于 ERP 系统

知微行易主要服务于新能源、建材、能源及高端制造领域的中国 500 强的大型企业客户。当前很多企业使用的是传统的 ERP 系统，系统主要依靠人工操作，无法主动思考和优化，即使部分系统增加了 AI 功能，也多是外挂式的局部优化，无法从根本上改变业务运营模式。因此带来了诸多技术问题：

- ◆ **响应迟缓**  
面对业务流程变更，传统 ERP 定制开发周期长、成本高，难以敏捷响应业务变化。
- ◆ **决策滞后**  
业务决策严重依赖人工经验，数据无法被实时、智能地分析并转化为行动。

- ◆ **效率瓶颈**  
AI 与业务流程脱节，例如 AI 聊天机器人生成的工单仍需大量人工介入，优化效果有限。

为了突破传统 ERP 系统正在面临的发展“瓶颈”，构建一个“端到端智能辅助驾驶”模式的 AI 原生企业智能运营系统，并确保系统能应对百亿级数据，实现毫秒响应和高可用性，知微行易选择与亚马逊云科技合作，依托亚马逊云科技在云原生技术、高性能计算及全面的人工智能和机器学习服务（如 Amazon SageMaker、Amazon Bedrock）等领域的强大技术实力，为大规模的 AI 工作负载提供坚实的技术底座，将 Agentic AI 服务能力深入植入 ERP 系统。

## 解决方案 亚马逊云科技为 Agentic AI 服务构建提供全链路的服务

知微行易基于亚马逊云科技提供的云框架再到 AI 模型工具链的完整服务，构建了“笃行（Duxing）”。该系统采用云原生的组装式架构，业务模块通过 Amazon EKS 实现了容器化部署和弹性伸缩。为满足海外业务访问和集成基础大模型的需求，知微行易使用 Amazon Bedrock，结合自身的业务经验，构建了面向 ERP、MES 场景的领域 Agent，同时依托亚马逊云科技的安全的服务与全球合规体系，确保了国际化服务的顺利落地：

- ◆ **加速 Agent 开发**  
面对海外业务，知微行易基于 Amazon Bedrock，让系统具备了 Agentic AI 能力的核心服务，通过调用和整合强大的基础模型，加速了“排产助手”等领域 Agent 的开发。

- ◆ **确保运行可靠性**  
“笃行”以云原生架构为基石，借助 Amazon EKS 提供的服务，以容器化方式保障了客户业务 7x24 小时不间断运行所需的弹性、韧性和可观测性。
- ◆ **提高数据稳定性**  
“笃行”采用了 Amazon RDS 提供的高可用、高扩展性的企业级关系型数据库支撑，有效确保了数据的稳定性与处理效率。

亚马逊云科技为我们提供了将颠覆性 AI 原生理念转化为服务世界 500 强客户的企业级产品的强大引擎，帮助我们提升了 50% 以上的交付效率，并赢得了与行业巨头竞争的信心与实力。

——谢少毅  
知微行易（上海）智能科技有限公司 创始人

在亚马逊云科技的帮助下，知微行易构建的“笃行（Duxing）”不仅能支撑数十个领域 Agent 的实时在线推理，还能为头部动力电池制造商等客户稳定处理海量的生产、库存、财务等数据事务。该系统不同业务模块内嵌的领域 Agent，可通过理解业务上下文，主动为用户推荐任务，生成流程。在 AI 在排产、补货等不同业务场景中，AI 既能提供决策建议，还能自动执行并根据反馈结果进行自学习和优化，真正形成 AI 原生执行的闭环。

## 业务成果 世界级技术底座支撑让 Agentic AI 服务构建快人一步

当前，系统在亚马逊云科技宁夏和日本区域上线后持续平稳运行，实现了 7×24 小时不间断作业。依托亚马逊云科技提供的世界级“技术底座”，知微行易不仅能将宝贵的专家资源聚焦于 AI 与业务场景的深度融合这一核心服务构建上，还能确保运行的稳定性、安全性，从而有效提升品牌的全球形象，让产品在交付效率、运维成本、稳定性等方面获得了更强的竞争优势：

- ◆ **高效交付**  
结合亚马逊云科技的敏捷基础设施，知微行易采用创新交付模式，可实现极速部署，交付效率提升 50% 以上，项目周期从数月缩短至数天，远超传统软件厂商的交付速度。
- ◆ **降低成本**  
借助亚马逊云科技提供的弹性、按需付费的云服务级强大的 AI 工具链，从根本上改变了知微行易的成本结构、研发模式和业务策略，既降低了 3 倍的直接投资，还减少了 70% 的运维成本。

- ◆ **高稳定性**  
面对海量数据和高并发请求，系统持续保持高性能和稳定性；此外还能快速调整和发布新功能，让客户体验到系统的“持续进化”。

基于此次合作，知微行易认为应大胆拥抱云原生和 AI 原生，将自身核心优势与云的强大能力做乘法，充分利用云服务，而不仅是把云当成虚拟机，这可以让业务创新事半功倍。同时将 AI 作为架构的核心，并在创业初期就基于云上 AI 服务来设计核心业务流程，可构建竞争壁垒，形成更强的竞争优势。

未来，知微行易会在海外业务中继续深化对 Amazon Bedrock 的应用，探索更前沿的 Agentic AI 技术，让系统中的“领域 Agent”更智能。而且，知微行易今年 8 月正式使用 Amazon Q 等服务，进一步提升系统的自然语言交互能力。此外，我们也将利用亚马逊云科技的全球节点和亚马逊云科技 Marketplace，支持全球化业务拓展。



### 关于知微行易

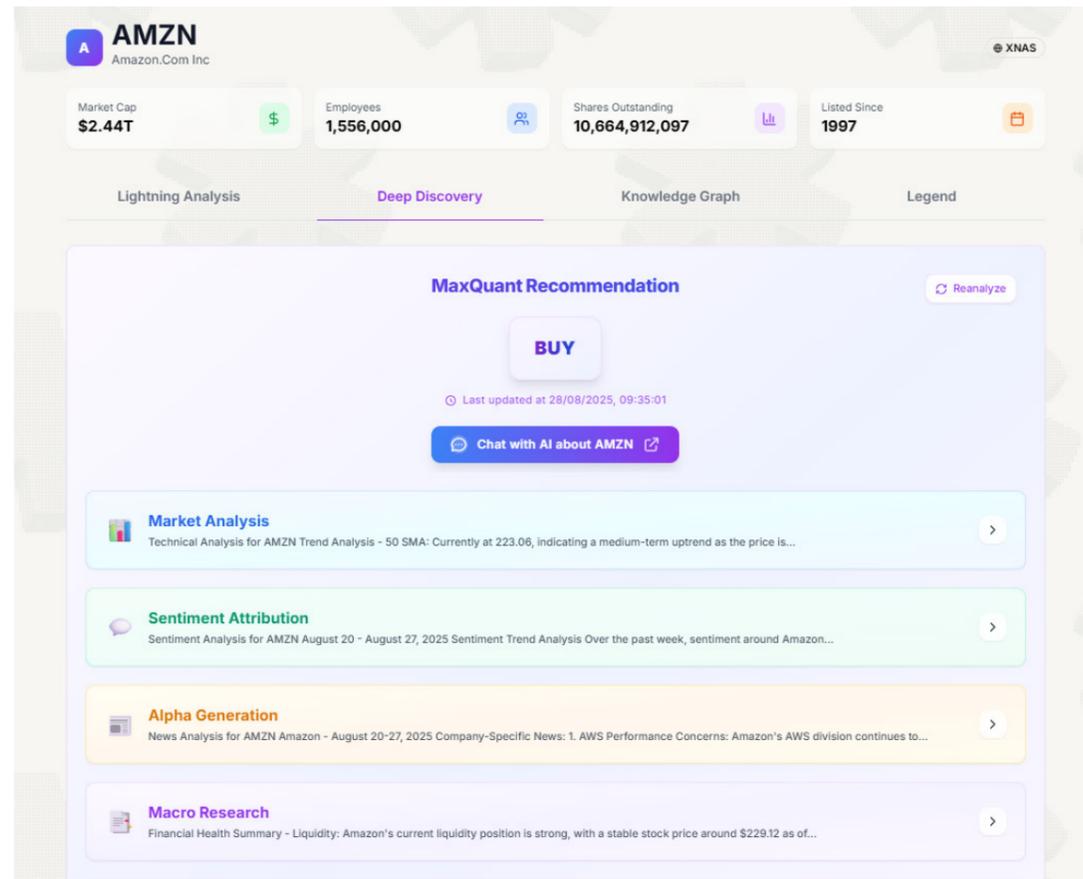
知微行易是专注于服务中国 500 强企业的智能企业运营专家。知微行易依托自研的 AI 原生系统“笃行（Duxing）”，为大型企业提供包括 DxERP（经营管理）和 DxMES（生产管理）在内的智能企业管理软件及交付服务，引领企业迈向由 AI 驱动的全新数字化形态。

# AgentCore 上阵 MaxQuant 让交易员告别 996

MaxQuant 是一个面向金融领域的 AI 基础设施系统，旨在构建新一代智能投研与交易决策系统，重新定义投资的未来，为每一次投资决策注入非同寻常的智能。

借助 Amazon Bedrock AgentCore，MaxQuant 高效实现了 Multi-Agent 编排与应用，将核心业务模块“投资建议助手”的部署周期由 4 周缩短至 1 周，加速企业级和消费级客户的量化策略创新与用户体验优化。

目前，MaxQuant 使用的亚马逊云科技产品与解决方案包括：Amazon Bedrock AgentCore、Amazon Lambda、Amazon EventBridge、Amazon DynamoDB、Amazon API Gateway 等。



\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 引入 Agentic AI 让金融交易更透明

传统金融交易往往存在门槛高、透明度低、资产跨度窄等痛点，散户与初学者往往需要掌握一定的代码和数据工程基础，才能完成策略的设计与执行，而在面对复杂的新闻与行情解读时，即便是有经验的投资者，也常因信息差而偏离最佳路径。MaxQuant 团队认为，这些结构性问题压缩了交易的空间，在限制创造力的同时，也影响着盈利的上限。

为此，MaxQuant 决定借助 Agentic AI 技术弥合交易的不对称性，探索更加开放及高效的量化交易模式，实现更广阔范围内的金融普惠，打造普通人也能轻松上手的量化交易服务。然而，Agentic AI 有着环境碎片化的特点，不同 Agent 的运行环境、上下文、依赖与调用协议各异，如果自行开发，可能会消耗大量时间用于编排、容错与监控。另一方面，金融交易通常波动剧烈，系统必须在具备高峰弹性的同时，符合金融场景的合规与可追溯要求，这对系统架构提出了严峻挑战。

最重要的是，市面上缺乏适合金融交易场景的 Agentic AI 基础设施：很多方案偏重模型本身却忽视工程层面的可扩展性，或是提供算力资源却难以支撑

Multi-Agent 的编排与运行。同时，Agentic AI 基础设施演进节奏相当快，对 MaxQuant 这样的初创团队来说，如果把大量精力投入到底层环境搭建以及后续的跟踪升级与维护，就需要专业团队与大量人力资源投入，势必会延误产品迭代节奏，无法在“快鱼吃慢鱼”的竞争环境中崭露头角。因此，团队迫切需要能够高效支撑 Multi-Agent 的弹性、合规的 Agentic AI 基础架构。

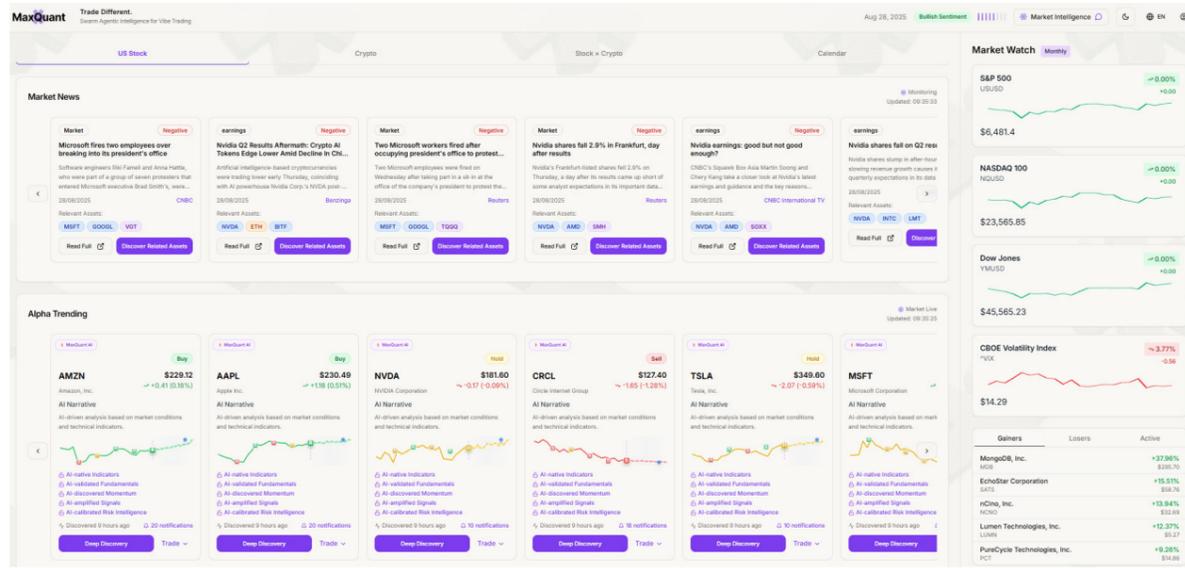
自 Amazon Bedrock 推出以来，MaxQuant 就持续关注它的发展与更新。2025 年亚马逊云科技纽约峰会上，Amazon Bedrock AgentCore 正式发布，MaxQuant 立刻意识到它的运行时 (Runtime) 能力能够与现有产品及业务实现有机融合，解决 Multi-Agent 编排与上下文统一等工程难题。

同时，亚马逊云科技也能满足 MaxQuant 对于金融场景中高弹性、高合规等方面的需求。于是，MaxQuant 积极与亚马逊云科技的架构师、工程师建立起紧密沟通，并决定将全部业务迁移至亚马逊云科技上运行，打造 Agentic AI 驱动的新一代智能投研与交易决策系统。

不同 Agent 运行环境、配置、上下文存储方式及依赖都各不相同，后期集成成本、团队协作成本及部署成本非常大。同时作为金融场景产品，我们对于数据吞吐，以及权限控制、运维及审计的要求也都非常高。亚马逊云科技不仅提供 Amazon Bedrock AgentCore，同时还能满足我们在系统稳定性、安全合规、可扩展性等方面的需求，是我们业务的优秀选择。

—— 曾境烽  
MaxQuant CEO

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。



## 解决方案 从技术底座到协作拓展的全面支撑加速初创企业成长

### 01. Amazon Bedrock AgentCore, 让研发回归业务创新本身

Amazon Bedrock AgentCore 提供一整套部署并高效运行 AI Agents 的能力，帮助开发者打通 AI Agents 从概念验证到生产部署之间的关键环节，快速安全地大规模部署和运行 AI Agents。Amazon Bedrock AgentCore 现已支持提供覆盖运行时、记忆、身份验证、网关、代码解释器、浏览器工具、可视仪表盘七大核心 AI Agents 服务。

MaxQuant 最初在量化交易业务中尝试使用 Amazon Bedrock AgentCore 运行时能力，将其应用于基础

的数据解析与策略回测等环节。相比传统架构，Amazon Bedrock AgentCore 运行时无需长期维护底层环境，能够更灵活地支持高频场景下的计算需求。随着合作的不断深入，MaxQuant 逐步将更多复杂逻辑，包括在交易前生成更细粒度的客户画像，或是在交易中对多策略组合进行自动优化等，迁移到 Amazon Bedrock AgentCore 的其他子模块上，通过高效完成数据读取、特征提取、模型调用、策略评估等环节的协作执行，实现了从分析到执行的自动化闭环，研发主线得以回归到策略与体验本身。同时，团队也无需为每次 Agent 扩展承担额外的编排成本，系统的可维护性与可扩展性显著提升。

使用 Amazon Bedrock AgentCore 之后我们无需再重写任务触发逻辑，只需声明关系、上下文传递策略即可实现 Agent 应用，避免了很多‘重复造轮子’的工作，可以把更多的精力放在业务场景和功能性拓展。

——张维桐  
MaxQuant 首席科学家

### 02. 轻重任务解耦处理，负担更低、系统更敏捷

在将复杂的多步推理与策略优化交由 Amazon Bedrock AgentCore 执行的同时，MaxQuant 也将缓存刷新、批量数据清理、日志处理、状态同步等轻量化逻辑封装为 Amazon Lambda 函数，并通过 Amazon EventBridge 进行触发。这种将任务逻辑按轻重解耦的方式，使 Amazon Bedrock AgentCore 能够专注于高价值业务处理，更好地释放 Agentic AI 体系价值。

同时，借助 Amazon EventBridge，MaxQuant 能够灵活捕获来自不同系统的事件并触发相应的 Amazon Lambda 函数，实现跨模块的自动化协作。基于事件驱动型架构，Amazon Lambda 具备天然的弹性与隔离性，在高并发下场景可以按需扩缩轻量任务，从而避免主流流程阻塞。拓展性方面，研发团队在新增功能或需求时，只需新增一个函数即可无缝接入既有流程，实现“改一处、全局可用”，显著降低了运维负担、提升了系统的敏捷性。

### 03. 技术、资源、拓展齐助力，铺就成长之路

MaxQuant 是一家年轻的创业公司，亚马逊云科技不仅为 MaxQuant 提供了“Amazon Bedrock AgentCore × Serverless × 事件驱动架构”的技术底座，还带来了成长路径上的多方助力：

#### ◆ 在技术层面

亚马逊云科技解决方案架构师团队为 MaxQuant 提供了持续的指导，帮助其在 Amazon Bedrock AgentCore 与 Serverless× 事件驱动架构结合的探索中避免早期风险，使研发团队能够更专注于量化交易的业务逻辑与产品体验。

#### ◆ 在资源层面

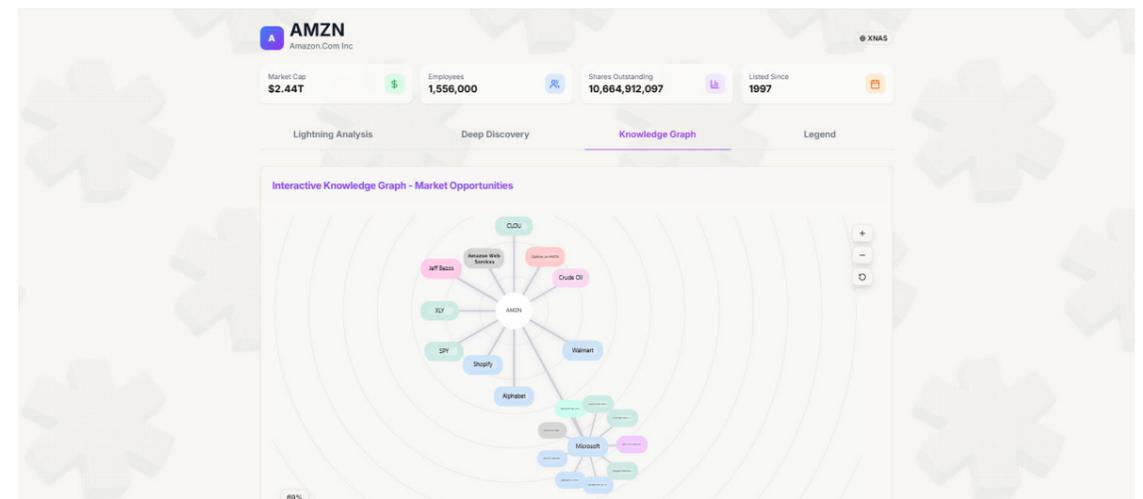
亚马逊云科技为初创企业提供的创业加速器支持计划为 MaxQuant 提供了宝贵的云资源额度，提升了快速试错与功能迭代的节奏。

#### ◆ 在拓展层面

亚马逊云科技为 MaxQuant 连接了更多合作伙伴提供潜在的合作机会，并积极提供进入亚马逊云科技 Marketplace 的路径指导，帮助 MaxQuant 获取更大范围内接触潜在客户与投资人的机会，提升触达效率。

同时，凭借亚马逊云科技全球合规与基础设施布局，MaxQuant 在国际化拓展过程中能够以更低的技术门槛、更高的专业度赢得客户信任。这些支持不仅让 MaxQuant 在早期就具备了稳定的技术与背书，也帮助其在研发与业务之间形成良性循环，加速了整体成长。

曾境烽提到：“很荣幸我们入围了今年的亚马逊云科技创业加速器，为我们拓展了更多潜在融资机会，让更多人知道了 MaxQuant 这款产品。”



\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

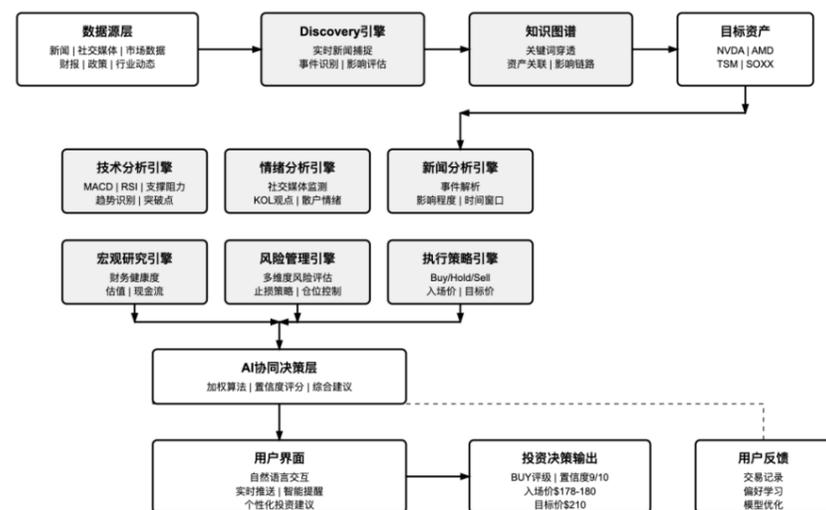
## 业务成果 Amazon Bedrock AgentCore 加速 MaxQuant 量化产品落地

在亚马逊云科技提供的 Amazon Bedrock AgentCore 等产品及服务帮助下，MaxQuant 成功将业务核心模块的开发周期从 4 周缩短至 1 周，实现了量化交易中数据解析、策略回测、客户画像、策略组合优化等服务的高效构建与应用。结合亚马逊云科技引领业界的安全合规实践，MaxQuant 在提升研发与交付效率的同时，也为金融行业客户带来了更高的安全性与信任感。在近期的 MaxQuant 伦敦线下集会中，多位量化交易员、策略师均对这套系统表现出浓厚兴趣，侧面反映出该方案的技术突破与巨大价值。

“目前在金融场景中，对 Multi-Agent 系统中单个 Agent 的运行支持还相对空白。基于亚马逊云科技所提供的 Amazon Bedrock AgentCore 运行时支撑，我们在这方面迈出了第一步，能够在既有业务框架上快速集成 AI Agents，为我们的客户构建下一代差异化 Agent 服务。”张维桐表示。

当前获得认可的同时，面对 Multi-Agent 体系在金融场景中的应用前景，MaxQuant 也给出了自己的思考。张维桐表示：“Agentic Trader 将成为首批实现完全自治运行的 Agent，具备超越人类的执行效率、覆盖广度与处理速度，能够在跨境、跨机构、跨资产的体系中无缝协作。人类将从繁琐的操作以及注意力中解放，转而承担价值观设定与治理监督的角色。MaxQuant 验证的不仅是 Agentic 策略收益率，更是 Agent 在复杂社会制度与实时合规环境下的稳健性与自主性。我们期待与更多像亚马逊云科技这样的同道者携手，共同迎接崭新的金融智能未来。”

未来，MaxQuant 希望继续携手亚马逊云科技，进一步深化 Amazon Bedrock AgentCore 服务应用，持续在金融场景下探索多 Agent 系统的差异化应用路径，在单标分析基础上推出面向股票、期权及 RWA 的真实世界资产等新兴数字资产的跨资产组合与对冲管理，继续朝着“打造基于 AI Agents 框架的一站式量化交易服务”愿景迈进。



MaxQuant

### 关于 MaxQuant

MaxQuant 是全球首个面向金融领域的 AI 基础设施系统，旨在构建新一代智能投研与交易决策系统，重新定义投资的未来，为每一次投资决策注入非同寻常的智能。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 依托 Agentic AI，Bika.ai 为“一人公司”提供完整的 AI 组织管理平台

Bika.ai 针对一人公司 (One-Person Company) 提供完整 AI 组织管理平台，从企业管理视角出发，提出一个新的角色概念“AI Organizer”

通过 AI Organizer 来组织管理各个 AI Agent，实现协同管理，为全球中小企业和个人创业者提供团队构建、自动化流程编排、数据库驱动应用搭建、营销与销售管理等服务。

目前，Bika.ai 采用的亚马逊云科技主要服务包括：Amazon Bedrock（仅海外业务使用）、Amazon Bedrock AgentCore Browser Tool、Amazon SageMaker、Amazon EC2、Amazon S3、Amazon Lambda、Amazon DynamoDB、Amazon Relational Database Service 等服务。



部署与扩展成本降低 30% 以上



构建与部署周期缩短约 40%



Agentcore Browser Tool 提升节省 90% 关键工具接入时间



\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 借助 Agentic AI 能力，打破传统 AI 工具使用瓶颈

当前“AI+ 自动化”正在成为一种标配，但对于创业者来说，应用琳琅满目的 AI 工具，只是从独立工作者变成了“流水工头”，任务分散、工具割裂等问题给创业者带来多重应用难题：

### ◆ 工具分散

写作助手、绘图工具、自动化平台各管一摊，文档、数据库、沟通系统互不兼容，信息孤岛难以打。

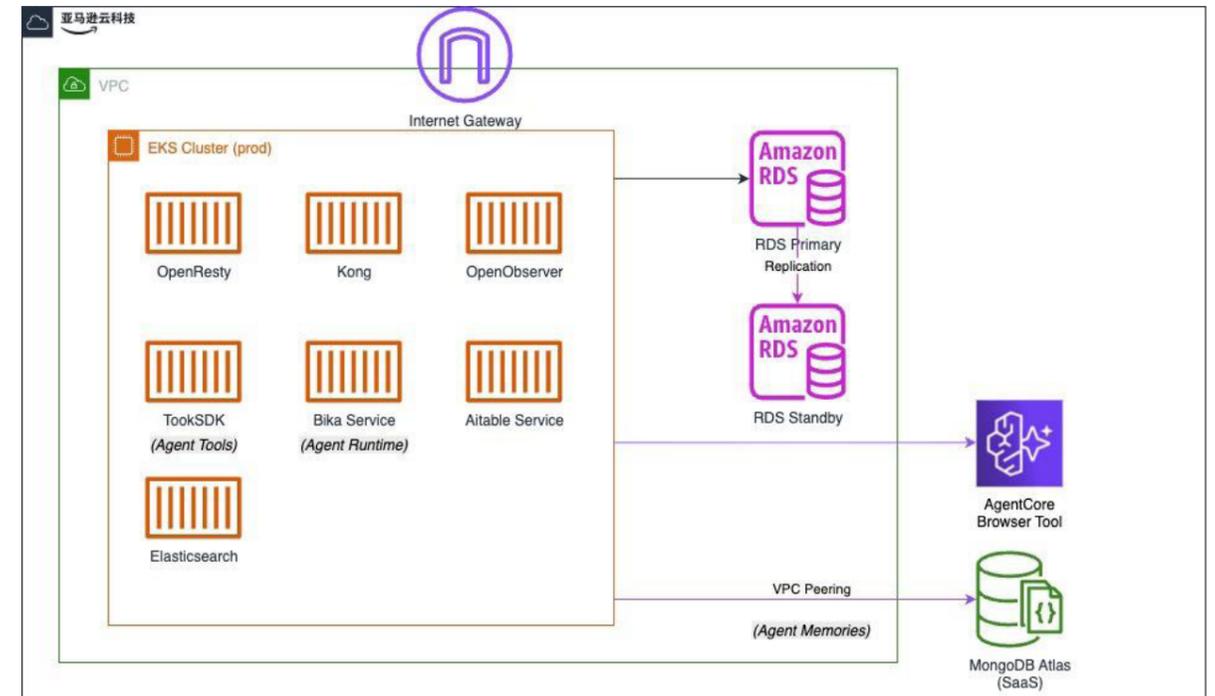
### ◆ 学习和维护成本高

自由职业者或小团队需要同时订阅多个服务，不同平台逻辑差别很大，连接一断就要重新排查。

### ◆ 效率较低

AI 虽然能干活，但大多数时候还是要手动操作、重新触发。

为了让创业者能像使用即时通讯软件一样管理 AI 团队，帮助创业者解决了任务分散、工具割裂等应用难题，Bika.ai 应用而生。Bika.ai 集成了 10000+ 应用，在部署之初，为确保 Bika.ai 运行的稳定性，可扩展性，并能满足未来 2 年公司全球扩展等需求，选择与亚马逊云科技合作，主要因为亚马逊云科技的云服务覆盖全球多个区域，且具有完善人工智能和机器学习的服务生态，并能给企业提供强大的技术支持，不仅能满足 Bika.ai 发展所需的大模型计算能力、存储与数据库等需求，还能为其 AI 模型运行提供高速的网络和**打造高效 Agent 所需的核心工具**。



## 解决方案 亚马逊云科技助力 Bika.ai 实现“AI 工具”到“AI 组织平台”的跨越

Bika.ai 团队将核心服务部署至亚马逊云科技的云上后，可以稳定地为全球用户提供 Agentic AI 协作、数据库驱动应用和营销自动化等服务。而且依托亚马逊云科技提供的 Agent 应用基础设施、无服务器架构、数据库与存储服务，可轻松承载应用运行所需的各类负载：

### ◆ 支撑数千客户并发

Bika.ai 在亚马逊云科技的云上部署了多 Agent 框架后，可满足数千用户并发的需求；Bika.ai 的 Agent 应用集成了基于无服务器的浏览器工具 AgentCore Browser Tool，支持 AI Agents 在完全托管的环境中进行大规模、低延迟的网页交互。

### ◆ 承载百万级数据量的应用

基于 Amazon Relational Database Service、Amazon DynamoDB、Amazon S3 等数据库与存储服务，可承载 CRM 和 ERP 应用的百万级数据量。

### ◆ 运行自动化工作流

依托亚马逊云科技提供的 Serverless 架构（Amazon Lambda），Bika.ai 能够在无需预置或长期维护服务器的情况下，快速运行和调度各类自动化工作流。通过事件驱动的函数执行模式，工作流能够根据用户行为、数据库变化或外部触发器自动启动，并在毫秒级完成资源调度与任务处理。这种架构不仅保证了系统的高弹性与高可用性，还能实现“按需计费”，帮助团队显著降低基础设施运维成本。同时，结合 Amazon S3 与 Amazon DynamoDB 的存储与状态管理能力，Bika.ai 的多 Agent 工作流可实现跨服务协同与复杂任务编排，灵活支撑从营销自动化到客户关系管理的多样化业务场景。

亚马逊云科技提供的丰富基础设施服务与开发部署技术支持，能让我们团队以更精简的规模支持全球服务，并帮助我们打造了真正“一人公司”的 Bika.ai 平台。

— Kelly

AI Organizer Company CEO & Co-Founder

## 业务成果 Agentic AI 技术加持，彻底颠覆创业者的工作方式

Bika.ai 采用了亚马逊科技 Serverless 架构和人工智能服务后，既降低了团队的技术负担，可以让团队无需关注底层运维，更专注于产品创新；还优化了在部署与扩展方面的成本，能支持百万级并发扩展：

### ◆ 支持智能体网页查询和信息检索

Amazon Bedrock AgentCore 的浏览器工具提供了企业级安全的浏览器运行时支持 Agents 自助处理各类网页任务。其截图功能使 Agents 能够解析动态内容和导航网页，高效地完成网页查询和信息检索等复杂的工作流程。

### ◆ 降低部署成本，提高并发扩展能力

依托亚马逊科技提供的基础设施及存储等服务，让部署与扩展成本降低了 30% 以上的同时，还能支持百万级的并发扩展。

### ◆ 加速团队创新，缩短部署周期

基于亚马逊科技提供的 Serverless 架构以及技术支持与培训，加速了 Bika.ai 创新的步伐，缩短了 40% 的构建与部署周期。浏览器工具的实时可视化监控和会话回放功能便利了 Agent Web 交互功能开发过程中的调试和审计。

未来，Bika.ai 将在海外业务中，采用 Amazon Bedrock、Amazon SageMaker 等更多 Agentic AI 相关的服务，以深入探索多模型协作；并依托亚马逊科技覆盖全球的云服务，将 Bika.ai 扩展至全球更多可用区，从而降低跨区域延迟。此外，Bika.ai 团队还将建立 Agentic AI 模板商店，为更多开发者和企业提供云端共享服务。一步降低当地客户的访问延迟。



### 关于 Bika.ai

Bika.ai 旨在帮助个人创业者和小型企业通过 AI 组建高效的“虚拟公司”，让 AI 不仅能执行任务，还能调度，组织和写作，为全球中小企业以及个人创业者提供 **Agentic AI 团队构建、自动化流程编排、数据库驱动应用、营销与销售管理** 等服务。目前，Bika.ai 已经在欧美积累了数万用户，其愿景是打造 Level-5 AI 组织者 --- 超越助手，成为统筹人类与 AI 协作的智能操作系统。

## 从“人力依赖”到“AI 自主” Sandwich Lab 实现全链路广告智能化

Sandwich Lab 定位为“商业智能 Agent”实验室，致力于构建创新型的数字化增长系统。Lexi 是该团队打造的人工智能 Meta 广告营销 Agent。

此前的行业解决方案缺乏高效自动化，依赖投手的专家经验，导致准入成本较高。与亚马逊科技合作后，Sandwich Lab 旗下产品 Lexi 通过 Agentic AI，构建了对用户投放网址的多模态分析处理能力，以及智能投放的能力，满足了投放前广告内容的自动化生产、投后效果追踪，实现用户投放广告效率提升。



海外业务通过 Amazon Bedrock 接入先进的系列模型实现广告生成采纳率相对提升 6%



托管计算服务实现弹性自动扩缩容计算成本下降 25%



构建 CI/CD 流水线研发交付效率提升 20%

目前 Sandwich Lab 采用的亚马逊科技主要服务包括：Amazon Bedrock（Sandwich Lab 仅海外业务使用 Amazon Bedrock）、Amazon Q、Amazon Aurora、Amazon RDS、Amazon DynamoDB、Amazon ECS、Amazon Glue、Amazon Redshift、Amazon DMS、Amazon Athena、Amazon QuickSight、Amazon S3、Amazon ECR、Amazon CDK 等。



\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 通过 Agentic AI 技术加速提升广告成效

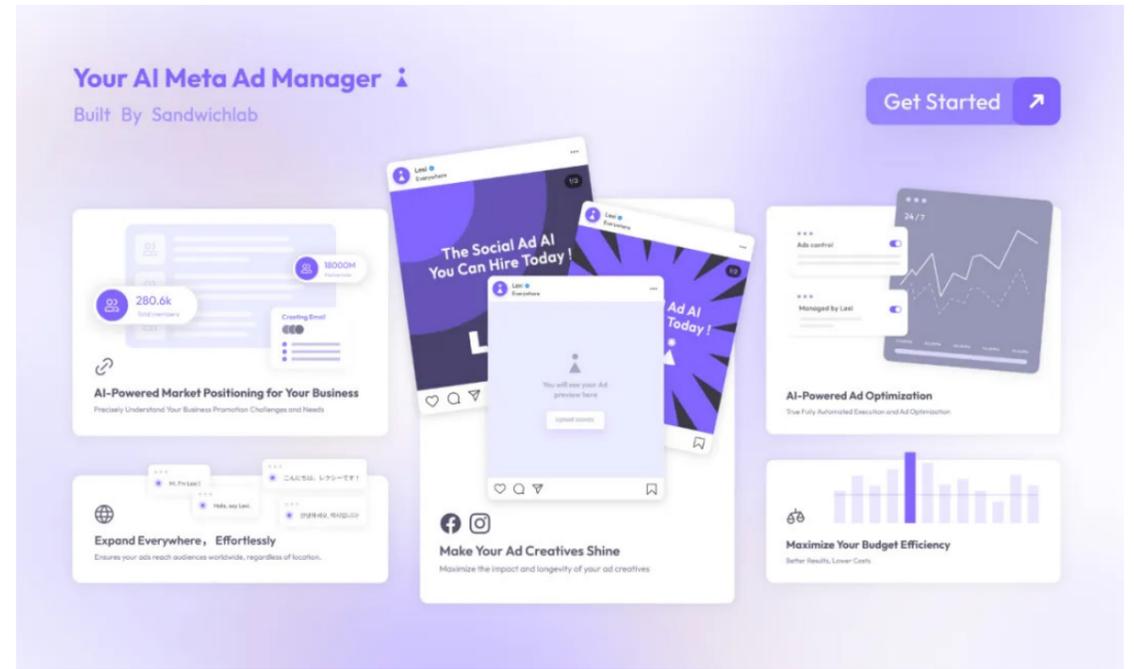
作为“商业智能 Agent”实验室，Sandwich Lab 致力于构建创新性的数字化增长系统。通过融合生成式 AI、行为数据分析、内容策略算法及跨平台广告优化技术，打造出覆盖品牌定位 - 营销 - 转化全链路的智能解决方案，将 Agentic AI 从“辅助工具”升级为真正的“营收增长引擎”。目前业务已覆盖全球 90 多个国家，聚焦北美、东南亚、非洲等新兴区域，展现出强大的跨境服务能力。

Sandwich Lab 的首发产品 Lexi，其颠覆性在于首创“AI Agent Shadowing Mode”（数字影子模式）。企业主仅需设定增长目标，系统即可自动完成市场调研、人群定位、文案生成、视觉设计、广告投放及效果优化的全闭环执行，构建“意图→结果”的无人化链路。

产品的设计初衷是打造由 Agentic AI 驱动的 Lexi，提供 AI 营销专家来解决海外用户获客的需求。但此前行业的解决方案缺乏高效自动化，只能依赖投放人员的经验，进而导致：

- ◆ 用户的准入门槛高，成本高，无法在全行业快速普及 Lexi 产品；
- ◆ 高度依赖人的决策，需要全天 24 小时支持，人员成本高；
- ◆ 专业投放人员的经验无法系统化传承，得不到延续。

经过细致的评估后，Sandwich Lab 选择与亚马逊云科技合作，依托亚马逊云科技覆盖广泛的基础设施、先进的 Agentic AI 能力，将 Agentic AI 应用到 Lexi 产品中，进而提升海外用户投放广告的效率。



## 解决方案 基于 Amazon Bedrock 等服务构建由 Agentic AI 驱动的 Lexi 加速海外业务发展

此次合作，Sandwich Lab 更看重亚马逊云科技庞大的业务体量，稳定性有保障以及不断迭代、易用的 Agentic AI 明星产品。同时，亚马逊云科技能为初创企业提供全生命周期的资源支持。

- ◆ **Amazon Bedrock 是亚马逊云科技一项完全托管的服务。**通过统一的 API 接口提供来自 AI21 Labs、Anthropic、Stability AI、DeepSeek-R1、OpenAI gpt-oss-120b 和 gpt-oss-20b、Anthropic 最新一代模型以及亚马逊等人工智能公司的超过 100 种自研及第三方模型，用于构建具有安全性、隐私性和负责任的人工智能应用程序。

使用 Amazon Bedrock，Sandwich Lab 可同时调用 DeepSeek-R1、SD XL 等多种模型，通过多模型战略规避技术风险，实现在不同应用场景下使用最合适的模型。基于 Amazon Bedrock 能力，Lexi 广

泛使用了 Amazon Bedrock 上的先进模型来打造 Agentic AI 产品，更好地服务海外业务。

- ◆ **Sandwich Lab 广泛使用了亚马逊云科技的基础设施服务。**“通过 Amazon CDK 管理基础资源，基于 Amazon Aurora、Amazon RDS 和 Amazon DynamoDB 解决存储需求，同时大量使用 Amazon S3 存储，并借助 CDN 加速资源访问。
- ◆ Sandwich Lab 采用多账户的模式管理研发环境，通过 Amazon CodePipeline 和 Amazon ECR 构建 CI/CD 研发流程；基于 Amazon ECS 构建在线计算服务，运用 Amazon Glue、Amazon Redshift 等工具处理数据分析和加工。其中，Amazon ECS 的自动伸缩模式，能帮助 Sandwich Lab 摆脱繁重的运维工作，让研发更加敏捷。

## 业务成果 Sandwich Lab 加速 Agentic AI 产品敏捷开发更好地服务海外业务

亚马逊云科技是 Sandwich Lab 产品研发和服务中不可或缺的部分。Sandwich Lab 在产品迭代之初就选择了亚马逊云科技，通过亚马逊云科技的云产品加速产品研发建设，打造 Agentic AI 产品功能更好地服务国际化业务。

- ◆ **引入 Agentic AI 能力，提升广告投放效率**  
通过 Agentic AI，Lexi 构建了对用户投放网址的多模态分析处理能力，以及智能投放的能力，满足了投放前广告内容的自动化生产、投放后效果追踪，实现用户投放广告效率的提升。
- ◆ **加速 Agentic AI 产品敏捷开发**  
借助 Amazon CDK 部署和管理资源，像 Amazon DynamoDB 这类有状态的基础服务可托管在亚马逊云科技上，结合 Serverless 的按量付费模式，能高效实现敏捷开发。对于后续遇到的复杂配置问题，直接通过 Amazon Q 自助处理即可解决。

- ◆ **让产品以更短路径与 Agentic AI 服务深度集成**  
2025 年 8 月初，Sandwich Lab 正式加入亚马逊云科技合作伙伴网络（APN）软件路径合作伙伴计划，并同步启动基础技术审核，全面验证其 AI 驱动广告系统在安全性、可靠性及性能上的最佳实践。依托 APN 提供的架构完善框架（Well-Architected Framework）与技术加速资源，Sandwich Lab 正以最短路径完成与 Amazon Bedrock、Amazon Q 等生成式 AI 服务的深度集成，为下一步申请亚马逊云科技生成式 AI Competency 奠定技术基线。

- ◆ **全球推广 Agentic AI 产品，一键触达海外中小企业市场**  
与此同时，Sandwich Lab 团队正筹备将旗舰产品 Lexi 上架亚马逊云科技 Marketplace，借助 APN 全球 200 多个国家/地区、数百万活跃客户的渠道网络，一键触达海外中小企业群体。通过 APN 的联合销售、市场营销基金及规模化交付支持，Sandwich Lab 不仅能以更低成本快速扩张业务版图，更将把“AI 即营收引擎”的创新范式带给全球客户，重塑海外数字营销的效率与可能性。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

我们在产品研发之初就选择了亚马逊云科技，它能很好地支撑产品研发，让我们具备了大型公司同等的基础设施，加快了产品研发效率，尽快将 Agentic AI 能力应用到 Lexi 产品当中，满足我们的海外业务需求。同时，亚马逊云科技 Startup 团队也给予了我们充分的辅导和帮助。

——刘尚楠 Sandwich Lab

超过

200,000+

的注册用户

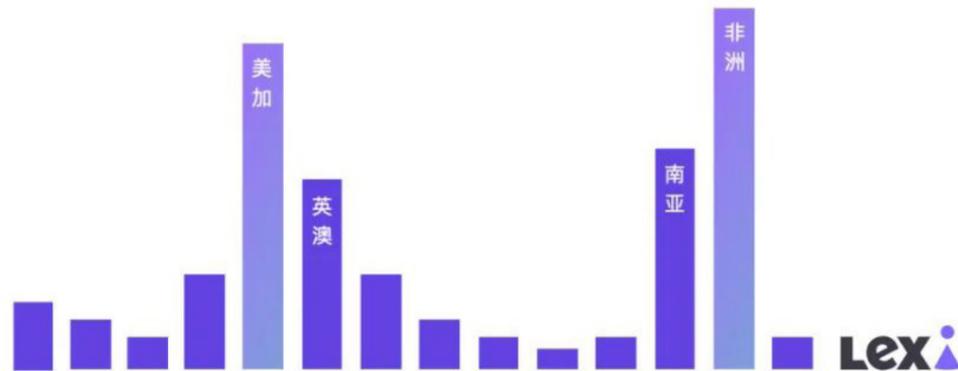
来自

90+

国家或地区

涵盖

电商 本地服务商  
数字服务提供商 ...  
等多个行业



## 未来展望

未来，Sandwich Lab 将继续深化与亚马逊云科技的合作，依托其稳定的基础设施与持续迭代的 Agentic AI 技术体系，不断拓展海外业务版图。同时，团队将积极探索并应用 Amazon Bedrock、Amazon Q 等相关产品与服务，持续强化 Lexi 在海外业务中的 Agentic AI 能力，优化广告投放的自动化与智能化水平，为全球更多中小企业提供更优质的广告营销解决方案，重塑全球中小企业增长引擎。

### 关于 Sandwich Lab

Sandwich Lab 定位为“商业智能 Agent”实验室，致力于构建创新性的数字化增长系统。通过融合生成式 AI、行为数据分析、内容策略算法及跨平台广告优化技术，公司打造出覆盖品牌定位 - 营销 - 转化全链路的智能解决方案，将 AI 从“辅助工具”升级为真正的“营收增长引擎”。

## 蔻町科技如何做到无需代码基础 全栈软件自动生成

蔻町科技（上海）有限公司（以下简称“蔻町科技”），主要从事代码大模型深度自研，实现端到端的软件生成。

目前公司预训练了全新一代 66B PLE 模型，并在此基础上推出了新一代端到端软件生成系统 AutoCoder.cc。该产品为全球首款端到端自动交付生产级软件的生成系统，用户无需代码基础，通过自然语言即可生成完整的软件，包含前端、后端、数据库等。产品自 7 月全球上线以来，累计新增用户数十万。



模型训练成本降低



产品上线时间缩短

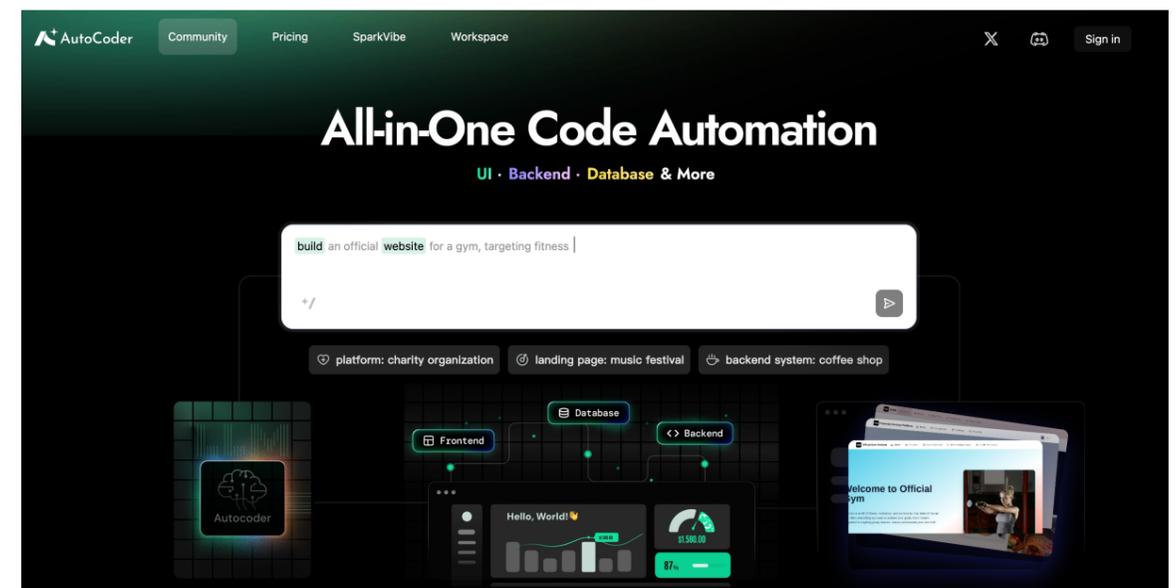


品牌曝光度提升



潜在客户增长提升

目前，蔻町科技采用的亚马逊云科技主要服务包括：**Amazon Bedrock（仅海外业务使用）、Amazon Q、Amazon EC2、Amazon EKS、Amazon EFS、Amazon WAF、Amazon Cloudfront、Amazon Lambda 等服务。**



\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 依托 Agentic AI 突破三重发展难题，激活全球发展潜能

蔻町科技的业务覆盖 C 端、B 端软件生成等多个业务场景，希望能基于生成式 AI 技术实现软件端到端自动生成，推动软件制作的技术平权，以加速业务在欧美等海外业务的布局，但是在应用 Agentic AI 前，蔻町科技面临开发链条覆盖不全、生成软件灵活性不足等技术问题：

- ◆ **开发链覆盖度不足**  
软件生成系统仅覆盖前端开发，无法覆盖后端、数据库等软件开发环节。
- ◆ **生成式效果无法快速改善和迭代**  
难以实现复杂任务的端到端软件生成。
- ◆ **生成软件的灵活性不足**  
生成效果依赖于模型的首次生成效果，难以按需修改。

蔻町科技综合评估全球业务覆盖、可扩展性、合规、成本等一系列核心要素后，发现**亚马逊科技提供的服务无论在 AI 基础设施的深度和广度、全球部署经验，还是对现有工作负载的兼容性方面都具有更强的优势**，可以更好地支持蔻町科技未来的海外业务布局。

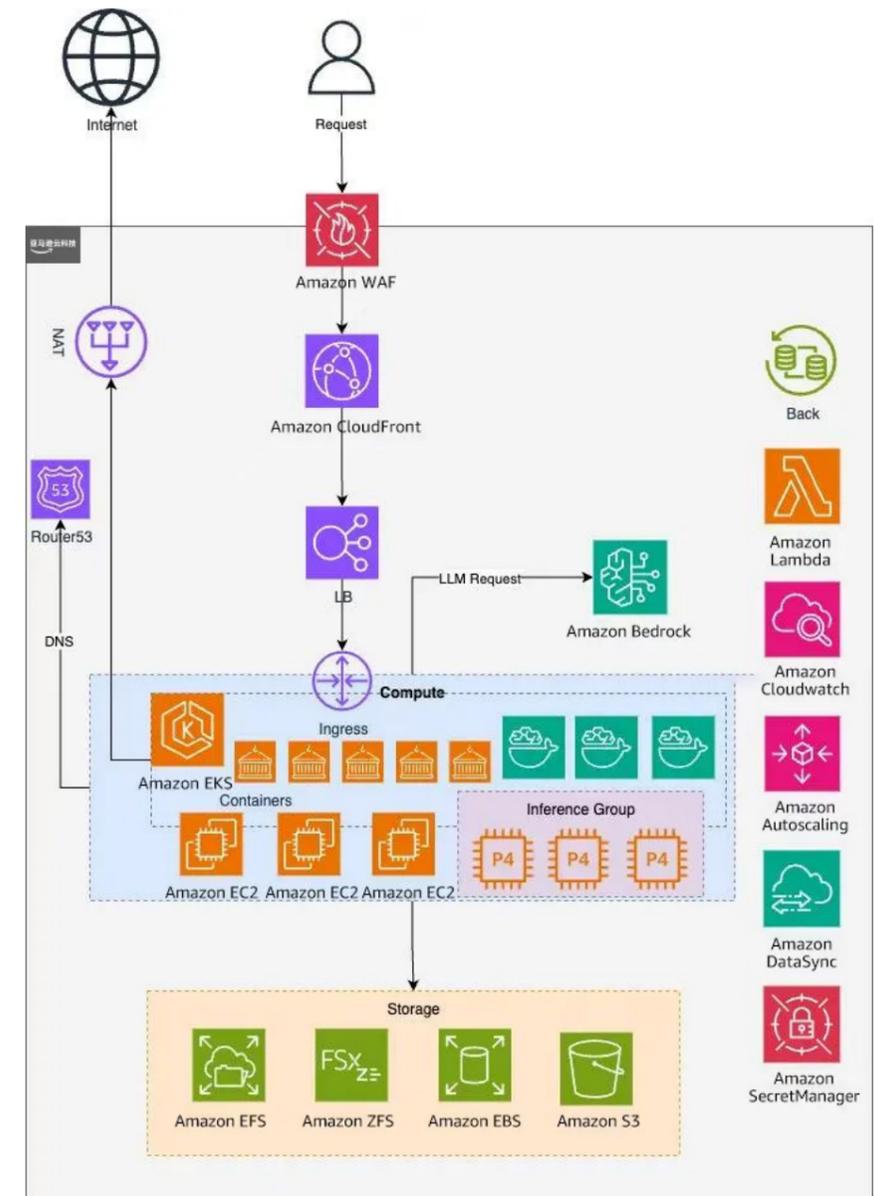
蔻町科技在海外部署过程中，不仅能**依托 Amazon Bedrock 安全可靠地接入业界前沿的大语言模型，实现多样化 AI 能力的灵活集成**，还能将底层模型的运维与扩展交给亚马逊科技托管，让团队能够聚焦于业务创新与应用落地。而且，结合 Amazon EKS 与高性能 GPU 实例的架构方案，蔻町科技在应对大规模并发请求时，可以同时兼顾性能与稳定性，为企业级 LLM 应用提供了高效、可扩展且更具成本优势的技术支撑。

## 解决方案 依托强大的云底座全面满足自动化软件构建需求

亚马逊科技提供的各类云上服务，为蔻町科技“业务 + 模型协同”的产品战略提供了坚实的技术底座，还显著简化了日常研发工作。同时在模型研发与测试环节，蔻町科技还能基于大规模弹性算力，以高性价比的方式支撑模型迭代、参数调优与实验探索。该模式不仅有效降低了整体计算成本，还提升了研发效率，使团队能更灵活开展高频次的模型优化与创新：

- ◆ **弹性拓展能力**  
以 Amazon EKS 作为计算资源调度与管理的核心系统，实现了 AI 应用的高效部署、自动化编排和弹性扩展，显著提升了整体运行效率与系统可靠性。
- ◆ **强大的算力支撑**  
在模型研发与测试阶段，蔻町科技结合 Amazon EC2 Spot 实例提供的大规模弹性算力，以高性价比的方式支撑了模型迭代、参数调优与实验探索，有效降低了计算成本并加快了研发周期。

- ◆ **高性能实例与存储服务**  
依托亚马逊科技提供的高性能实例与存储服务，确保了蔻町科技复杂计算任务运行的稳定性，并满足蔻町科技高吞吐量的要求。
- ◆ **确保数据安全合规**  
面向欧美区域，蔻町科技通过 Amazon KMS 实现数据全程加密，并在 VPC 子网、网络 ACL 与安全组层面进行了精细化访问控制，确保数据传输与存储的安全性。同时，蔻町科技还希望借助 Amazon API Gateway、Amazon WAF 与 Amazon PrivateLink，建立受控的数据访问路径，实现严格的网络隔离，以满足 GDPR 与 CCPA 等合规要求。



通过 Amazon EC2 Spot 实例提供的大规模弹性算力，不仅让我们能以高性价比的方式支撑模型迭代、参数调优与实验探索，还有效降低了整体计算成本，而且团队能灵活开展高频次的优化与创新，让产品响应速度提升了 50%。

——陈秋武、宿文  
蔻町科技（上海）有限公司创始人

通过以上架构设计，团队能够在保证性能的前提下，以灵活、低成本的方式实现高频次的模型优化与创新，以满足 Agentic AI 对高性能计算、灵活调度和低成本迭代的核心需求。此外，**蔻町科技的团队还依托亚马逊科技提供的免费培训资源与官方文档，快速掌握了 Agentic AI 功能的应用**，有效提升了自动化任务处理的效率与智能化水平。

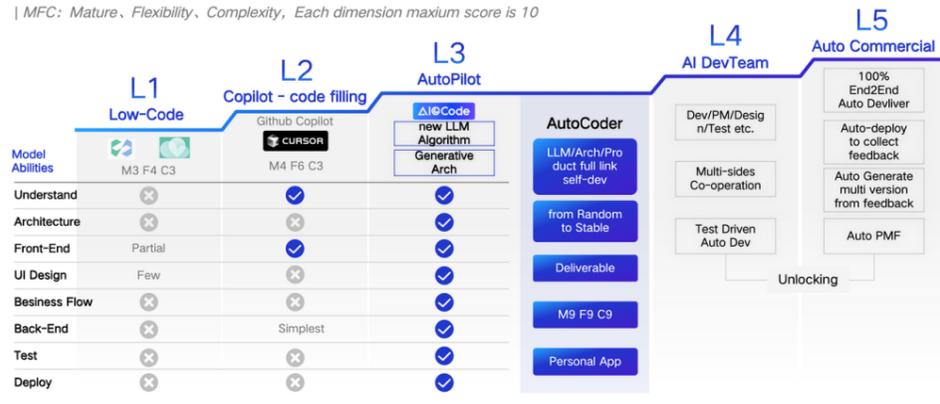
\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## L1-L5 of AI DevTeam for End2End software deliver

Five stages in AI DevTeam, Only by achieving collaborative optim across MFC can a truly End2End Software be Auto built.

| MFC: Mature, Flexibility, Complexity, Each dimension maximum score is 10



## 生成式 AI 助力 Notta 带飞全新会议体验：一键生成会议总结和会议纪要 58 种语言无障碍参会

Notta PTE. LTD (以下简称 Notta) 主要业务为会议转录工具 Notta 及其相关产品的开发和运营，其业务范围覆盖全球，并在日本和美国等地取得了良好的客户反响。

Notta PTE. LTD 主要业务为会议转录工具 Notta 及其相关产品的开发和运营。面向日本、美国等国用户，Notta 采用亚马逊科技 Amazon Bedrock 等生成式 AI 服务，一键生成会议总结和会议纪要，可支持 58 种语言的参会者无障碍参会。合作后，Notta 产品的版本开发迭代周期由 2 周变成了 1 周，总体成本节省 50%，3 年内已帮助数百万客户提升会议决策效率。

2 个月

2 个月完成生成式 AI 在智能提单和智能指标分析场景的落地

准确率超过 90%

简单场景中的自然语言查询，第一轮答复准确率超过 90%

准确率高达 100%

借助 Agents，复杂场景中的自然语言查询答复准确率可达 100%

截至 2024 年 8 月，Notta 已被 4000 家企业引入，累计用户数超过 500 万人。在日本、美国等国开展业务时，Notta 通过采用亚马逊科技 Amazon Bedrock 等生成式 AI 服务，根据会议转录文字自动生成会议纪要、会议总结，实现多语种会议翻译，构建了内部客服机器人 Notta AI，并探索将 Agent (智能体) 应用于敏捷项目管理、销售管理会议等场景。

面向日本、美国的用户提供服务时，Notta 目前采用的亚马逊科技服务包括：Amazon Bedrock、Amazon API Gateway、Amazon Lambda、Amazon Simple Storage Service (Amazon S3)、Amazon CloudFront、Amazon Certificate Manager、Amazon Secrets Manager、Amazon DynamoDB、Amazon Simple Notification Service (Amazon SNS)、Amazon Simple Queue Service (Amazon SQS)、Amazon Simple Email Service (Amazon SES)、Amazon Polly、Application Load Balancer、Amazon Elastic Compute Cloud (Amazon EC2)、Amazon Aurora、Amazon Elastic Block Store (Amazon EBS)、Amazon Elastic File System (Amazon EFS) 等。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 业务成果 将 Agentic AI 融入软件开发全流程为海外业务拓展注入新活力

基于亚马逊科技提供的强大 AI 基础设施和全球网络支持，蔻町科技有效缩短了模型的开发与部署周期，同时还借助分布式训练能力，减少了模型训练时间，并降低了相应成本，并通过持续迭代，让其产品具备了更多竞争优势：

### 提高竞争优势

蔻町科技在面向海外市场进行服务部署过程中，借助 Amazon Bedrock 等服务，在端到端网站应用生成的过程中，简化了前端页面代码逻辑的生成。同时通过与其自身模型深度融合，实现了端到端网站和应用完整代码的生成。

### 降低成本投入

依托亚马逊科技提供的高效计算资源和一系列的自动化工具，让训练模型和基础服务的成本降低了约 30%。

### 提升潜在业务增长

依托亚马逊科技丰富的技术解决方案和资源，让产品的响应速度提升了 50%，从而大大优化了用户的使用体验，而且还提升了 150% 的品牌曝光度和 25% 的潜在客户转化。

蔻町科技(上海)有限公司创始人&CEO 宿文认为：“亚马逊科技帮我们提升了运营效率、有效节约了成本，并让我们的产品有了更强大的业务竞争力，是我们扬帆出海之行的最好伙伴。”

未来，蔻町科技将聚焦产品业务拓展与模型升级等方面，持续深化与亚马逊科技的合作。产品方面，蔻町科技将通过共创社区，共拓广阔业务范围，预计在下一季度，将实现用户生成的软件直接在亚马逊科技上部署。模型方面，蔻町科技则计划通过与亚马逊科技的合作，进行更大参数的模型预训练，以支持产品的迭代升级。



### 关于蔻町科技

蔻町科技从 2023 年年底开始研发，主要从事代码大模型深度自研，实现端到端的软件生成。目前公司预训练了全新一代 66B PLE 模型，并在此基础上推出了新一代端到端软件生成系统 AutoCoder.cc。该产品为全球首款端到端自动交付生产级别软件的生成系统，用户无需代码基础，通过自然语言即可生成完整的软件，包含前端、后端、数据库等。AutoCoder.cc 自 7 月全球上线以来，累计新增用户数十万。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 更好地进行会议管理，生成式 AI 提升会议决策效率

作为一个 AI 原生的会议转录工具，Notta 旨在帮助全球用户提高会议效率，支持多种语言的转录、多语言会议实时翻译、一键总结和协作分享等功能。随着生成式 AI 和 LLM (Large Language Model, 大语言模型) 的技术更迭和爆发，Notta 致力于拥抱最新技术，用生成式 AI 赋能业务，帮助企业更好地进行会议管理，在会前、会中和会后帮助客户提升会议决策效率。

### ◆ Notta 生成式 AI 技术的部署聚焦于

1. 选择合适的 LLM 提升会议输出效率和精度：如何选择合适的模型，将强大的多模态 LLM 能力适配到会议的不同环节，提升会议转录文字生成纪要和翻译的效率与精度；
2. 企业会议引入 Agent 探索新应用：如何借助 Agent 自动捕捉会议中和公司内外数据的关键信息，对接工作流，探索企业会议场景的应用落地；
3. 安全合规保障：如何解决业务全球化和生成式 AI 应用中的隐私保护和合规问题。

Notta 跟亚马逊科技的合作主要面向日本、美国等市场，前两个阶段 Notta 将业务整体迁移到亚马逊云科技上，并大量采用数据库、人工智能和机器学习等技术服务，今年双方在生成式 AI 上的合作顺理成章。

通过亚马逊科技 Amazon Bedrock 访问业界领先的大语言模型，让 Notta 的生成式 AI 应用高效落地，产品的版本开发迭代周期由 2 周变成了 1 周，总体成本节省 50%，3 年内已帮助数百万客户提升会议决策效率和会议目标达成率。基于亚马逊科技完善且稳定的云服务以及强大的技术、商务支持，让 Notta 能够始终聚焦自己的核心业务，极大促进了我们的业务发展。更重要的是，作为一家服务全球客户的公司，数据安全和合规性对我们至关重要。亚马逊科技的全球本地化数据中心战略让我们能够轻松满足各个市场的数据保护要求。例如，我们可以确保日本用户的数据存储在日本的亚马逊云科技数据中心，美国用户的数据存储在美国的数据中心。这不仅帮助我们遵守了各地的数据保护法规，还显著提高了服务性能，减少了数据访问延迟。

—— 代同杰  
NOTTA PTE. LTD CTO

## 解决方案 通过 Amazon Bedrock 调用大语言模型，聚焦会议管理的三类场景

Notta 从生成式 AI 兴起早期就开始利用 LLM 能力，2024 年 4 月，通过 Amazon Bedrock 调用业界领先 LLM，聚焦于会议管理的三类场景，正式上线了生成式 AI 新应用。

### 场景一：基于会议转录文本，高效、准确地生成会议总结和会议纪要

- ◆ Notta 可实现边录音边转录，语音实时转文字，听到即看到；也可将录音文件或音频会后导入并转为文字，并支持边听边改，轻松编辑；支持一键分享音频，可将转录文本导出为 PDF、DOCX、TXT、SRT 等文件格式，分享至其它应用。

在应用 LLM 之前，Notta 仅能通过 Amazon Comprehend 提取会议关键词，无法输出完整的会议总结。采用 Amazon Bedrock 调用 LLM 之后，Notta 可基于转录文本借助大模型对会上的各方发言提炼关键信息，高效生成会议总结、会议纪要等。为方便参会人员快速精准获取会议关键信息，亚马逊科技技术专家帮助 Notta 团队优化会议场景的 Prompt (提示词)，通过加入 xml 标签来区分不同变量。例如，通过 <JSON> 标签标记格式内容，通过 <context> 标签标记正文，显著提升模型的响应精度，并在 Prompt 中增加语言参数，有效解决响应语言与用户期望语言不一致的问题。经过反复优化，Notta 生成的会议总结和会议纪要得到了众多使用者的认可，不少用户认为 Notta 自动总结的内容甚至比转录原文更准确。

### 场景二：选择合适 LLM，高效、精准、低成本实现多语种会议实时翻译和会后翻译

Notta 采用亚马逊科技的 Amazon Bedrock，基于不同的场景需求，通过 Amazon Bedrock API 调用业界领先的与场景匹配的 LLM 模型。

- ◆ 在实时翻译场景中  
在 Amazon Bedrock 上选择实时性要求高、低时延的 LLM，成本效益更好。

- ◆ 在会后翻译场景中  
将音频文件上传到 Amazon EC2 G5 实例运行 Notta，并调用 Notta 的自研模型 Whisper。Amazon EC2 G5 实例是基于英伟达 GPU 的最新一代实例，与 Amazon EC2 G4dn 实例相比，G5 实例可提升 3 倍的性能，且将机器学习推理的性价比提高 40%，还可为机器学习训练的性能提升 3.3 倍；与 Amazon EC2 P3 实例相比，G5 实例可将训练成本降低 15%。G5 实例帮助 Notta 获得高性能的基础设施，Notta 就可以高效构建自有模型 Whisper。

- ◆ 在多语种会议场景中  
Notta 以前采用的是第三方翻译服务，翻译的准确度并不稳定。通过 Amazon Bedrock 访问业界领先 LLM 后，Notta 可支持实时的双语互译（日语、英语、中文简体、中文繁体、韩语、印度尼西亚语、西班牙语、葡萄牙语、爪哇语、尼泊尔语、越南语等 11 种语言）和 58 种语言的会后翻译。更重要的是，生成式 AI 的引入极大地提升了翻译准确度，给多语言会议带来无碍体验。

### 场景三：进行会议 Agent、聊天机器人等新功能探索

Notta 借助 Agent 打造知识共享底座，持续创新推出生成式 AI 在会议管理领域的新应用，目前 Notta 先行探索了两类应用：

- ◆ 会议 Agent  
Agent 自动感知并提取会议的关键信息，识别会者的意图，根据识别的结果调用相关工具。目前在销售会议、敏捷项目管理两类场景同步应用 Agent 自动感知会议关键信息。比如，在销售会议中，Agent 可以对接工作流，判断下一步动作，销售人员可以通过 Meeting Agent 发送邮件邀约、分析和跟踪会议内容、调用第三方客户关系管理软件服务提供商的 API 与销售流程做对接，甚至辅助完成销售中的所有步骤；在敏捷项目管理场景中，通过 Agent 调用第三方敏捷项目管理工具的 API，将会议内容与敏捷开发流程做对接。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

◆ 聊天机器人 Notta AI

可以直接与会议纪要、会议内容对话，快速查询、获取会议内容。会议期间，与会者可以基于多份会议文档进行讨论。为了提高内容搜索效率，Notta 还采用基于 OpenSearch 实现 RAG (Retrieval Augmented Generation, 检索增强生成)，利用 OpenSearch Service 的向量数据库功能，让

Notta 的用户可以根据关键字进行内容搜索，也可以进行语义搜索。在聊天机器人回答客户问题时，Notta AI 可高效查找和获取产品说明书、产品功能介绍、历史记录等，节省客服搜索和响应时间；Notta AI 还加入了反馈机制，对于知识库中可能存在的无效、老旧知识进行反馈和更新。

◆ 亚马逊科技全球本地化数据中心确保数据合规，保障客户符合严苛安全合规需求

通过亚马逊科技遍布全球的数据中心网络，确保客户数据的存储位置符合各地区的数据安全要求和法规。

在业务所在地区存储数据	满足当地安全合规认证
借助亚马逊科技遍布全球的数据中心网络，确保不同地区的用户数据严格存储在其所在地区的亚马逊科技数据中心。例如，日本用户的数据存储在亚马逊科技的日本数据中心，美国用户的数据则存储在亚马逊科技的美国数据中心；	亚马逊科技帮助推动 Notta 通过 ISO 27001 信息安全管理体系、GDPR(EU General Data Protection Regulation, 欧盟通用数据保护条例)、日本的 APPI(Act on the Protection of Personal Information, 个人信息保护法) 等安全合规认证；
降低数据访问延迟	保障数据安全
亚马逊科技的区域化存储策略不仅帮助 Notta 满足各国严格的数据保护法规，还显著降低了数据访问延迟，提升了整体服务性能；	Notta 还采用 Amazon WAF, Amazon GuardDuty, Amazon Inspector, Amazon IAM, Amazon Secrets Manager, Amazon Cloud Trail 等服务帮助 Notta 保障数据加密和安全存储等，使得 Notta 满足 RPM(Request Per Minute, 每分钟请求数)、SLA(Service-Level Agreement, 服务等级协议) 以及其它安全需求。

此外，亚马逊科技解决方案架构师团队向 Notta 分享领域驱动设计方法论，赋能 Notta 团队进行微服务改造，把单体应用变成微服务的应用；亚马逊科技解决方案架构师团队还为客户引进了容器化应用改造，提高系统的可用性。

业务成果

生成式 AI 助力产品的版本开发迭代周期缩短一半，总体成本节省 50%，3 年内用户达百万级

◆ 产品迭代周期从 2 周变为 1 周，实现一键生成会议总结和会议纪要，58 种语言无障碍参会

应用亚马逊科技服务后，Notta 的开发效率得到极大提升，产品的版本开发迭代周期由 2 周变成了 1 周，产品发布速度提升了 1 倍，Notta 的生成式 AI 功能在一个月内得部署成功，实现一键生成会议总结和会议纪要，可支持 58 种语言无障碍参会。同时，亚马逊科技解决了从基础设施、生成式 AI 应用、安全合规到架构现代化等问题，Notta 能够更加聚焦自身业务创新，实现全球用户和营收数倍增长。

◆ 生成式 AI 应用和云基础设施成本均得以降低，帮助 Notta 总体成本节省 50%

采用 Amazon Bedrock 访问业界领先 LLM，对比前期采用其它 LLM 的成本降低约 30%。引入 LLM 进行翻译后，成本仅为之前采用第三方翻译服务的成本的 1/7。在云基础设施方面，Notta 使用 Amazon EC2 预留实例，结合 Amazon EC2 Spot 实例和 On-Demand (按需定价) 实例，计算成本节省 30%~40%。综合以上，总体成本节省 50%。

◆ 3 年内用户数量达百万级，位列 App Store 人工智能文字转录服务类别排名第一

借助亚马逊科技服务构建应用驱动，Notta 的全球用户数倍增长，3 年内用户数量达到百万级。得益于其高效的会议转录，适配日本市场、98% 准确率的文本精度，在 2023 年日本市场研究机构 JMR 的研究报告中，Notta 被评为 Apple 公司 App Store 人工智能文字服务类别排名第一的 App。

随着 Notta 业务的发展，公司正将业务重心从日本拓宽至全球。同时，除了会议转录、总结及纪要、会议翻译等功能外，将基于亚马逊科技 Amazon Bedrock 继续探索包括会议 Agent、聊天机器人在内的 Agent 在企业会议领域的应用，实现公司的愿景——通过对话发现无限价值。



关于 NOTTA PTE. LTD

Notta PTE. LTD 主要业务为 Notta 及其相关产品的开发和运营，业务范围覆盖全球，并在日本和美国市场取得了客户的良好反响。Notta 可通过高精度的 AI 和声音识别引擎进行文字转写和总结，覆盖从会议日程预约到会议纪要生成并共享的全流程，大幅提高会议的生产力。截至 2024 年 8 月，Notta 已被 4000 家企业引入，累计用户数超过 500 万人。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 包小盒如何做到包装设计秒定稿，产品开模低成本

包小盒专注为全球客户提供专业高效的在线包装设计工具，旗下 Pacdora 拥有创新且成熟的 3D 数字包装结构设计与渲染仿真技术，为企业提供直达生产的全链路数字化解决方案。

亚马逊云科技为包小盒提供了 Serverless 架构以及多样化的生成式 AI 服务，帮助包小盒提升开发与运营效率，加速海外业务拓展与创新。

目前，包小盒使用的亚马逊云科技产品与解决方案包括：Amazon Bedrock、Amazon Q Developer、Amazon CloudFront、Amazon Lambda、Amazon Elastic Compute Cloud (Amazon EC2) 等。



一般任务平均导出时间



用户付费率提升



完成业务数据精准收集

### 机会 业务出海驱动合作深化，加速构建全球交付能力

近年来，随着新兴消费品牌快速崛起，小批量上新、快速迭代成为主流销售模式，推动了包装设计服务的智能化、自动化发展。相比传统依赖人工设计、周期长、成本高的方式，品牌客户更希望在产品快速上架前就能获得便捷高效、低门槛的包装创意工具与打样服务。包小盒正是在这一背景下应运而生，致力于通过技术手段降低包装设计的使用门槛，实现从图纸设计到 3D 渲染、在线打样、快速交付的全流程数字化。在产品能力方面，包小盒旗下的 Pacdora 支持 3D 包装的实时生成与可视化预览，使设计师无需实物打样即可模拟不同使用场景下的包装效果，显著提升提案效率与客户下单转化率。对于网站客户与包装工厂，Pacdora 还能自动输出匹配尺寸的刀线文件，减少重复建模与人工制图工作，提升整体生产协同效率。用户调研显示，包小盒在缩短包装设计交付周期、提升沟通效率与促成下单转化等方面带来了显著价值。

考虑到海外客户在设计效率、协同体验等方面更强烈的定制需求与付费意愿，包小盒自 2022 年起启动出海布局，而亚马逊云科技以良好的品牌口碑、覆盖全球的基础设施以及出色的服务响应能力，成为包小盒海外业务拓展的首选。面对多区域用户访问延迟、资源调度复杂等挑战，包小盒借助 Amazon CloudFront 优化内容分发路径，在欧美、东南亚等主要地区显著提升了访问性能，为其生成式 AI 服务的全球交付奠定了技术基础。

同时，亚马逊云科技也积极围绕网络安全、数据合规与风险识别展开分享，通过一系列培训，帮助包小盒团队建立起清晰的安全合规意识。在亚马逊云科技帮助下，包小盒对公司内部潜在风险进行了排查，进一步提升了安全防护意识与治理能力，同时还完成了对美国、欧盟等地的法规要求梳理，仅花费 3 个月即建立起相应的数据本地存储与访问策略，为出海业务提供了有力支撑。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

基于双方在全球部署、系统安全合规等方面的长期合作，包小盒在启动生成式 AI 探索时，选择继续依托亚马逊云科技的技术与服务，优化系统架构并将生成式 AI 能力引入业务流程中的多个环节，持续提升产品的智能化能力与交互体验。

亚马逊云科技为我们提供了关于安全、网络和数据合规等方面的系统培训，这对我们这样的初创企业非常关键，帮助团队快速建立起清晰的认知与规范体系。

——李法政  
包小盒联合创始人

### 解决方案 云计算 + 生成式 AI，包小盒打造秒级响应的智能设计引擎

#### Amazon Lambda 弹性调度：全球任务处理提速至十数秒内

Pacdora 是包小盒的核心业务模块，用户可在此完成从包装模板选择、参数配置、设计编辑到效果预览、结果生成的完整设计流程。其中，将设计内容导出为 PDF 文件以便后续用于提案、审阅或生产对接，是整个包装设计流程中最后且尤为重要的一步。随着网站用户量持续增长，该功能的调用频率不断上升，而北美、欧洲、东南亚等地区的访问高峰呈现出时间交错、分布不均匀等特征，原有基于消息队列的处理架构难以灵活应对这一压力，任务堆积、响应延迟等问题时有发生。

运用 Serverless 架构后，包小盒显著提升了任务处理效率，为全球用户带来了更加稳定流畅的使用体验。目前，一般任务的结果导出时间可控制在十数秒内。

我们早期采用消息队列进行处理，高峰时段用户可能需要较长时间等待。引入 Amazon Lambda 后，系统的计算压力由亚马逊云科技弹性承载，无论并发请求数量是多少，都能够实现高效分发与稳定响应。

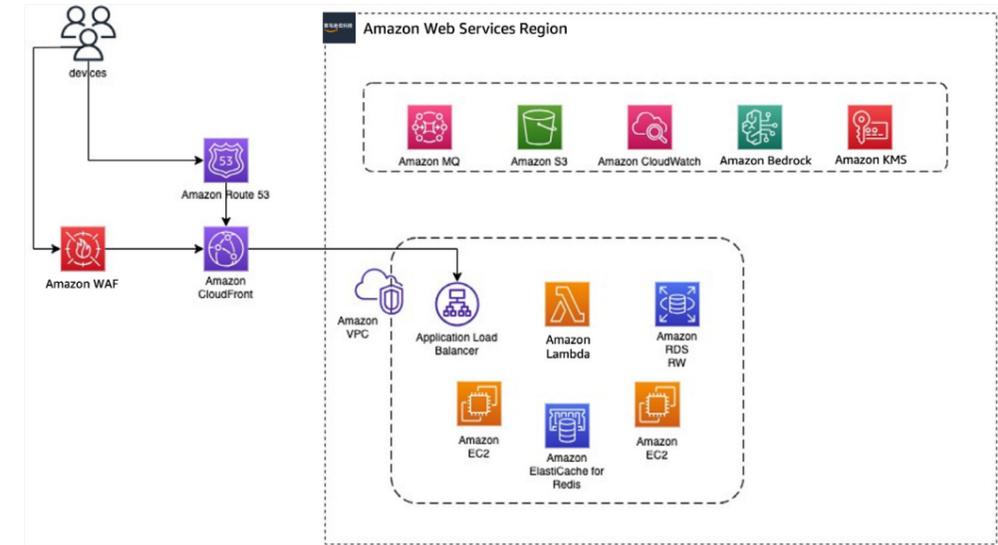
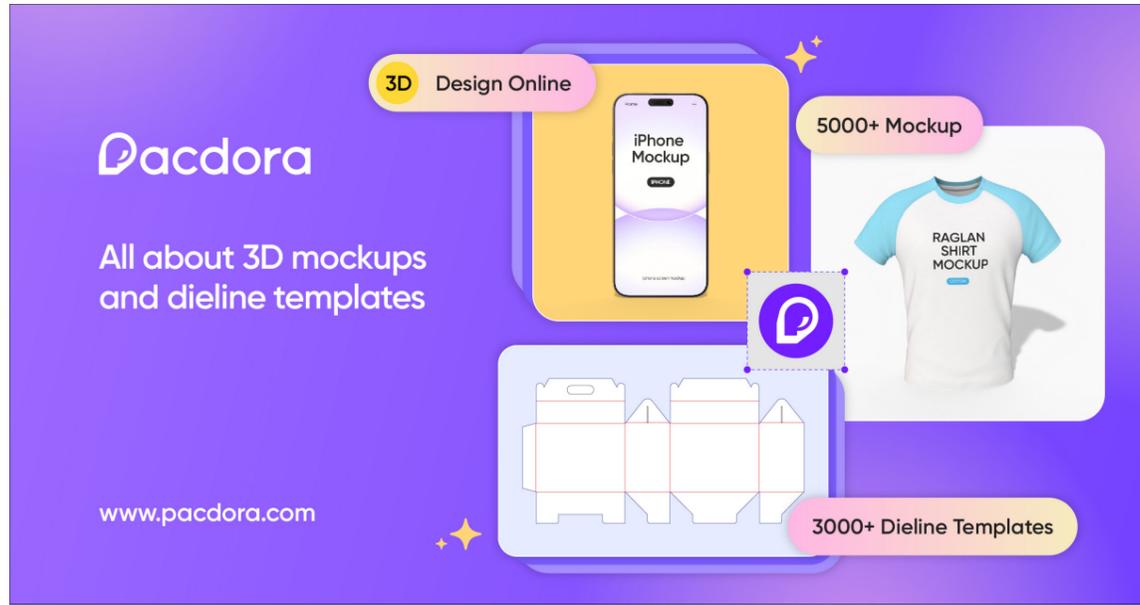
——李法政  
包小盒联合创始人

#### 生成式 AI 加持，驱动设计体验增长、运维效率提升

在稳定支撑全球任务执行的基础上，包小盒持续探索如何以生成式 AI 进一步优化包装设计流程。为此，团队搭建了试验工具 Packify.ai，尝试以更智能、便捷的方式帮助用户清晰表达设计意图、释放灵感。

经由 Amazon Bedrock 接入 DeepSeek 模型作为支持，包小盒以全新的智能对话式交互，替代了传统的图形界面操作。Packify.ai 工具将通过聊天方式，逐步引导用户明确包装类型、风格偏好、色调倾向等核心设计要素，并自动将用户需求转化为结构化设计指令，交由包小盒自研视觉模型完成设计图像生成。相比以往，该方式更符合用户的自然表达习惯，显著提升了设计过程的完整性与交互体验。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。



包小盒基于亚马逊云科技的系统架构示意图

**业务成果** 研发提效、体验创新，加速全球业务增长

在应用生成式 AI 能力的过程中，亚马逊云科技始终为包小盒提供稳定的底层算力支持。2024 年底，包小盒上线“AI 背景图设计”功能，获得用户高度评价、短时间内 GPU 渲染请求激增。面对突增的算力需求，亚马逊云科技快速响应，持续协助包小盒将初期部署的 2 台 GPU 节点平滑扩展至 15 台，以弹性的资源调度能力实现“用户无感知”扩容，有力保障了新业务的运行。包小盒数据调研显示，使用“AI 背景图设计”功能的用户，其付费率相较其他用户提升了 20%-30%。

持续引入生成式 AI 优化产品体验的同时，包小盒也将 AI 能力扩展至研发与运维流程，借助 Amazon Q Developer 辅助开发人员快速生成运维所需的命令行脚本与自动化代码（如 Python 脚本或 Batch 指令），内容范围包括通过 API 调用亚马逊云科技服务、批量服务器管理等常见任务。运维团队仅需通过自然语言提问即可获取所需脚本或操作指令，大幅减少了人工查找文档或重复编写代码的工作量。

上线这一功能后用户反响远超预期，短时间内 GPU 资源迅速耗尽。亚马逊云科技在资源扩展方面提供了高效且持续的支持，帮助我们在两个月内将初始部署的 2 台 GPU 服务器平滑扩展至 15 台，确保了新功能在高访问量下依然能够稳定运行并持续交付良好体验。

——李法政  
包小盒联合创始人

在亚马逊云科技覆盖全球的基础设施、稳定高效的技术架构以及丰富的生成式 AI 服务支持下，包小盒获得了开发效率与用户体验的双重提升，能够以更快的速度完成业务验证与新功能上线，为业务持续创新与全球拓展奠定了坚实基础。

通过使用亚马逊云科技的服务，我们显著减轻了非核心事务上的管理压力。这正是我们采用云服务的初衷——依托可靠的云技术伙伴降低运营负担，专注产品创新与业务拓展。

——李法政  
包小盒联合创始人

未来，包小盒希望携手亚马逊云科技进一步拓展日本等更广阔的海外新兴地区，结合亚马逊云科技在当地的基础设施与合作网络，加速产品本地化落地与用户触达。双方还将围绕更多生成式 AI 能力与智能设计场景展开联合创新，持续释放云计算与生成式 AI 融合带来的业务价值，为全球用户带来更加智能、高效的设计解决方案。

**以数据驱动精准迭代，构建更高效的全球分发机制**

为验证新功能在不同国家与地区的适配性与用户接受度，包小盒在亚马逊云科技帮助下构建了面向全球的灰度测试机制。依托 Amazon CloudFront 提供的地域分发与访问控制能力，包小盒团队可精准指定位于代表性国家的 5%-10% 用户群体进行功能试点，确保所回收数据具备统计学价值。在为期一周的调研后，包小盒就能完成对业务数据与关键指标的收集与整理，为后续功能优化与全球推广提供重要的决策依据。

**关于包小盒**

包小盒成立于 2020 年，专注为全球客户提供专业高效的在线包装设计工具，目前已成功服务超过 500 万用户。包小盒研发的 Pacdora 是一款简单、高效和专业的在线 3D 包装设计工具，以创新且成熟的 3D 数字包装结构 / 模型、云设计与渲染仿真技术，为企业提供在线营销、3D 设计、直达生产的全链路数字化解决方案，让包装设计“所见即所得”。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

# 02

## 传统软件企业

融入 Agentic AI  
开启智能化发展新征程



### 从“敲命令”到“说需求” 合合信息如何借助 Agentic AI 让运维化繁为简

上海合合信息科技股份有限公司始终致力以 AI 技术创新赋能，打造多模态大模型文本智能技术，向全球 C 端用户和多元行业 B 端客户提供“真有用、真好用”的产品。公司旗下 C 端 APP 产品扫描全能王、名片全能王等产品服务了全球超过 200 多个国家和地区的亿级用户；B 端产品已在近 30 个行业场景中落地。

合合信息 2022 年加入亚马逊科技合作伙伴网络 APN，与亚马逊科技在品牌、业务、销售协同、技术和创新上均有深度的合作。合合信息构建的 Chaterm 等 Agentic AI 服务依托亚马逊科技提供的 Amazon Bedrock（仅海外业务使用）等服务，可以更灵活地调用大模型，从而打破了传统运维服务瓶颈，帮助企业大幅提升运维效率，降低运维工作的复杂性。



Github star  
数量超过 1100



代码提交次数  
达 2100+



Fork 达到  
160+



开源社区贡献者  
已超 20 人

合合信息旗下产品 Chaterm 开源数据

#### 机会 依托 Agentic AI 打造智能、高效的运维服务工具

合合信息旗下的主要产品及服务皆由公司自主研发，希望通过长期持续的技术创新，能让“机器像人类一样阅读”，从而进一步提升人类的效率。当前，合合信息在复杂场景文字识别、智能图像处理、自然语言处理、知识图谱及大数据挖掘等核心技术领域均位于行业前列。但当前云管理和运维中 DevOps 实践需要开发者管理成百上千台服务器和容器，由于智能化程度低，让运维复杂度急剧提升，导致运维效率下降，在批量操作、故障排查等方面都给运维的团队带来了巨大技术挑战：

##### ◆ 批量操作繁琐

在大规模分布式系统中，运维工程师经常需要在数百台服务器上执行相同操作。虽然传统的 Amazon Systems Manager Agent (SSM Agent) 也能实现对集群机器的批量处理，但缺少大模型的加持，其智能化程度难以满足日常工作的需求。

##### ◆ 知识门槛高

运维技术栈的深度和广度构成了显著的知识壁垒。运维人员需要熟练掌握多种命令行工具、脚本语言、正则表达式和系统配置知识。这种从操作系统内核到应用层的全栈知识要求，使得新手工程师至少要半年以上的实践才能应对常规问题。

##### ◆ 故障排查复杂

在微服务架构中，故障排查演变为复杂的分布式追踪难题。当用户报告问题时，运维人员需要通过 ELK Stack 检索横跨 API 网关、订单服务、支付服务的数百条日志，结合 Jaeger 追踪 ID 关联调用链。这种跨服务、跨组件的日志关联分析，往往需要资深工程师耗费数小时才能定位到具体问题。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

为更好地解决上述问题，构建更智能、高效的解决方案，合合信息选择基于亚马逊云科技的产品基础架构作为底座，借助 Amazon Bedrock 调用大模型的能力，面向海外的运维需求，开发了 SRE Agent Chaterm。

Chaterm 是一款由 AI Agent 驱动的智能终端工具，它可将 AI 能力与传统的终端功能相结合。该工具的核心理念在于通过自然语言交互，革新开发者与终端的互动方式。用户使用自然语言就能进行交互，能简化复杂的终端操作，无需记忆不同操作系统中的复杂命令语法，就能高效完成日常运维工作。

## 解决方案 依托 Amazon Bedrock 构建 Agentic AI 应用，突破传统运维流程

Chaterm 将 AI 从简单的命令生成器提升为真正的运维助手，不仅能提供 AI 对话和终端命令执行功能，还具备 Agentic AI 的自动化能力，可以通过自然语言设定目标，由 AI 自动规划，并一步一步执行，最终达成需要处理的任务或完成需要修复的故障。无论在系统提示工程的技术优化，还是任务规划、上下文理解等方面，依托 Amazon Bedrock 构建的开源 Agentic AI 工具 Chaterm 都获得了重大技术突破：

### ◆ 系统提示工程的优化

Chaterm 基于精心设计的系统提示，被定位为“拥有 20 年经验的资深运维专家”，具备网络安全、故障排查、性能优化等多方面专业知识以及强大的解决问题的能力，这种定位使其能够从专业运维人员的视角思考问题，提供更符合最佳实践的解决方案。

### ◆ 上下文感知与状态管理

合合信息对 Chaterm 的上下文感知和状态管理进行了深度优化。在维护执行过程中，可通过上下文信息，确保后续操作能够基于之前的结果进行，并能够感知上下文窗口进行提醒以及合适的防溢出机制。在任务管理方面，Chaterm 支持任务的恢复与继续。

### ◆ 自适应执行与错误恢复机制

与简单的脚本执行不同，Agent 具备自适应能力，能够根据每步执行结果动态调整后续计划。当遇到错误时，Agent 会尝试理解错误原因，提供修复方案，并在必要时调整执行路径。

### ◆ 优化推理速度

Chaterm 利用 Amazon Bedrock 提供的多项技术优化推理速度，不仅提升了海外云端资源的管理效率，还对常用提示进行缓存，进一步降低延迟和成本。在运维场景中，系统提示和工具定义等静态内容占据了大量 Token，通过缓存这些内容，可以显著减少每次请求的首个令牌生成时间 (TTFT)，从而整体优化和提升推理速度。

此外，Chaterm 还基于 Amazon EKS 服务构建了 Agent 的实时可观测性系统，通过全链路记录 Agent 执行过程，可实现对 Agent 行为的全面监控、评估和持续优化。

依托亚马逊云科技的 Amazon Bedrock、Amazon KMS 等的服务，既确保了 Chaterm 功能的丰富性，为未来探索更多 AI Agent 提供了更多可能性，还有效确保了运行的安全性与可靠性。

——段炼  
Chaterm 产品负责人



## 业务成果 Agentic AI 重塑运维工作流程让运维工作化繁为简

当前，Chaterm 已经开源，开源后 Github star 数量超过 1100，Fork 达到 160+，代码提交次数达 2100+ 次，开源社区的共享者已超过 20 位，这不仅展示了合合信息在人工智能和大数据方面的技术实力，还全面打破了传统运维工作方式，从技术门槛、高效运维等方面突破了工程师面临的核心痛点：

### ◆ 降低技术门槛

将 Agentic AI 技术融入终端后，用户无需再记忆复杂的命令语法，只要使用自然语言即可完成日常的运维工作。

### ◆ 提升上下文理解能力

合合信息依托于 Amazon Bedrock，构建了面向海外云端资源的管理工具 Chaterm，借助大模型的强大的语义的理解能力，能够清晰理解输出的信息，即使信息有误，也能通过自动分析在大量的系统日志中定位问题。

### ◆ 实现高效批量操作

面对大量重复任务，Chaterm 可以同时多个不同的主机或操作系统上统筹任务，并行执行操作指令，从而大大提升了运维效率。

Chaterm 作为 AI 运维工具，未来，合合信息将依托 Amazon Bedrock 在海外陆续推出语音控制、自动化场景编排等功能，并扩展对多云环境的原生支持能力，为用户打造统一的多云管理界面和体验。



INTSIG 合合信息

### 关于合合信息

上海合合信息科技股份有限公司始终致力以 AI 技术创新赋能，打造多模态大模型文本智能技术，向全球 C 端用户和多元行业 B 端客户提供“真有用、真好用”的产品。合合信息一直坚持“让世界更高效”的使命，依托合合信息研发队伍，通过不断的技术开发，解决用户需求，帮助企业转型数字化，通过技术创新为全球个人用户和企业客户提升效率。合合信息 2022 年加入亚马逊云科技合作伙伴网络 APN，是亚马逊云科技 Advaned 级合作伙伴。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 回复准确率超 87%，大模型幻觉降低 90%

### Agentic AI 重塑智齿科技智能客服体验

北京智齿博创科技有限公司（以下简称“智齿科技”），一体化客户联络解决方案提供商。面向全球企业，提供基于「客户联络」场景的一体化解决方案。

智齿科技为 **OPPO、三星、惠普、Opay、京东到家、屈臣氏** 等全球上万家头部企业提供了服务，涵盖零售、教育、金融、企业服务、生活消费、运营商、政府以及公共事业机构等多个领域。

目前，智齿科技采用的亚马逊云科技主要服务包括：Amazon Bedrock（智齿科技仅海外业务使用 Amazon Bedrock）、Amazon OpenSearch、Amazon Comprehend、Amazon Translate、Amazon RDS、Amazon EMR 等。



基于 Agentic AI 的智能客服大模型幻觉降低 90%



基于 Agentic AI 的智能客服人工介入降低 42%



基于 Agentic AI 的智能客服回复准确率超 87%



全托管服务，全程数据加密，数据可靠性提高 1 倍

## 机会 使用 Agentic AI 重塑智能客服产品为客户提供更准确的 AI 服务

智齿科技不断整合前沿的人工智能及大数据技术，已经构建形成呼叫中心、机器人「在线+语音」、人工在线客服、工单等多个软件产品，智能客服机器人作为企业运营中的新型解决方案，凭借高效、便捷、全天候的特点，正逐步成为企业提升客户满意度、优化运营成本的关键利器。

大语言模型的兴起，无疑为智能客服领域注入了新的活力。通过集成大语言模型，不仅显著提升了客服机器人的对话、理解和应变能力，还让智能客服机器人能够更加精准地捕捉客户意图，并通过为客户提供更个性化的服务，大幅提高了客户满意度和客户体验。但在智齿科技国际化开发和适配智能客服产品的过程中，遇到了复杂场景交互、超负荷工作量等挑战：

#### ◆ 多语言复杂场景下交互准确性难确保

国际化过程中，智能客服产品需要支持诸多语种方案、多渠道整合需求，以及确保复杂场景下的交互准确、可控。

#### ◆ 开发与测试工作量大幅增加

在早期基于机器学习的 NLP 模型基础上，不同语种需要进行单独的算法适配、功能开发、功能测试，带来了巨大的开发与测试工作量。

为了突破产品开发和适配中遇到的问题，重塑智能客服机器人产品，智齿科技选择与亚马逊云科技合作，主要因为亚马逊云科技的可用区遍及全球多地，而且提供的 Amazon Bedrock 等生成式 AI 服务能满足智齿科技海外客户并发处理和快速响应需求，并支持全球多渠道海量文本和语音对话，可与海外客户即时交流中提供快速、高效、准确的 AI 服务。此外，亚马逊云科技在安全合规领域拥有丰富的解决方案，可确保智能客服产品在国际化过程中运营的合规性。

## 解决方案 借助 Amazon Bedrock 构建高可靠高可用的 Agent 解决方案

针对海外的智能客服产品，依托 Amazon Bedrock 等服务，助力智齿科技基于 Agentic AI 重塑了智能客服机器人产品，不仅能更好地适配多语言和国际化场景，还开发了多重坐席辅助功能：

#### ◆ 提高智能客服产品的多语言服务水平

智齿科技基于 Agentic AI 重塑智能客服机器人产品后，智能客服机器人的对话变得更拟真，理解也更精准。同时也简化了搭建流程的复杂度，能支持更多渠道，并在工作量可控的前提下，实现中文、英语、葡萄牙语等 19 种多语言与国际化场景的适配。

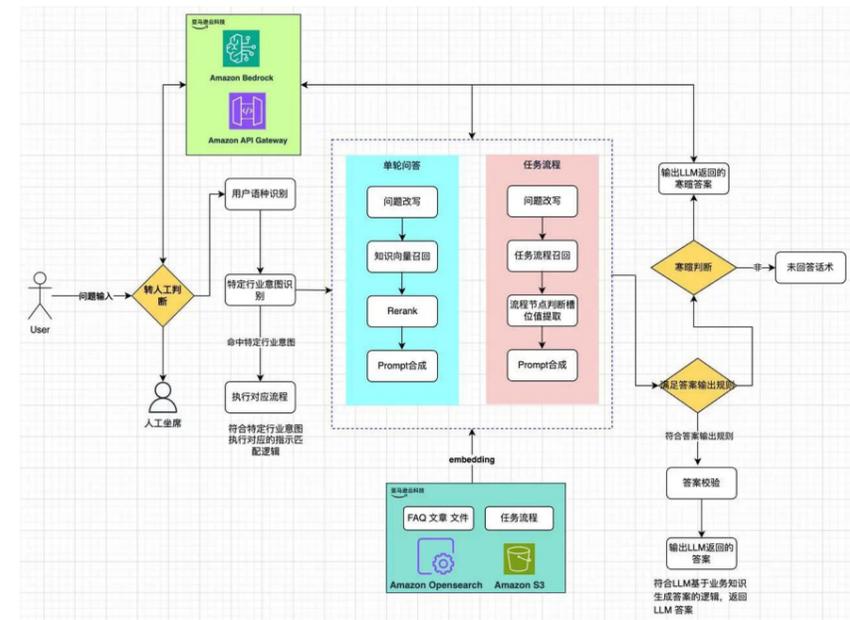
#### ◆ 增加坐席智能辅助功能

基于 Agentic AI，智齿科技针对智能服务总结、会话辅助等场景，开发了多款针对客服坐席的辅助功能。

智齿科技依托亚马逊云科技提供的 Agentic AI 服务构建的智能客服机器人作为企业运营中的新型解决方案，凭借高效、便捷、全天候的优势，已成为帮助不同行业企业提升客户满意度、优化运营成本的关键利器。

在打造智能客服服务的过程中，依托亚马逊云科技的云上基础设施，海外智能客服产品能够基于 LLM 通过自然语言识别客户的对话意图，同时自主识别不同任务，通过分别调用后端的不同流程，让融入 Agentic AI 的智能客服产品能提供全自动、高精度、个性化的体验交互。

此外，智齿科技还借助 Amazon Native Solution 实现了不同工作负载的自动扩容功能。同时，采用亚马逊云科技提供的托管服务，实现了全程数据加密，降低安全风险，让数据可靠性提高了一倍。



亚马逊云科技提供的 Amazon Bedrock 等生成式 AI 服务，帮助我们突破了产品海外适配过程中多语言复杂场景交互等难题，重塑了智能客服产品，能让智能客服在与海外客户即时交流中提供快速、高效、准确的 AI 服务。

——段鑫龙  
资深行业咨询专家

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 业务成果 基于 Agentic AI 让大模型幻觉问题减少 90% 大幅提升客户体验

智齿科技依托亚马逊云科技的 Amazon Bedrock 构建的海外智能客服，不仅让智能客服的回复更接近人工客服，还减少了 90% 的大模型幻觉问题。应用 Agentic AI 后，智能客服回复的准确率超过 87%，人工介入降低 42%，持续优化并大幅提升客户体验。

### ◆ 提升互动效率和问题解决率

通过抽象意图识别、智能变量信息识别，对访客发出的图片、视频等多模态内容更好地理解用户问题，借助矛盾澄清、信息追问等方式提升互动效率，最终提升问题解决率。

### ◆ 实时响应，提升客户满意度

在咨询中，越来越多用户渴望获得情感共鸣与个性化互动。依托 Agentic AI 技术，智齿科技的海外智能客服能以贴近品牌情感的虚拟人设出场，可结合用户的行为数据与标签，实现实时响应，并给予深度陪伴，从而有效提升了用户满意度。

通过此次与亚马逊云科技合作，进一步强化了智齿科技的技术决策和开发团队的技术氛围与技术文化：

### ◆ 技术决策者

优先推动 IaC 模板化、零信任架构，结合 Well-Architected 框架定期评估架构风险，确保技术投资与业务目标对齐。

### ◆ 开发团队

采用 GitOps 和 CI/CD 流水线提升交付效率，通过 CloudWatch 实现全链路监控，同时灵活运用 Spot 实例和节省计划降低成本。

随着双方合作的持续深入，2022 年 11 月智齿科技加入了亚马逊云科技合作伙伴网络，成为了 ISV 合作伙伴，随后将其产品上架亚马逊云科技中国区和海外区，以扩大客户范围、加快交易达成。通过亚马逊云科技合作伙伴网络提供的培训、赋能计划、专业化认证和资源等，智齿科技不断深化和多样化产品，为客户提供差异化价值。

## 未来展望

未来，智齿科技将继续深化与亚马逊云科技的合作，以 AI 引擎为核心，同时驱动 AI Agent、AI Copilot、AI Insight 三大模块联合应用，打造生成式 AI 驱动的全球客户联络中心，帮助全球更多企业打通服务、运营、洞察管理的闭环。

### 关于智齿科技

智齿科技，是一家面向全球企业，提供基于「客户联络中心」场景的一体化解决方案，包括公域 + 私域、营销 + 服务、软件 + BPO 的三维一体化。智齿科技不断整合前沿的人工智能及大数据技术，已构建形成呼叫中心、机器人「在线 + 语音」、人工在线客服、工单、智齿智客 SCRM、WhatsApp SCRM 等多个软件产品，并提供 BPO 服务管理外包。2014 年至今，智齿科技已经发展超过 2000 人的规模，覆盖中国、东南亚、北美等区域。

智齿科技

## 传统 CRM 工具变身“智能业务伙伴” 销售易这样做

北京仁科互动网络技术有限公司旗下的销售易主要为企业提供基于云的、智能化的、一体化的 CRM 系统。它不仅仅是一个管理客户资料的软件，更是一个以客户为中心的商业运营系统，旨在通过数字化技术，重塑企业的营销、销售和服务流程，最终驱动业务增长。

目前，销售易采用的亚马逊云科技主要服务包括：Amazon Bedrock（仅海外业务使用）、Amazon SageMaker、Amazon Lambda 等服务。



## 机会 为传统 CRM 系统注入 Agentic AI 技术打通各环节升级壁垒

销售易基于深度智能的技术底座，强大的体系协同能力，灵活可扩展的 PaaS 系统，使其持续获得国际权威认可。销售易致力于通过移动、社交、AI、大数据、物联网等新型互联网技术，打造符合数字化时代下企业需求的营销服全场景智能 CRM 产品，将企业同客户互动的全过程数字化、智能化，帮助企业转型为真正以客户为中心的数字化运营组织，实现业绩的可持续增长。但面对生成式 AI 的发展，销售易亟需基于 Agentic AI 等技术突破产品在销售、营销等场景中的技术瓶颈：

### ◆ 销售环节

数据录入工作量大，给团队带来巨大负担，同时也无法为企业提供深入的分析，只能做简单的现象描述。

### ◆ 营销环节

原有的线索评分规则僵化，难以实现对线索的动态洞察。

### ◆ 客户服务环节

面对大量的数据，客服人员在服务过程中，难以快速找到最新的数据。

在部署 Agentic AI 前，销售易的技术团队经过多重对比，最终选择亚马逊云科技，主要因为其能提供从底层算力到顶层 AI 模型的全套服务，且其全球基础设施能完美满足销售易国内外客户对数据合规性和低延迟的需求。



**业务成果** 融合 Agentic AI 为企业客户打造全新工作流程

销售易团队使用亚马逊云科技提供的 Agentic AI 等技术服务，获得的远不止 IT 资源的替代升级，最大收益在于获得了驱动产品核心智能化的“加速器”，使产品在激烈的业务竞争中脱颖而出。此外，销售易通过融入 Agentic AI，还在品牌信誉的增强、全球业务的拓展、组织效能的优化以及未来适应性构建等方面获得了多重无形收益。这些收益将销售易从一家传统的 CRM 软件公司，升级为一家基于云和 AI 的、具备全球服务能力的现代化智能服务商。从客户视角看，销售易融入了 Agentic AI 服务后，减轻了企业的工作负担，加快了决策速度，全面颠覆了企业原来使用 CRM 的体验：

- ◆ **体验更轻松**  
改变了原来手动录入的模式，减少了繁琐的重复性人工工作，实现了自动化业务流程。
- ◆ **决策更前瞻**  
现在系统可根据捕捉到的信息，进行深入的数据分析，将“被动响应”变为“主动预警”。
- ◆ **使用更简单**  
原来操作相对繁琐，融入 Agentic AI 后，用户通过“自然交互”这种更便捷的方式即可完成操作。

使用亚马逊云科技提供基础设施、Agentic AI 等技术服务从根本上重塑了销售易的产品能力、商业模式和业务格局。未来，面向海外业务的需求，销售易团队将与亚马逊云科技展开更深入的合作，通过 Amazon Bedrock 的大模型能力，进一步重塑 CRM 的边界，将传统流程工具升级为“AI 驱动的增长 Agent”。

**Neocrm 销售易**

» **关于销售易**

销售易隶属于北京仁科互动网络技术有限公司，致力于通过移动、社交、AI、大数据、物联网等新型互联网技术，打造符合数字化时代企业需求的营销、销售和服务全场景智能 CRM 产品，旨在为企业提供一套基于云的、智能化的、一体化的 CRM 系统，进而通过数字化技术，重塑企业的营销、销售与服务流程，最终驱动业务增长。

**解决方案** 从性能到安全亚马逊云科技全链路赋能销售易智能化升级

销售易团队通过整合亚马逊云科技的服务，针对海外业务需求，借助 Amazon Bedrock、Amazon SageMaker 打造的 NeoAI，不仅优化了软件的性能、还降低了硬件等成本投入，此外还进一步加强了数据的安全性，让销售易获得了巨大的性能提升：

- ◆ **敏捷创新**  
针对海外用户需求，依托 Amazon Bedrock、Amazon SageMaker 等服务，快速集成和实验最新 AI 模型，大幅缩短了新功能上市时间。

- ◆ **可靠性与可扩展性**  
亚马逊云科技提供的基础设施既能确保 NeoAI 能够随客户数量增长进行灵活扩展，还能保持 7x24 小时稳定运行。
- ◆ **成本效益**  
销售易采用亚马逊云科技 Serverless 架构后，只需按实际使用量付费，避免了 AI 项目前期巨大的硬件投入。
- ◆ **企业级安全**  
亚马逊云科技拥有完善的安全体系，帮助销售易构建了让中大型企业更放心、更可信的 AI 环境。

通过集成亚马逊云科技的多项服务，销售易将其 CRM 从被动的记录工具转变为主动的智能业务伙伴，为客户带来了工作流程的极致自动化与决策洞察的前瞻价值，构筑了坚实的竞争壁垒。

——涂康康  
北京仁科互动网络技术有限公司 生态总监

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

# 不止快，更能省！特赞科技运用 Agentic AI 破解商业分析 4 大痛点

特赞（上海）信息科技有限公司 Tezign（以下简称：特赞科技）是全球备受企业信赖的“内容+人工智能”独角兽，致力于通过生成式 AI 技术助力企业实现更高效、更优质的内容管理、积累、生产、分析洞察。

目前特赞科技采用的亚马逊科技主要服务包括：Amazon Bedrock、Amazon EKS、Amazon Aurora Global Database。（特赞科技仅海外业务使用 Amazon Bedrock）

6-9 个月

基于 Amazon Bedrock 上的先进模型服务于海外业务的产品研发周期缩短时长

10-12 分钟

即可基于 Agentic AI 交付高质量分析报告

## 机会 依托 Agentic AI 突破 4 大挑战为商业研究提供更多可能

特赞科技累计申请 160+AIGC 发明专利、50 项软件著作权，完成 5 项算法备案，是工信部国家级工业设计中心、上海市经信委高质量产业发展的人工智能专项企业，其科技成果被认定为“国内领先、多项指标国际先进、填补国内空白”。多次上榜 Gartner、Forrester、Bloomberg、Fast Company 等国际榜单。

多年来特赞科技一直致力于开发企业级内容人工智能系统（Enterprise AI+Content Platform），打造了一站式创意内容服务方案、企业数据管理等多款解决方案，可为企业更高效、优质的内容生产、管理、分发服务。

此外，2025 年 4 月，特赞科技发布了首个商业研究 Agent 框架 atypica.AI。该 Agent 框架使用了 Amazon Bedrock 上的先进模型作为底层大模型之一。该 Agent 上线后，特赞科技发现应用了 Agentic AI 后，有效解决了企业客户在进行深度商业研讨时面临的 4 大痛点：

- ◆ **时间成本高昂 (Time-Consuming)**  
一份高质量商业分析报告，如行业进入可行性分析或竞品深度洞察，通常需要分析师花费数天甚至数周的时间进行资料搜集、数据处理、信息整合和报告撰写。
- ◆ **人力成本高昂 (High Cost)**  
依赖资深商业分析师团队意味着高昂的人力成本，这使得深度研究成为一种奢侈品，很多有价值的探索性问题因为预算限制而无法被执行。
- ◆ **难以规模化 (Lack of Scalability)**  
人工研究的产出和质量与分析师的个人能力和精力强相关，很难在短时间内响应大量、并发的研究需求，也难以保证不同报告之间洞察力度的绝对一致性。
- ◆ **信息处理瓶颈 (Data Overload)**  
面对海量的非结构化数据（如行业新闻、财报、社交媒体讨论、用户评论等），分析师手动筛选和提炼有效信息变得越来越低效，容易遗漏关键信号。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

特赞科技选择与亚马逊科技合作，主要因为 Amazon Bedrock 上的先进模型是广泛可用的混合推理模型，并能在 Amazon Bedrock 中轻松切换快速响应与深度分析。从成本方面看，Amazon Bedrock 的先进模型推理成本可调，可根据需求控制模型在推理成本、速度与性能等方面的平衡。

## 解决方案 借助 Amazon Bedrock 构建高可靠高可用的 Agent 解决方案

特赞科技 atypica.AI 模型使用了 Amazon Bedrock 先进模型，并构建在基于亚马逊科技的现代化、高可用的云原生架构之上，无论在 AI 引擎、基础设施、容器化编排，还是数据库和安全性方面都获得了亚马逊科技提供的强大技术支持，有效保障了 Agent 运行的稳定性、连续性与安全性：

- ◆ **核心 AI 引擎**  
atypica.AI 的“大脑”主要通过调用 Amazon Bedrock 服务实现。Agent 利用 Amazon Bedrock 上的大模型强大的长文本理解和深度推理能力，可对搜集到的信息进行交叉分析、提炼观点、识别趋势并生成高质量分析内容，形成独到的商业洞察。
- ◆ **容器化与编排**  
特赞科技将整个应用都进行了容器化，并部署在 Amazon EKS (Amazon Elastic Kubernetes Service) 上，借助标准化的应用模板（如 Helm Charts），并通过 IaC 工具，实现了高效、自动化的 CI/CD 流程。
- ◆ **全球数据库架构**  
为了支撑 AI Agent 进行高效的全球信息检索，并为全球低延迟的服务，特赞科技采用了 Amazon Aurora Global Database。特赞科技依托这项亚马逊科技特有的功能，可部署一个跨越多个亚马逊科技区域的单一数据库。其中主区域主要处理写操作，而全球各地的只读副本能以低于 1 秒的延迟进行数据同步，以确保用户能快速访问最新的数据，实现近乎实时的全球洞察。
- ◆ **基础设施即代码 (IaC)**  
特赞科技全面拥抱 IaC 理念，使用 Pulumi 等工具来定义和管理所有的云资源。这使得特赞科技的整个架构（从 VPC 网络到 EKS 集群再到应用部署）都是可版本化、可审计和可复现的，极大提升了部署的一致性和可靠性。
- ◆ **安全与权限**  
特赞科技使用 IAM Roles for Service Accounts (IRSA)，为 Amazon EKS 中的每个服务 Pod 分配精细化的、临时的访问权限，避免了任何硬编码的密钥，确保可严格遵循最佳安全实践。

亚马逊科技通过其强大的 Amazon Bedrock 和 Amazon EKS 服务，让我们能专注于核心 AI 算法创新，将过去需要数周的深度商业研究，浓缩为 10 分钟的 AI 自主分析，为企业决策带来了前所未有的敏捷性与洞察力。

——丁鑫栋  
特赞科技 Tezign CTO  
atypica.AI 产品技术负责人

atypica.AI 自 2025 年 4 月发布以来，目前正在为快消、美妆、科技、咨询等多个行业的头部企业客户提供服务，客户在可行性测试、竞品分析、营销策略规划、消费者研究、新品共创等典型场景中使用该 Agent 时，只需提出商业研究问题，atypica.AI 就可以通过 10-20 分钟的“长推理”生成清晰完整有洞察的商业研究报告。

除了 atypica.AI，特赞科技还将 Agent 的理念应用于了内容资产管理和内容效果分析中：

- ◆ **内容资产管理 (DAM)**  
AI Agent 可以自动分析和标注上传的图片、视频素材，打上业务维度、品牌元素、场景情绪等多维度标签，极大提升了资产的检索和复用效率。
- ◆ **内容效果分析 (Content Hub)**  
Agent 可以持续追踪内容在不同渠道的效果数据，结合内容标签进行归因分析，并主动提出优化建议新的内容创意方向。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

业务成果

Agentic AI 重塑工作模式实现成本与性能的绝佳平衡

特赞科技与亚马逊科技合作后，依托 Amazon Bedrock 可将复杂的模型调用封装成了简单的 API，后端开发人员可以通过 Amazon SDK 轻松集成，**极大地降低了 AI 应用的开发门槛。**

**在自动化部署方面**，使用 Amazon EKS，通过 Pulumi 等 IaC 工具建立了一套成熟的 GitOps 流程。这使得特赞科技能够将整个云基础设施以代码的形式进行版本控制和管理。开发人员只需提交代码，CI/CD 流水线就会自动完成构建、测试和向 Kubernetes 集群的滚动更新。相比于手动配置或传统脚本，这套体系的效率和可靠性实现了数量级的提升。

**在敏捷性方面**，借助 Amazon Bedrock 发布新的模型或更新时，几乎可以在当天完成 atypica.AI 的适配和测试，这种“即插即用”的体验对于保持特赞科技在 AI 领域的竞争力至关重要。在合作过程中，特赞科技还获得了亚马逊科技专业团队提供的专业架构建议和最佳实践指导，这让其减少了不必要的试错过程。

特赞科技应用 Amazon Bedrock 等亚马逊科技的生成式 AI 的服务后，不仅加快了产品上市周期，优化了成本与性能，为未来的业务发展奠定了坚实的基础。

**加速产品上市**  
Amazon Bedrock 是特赞科技能够快速推出 atypica.AI 的关键催化剂，让其免去了自建和维护复杂的大模型推理基础设施的巨大投入，还可以直接、按需、可扩展地使用业界前沿的模型。这使得特赞科技的研发团队能将 100% 的精力聚焦于核心优势——Creative Reasoning Agent 框架和业务逻辑的开发上，从而将产品研发周期缩短了 6-9 个月。

**优化成本与性能**  
Amazon Bedrock 提供的多种模型选择和按需付费模式，使特赞科技能够为不同的研究任务匹配最合适的模型，实现了成本和性能的绝佳平衡。

**提升创新能力**  
借助 Amazon EKS 和 Amazon Bedrock 等托管服务，让特赞科技的工程师从繁重的底层运维工作中解放出来，进行更多前沿 AI 技术的实验和迭代。

此外，随着用户量的增长，基于 Amazon EKS 和 Amazon Bedrock 可以自动、无缝地扩展的架构，特赞科技能轻松应对流量“洪峰”。同时依托亚马逊科技的安全合规体系，特赞科技可以向企业客户确保其商业查询和数据的最高级别安全与保密。

关于特赞科技

特赞（上海）信息科技有限公司 Tezign（以下简称：特赞科技）是全球备受企业信赖的内容 + 人工智能独角兽，致力于通过生成式 AI 技术助力企业实现更高效、更优质的内容管理、积累、生产、分析洞察。已服务 200 多家国内外大型企业，覆盖快消、美妆、鞋服、医疗、工业制造等行业，为全球数百家企业搭建“内容 + 人工智能”基础设施。内容创作者社区聚集 10 万内容创作者，生产 15 万 + 内容资产，专有 AIGC 模型调用日均百万级，积累了 10 亿 + 企业 AIGC 的数据集规模。



跨境物流再提速，订单服务效率提升超 50%  
WallTech 借助 Agentic AI 实现业务增长

上海沃行信息技术有限公司 WallTech 于 2011 年创立，是业界知名的跨境物流 SaaS 云服务研发和运营服务商，目前用户数量超十万，覆盖全球 58 个国家。

WallTech 的海外业务基于亚马逊科技的 Amazon Bedrock，接入 Amazon Bedrock 上的先进模型 API，落地了智能国际物流助手、智能文档识别与结构化录入等解决方案，整体物流订单服务效率提升超 50%。同时基于 Amazon Q Business、Amazon Bedrock Agents 构建了智能客服与知识库问答，实现 7×24 小时信息响应。引入 Agentic AI 后 WallTech 整体开发周期缩短了三分之一，硬件成本节省约 20%，人工成本降低 50%。



\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机遇 自动化智能化水平亟待提升借 Agentic AI 补行业发展短板

WallTech 致力于通过数字化和 AI 技术重塑全球物流与供应链管理，让跨境贸易更高效、更智能。WallTech 提供的国际货代物流云服务 CargoWare 和跨境电商物流协同云服务 eTower，覆盖从销售、操作到财务和管理的全流程，广泛应用于国际货运代理、跨境电商物流、出海品牌企业国际物流管理的端到端全链路管理场景。

公司业务涉及全球，以欧美、东南亚市场为主逐渐辐射全球跨境电商物流领域，帮助客户实现智能化转型与增长。然而，伴随业务的快速发展，自动化与智能化水平不足、数据整合分析能力薄弱等痛点，正成为制约企业发展的关键瓶颈：

### 01. 公司内部挑战

客户数量高速增长带来大量定制化需求，导致开发周期延长、开发资源持续紧张。

### 02. 产品行业痛点

#### ◆ 流程高度依赖人工

大量重复、规则明确的任务需由员工手动处理，易出错且效率低。

#### ◆ 决策响应不及时

面对突发事件或需要实时响应的场景，传统流程响应慢，难以及时处理。

#### ◆ 信息利用率低

各类非结构化数据（如文档、邮件、工单等）难以高效识别和分析，知识难以沉淀复用。

经过综合评估，**亚马逊云科技凭借全球范围内广泛的基础设施、可靠的安全性和扩展性，以及丰富的工具和解决方案在众多选项中脱颖而出。**WallTech 选择与亚马逊云科技合作还考虑到其灵活的定价和优质的技术支持，满足多样化需求的同时确保合规性。

得益于亚马逊云科技全球广泛的分布，我们可以为各个国家和地区的畅享本地服务的速度和世界范围内的数据同步与互联互通。

——公剑  
上海沃行信息技术有限公司 CTO

## 解决方案 以 Agentic AI 破局业务增长诉求

通过与亚马逊云科技合作，WallTech 海外业务基于 Amazon Bedrock，并通过接入 Amazon Bedrock 上的先进模型 API，在 Agentic AI 应用方面，落地了智能国际物流助手、智能文档识别与结构化录入等核心解决方案。

### 01. 智能国际物流助手

结合 Amazon Bedrock 上的先进模型通用模型与 WallTech 在国际物流领域的强知识库，打造了覆盖在线订舱、智能询价、运价查询、订单追踪等功能的物流 AI 助手，帮助客户实现从询价到下单、履约的端到端自动化，整体物流订单服务效率提升**超过 50%**。

### 02. 智能文档识别与结构化录入

在单证处理过程中，WallTech 将传统 OCR 识别与生成式 AI 自然语言理解结合，针对提单、发票、装箱单、报关单等非结构化文件，自动识别并录入至系统结构化表单，识别准确率由原来的 85% 提升到 **98%**，减少人工校验工作量，提升操作人员效率。

为更好地解决响应时间不及时、信息利用率低等问题，WallTech 还借助亚马逊云科技的 Agentic AI 服务打造了创新解决方案，不仅增加了企业信息的利用率，实现了 **7×24 小时** 响应，还实现了 IT 运维与物流订单业务的自动化，提升了企业的业务效率：

#### ◆ 智能客服与知识库问答

通过 Amazon Q Business 和 Amazon Bedrock Agents，自动响应用户咨询、检索和总结内部知识库信息，实现 7×24 小时信息响应。

#### ◆ IT 运维与自动化响应

利用 Agentic AI 完成故障检测、事件溯源与自动修复，自动提交变更、分析日志，提升系统可用性。

#### ◆ 物流订单业务流程自动化

Agentic AI 集成内外部数据系统，自动分派物流订单、订单履约协同，并可实现监控与溯源。

在解决业务需求和挑战方面，亚马逊云科技为 WallTech 带来了显著的价值和改变。**直观的体验就是产品上线更快、功能更新更频繁，AI 智能服务的可用性和稳定性也更有保障。**

## 业务成果 依托 Agentic AI 实现运维安全双提升 + 人力成本锐减

通过与亚马逊云科技合作，并引入了 Agentic AI 后，WallTech 在 AI 应用开发、硬件成本控制等方面获得了多重收益：

- 1 开发周期缩短了三分之一，AI 应用可以更快上线，及时满足客户需求。
- 2 硬件成本节省约 20%，无需投入自建算力资源。
- 3 依托亚马逊云科技提供的自动化运维与安全管理能力，即使服务器和调用量增加，也无需增加运维人员数量，有效降低超 50% 的人力成本。

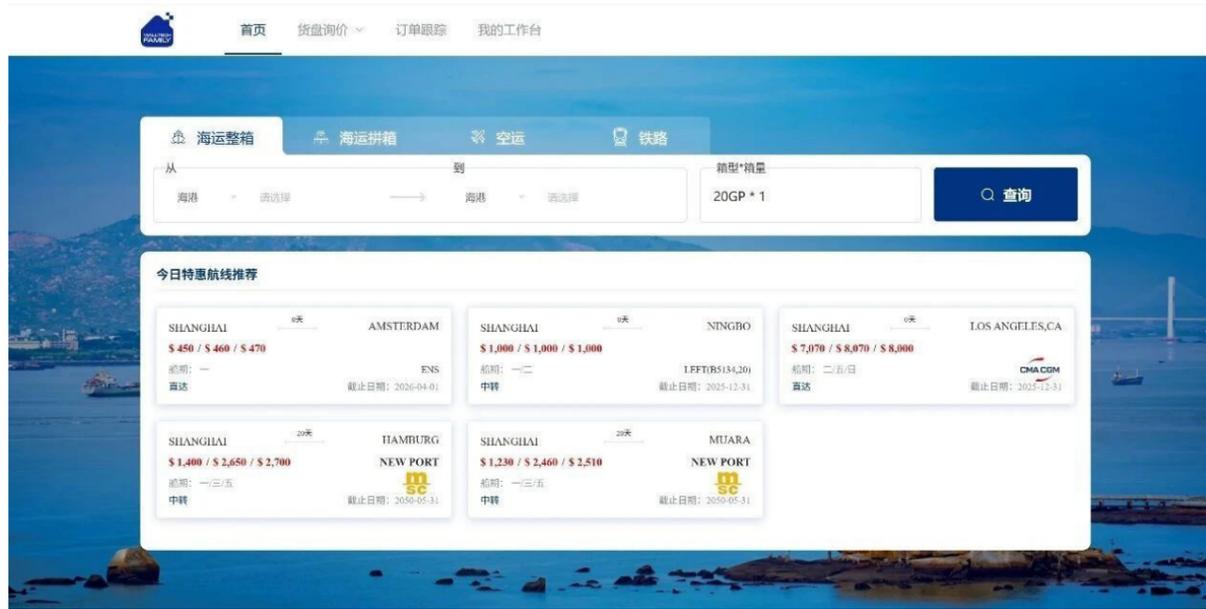
除了这些可量化的收益，基于 Amazon Q 等解决方案构建的 Agentic AI 应用，让 WallTech 无论在业务流程创新与自动化方面，还是在数据与系统整合方面，都获得巨大的改善：

**在业务流程创新与自动化方面**，利用亚马逊云科技的 Agentic AI 与自动化工具，大量减少了人工介入，实现了业务流程的全流程自动化，包括智能客服、工单自动分派、运维故障响应等，使响应速度和准确性提升。

**在数据与系统整合方面**，亚马逊云科技支持多源异构数据的高效整合和自动化处理，打破了“数据孤岛”，让跨系统、跨部门数据流转和分析更加顺畅，提升了数据驱动决策能力。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。



## 基于亚马逊云科技，金蝶利用生成式 AI 优化 ERP 系统智能提单和指标分析流程，重塑用户体验

金蝶国际软件集团有限公司（以下简称金蝶）是全球知名的企业管理云 SaaS 公司，成立于 1993 年，已为世界范围内超过 740 万家企业、政府组织提供企业管理云产品及服务。应用亚马逊云科技的生成式 AI 服务，金蝶面向海外业务构建智能提单系统和智能指标分析系统，2 个月完成生成式 AI 功能落地，简单场景中的自然语言问询准确率超过 90%，借助 Agents，复杂场景中的自然语言查询答复准确率可达 100%。

金蝶国际软件集团有限公司是全球知名的企业管理云 SaaS 公司，为进一步提升海外用户的 ERP 使用体验，金蝶采用亚马逊云科技的生成式 AI 服务，利用 Agents for Amazon Bedrock 构建智能提单系统和智能指标分析系统，2 个月完成生成式 AI 功能落地，以自然语言交互简化用户提单，优化智能指标查询业务流程。

**2 个月**

2 个月完成生成式 AI 在智能提单和智能指标分析场景的落地

**准确率超过 90%**

简单场景中的自然语言查询，第一轮答复准确率超过 90%

**准确率高达 100%**

借助 Agents，复杂场景中的自然语言查询答复准确率可达 100%

目前，金蝶在服务其海外客户时采用的亚马逊云科技产品包括：Agents for Amazon Bedrock、Amazon Bedrock 中 Anthropic 的 Amazon Bedrock 中的 Cohere Embed - Multilingual、Amazon OpenSearch Service、Amazon Lambda 等。

## 未来展望

为了可以不断创新并扩大客户覆盖范围，以推动盈利增长，Walltech 加入了亚马逊云科技合作伙伴网络 (APN)，成为了 ISV 合作伙伴。通过 APN 能力认证体系赋能，Walltech 成功认证了 Amazon Linux, Amazon RDS, Amazon Privatelink service ready，充分体现了 Walltech 与亚马逊云科技服务的深度集成和差异化能力。借助亚马逊云科技 APN 合作伙伴网络，双方将继续在制造业，零售快销，物流服务等行业开展联合销售合作，共同拓展市场。

未来，WallTech 希望通过亚马逊云科技强大 Agentic AI 技术以及云基础和服务，连接跨境物流供应链所有环节，构建具有韧性、高协同性、可持续性的物流控制塔，强化数智化系统和应用，让数据产生更大价值。与此同时，Walltech 也会和亚马逊云科技一起为客户提供更多个性化的、创新的解决方案，创造更大的客户价值。

### 关于 WallTech

WallTech 致力重塑国际物流软件云服务，通过十余年行业积累构建了以国际货代物流云服务 CargoWare 和跨境电商物流协同云服务 eTower 为基础的云服务产品矩阵，将先进的技术与跨境物流行业丰富的经验紧密融合，赋能传统中国国际物流企业数字化改造升级，有效提高业务运行效率并最大化降低企业经营成本。



\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 机会 生成式 AI 重塑 ERP 软件领域，自然语言对话改善用户体验

生成式 AI 问世以来，给几乎所有行业都带来巨大影响，ERP（Enterprise Resource Planning，企业资源计划）领域也在其中，业界普遍期待借助生成式 AI 来革新 ERP 的用户体验。

◆ 设想一位刚入职不久的财务专员，要为部门新采购的办公设备进行报销，由于设备类型、采购方式、供应商条款、单据种类等不同，可能涉及系统中数十个单据，以及数百个页签，面对 ERP 系统庞杂的界面和功能，财务专员一时无从下手，难道只能默默打开百科全书式的 ERP 系统用户手册一步步按章操作？

再设想一位企业高管，在外与客户进行商务沟通时，想要即时了解上个季度某子公司某项产品的研发投入和同比增长情况，而系统并无现成数据参考，是不是希望有一个自助式的查询工具，能够便捷高效、随时随地了解具体各类业财数据？

在生成式 AI 时代，金蝶提出了“ALL IN AI”的战略口号，认为生成式 AI 将在 ERP 领域重塑用户体验、重塑企业的业务流程和重塑企业管理者的决策。金蝶对其海外业务 ERP 典型使用场景经过诊断、评估后，决定先聚焦智能提单和智能指标分析两大场景。

要在金蝶的业务场景中应用生成式 AI 精准匹配客户需求，依靠基础模型本身远远不够，还需要集成公司内部系统、API 以及外部系统的数据。Agents（代理）技术出现后，可以从专有数据源检索所需信息并提供准确且相关的响应，精准捕捉用户意图，十分适合上述两大场景。

◆ 然而在实践过程中，金蝶遇到了两个主要挑战：

1. Agents 在企业管理软件中的成功用例鲜少，需要创新和探索；
2. AI 在企业生产场景中的准确率难以预测，需要实际尝试。

亚马逊云科技与金蝶的合作自 2015 年就已经开始，金蝶的企业管理 SaaS 金蝶云·星瀚、金蝶云·星空和金蝶云·星辰、EAS Cloud 等多款产品均已在亚马逊云科技上线；此次生成式 AI 领域的合作，意味着双方合作进入到 2.0 阶段，金蝶的海外业务出现生成式 AI 应用需求后，亚马逊云科技团队立即与金蝶接洽快速进入产品的 POC (Proof of Concept, 概念验证) 阶段。

金蝶集团智能财务产品经理方首宇对亚马逊云科技的产品与服务积极认可：“在调研市场上一系列产品后，我们发现亚马逊云科技的生成式 AI 服务清晰易懂，Agents 开发工具成熟、易操作，服务团队专业高效。基于这些优势，我们决定与亚马逊云科技展开合作，在生成式 AI 领域共创。”

亚马逊云科技提供的 Amazon Bedrock 等生成式 AI 开发工具调用方便，产品迭代速度快，同时提供了构建生成式 AI 应用所需的一系列功能，帮助客户简化开发，同时维护客户数据隐私和安全；在开发过程中，还能提供及时的检测跟踪、快速的调试功能，让开发更高效、智能，使得金蝶面向海外业务的生成式 AI 应用迅速落地。

——方首宇  
金蝶集团智能财务产品经理

## 解决方案

## 基于 Agents for Amazon Bedrock 高效落地智能提单、智能指标分析两大场景

◆ Agents for Amazon Bedrock 助力实现智能提单、智能指标分析两大场景从 0 到 1 快速构建

亚马逊云科技 Amazon Bedrock，拥有丰富的模型选择、差异化模型定制、应用集成等特性，让金蝶在其海外业务中能够轻松访问各种主流模型，同时满足了企业级安全性需求。金蝶还在其海外业务中应用亚马逊云科技 Agents for Amazon Bedrock，基于金蝶的 API 和知识库，快速构建专用的智能体应用。Agents for Amazon Bedrock 能够与无服务器服务 Amazon Lambda 和多种向量数据库自动集成，帮助金蝶快速实现生成式 AI 在 ERP 软件应用中的落地。

**场景一：智能提单场景，让用户从复杂系统操作转变为便捷的自然语言对话交互**

金蝶利用亚马逊云科技 Agents for Amazon Bedrock 对其面向海外客户的提单系统进行优化，构建智能提单系统，将原有 ERP 中基于 GUI (Graphical User Interface, 图形用户界面) 的表单界面提升为 LUI (Language User Interface, 自然语言用户界面)。用户通过自然语言描述后，生成式 AI 能够帮助分析和推导其意图，并给出操作提示，用户按照提示即可简便发起报销、借款、出差等各类单据提交。方首宇说到：“在基于亚马逊云科技面向海外客户构建智能提单系统后，用户只需告诉系统自己的职位以及想要完成的任务，系统就能根据相应规则作出推理，给出相应引导，并反馈给用户。”

**场景二：智能指标分析场景，企业管理者无需在繁杂的 BI 报表中寻找数据，而是通过即时对话查询指标和获取精准结果**

除智能提单外，Agents 允许用户进行实时的补充和澄清，还可应用于智能指标分析场景。金蝶应用亚马逊云科技 Agents for Amazon Bedrock 面向海外客户构建智能指标分析系统，企业管理者可与线上专属 AI 助理进行自然语言对话，获得即时、

精准的指标数据查询，避免了在繁杂的 BI 报表中寻找数据的烦恼。在海外业务场景中，与用户确认意图后，金蝶 AI 助理自主拆解任务，并学习企业的知识库，匹配推荐相关度最高的指标。企业管理者就像拥有了 7x24 小时的 1v1 专属财务助手，轻松查询各类财务指标对比分析，重塑了企业的分析和决策流程。

◆ 丰富且强大的生成式 AI 工具提升 ERP 系统智能化水平

除了 Agents for Amazon Bedrock，金蝶在其海外业务中还采用了一系列亚马逊云科技的其它服务。智能指标分析场景涉及超过 200 个指标大类、超过 4000 个指标子类，数据组成异常复杂，为此，金蝶通过 Amazon Bedrock 调用 Cohere Embed-Multilingual 向量模型，以及采用 Amazon OpenSearch Service 向量库完成指标的向量化、存储和语义查询，通过语义匹配准确获取客户需要查询的指标，快速、精确地返回查询结果。

面向海外客户，金蝶利用 Amazon Bedrock 上的先进模型完成自然语言理解、任务拆解和参数提取。通过在 Agent 工具中自建基于 Amazon Bedrock 上先进模型的分类器和命名实体识别 (NER) 提取器，准确获取用户意图，以及历史会话中的参数信息。测试数据表明，在 Amazon Bedrock 先进模型中提供的中文理解和信息提取能力准确率达 97%。同时，由于历史信息可通过返回值传递，直接自动填充相关参数，避免信息重复收集，使得用户交互方式更加简便。

在面向海外客户的智能指标分析场景中，为了实现更精准的运营情况查询，金蝶在 Agents for Amazon Bedrock 之上又引入了 Text2SQL，该技术能够通过自然语言处理，将用户的文字查询转化为数据库中的结构化 SQL 查询语句，从而更灵活地获得所需数据。通过这种方式，金蝶进一步提升了查询的灵活性，确保用户能够迅速且准确地获得所需的指标和数据。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

◆ 亚马逊科技服务团队全方位助力金蝶生成式 AI 功能的快速落地

亚马逊科技的服务团队为金蝶提供了全方位的专业支持。在需求讨论时，亚马逊科技的服务团队与金蝶 AI 团队进行交流，提供试用早期选型；在技术难题出现时，亚马逊科技的技术专家提供即时响应和支持；在项目推进过程中，亚马逊科技为金蝶提供一系列技术培训和 Workshop 研讨等活动，帮助金蝶快速应用前沿技术。方首宇评价道：“亚马逊科技的技术团队拥有丰富的生成式 AI 应用经验和专业知识，从产品落地和技术攻关的角度来看，我们对亚马逊科技的支持力度和专业性十分满意。”

业务成果

2 个月完成生成式 AI 功能落地，借助 Agents 查询答复准确率可达 100

◆ 2 个月完成生成式 AI 应用高效部署

2024 年 1 月，经过系统评估，金蝶在一周内就完成了面向海外客户的智能提单系统的 POC 验证，并开始面向企业客户推广，获得客户认可。方首宇表示：“截至 2023 年 12 月底，业内开放的 Agents 技术较为稀缺，亚马逊科技的 Agents for Amazon Bedrock 成熟且易于上手。试用亚马逊科技的开发工具和平台后，我们迅速实现了 Agents 的早期应用场景。

到了 2024 年 5 月，市场上虽然出现了众多厂商的 Agents 框架，但评估发现，亚马逊科技的 Agents for Amazon Bedrock 在成熟度和易用性方面依然表现优异。亚马逊科技提供的生成式 AI 开发工具调用方便，产品迭代速度快，前端框架和页面都很清晰；在开发过程中，提供及时的检测跟踪、快速的调试功能，让开发更高效、智能，使得金蝶面向海外业务的生成式 AI 应用迅速落地。”利用亚马逊科技的生成式 AI 产品和服务，金蝶 2 个月成功完成了生成式 AI 功能的落地。

◆ 简化交互，提升用户 ERP 系统提单体验

面向海外客户的智能提单系统落地后，想要发起 ERP 提单流程的用户，再也不用在 ERP 系统上百

个应用、每个应用数百个页签的复杂界面无所适从，仅仅通过自然语言对话描述，就能自动发起报销、借款、出差等各类单据提交，显著简化了用户与系统的交互方式，极大提升了用户提单体验。

◆ 优化流程，借助 Agents 技术智能指标查询准确率可达 100%

采用生成式 AI 技术后，通过面向海外客户的智能指标分析系统，企业管理者能够随时随地查询企业财务状况和数据，且查询的准确率显著提升。在金蝶的实践中，针对简单场景的查询，第一轮答复准确率超过 90%；针对复杂场景的查询，例如带有指标趋势计算的聚合逻辑，第一轮答复准确率为 70% ~ 80%；借助 Agents 模式与用户进行对话后，如补充具体数据、来源细节后，第二轮答复准确率可达 100%。通过将生成式 AI 技术融入了企业的业务流程，优化了企业使用 ERP 系统的方式和过程，进一步提升企业运营效率。

未来，金蝶将与亚马逊科技紧密合作，进一步探索将生成式 AI 技术应用于企业管理云产品及服务。双方将共同致力于打造便捷、高效、实用的解决方案，提升用户体验，持续推动生成式 AI 应用于企业转型和创新浪潮。



关于金蝶

金蝶是全球知名的企业管理云 SaaS 公司，成立于 1993 年，致力成为“最值得托付的企业服务平台”，已为世界范围内超过 740 万家企业、政府组织提供企业管理云产品及服务。

“可控”更“懂你”  
携手乐鑫解锁智能硬件出海新密钥

在智能家居进入“场景化 + 智能化”的关键期，亚马逊科技合作伙伴乐鑫科技 (Espressif) 与亚马逊科技合作，面向出海智能品牌与开发者，打造了完整的 Agentic AI 软硬件开发套件，该套件结合乐鑫 ESP RainMaker，乐鑫硬件与 Amazon Bedrock (仅服务于海外业务) 的能力，提供了端到端的联合解决方案。

ESP RainMaker 是乐鑫基于亚马逊科技 Serverless 架构和云原生服务打造的 AIoT 云系统，它大大降低了物联网设备的开发和维护成本并提供了良好的用户体验，为客户云私有化部署提供一站式的 AIoT 解决方案。Agentic AI 软硬件开发套件，是基于 ESP RainMaker 的创新升级。

许多智能硬件品牌希望借助 AI 增强用户交互体验，让设备不仅“可控”，更“懂你”。然而在实际应用中，他们普遍面临三大挑战：研发与运营成本高昂、对用户隐私和数据安全的担忧，以及部署架构和集成流程的复杂性。尤其是在跨设备场景，如照明联动、情境感知和个性化能耗优化中，实现设备端、云端与模型推理的完整闭环，对大多数品牌而言仍是一个高门槛任务。借助乐鑫与亚马逊的联合解决方案，品牌客户能够深入探索智能化场景，**利用生成式 AI 解决终端智能设备语音接入复杂、智能化能力不足的挑战，从而为智能设备和应用带来更加自然、高效的 AI 体验。**



\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

## 硬件 Agent (智能体) 如何听懂、看懂并感知世界

Agentic AI的发展为提升硬件的智能提供了无限可能,许多品牌商会选择乐鑫的多种 SoC 包括 ESP32-S3、ESP32-P4 等进行智能设备的开发。**联合解决方案正是将 Agentic AI 能力嵌入到端侧硬件系统中,帮助品牌商快速开发智能产品。**其技术亮点包括:支持搭载 ESP-Brookesia 框架实现整体 UI 的构建与渲染;整合 Amazon Nova Sonic 模型与乐鑫全新音视频框架 ESP-GMF,集成多项端侧优化功能;支持全双工语音交互、多模态识别与 Agent 控制;支持离线语音唤醒与声源定位结合电机控制,可实现 180°范围内的方向跟踪。

基于此,乐鑫推出了新一代 AI 萌宠——EchoEar (喵伴)。融合 ESP RainMaker 与 Amazon Nova Sonic 语言模型,提供了 AI 萌宠的新范式,EchoEar 具备智

能对讲能力,可主动识别用户的意图与情绪变化,实现更具陪伴感的个性化体验。设备交互更自然、更具沉浸感,可作为智能家居系统的本地中枢,远程控制家庭设备。



## 在亚马逊科技上提供端到端的 Agentic AI + IoT 解决方案

乐鑫携手亚马逊科技推出的 Agentic AI 套件,涵盖乐鑫端侧 AI 硬件开发套件、ESP RainMaker AIoT 云系统,以及 Amazon Bedrock 服务。在云端,**通过将 Amazon Bedrock 和 ESP RainMaker AIoT 云系统结合,提供了从基础的设备互联到高阶 Agent 的应用和管理等一系列能力。**包括,以 Amazon Nova Sonic speech-to-speech 模型为基础的语音交互;标准的

设备侧接口、音色配置、提示词和角色管理功能;扩展多种可能行的 MCP server 的接入支持;通过“RainMaker”的融合集成,实现复杂的设备控制等。品牌商还可以二次开发来自定义软硬件功能,构建知识库或自定义 Agent。加速个性化的智能产品创新落地。



## 从概念到量产 帮助智能硬件开发按下“快进键”

借助乐鑫提供的开发板及整合亚马逊科技生成式 AI 技术的 Agentic AI 软硬件开发套件和整体解决方案,智能硬件开发者能够轻松获得强大的可管理性、安全性、隐私与合规保障以及高度可定制的能力支持。**开发者可更专注于产品场景与外观设计,无需深入底层技术研发,从而缩短研发周期与上市时间,更快的将自己的产品推出。**这一合作旨在为品牌客户打造从试验到量产的“绿色通道”

## 安全与合规 为智能硬件企业出海保驾护航

乐鑫支持客户私有化部署 ESP RainMaker 以及 Agentic AI 软硬件开发套件和整体解决方案,**结合亚马逊科技全球分布的云基础设施和安全合规的托管服务,能够满足智能硬件出海客户严苛的隐私保护、合规要求和定制化需求。**此外,方案支持个性化定制,可扩展到不同的垂直领域业务场景。

展望未来,乐鑫将持续深化与亚马逊科技的合作,融合自身在芯片、模组、开发板及 SDK 等硬件领域的技术优势,与亚马逊科技的云端创新能力特别是 Agentic AI 能力结合,共同打造更易用、更丰富的解决方案。双方将致力于推动跨行业的深度融合与协同创新,让用户切实感受到人工智能与物联网融合所带来的智能便捷与美好生活。



### 关于乐鑫

乐鑫信息科技 (688018.SH) 率先开创全栈 AIoT 云系统,驱动新一代智能、低功耗物联网与边缘 AI 设备的持续演进。其以开源技术为核心,赋能从消费电子到工业物联网 (IIoT) 等领域构建可扩展、安全且高性价比的智能应用,正在重塑各行业的技术格局。通过在 SoC 中深度融合计算与连接能力,乐鑫重构物联网体系,推动智能设备大规模普及。作为全球半导体创新领导者,乐鑫提供集 AI、自研操作系统、先进安全机制与云端连接能力于一体的高性能 RISC-V SoC 与整体解决方案,持续引领智能互联系统的未来发展。

### 关于 EchoEar (喵伴)

EchoEar (喵伴) 主控采用乐鑫 ESP32-S3-WROOM-1-N32R8 模组,支持 2.4 GHz Wi-Fi 和 Bluetooth 5 (LE) 无线连接。存储方面,整机具备 8 MB PSRAM、32 MB Flash 的存储空间,同时还配备了一个可支持高达 32 GB 的 microSD 卡插槽,满足语音交互和多媒体处理需求。配备 1.85 英寸圆形触摸屏 (360×360 分辨率) 与 ESP32-S3 原生触摸传感器,提供直观丰富的交互体验。音频方面,EchoEar 内置 2W 扬声器和双麦克风阵列,支持本地语音唤醒和声源定位。电源系统兼容 5V DC 和 3.7V 700mAh 锂电池供电。此外,还集成一个 USB-C 接口支持供电与编程下载,同时预留 Pogopin 接口方便功能拓展。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营,具体信息以中国区域官网为准。

\* 前述特定亚马逊科技生成式人工智能相关的服务目前在亚马逊科技海外区域可用。亚马逊科技中国区域相关云服务由西云数据和光环新网运营,具体信息以中国区域官网为准。

## Agent Facing Analytics, SelectDB 打造智能化实时数据分析

北京飞轮数据科技有限公司（以下简称“飞轮科技”），是一家基于开源分析型数据库 Apache Doris 的商业化公司，由 Apache Doris 原创团队于 2022 年 1 月创建，主要面向全球提供实时数据仓库的产品与解决方案 SelectDB，以满足各类场景的实时数据分析需求。

其产品在 2022 年 10 月登顶全球分析型数据库测评榜单 ClickBench，在多种场景下，性能全球排名第一。当前飞轮科技的业务主要以国内中大型企业为核心，未来三年，飞轮科技将聚焦跨境电商、新能源、跨境金融等出海企业的实时数据需求，加速拓展东南亚、北美等海外市场。

目前，飞轮科技采用的亚马逊云科技主要服务包括：Amazon SageMaker、Amazon EC2、Amazon S3、Amazon MSK、Amazon Lambda 等服务。



数据分析周期缩短 90%



数据分析和创新速度提升 10 倍



数据分析不再局限于办公场所，更多移动场景和业务人员参与

### 机会 将 Agentic AI 与数据库深度融合，为客户打造智能的数据分析服务

传统数据分析如 BI 依赖数据分析师和工程师开发 SQL 来完成，一个典型的数据分析需求要经过需求沟通、研发排期、开发测试、上线应用等多个步骤，通常面临如下一些挑战：

#### ◆ 时间和人力成本高、效率低

期间需要业务、开发、测试、运维等多个团队协作迭代几轮才能完成，时间周期往往需要一周甚至一个月。

#### ◆ 数据分析的场景受限

由于需要多团队配合，大多数的数据分析局限在办公室环境中，不能充分利用移动互联网的优势满足随时随地的数据分析需求。

#### ◆ 数据分析和创新想法被抑制

由于数据分析的成本高、效率低，很多有价值的数据分析受限于时间和资源的限制不得不延期甚至搁置，好的创新想法被隐形抑制，导致企业创新速度降低。

为解决上述问题，飞轮科技选择与亚马逊云科技合作，主要因为亚马逊云科技在全球化能力、AI 工具链的深度适配和合作伙伴资源等方面，具有显著优势。依托亚马逊云科技在人工智能和大模型领域成熟的工具，飞轮科技快速将 Agentic AI 深度融入 SelectDB 产品全链路，实现了“数据存储 - 查询 - 分析 - 洞察”全流程智能化。

### 解决方案

### Agent Facing Analytics, 全面提高数据分析的效率

双方达成合作后，飞轮科技构建面向 AI Agent 的实时数据分析：Agent Facing Analytics，打破了原来的高技术门槛、低建模效率等技术瓶颈，重塑了数据分析的服务模式。

#### ◆ ChatBI

用户可通过自然语言（如“查询过去 7 天华南地区女装类目的订单量及客单价 Top5”）发起请求，AI Agent 自动解析意图、生成最优 SQL、执行查询并返回可视化结果，无需人工编码，查询效率提升 80%，非技术人员数据获取周期从“周级”缩至“分钟级”。

#### ◆ Agent 自动化数据分析

用户发起数据分析需求，AI Agent 自动识别数据的分析目标、拆解出子任务和分析步骤，向 SelectDB 数据库发起查询请求，以图标等方式向用户展示结果。

在这些新的数据分析模式中，亚马逊云科技提供了开箱即用的大模型调用服务、Agent 开发平台，可以很便捷高效地构建各种数据分析 Agent。而大量 AI Agent 发起数据库查询请求，也给数据库带来更大的挑战：

#### ◆ 数据实时性要求更高

Agent 能自动快速发起高频的查询，以往人几小时几天一次的查询变成几分钟甚至几秒钟一次，有些 Agent 需要基于最新数据做即时决策，比如风控 Agent 需要在秒级时间内分析用户实时交易行为、

历史信用记录、当前账户状态等信息，判断是否存在欺诈风险。因此需要最新的数据才有意义，数据的延迟要求从小时和天级提高到秒级。

#### ◆ 查询响应要求更快

Agent 需要快速完成多步任务，比人对查询延迟的忍耐度更低，通常需要查询延迟在秒级甚至亚秒级。

#### ◆ 查询并发更高

人类数据分析的并发往往收到使用人数的限制，而 AI Agent 的数量则远超人类分析师的数量，以电商平台场景为例：客服 Agent 处理客户咨询，推荐 Agent 生成个性化推荐，库存 Agent 监控库存状态，价格 Agent 动态调整价格策略，营销 Agent 优化广告投放，风控 Agent 实时检测异常交易。每个 Agent 都持续查询数据库，查询频率和复杂度远超传统应用，因此查询并发会激增到每秒成千上万次请求。

#### ◆ 查询未经过细致调优

AI 生成的 SQL 未经过人工分析和优化，可能存在不够高效的问题，因此需要数据库查询优化器能够自动优化改写，提升查询效率避免资源浪费。

#### ◆ 查询的数据更加广泛

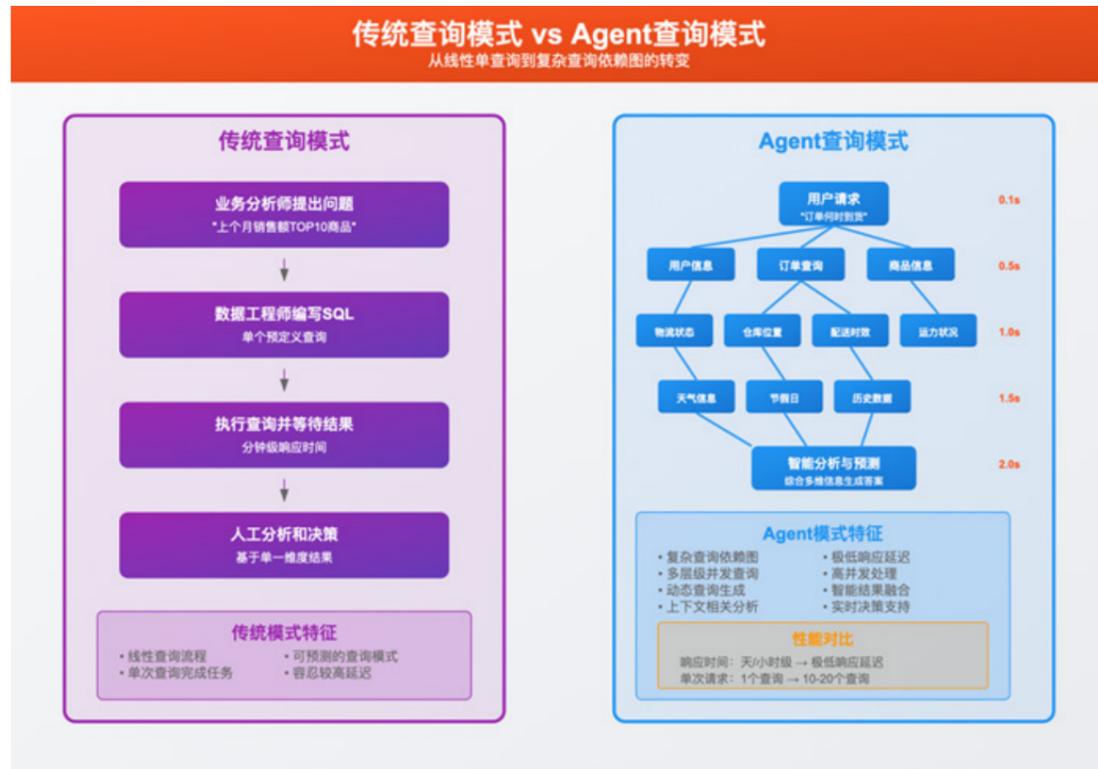
不再局限于数据库本身的数据，还需要数据湖中的跨业务数据、企业知识库中的非结构化数据等。

亚马逊云科技不仅为 SelectDB 提供了稳定、高效的全球化云基础设施，更通过成熟的 AI 工具链与出海生态，加速了我们“云原生转型”与“Agentic AI 落地”，提升了海外业务扩张效率、降低了研发成本，成为我们从“国内领先”迈向“全球布局”的核心支撑。

——肖康  
北京飞轮数据科技有限公司联合创始人 & 产品 VP

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。



## 业务成果 借力 Agentic AI 提升数据分析能力，加速业务创新

飞轮科技借助亚马逊云科技提供的一系列服务，打造的 Agent Facing Analytics 能力，为企业数据分析带来智能化和高效率。

同时，通过与亚马逊云科技合作，飞轮科技在构建海外服务的过程中还有机会优先应用 Amazon Bedrock、Amazon SageMaker Studio Lab 等创新服务，以进一步增强其 Agentic AI 方案的竞争优势。未来，飞轮科技将持续深化与亚马逊云科技的合作，重点推进三大计划：

- ◆ **数据分析成本大幅降低，效率大幅提升**  
常见的数据分析需求可以通过自然语言对话和多轮迭代完成，数据分析需求的满足周期从周级降低到分钟级，时间降低 90% 以上。
- ◆ **数据分析的场景更多元化**  
不同角色的用户可以自助完成数据分析，而且不再受限于办公室环境，在拜访客户的路上、准备会议前一小时，都可以快速发起相关数据分析，数据的价值被放大。
- ◆ **加速创新的频率和速度**  
更多有价值的数据分析可以快速尝试，加速好的想法验证，通过快速试错和迭代，加速创新的频率和速度 10 倍以上。

- ◆ **AI 能力升级计划**  
海外服务计划采用 Amazon Bedrock，集成多模态大模型（如文本、图片、表格），增强 Agentic AI 的“跨模态数据分析”能力；同时探索 Amazon SageMaker JumpStart，快速复用行业预训练模型，缩短 AI 功能上线周期；
- ◆ **金融行业服务优化**  
计划采用 Amazon WAF（Web 应用防火墙）和 Amazon Macie（数据安全服务），增强 SelectDB 的数据安全防护，满足金融客户的高合规要求；
- ◆ **全球性能优化计划**  
计划依托亚马逊云科技在北美（弗吉尼亚）、欧洲（法兰克福）区域新增的 SelectDB 集群节点，进一步降低当地客户的访问延迟。

作为一款实时分析型数据库，SelectDB 经过多年在客户大规模生产时间中的打磨，能够应对 Agent Facing Analytics 的挑战：

- ◆ 通过实时导入、主键模型实时更新满足秒级数据实时性的要求。
- ◆ 通过 MPP 分布式架构和向量化执行引擎提供最快的查询速度。
- ◆ 通过分区分桶、主键索引、倒排索引等提升并发查询处理能力。
- ◆ 通过 RBO、CBO、HBO 结合的智能查询优化器自动对 AI 生成的 SQL 进行优化提升查询效率；
- ◆ 基于 Lakehouse 湖仓一体查查询引擎，能够查询 Iceberg、Hudi、DeltaLake、Paimon、Hive 等数据湖中的数据，而且查询速度比 Trino 和 Presto 快 3 倍。SelectDB 还支持向量和文本检索混合检索，能够对文本、图片、音视频等非结构化数据进行快速检索。
- ◆ 基于 SelectDB 还可以构建高性能、低成本的可观测性平台，对 AI Agent 和应用进行全生命周期可观测，加速开发效率、提升持续运行稳定性。



### 关于飞轮科技

北京飞轮数据科技有限公司创立于 2022 年，成立以来，飞轮科技在积极投身 Apache Doris 开源社区建设的同时，自主研发了现代化实时数据仓库 SelectDB，面向全球提供多种部署形态的实时数据仓库产品 SelectDB 与相关解决方案，满足大规模实时数据场景下的极速查询分析需求。

飞轮科技总部位于北京，在上海、深圳、广州、西安、美国、新加坡等海内外 9 地设有研发机构。创始团队来自于原百度智能云初创人员和 Apache Doris 项目核心成员，公司 70% 员工为技术人员，且均来自于全球顶级数据库、云计算和互联网企业，拥有深厚的技术研发和服务经验。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。

亚马逊云科技 FOR SOFTWARE  
AND TECHNOLOGY



扫码获取白皮书电子版

\* 前述特定亚马逊云科技生成式人工智能相关的服务目前在亚马逊云科技海外区域可用。  
亚马逊云科技中国区域相关云服务由西云数据和光环新网运营，具体信息以中国区域官网为准。