

2025年中国生成式AI最佳实践

—— Agentic AI下的任务重构、决策与创新

生成式AI、Agentic AI、场景化、任务导向

报告说明

沙利文联合头豹研究院谨此发布中国生成式AI系列报告之《2025年中国生成式AI行业最佳应用实践》。本报告旨在梳理生成式AI技术发展动向，明晰各行业对于生成式AI的需求，并基于完整的评选指标与流程，筛选出各行业的生成式AI最佳应用实践。

沙利文联合头豹研究院对AI基础设施行业进行了下游用户体验调查。受访者主要来自基建、金融、医疗健康、互联网等行业。

本报告提供的生成式AI在各行业的发展现状亦反映出生成式AI的整体发展动向。报告最终对最佳应用实践的评选判断基于各厂商向沙利文提报案例，且仅适用于本年度中国生成式AI应用案例发展周期。

本报告所有图、表、文字中的数据均源自弗若斯特沙利文咨询（中国）及头豹研究院调查，数据均采用四舍五入，小数计一位。

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系弗若斯特沙利文及头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经弗若斯特沙利文及头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，弗若斯特沙利文及头豹研究院保留采取法律措施、追究相关人员责任的权利。弗若斯特沙利文及头豹研究院开展的所有商业活动均使用“弗若斯特沙利文”、“沙利文”、“头豹研究院”或“头豹”的商号、商标，弗若斯特沙利文及头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表弗若斯特沙利文或头豹研究院开展商业活动。

研究框架（1/2）

◆ 生成式AI趋势洞察	6
• 技术优化与成本降低	7
• Agentic AI 创造新价值	8
• 合成数据推动大模型迭代与落地	9
• AI安全治理体系持续完善	10
◆ 生成式AI海外优质案例分析	11
• 生成式AI海外优质案例评价维度分析	12
• Google Cloud × Seven-Eleven Japan	13
• 亚马逊云科技 × Deye	14
• 亚马逊云科技 × TVCMALL	15
• Azure AI Foundry × ASOS	16
• Azure × Sage Freight	17
◆ 生成式AI最佳实践案例评选流程及维度解析	18
• 最佳案例评选流程与方法	19
• 最佳案例评选维度解析	20
• 最佳案例2025重点维度解析	21
◆ 中国生成式AI行业最佳应用实践方案	22
• 生成式AI行业应用实践情况分析	23
• 生成式AI功能投入分析	24
• 中国生成式AI行业最佳应用实践图谱	25
• 交通基建行业的场景需求与挑战	26
• 交通基建行业最佳应用实践	27
• 金融行业的场景需求与挑战	31
• 金融行业最佳应用实践	32

研究框架（2/2）

• 医疗健康行业的场景需求与挑战	-----	36
• 医疗健康行业最佳应用实践	-----	37
• 汽车行业的场景需求与挑战	-----	41
• 汽车行业最佳应用实践	-----	42
• 能源与工业行业的场景需求与挑战	-----	46
• 能源与工业行业最佳应用实践	-----	47
• 娱乐行业的场景需求与挑战	-----	51
• 娱乐行业最佳应用实践	-----	52
• 企业服务行业的场景需求与挑战	-----	56
• 企业服务行业最佳应用实践	-----	57
• 互联网行业的场景需求与挑战	-----	61
• 互联网行业最佳应用实践	-----	62



章节一 生成式AI趋势洞察

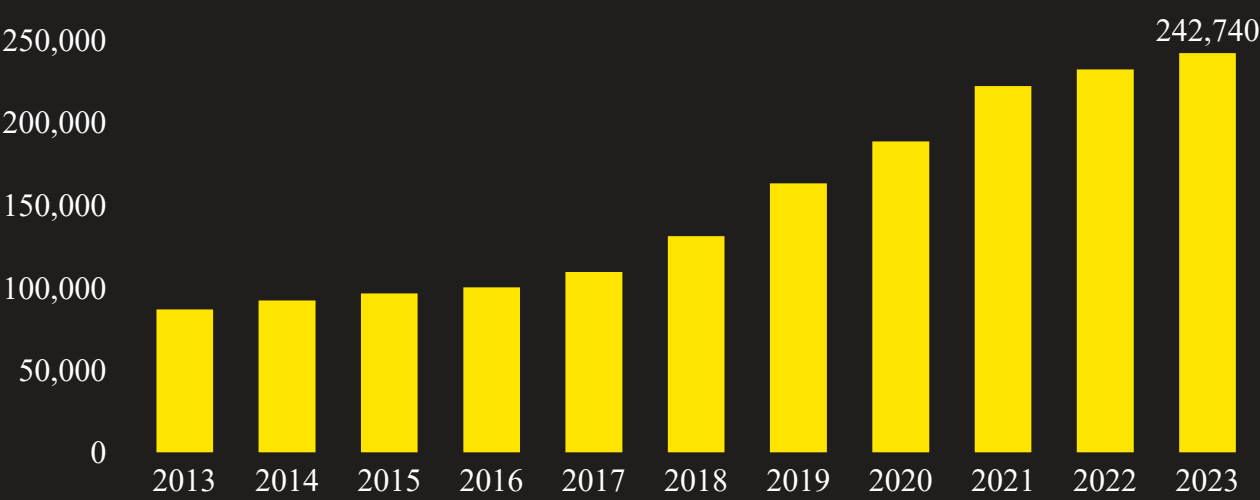
- ❑ 技术优化与成本降低
 - ❑ Agentic AI创造新价值
 - ❑ 合成数据推动大模型迭代与落地
 - ❑ AI安全治理体系持续完善
-

1.1 生成式AI技术优化与成本降低

得益于人工智能研究的持续火热，人工智能论文发表总量持续增长，人工智能在严苛比较基准测试中的性能持续提升，并且实现特定性能水平的推理成本已显著下降。

□ 2013 年至 2023 年间，人工智能相关论文发表的总数翻了一番多，从 2013 年的约 10.2 万篇增至 2023 年的超过 24.2 万篇。过去一年间的增长率达 19.7%，这一增幅具有重要意义。计算机科学的众多领域，从硬件与软件工程到人机交互，如今均对生成式AI的发展都有所贡献。因此，观测到的增长现象反映出该人工智能更广泛且日益增强的关注度。

全球AI相关文献发表



来源：斯坦福大学，沙利文

- 2023 年，研究人员推出了 MMMU、GPQA 和 SWE-bench 等一系列新型比较基准，旨在测试前沿人工智能系统的极限。仅一年后，性能就大幅提升：MMMU、GPQA 和 SWE-bench 的得分分别提高了 18.8%、48.9% 和 67.3%。除这些比较基准，人工智能系统在生成高质量视频方面也取得了重大进展，在某些特定场景下，基于语言模型的智能体在时间受限的编程任务中甚至表现优于人类。
- **当前人工智能变得更加高效、经济和易用。** 依托小型模型能力跃升，执行GPT-3.5 级别的系统的推理成本在2022 年11 月至2024 年10 月间骤降280 多倍。硬件层面，年化成本降幅达 30%，能效年提升率达40%。同时各个厂商的模型API调用价格也持续下降，截止目前，常规模型的API输出价格已经达到10元人民币每百万Tokens以下甚至更低。

元/百万token	GLM-4.5	通义千问Max	Deepseek-R1	SenseNova-V6
输入价格	0.8	2.4	2	3
输出价格	2	9.6	8	9

来源：沙利文

1.2 Agentic AI创造新价值

尽管生成式AI（GenAI）本身已经极大地提升了内容生成的效率和质量，但Agentic AI作为其更高级的形态，通过引入自主性、目标导向和多步骤执行能力，在生成式AI的基础上创造了全新的、更深层次的价值。

什么是Agentic AI?

Agentic AI又称代理型AI，将大量宝贵的数据、能力和程序与自主系统相结合。与传统的生成式AI不同，后者被动地处理输入以产生固定的输出，代理型AI更像是一个合作者，它在业务环境中积极地朝着某个目标努力，而不仅仅是遵循指令。

这种协作能力使组织及其员工能够具备前所未有的效率和创造力水平，同时确保与组织目标保持一致。通过具备推理、感知环境、学习和适应的能力，智能体可以被赋予一个目标，然后独立执行复杂任务并解决问题以实现该目标，无需人类干预。

生成式AI进化革命



为什么Agentic AI是生成式AI的下一个阶段?



1.3 合成数据推动大模型迭代与落地

人工智能系统在算法上取得实质性改进的主要驱动力之一，是在越来越大的数据集上扩展模型及其训练。然而，随着互联网训练数据的日益枯竭，人们越来越担心这种扩展方法的可持续性以及数据瓶颈的可能性，使用合成数据如世界模型生成的数据来训练模型也被认为是解决潜在数据短缺的一种方案。

什么是世界模型？

世界模型（World Models, WMs）作为一种生成式AI模型，能够通过构建内部表征理解真实世界的动态规律（涵盖物理特性与空间属性），同时借助文本、图像、视频和运动数据等输入信息生成视频内容。通过从感官数据中学习表征和预测运动、力和空间关系等动态特性，实现对现实环境物理属性的理解并通过生成环境及动作，从而模拟、指导及实施决策。

合成数据特点及发展趋势

随着合成数据的普及，特别是人工智能生成的网络内容所占比例越来越大，未来的模型将不可避免地在非人类生成的材料上进行训练。虽然合成数据具有近乎无限供应的优势，但要有效地利用合成数据进行模型训练，还需要深入了解其对学习动态和学习效果的影响。数据集扩展的一种方法是数据增强，即通过修改真实数据（如图像倾斜或混合）在保留关键特征的前提下创造新的变化。合成数据生成和数据增强都为增强人工智能模型提供了机会。



合成数据应用场景



自动驾驶与机器人仿真：在真实世界中收集极端或危险场景（如车祸、行人突然闯入）的数据既困难又危险。

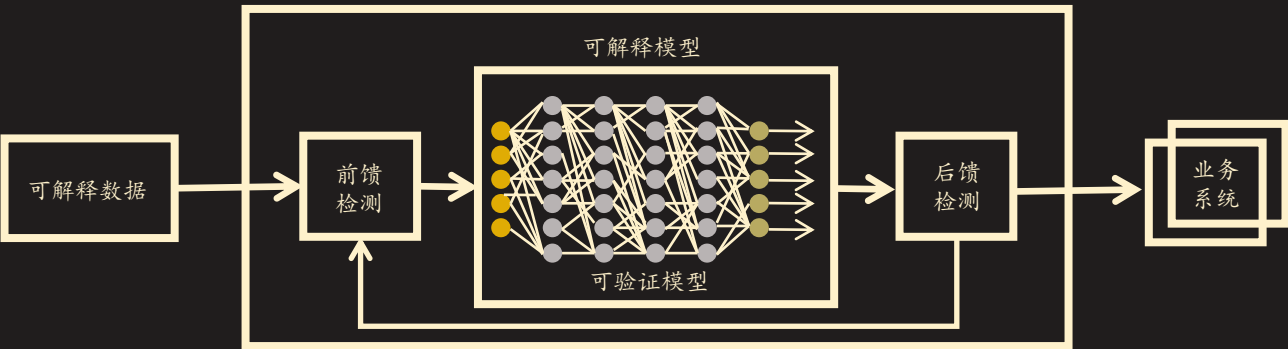
科学模拟及研究：在科学研究和工程领域，通过跨尺度的推演，合成数据帮助科学家和工程师更好地理解 and 掌控复杂系统。

游戏与VR：用于构建虚拟环境的逻辑与物理规则，使玩家能与动态场景互动。通过世界模型，系统可模拟重力、碰撞、光照等真实效果，提升沉浸感。

来源：沙利文

1.4 AI安全治理体系持续完善

AI安全风险不仅仅存在于理论分析，并且真实的存在于现今各种AI应用中，，AI系统在设计上面临五大安全挑战：软硬件的安全、数据完整性、模型保密性、模型鲁棒性、数据隐私，其中模型与数据的可解释性贯穿始终，保证AI应用安全。



- 可解释数据与前馈检测：输入数据需要具备可解释性，并在进入模型前经过前馈检测，以发现潜在的恶意或异常数据，防止攻击样本混入
- 可解释模型与可验证模型：AI模型本身不仅要具备较好的可解释性和健壮性，还需要通过验证机制来保障模型不会因对抗样本、后门攻击或数据污染而产生错误判断
- 后馈检测与业务反馈：在模型推理完成后，系统通过后馈检测机制再次验证输出结果的合理性，确保输出不会对业务系统造成风险。如果检测到异常，可以通过熔断、冗余等手段进行安全回退
- 业务系统集成：最终，AI推理结果才会被导入到业务系统中。通过这一整套架构，AI治理体系能在“数据—模型—业务”的全链路上实现安全保障

AI安全防御技术

	数据收集阶段	模型训练阶段	模型使用阶段
闪避攻击	对抗样本生成	网络蒸馏 对抗训练	对抗样本检测 输入重构 DNA模型验证
药饵攻击	训练数据过滤 回归分析	集成分析	
后门攻击		模型剪枝	输入预处理
窃取攻击	差分隐私	隐私聚合教师模型 模型水印	



章节二

生成式AI海外优质案例分析

- 生成式AI海外优质案例评价维度分析
 - 生成式AI海外优质案例分析
-

2.1 生成式AI海外优质案例评价维度分析

效率提升度

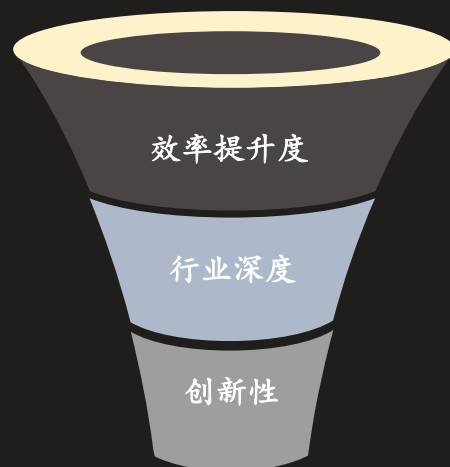
效率提升度指标包含成本效益降低、生产效率提升、资源利用率提升等

创新性

创新性指标包含技术创新、应用模式创新、模式创新等

行业深度

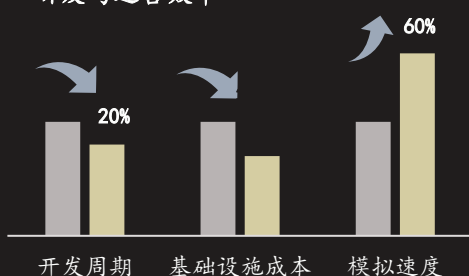
行业深度指标包含所处行业价值链覆盖程度、行业专业度、行业影响力等



- **海外优秀生成式AI应用实践：**亚马逊云科技与法拉利（Ferrari）合作，法拉利将其作为首选云服务商，利用其生成式AI技术实现多维度创新，有效实现客户个性化，并提高生产效率

效率提升度

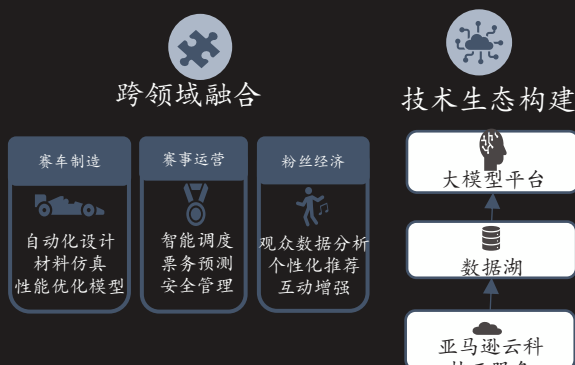
研发与运营效率



决策与响应速度



行业深度



创新性

应用亚马逊云科技生成式AI：辅助赛车及赛道设计、分析比赛策略、驱动粉丝互动

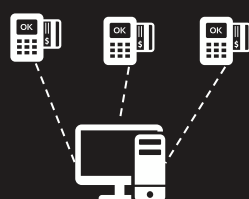
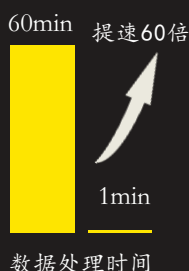


来源：沙利文

Google Cloud × Seven-Eleven Japan



日本 7-Eleven 与 Google Cloud 合作打造“Seven Central”数据平台和“7NOW”配送服务，实现门店数据的实时分析与可视化配送追踪，大幅提升运营效率和客户体验，推动便利店数字化转型。



门店POS机
数据规模达到21,000+



30,000



1,000

设计预期支持
门店×客流量数



跨
领
域
融
合

门店零售

数据整合
POS系统
终端联通

IT系统

中台搭建
数据流通
实时同步

运营协同

决策闭环
部门联动
响应提速



技
术
生
态
构
建

数据采集层：
Stream Analytics

存储层：
Cloud Spanner

分析层：
BigQuery

接口层：
Apigee



创新性

用云计算把全国便利店的分散数据“秒级化”，
用地图服务重塑最后一公里体验，
并尝试用生成式AI颠覆新品开发模式



来源：沙利文

Amazon Web Services × NBA

Inside the Game: NBA 的篮球智能平台

通过将Amazon Web Services AI 集成到其技术基础设施中，NBA 现在可以将数十亿个数据点转化为互动的球迷体验和直播期间共享的实时洞察。

亚马逊科技



效率提升度

全联盟数据处理能力大幅增强

高级统计实时生成

直播战术洞察响应更迅速

全球观赛体验更流畅

多场比赛并发处理更稳定

内容分发效率显著提升



创新性

NBA 与亚马逊科技合作，通过Amazon Web Services AI 帮助 NBA 首次实现智能化的实时赛事分析，生成投篮难度、防守影响与球员引力等全新高级统计数据。Play Finder 技术能够识别数千场比赛的战术模式，让球迷以前所未有的方式理解比赛。依托Amazon Web Services 的算力与平台能力，NBA 正将传统观赛体验升级为数据驱动的智能互动体验。



行业深度



跨领域融合



全球部署

数据加速
延迟优化
稳定互联



安全合规

合规保障
数据主权
安全可控



AI辅助运营

高级统计
综合评分
战术识别



技术生态构建

接入与集成层

全球加速层

数据处理分析层

Amazon Web Services × Mondelez International

亚马逊云科技



Mondelez International

Mondelez International 与 亚马逊云科技合作，将核心系统全面迁移至Amazon Web Services，并采用云原生数据库与高性能计算技术，整体提升了业务的敏捷性、工程效率与平台可靠性。

通过引入生成式 AI 工具和更现代化的云架构，Mondelez 的全球团队能够更快速地交付应用、优化运营流程。

行业深度



数字化转型



云基础设施

跨区域部署
稳定运行
高可用架构



数据库现代化

迁移至云原生
降低许可复杂度
优化读写性能



运营效率

开发流程加速
部署更便捷
运营效率提升



技术生态构建

资源迁移与底层架构层

云原生数据库与服务层

工程生产力与业务交付层



效率提升度

20分钟

服务器预置时间降至

60%

处理器可减少能耗高达

工程团队将工作效率提高至2到3倍



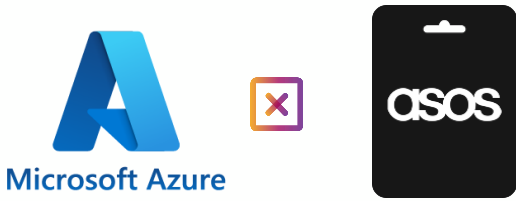
创新性

Mondelez International 的创新点在于利用Amazon Web Services的现代化工程体系重塑其IT运营方式。通过基础设施即代码实现服务器的快速、合规化部署，并使用 Amazon Q 引入 AI 辅助工程模式，使团队能够更快诊断问题、生成代码与优化架构。这些能力让原本高度依赖人工的配置、排查与交付流程变得更自动化、更智能化，形成了企业级的“AI 驱动运维与工程加速”框架，为全球业务的持续创新奠定了技术根基，并进一步增强数字竞争力。

来源：沙利文

Azure AI Foundry × ASOS

ASOS 与 Azure AI Foundry 合作，打造可与用户自然语言互动的 AI 风格顾问，提供个性化时尚推荐。该项目旨在提升购物体验 and 转化率，同时推动零售数字化和个性化服务创新。



效率提升度



创新性

150+
用户测试

偏见少

恶意“AI”越狱
高质量交互和输出

0

客户和品牌
安全警示信号

上手快速

内置资源减少自定义代码需求

快速构建了一个 AI 风格顾问（AI Stylist），用户可以与它进行自然语言互动，为他们推荐符合潮流趋势的时尚穿搭。这种“对话式推荐”模式植入了品牌语调与时尚洞察，不仅提升了购物体验，也开辟了智能推荐的新方式



行业深度



跨领域融合



时尚平台

上新提效
数据赋能
全球联动



对话式AI

实时响应
成本精简
内容统一



无文化偏见

语义调优
风格适配
人力释放



技术生态构建

模型接入层

内容生成与处理层

内容安全与性能保障层

来源：沙利文

Azure X Sage Freight



Sage Freight 是一家快速成长的物流公司，每周处理超过 3,000 笔交易。为解决传统财务系统在数据完整性、结账效率和实时可视性方面的瓶颈，公司与 Microsoft 合作，部署 Dynamics 365 Business Central 云端 ERP 系统，实现会计流程自动化和全面财务洞察，从而支撑业务扩张和运营效率提升。


效率提升度

21天  **12** 天
结账速度提高33%

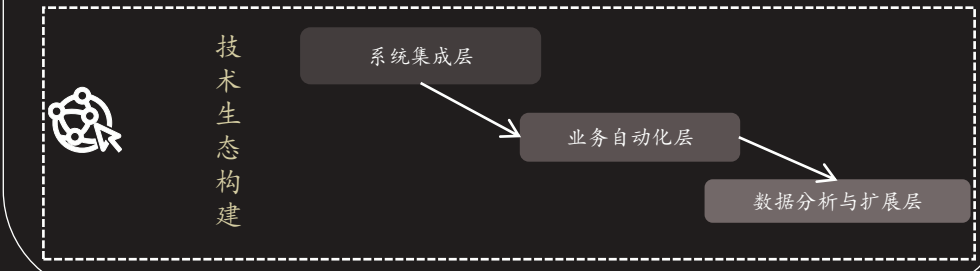
70%  **88** %
准时收款率提高 18%

3,000+ 交易综合数据/周


行业深度




创新性



Sage Freight 利用 Dynamics 365 Business Central 实现了财务和运营流程的全面数字化与自动化，将传统月末结账时间缩短 33%，应收账款按时率提高 18%。通过云端 ERP 与行业系统深度集成，公司不仅获得实时财务可视化，还能将团队精力更多投入到客户关系和业务增长上，形成了物流财务管理的新模式。



章节三

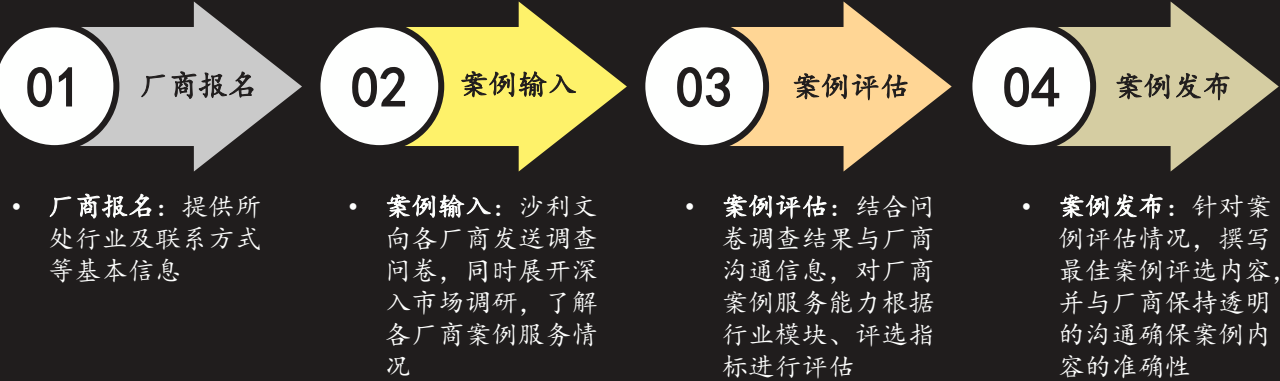
生成式AI最佳实践案例评选流程及维度解析

- ❑ 最佳案例评选流程与方法
- ❑ 最佳案例评选维度解析
- ❑ 最佳案例2025重点维度解析

3.1 最佳案例评选流程与方法

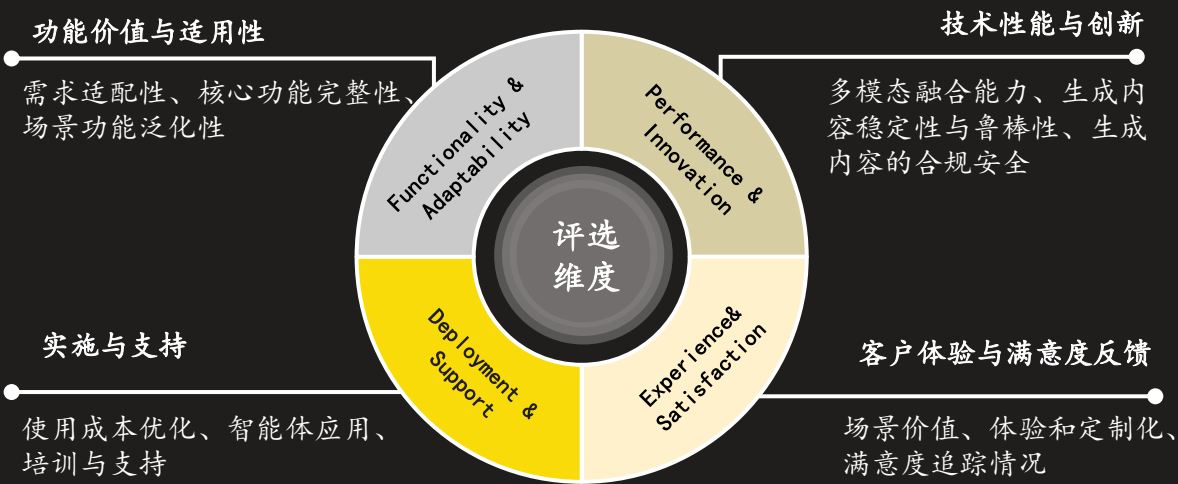
最佳案例评选流程为厂商报名参与、案例输入、案例评估以及最后的案例交付与发布。案例评选将结合行业交叉大数据与沙利文创新的全维度评选指标，客观公正地呈现生成式AI行业的最佳案例。

评选流程



方法论

沙利文融合了传统与创新的研究方法论，结合行业交叉大数据，通过多元化的调研方法与创新的全维度评估指标，客观公正地呈现生成式AI行业的最佳案例。



3.2 最佳案例评选维度解析

最佳案例评选维度围绕生成式AI最新发展趋势，结合功能价值与适用性、技术性能与创新、实施与支持以及客户体验与满意度反馈四个维度全面评估不同行业各模块的案例服务能力。

功能价值与适用性维度

- **需求适配性：**关注产品是否能精准匹配用户的业务需求和场景逻辑，评估在目标达成、收益实现、代价考量及战略对齐方面的表现，以确保其适应并解决市场需求。
- **核心功能完整性：**关注产品是否具备完整的核心功能模块，包括自然语言处理、图像生成、语音交互等核心模块协同性及成熟度，以满足实际业务闭环。
- **场景功能泛化性：**衡量产品在不同应用场景中的适应性和扩展能力，以实现跨业务场景的无缝适配，泛华能力决定了技术的复用价值与长期持续性，是衡量系统扩展性的关键指标。

技术性能与创新维度

- **多模态融合能力：**关注产品能否高效处理文本、图像、音视频等多模态数据，并实现跨模态内容生成能力，该能力是实现复杂场景智能化的核心支撑。
- **生成内容稳定性与鲁棒性：**关注系统在不同输入条件下输出内容的一致性与可靠性，关注在异常输入（如模糊指令、噪声数据）下的容错能力，确保生成结果始终符合预期，同时也关注生成质量提升方案
- **生成内容的合规安全：**聚焦内容生成过程中的法律、伦理与数据安全风险，关注模型合规情况，自动化监测及合规保障措施，同时也关注在异常场景下的容错能力和应急响应制度

实施与支持维度

- **使用成本优化：**关注部署、运行及维护生成式AI系统的全周期成本，包括硬件资源消耗、算法迭代费用及人力投入，关注轻量化模型、云原生架构等降本方案，以提升技术普及率和性价比。
- **智能体应用：**关注智能体在模拟人类决策、自动化流程中的实际效果，验证其在客服、运营等场景中的响应效率与准确性，以及对业务流程的优化贡献，关注智能体能力及其技术价值转化。
- **培训与支持：**关注用户的理解与操作门槛，关注厂商对客户实施建设后的持续赋能的资源和服务支持、实时技术支持及案例库，降低技术应用的复杂性并加速落地进程。

客户体验与满意度反馈维度

- **场景价值：**关注产品在具体业务场景中实际带来的提升、收益与价值，结合用户反馈与业务数据，评估技术对核心需求的贡献度，关注获得的长期支持与投入。
- **体验和定制化：**关注产品方案交互设计的友好性（如API易用性、界面直观性）及定制开发能力（如参数可调性、模块扩展性），关注满足增长需求的实际有效性与多系统融合性，关注能否通过全面支持满足企业增长需求。
- **满意度追踪情况：**关注产品在使用过程中不同使用场景与使用角色的满意度，关注对于客户满意度情况分析反馈，以及是否识别共性问题并驱动产品迭代，持续贴合用户需求及提升忠诚度。

3.2 最佳案例2025重点维度解析

在最佳案例2025评选维度中，我们在技术性能与创新维度中更关注多模态融合能力，并且在实施与支持维度更关注智能体应用情况。

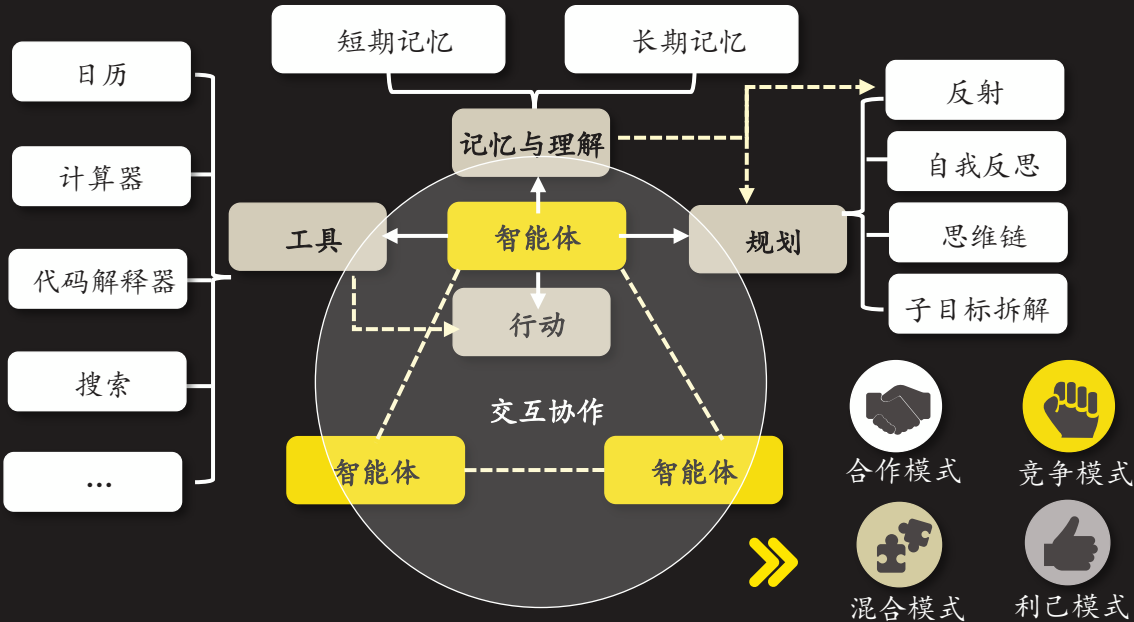
- 多模态融合能力：生成式AI模型应该能够通过统一的架构设计，能够同时接收和处理来自不同感官通道的信息。这要求模型具备强大的编码能力，将异构数据（如时序信息、像素、声波、字符）转化为统一的语义向量空间中的表示，从而实现信息的对齐与融合。

多模态交错数据预训练

模型通过大规模的多模态交错数据建立初步的跨模态关联

训练数据	多模态交错文档	双模态对应数据	针对特定场景构建的多模态数据集
训练目标	- 场景构建多模态序列，采用文本生成损失进行训练，帮助模型从多模态语境中生成相关文本。 - 针对具有多模态生成能力的模型，还需生成序列中的其他模态内容。		
其他训练设定	- 激活不同模块以支持模型架构需求，确保足够的可学习参数容量。 - 引入高质量的数据进一步提升模型表现。 - 可采用分阶段的预训练策略，逐步学习复杂的多模态关联。		

智能体应当具备四大能力：感知与理解：通过自然语言接口接收用户指令，理解语义、意图和上下文；任务分解与规划：将高层目标拆解为可执行的子任务序列，动态选择最优路径；工具调用与执行：调用API、数据库、代码解释器、外部系统等工具完成具体操作；反思与迭代：根据执行结果进行自我评估与修正，提升长期表现。



来源：沙利文

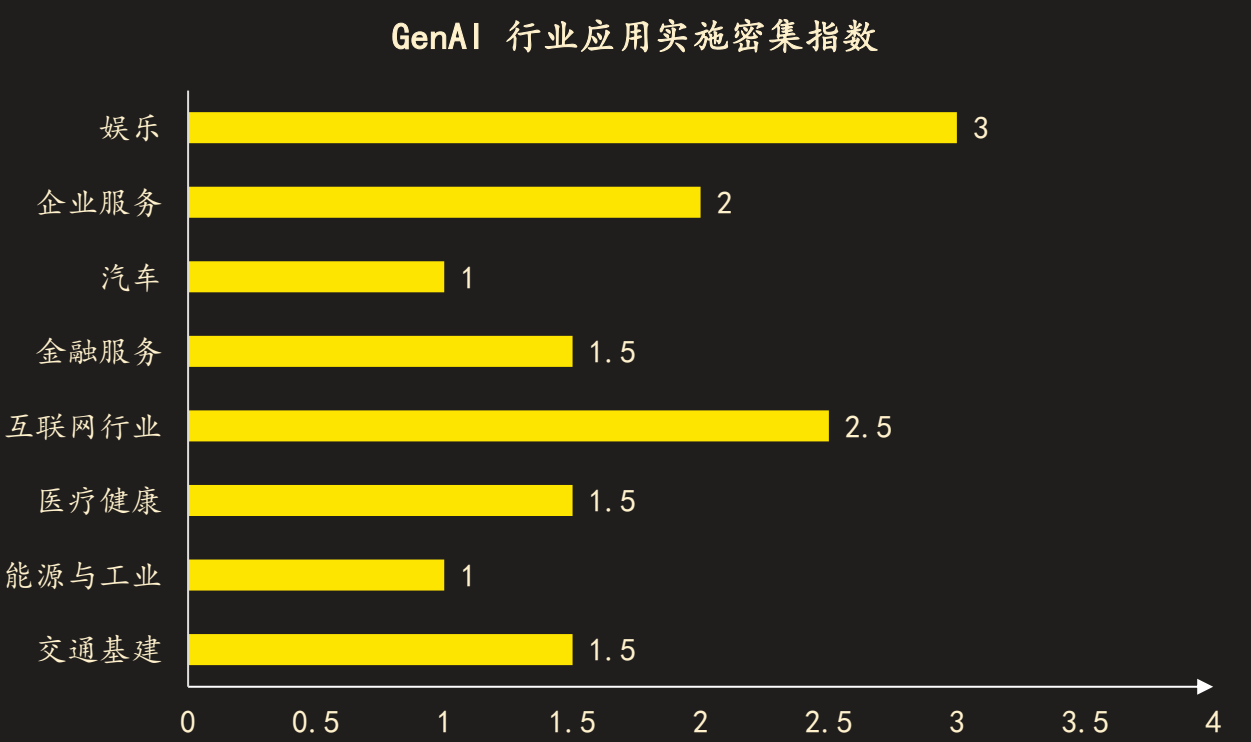


章节四 生成式AI最佳实践案例汇编

- 生成式AI行业应用实践情况分析
 - 生成式AI功能投入分析
 - 中国生成式AI行业最佳应用实践图谱
 - 各行业生成式AI场景需求与挑战
 - 各行业生成式AI最佳应用实践
-

4.1 生成式AI行业应用实践情况分析

在最佳案例2025评选维度中，我们观测到不同行业中生成式AI应用实践的关键信号与实施密集度。其中以娱乐行业生成式AI应用最为广泛，其次企业服务与互联网行业同样在生成式AI应用的探索中处于领先地位。



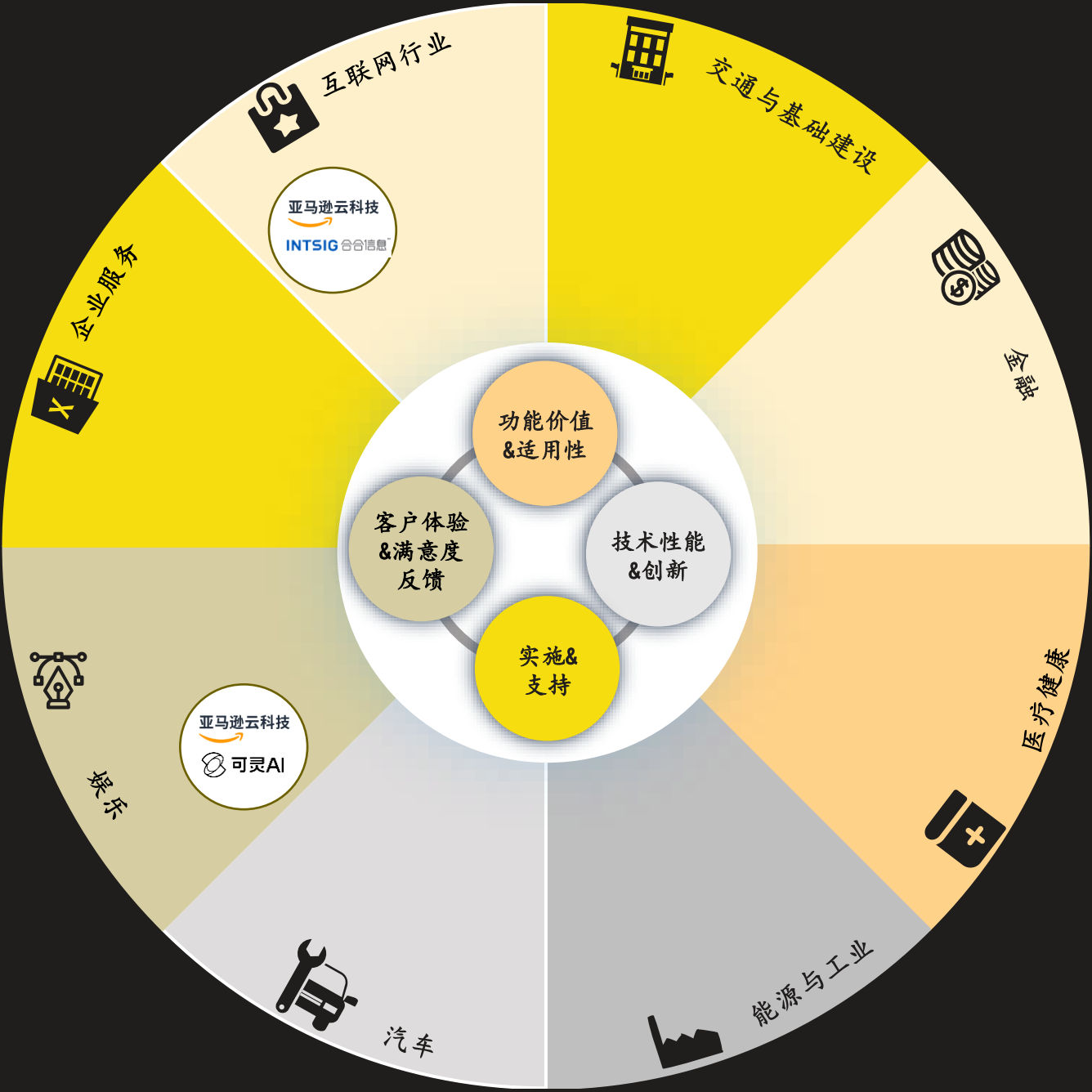
行业	关键信号
娱乐	生成式AI应用迅速崛起；文生图、文生视频；工作流程发生变化
互联网行业	AI 原生内容兴起；广告模式变化；传统巨头仍在增长
企业服务	效率提升；客户服务方式改变
医疗健康	进行文档/转录类试点；临床模型基本未变
金融服务	自动化客户支持；自动化文档处理
汽车	后台流程自动化；设计自动化
交通基建	维护类试点为主；工作流试点
能源与工业	新型系统建设；全产业链服务

4.2 生成式AI功能投入分析

- 当前市场中，约50%的GenAI预算被投入到销售与营销部门
- 尽管 50% 的生成式 AI 预算流向了销售与市场部门，但后台自动化往往能带来更高的投资回报率。
- 企业类型对投资分布有着显著影响：例如，制造商和医疗健康机构通常对销售与市场部门的投入极少，而更倾向于将资金重点投向运营领域；科技与媒体公司则往往优先考虑营销、内容创作及开发者效率工具；专业服务机构则更偏向于文档自动化及法律合规类工具。



4.3 生成式AI行业最佳应用实践综合图谱



来源：沙利文



娱乐行业的发展挑战与最佳应用实践

- 生成式AI技术赋能娱乐行业颠覆了传统的美术制作、游戏体验、运营以及营销模式，以资源超分、自动捏脸为例的领先技术相继应用于娱乐行业，推动行业的技术革新。
- 生成式AI赋能娱乐行业各流程阶段，其中产品开发阶段在生成式AI技术辅助下，加速了企业创作流程且大幅降低了美术资源创作成本；此外针对娱乐的版权问题风险，也可通过生成式AI进行检测与分析。
- 生成式AI广泛应用于娱乐行业的开发创作，但内容的创作除了普遍考量的隐私与数据风险外，大量依赖相关技术的短平快娱乐内容，可能降低行业门槛，挤占行业优质资源空间，给行业带来挑战。

中国娱乐行业最佳应用实践

亚马逊云科技 × 可灵AI

可灵AI

亚马逊云科技



客户痛点

内容创作耗时费力

专业制作投入过高

灵感常陷枯竭

视频质量难以把控

核心评价关键词

视频生成领域大模型

生成效率提升与体验革新

云端部署与高效扩展

效率提升与成本降低

功能价值与适用性维度

可灵AI为企业与个人提供一站式创作能力，支持从文字、图像到动态视频的自动生成，覆盖宣传片、广告物料、培训资料和品牌短片等多场景应用。其价值不仅在于替代传统拍摄与剪辑流程，还体现在跨行业的广泛适用性，已服务于广告、游戏、电商、零售、娱乐、音乐等多个领域。

技术性能与创新力

亚马逊云科技携手可灵打造行业解决方案。该方案融合可灵先进的多模态AI技术与Amazon Bedrock中的前沿大模型。实现了从文字或图像到动态视频的一站式转化，满足多样化的视频制作需求，并支持多语种输入。

实施与支持

可灵AI已在亚马逊云科技 Marketplace 上架，企业可便捷接入并快速落地。凭借亚马逊云科技的 Auto-Scaling、RDS 和 ElastiCache，方案能在高并发场景下保障稳定与安全；同时 Amazon CloudFront 提供全球加速分发，确保跨区域一致体验。

客户体验与满意度反馈

可灵AI与Amazon Bedrock帮助用户显著提升创作效率并降低成本，缩短制作周期的同时产出接近专业水准的内容。平台还利用深度学习与数据分析提供持续灵感与个性化建议，助力突破创意瓶颈。包括小米、百度、BlueFocus、三七互娱、4399、海艺在内的客户已采用并认可该方案，验证了其在效率、成本和体验上的优势。

解决客户痛点的核心价值

提升创作效率

可灵AI基于快手自研的视频生成大模型“可灵”和图像生成大模型“可图”，能快速生成高质量的视频和图片。创作者可以更快地迭代创意，减少拍摄、剪辑、后期制作的工作量，缩短生产周期。

降低创作成本

对于内容创作者来说，无需专业的设备和团队，通过可灵AI就能快速生成具有专业水准的内容，大幅节省设备采购、人员雇佣等成本，并显著提升创作效率。

提供创意支持

可灵AI平台通过深度学习算法和大数据分析，能够精准理解用户意图，提供个性化的创意建议和优化方案，为用户提供源源不断的灵感来源，帮助创作者突破创意瓶颈。

注：前述特定亚马逊云科技生成式人工智能管线服务目前在亚马逊云科技海外区域可用。亚马逊云科技的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。

FROST & SULLIVAN



互联网行业的发展挑战与最佳应用实践

- 生成式AI在互联网行业中的发展呈现出爆发式增长与深度变革的态势，其应用已从辅助工具演变为重塑产品形态、用户体验和商业模式的核心驱动力。
- AI技术的发展反向应用并重构互联网行业。互联网行业面临庞大的用户基数与高频次的用户交互，对高效推送与网络安全有较高需求。因此，AI大模型在互联网行业主要应用于个性化推荐与智能运维等场景。
- 此外，互联网行业凭借其海量数据、强大的技术迭代能力和直接的用户触达，成为生成式AI应用最广泛、最前沿的领域。
- 尽管生成式AI在互联网行业展现出巨大潜力，但其发展仍面临诸多挑战，包括生成内容的“幻觉”问题、数据安全和隐私泄露风险、版权归属模糊、算法偏见与伦理问题、高昂的算力成本以及用户信任缺失等。要实现可持续应用，企业必须在推动创新的同时，加强内容可靠性验证、数据安全防护、版权合规管理、伦理审查与透明度建设，平衡技术效率与社会责任，构建安全可信的AI应用生态。

中国互联网行业最佳应用实践

Chaterm——亚马逊云科技×合合信息

Chaterm



客户痛点

效率低下：批量操作繁琐，跨工具切换频繁，日常运维流程割裂

知识门槛高：技术栈复杂，新手学习曲线长，人才培养成本高

故障排查困难：分布式架构下日志分散，调用链复杂，定位问题耗时。

核心评价关键词

智能前瞻

无缝融合

敏捷赋能

降本增效

功能价值与适用性维度

其核心功能价值在于通过自然语言交互，让开发者能用自然语言对话直接管理与运维多云资源。它实现了操作的极致简化和智能自动化，大幅提升工程效率。在适用性上，它精准服务于开发者、运维工程师及技术管理者，尤其适用于面临混合多云管理复杂、运维成本高企、追求敏捷与DevOps极致效能的中大型企业与技术驱动型团队。

技术性能与创新力

Chaterm凭借其强大的AI Agent内核，实现了对异构云环境的高性能抽象与毫秒级响应，确保了复杂操作流程的流畅与稳定。其技术核心创新在于将大语言模型的意图理解能力与专业的云API操控深度结合，构建了真正的会话式、自主化云资源管理新范式。

实施与支持

在私有子网环境中，由于网络隔离，Agent模式可能面临连接限制。此时，Command模式成为更实用的选择。用户可以通过堡垒机连接到私有子网中的服务器，然后使用Chaterm的Command模式生成命令，由用户确认后在当前会话中执行。

客户体验与满意度反馈

增强团队协作功能，支持运维知识共享、操作审计、权限管理等企业级需求。团队成员可以共享自定义命令模板、运维脚本和最佳实践，形成组织级的运维知识库。

技术性能与创新力

实施与支持

功能价值与适用性

客户体验与满意度反馈

Chaterm——亚马逊云科技×合合信息

Chaterm 是合合信息联合亚马逊云科技打造的智能运维终端，旨在解决大规模云环境下批量操作繁琐、故障排查复杂和安全管控困难等痛点。它将 AI Agent 能力直接嵌入终端，支持自然语言交互与双模式操作（Command/Agent），既能辅助生成命令，也能自动规划执行任务。通过与亚马逊云科技云资源的深度适配，Chaterm 可高效管理成百上千台服务器，并结合零信任认证、权限管理和全生命周期审计，保障企业级安全。

Chaterm 介绍与架构设计



Chaterm是一款开源的AI智能终端工具，专为云资源管理和运维场景设计，通过自然语言交互革新了开发者与终端的互动方式。Chaterm采用现代化的分层架构设计，确保高性能、安全性和可扩展性。这种架构设计使其能够在提供强大功能的同时，确保高度的安全性和可扩展性，满足从个人开发者到大型企业团队的各种需求。

主要特点

Agent智能代理

Agent自主规划，自动完成复杂任务（可以基于日志回滚）

智能命令补全

基于个人操作习惯的，跨服务器，跨平台智能命令补全

语音输入控制

移动环境下，使用语音代替键盘输入，提高远程操作效率

全局语法高亮

基于用户个人配置的全局语法高亮，仅需设置一次，即可任意主机跟随

可视化VIM

在终端环境下实现可视化的语法高亮文件编辑体验，支持多种编程语言

零信任认证

支持企业级零信任安全认证，实现会话级的无感知安全认证

快捷命令/别名

设置全局快捷命令或Alias别名，简化复杂命令和批处理任务效率

统一授权管理

通过工作空间使用企业统一SSO账号管理企业不同组织间的数字资产

MCP协议支持

可低成本接入企业Notion，Github等知识库或工具，实现技能与知识扩展

注：前述特定亚马逊云科技生成式人工智能线管的服务目前在亚马逊云科技海外区域可用。亚马逊云科技的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。

方法论

- ◆ 头豹研究院布局中国市场，深入研究19大行业，持续跟踪532个垂直行业的市场变化，已沉淀超过100万行业研究价值数据元素，完成超过1万个独立的研究咨询项目。
- ◆ 头豹研究院依托中国活跃的经济环境，研究内容覆盖整个行业发展周期，伴随着行业内企业的创立，发展，扩张，到企业上市及上市后的成熟期，头豹各行业研究员积极探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业视野解读行业的沿革。
- ◆ 头豹研究院融合传统与新型的研究方法论，采用自主研发算法，结合行业交叉大数据，通过多元化调研方法，挖掘定量数据背后根因，剖析定性内容背后的逻辑，客观真实地阐述行业现状，前瞻性地预测行业未来发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 头豹研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 头豹研究院秉承匠心研究，砥砺前行的宗旨，以战略发展的视角分析行业，从执行落地的层面阐述观点，为每一位读者提供有深度有价值的研究报告。

法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何证券或基金投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告或证券研究报告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本报告所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本报告所载资料、意见及推测不一致的报告或文章。头豹均不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

主笔分析师

江林辉、蔚迪诺

☎ 17721418134

✉ chloe.jiang@frostchina.com

dino.wei@frostchina.com

深度研究小组负责人

李庆

☎ 13149946576

✉ livia.li@frostchina.com

 www.frostchina.com ; www.leadleo.com

 <https://space.bilibili.com/647223552>

 <https://weibo.com/u/7303360042>

©弗若斯特沙利文咨询（中国）

©头豹研究院

