

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
• AWS × Deye	14
• AWS × 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

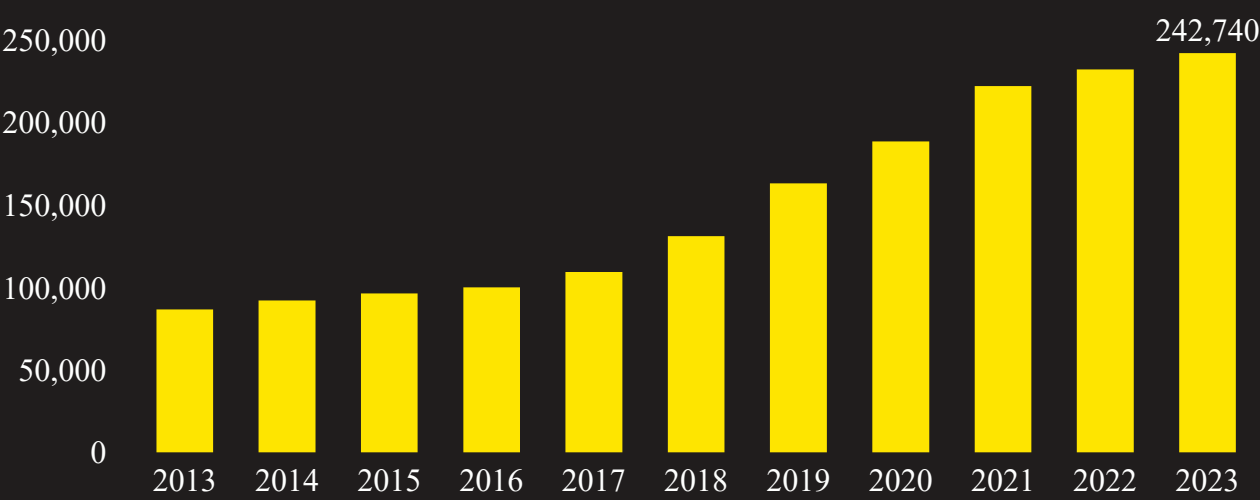
- ❑ 技术优化与成本降低
 - ❑ Agentic AI创造新价值
 - ❑ 合成数据推动大模型迭代与落地
 - ❑ AI安全治理体系持续完善
-

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？

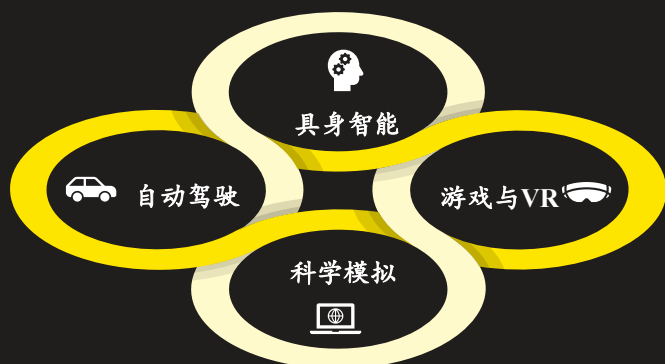
世界模型（World Models, WMs）作为一种生成式AI模型，能够通过构建内部表征理解真实世界的动态规律（涵盖物理特性与空间属性），同时借助文本、图像、视频和运动数据等输入信息生成视频内容。通过从感官数据中学习表征和预测运动、力和空间关系等动态特性，实现对现实环境物理属性的理解并通过生成环境及动作，从而模拟、指导及实施决策。

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。

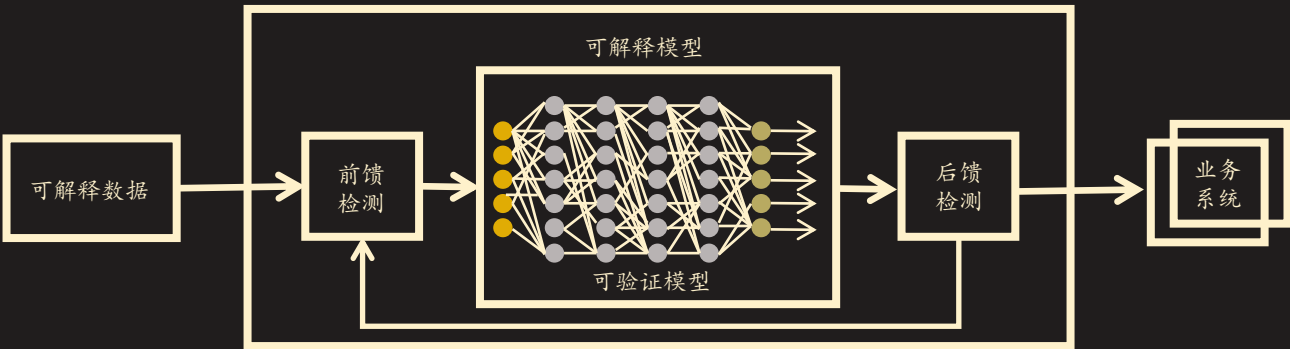
科学模拟及研究：在科学研究和工程领域，通过跨尺度的推演，合成数据帮助科学家和工程师更好地理解和掌控复杂系统。

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
-



头豹
LeadLeo

F R O S T & S U L L I V A N

沙利文

2.1 生成式AI海外优质案例评价维度分析

效率提升度

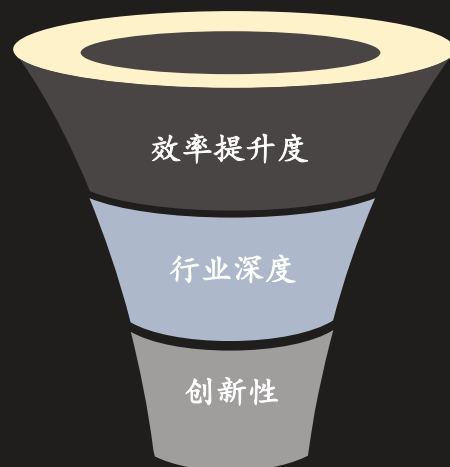
效率提升度指标包含成本效益降低、生产效率提升、资源利用率提升等

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

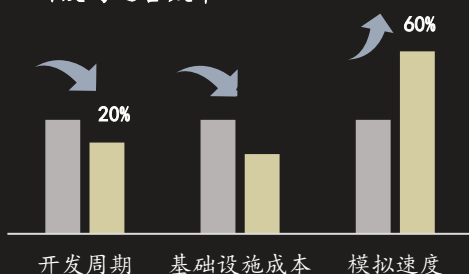
行业深度指标包含所处行业价值链覆盖程度、行业专业度、行业影响力等



- **海外优秀生成式AI应用实践：AWS 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AWS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

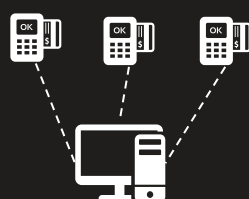
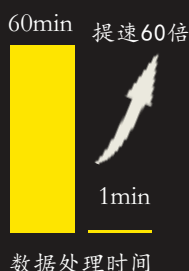


来源：沙利文

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



AWS × Deye

Deye 与 AWS 的合作，通过构建全球低延迟云架构、安全合规体系与 AI 驱动服务，实现“新能源设备环球联网、智能电价策略与多语种客服”融合创新，加速全球能源智慧化布局与数字化转型。



效率提升度



- 业务覆盖全球 110+ 国家
- 客服效率提升超 30%
- 跨区域业务延迟降至 0.2 秒
- API 响应时间大幅缩短
- 电费支出降低 30%

创新性

Deye 与 AWS 合作，通过全球云基础设施实现联网设备的实时互联，将跨区域延迟降至 0.2 秒。AI 助手基于电价预测和负载分析，提供“低买高卖”策略，帮助用户降低电费支出并优化储能回报。生成式 AI 智能客服支持 30 多种语言，自动处理大部分咨询，提高客服效率。整个系统的云端与 AI 深度融合，不仅优化运营效率，也推动 Deye 在全球能源智能化领域的技术领先。

行业深度



AWS X TVCMALL



TVCmall 与 AWS 合作，利用 Amazon Bedrock 等生成式 AI 技术，打造了一套智能多语言翻译与商品内容生成系统。该平台仅用 1 个月就成功上线，帮助企业将海量 SKU 快速上架，实现商品信息的多语种覆盖。项目实施后，商品上架效率提升约 30%，翻译成本降低约 40%，显著增强了 TVCMALL 在跨境电商领域的全球竞争力。

行业深度



跨领域融合



电商平台

上架提速
数据聚合
全球同步



多语言翻译

加速翻译
成本优化
内容标准化



商品渲染

渲染调色
纹理优化
释放人力



技术生态构建

数据与多模态采集层

生成式 AI 翻译与
内容生成层

业务集成与智能优化层

效率提升度

40%
翻译成本节省

30%
商品上架效率提升

1 个月完成 AI 智能商品
翻译方案开发与上线



创新性

TVCmall 借助 AWS 的生成式 AI 技术，将商品翻译与内容生成流程从人工为主转型为“AI 自动化+人工抽检”模式，大幅缩短了上架周期。通过这一应用创新，不仅实现了多语种商品快速上线，也有效降低了翻译与运营成本，推动跨境电商业务的全球化扩展。

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



头豹
LeadLeo

www.leadleo.com

400-072-5588

FROST & SULLIVAN

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

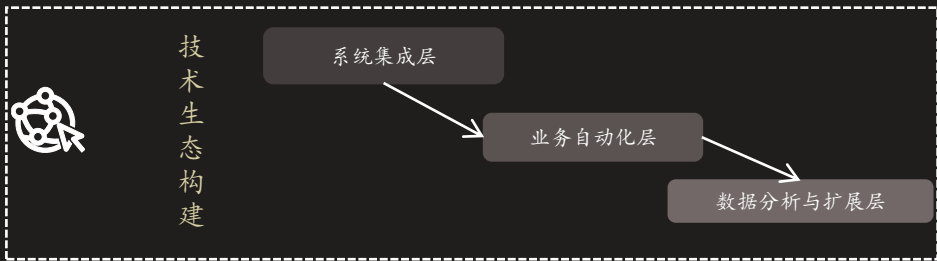
21天 **12**天
结账速度提高33%

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。

来源：沙利文



章节三

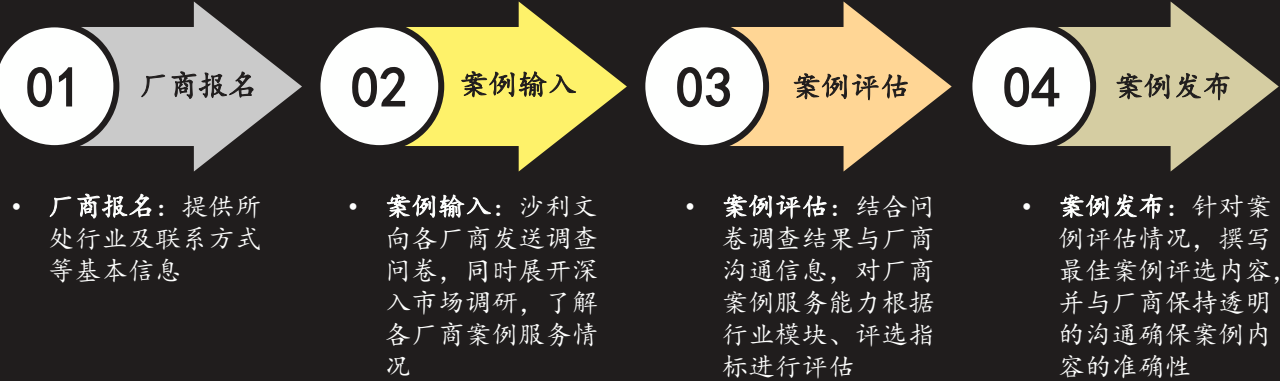
生成式AI最佳实践案例评选流程及维度解析

- 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

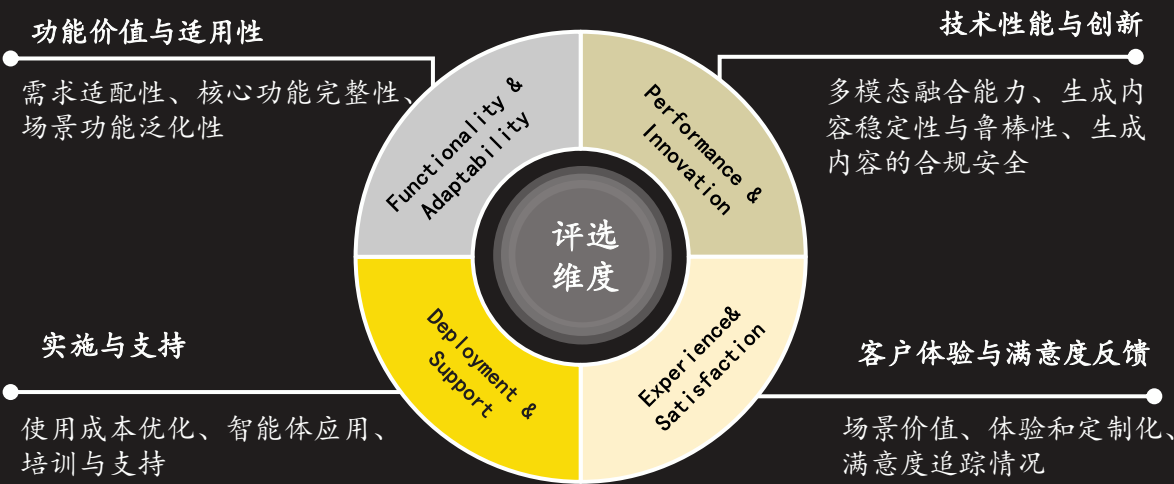
最佳案例评选流程为厂商报名参与、案例输入、案例评估以及最后的案例交付与发布。案例评选将结合行业交叉大数据与沙利文创新的全维度评选指标，客观公正地呈现生成式AI行业的最佳案例。

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。

来源：沙利文

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。

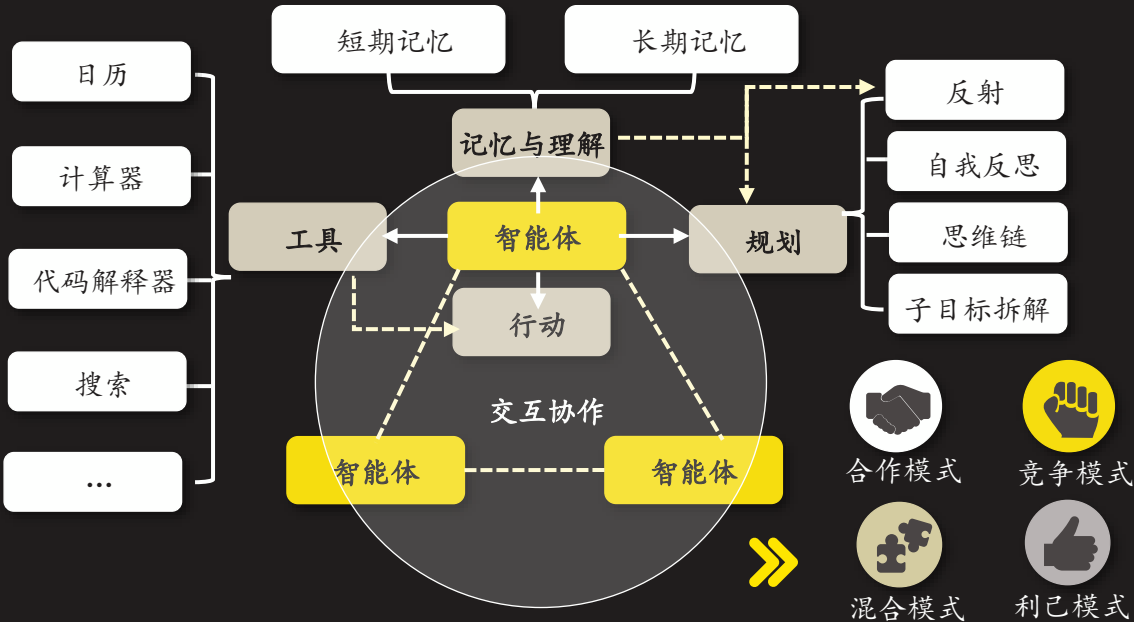
- **多模态融合能力：**生成式AI模型应该能够通过统一的架构设计，能够同时接收和处理来自不同感官通道的信息。这要求模型具备强大的编码能力，将异构数据（如时序信息、像素、声波、字符）转化为统一的语义向量空间中的表示，从而实现信息的对齐与融合。

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。



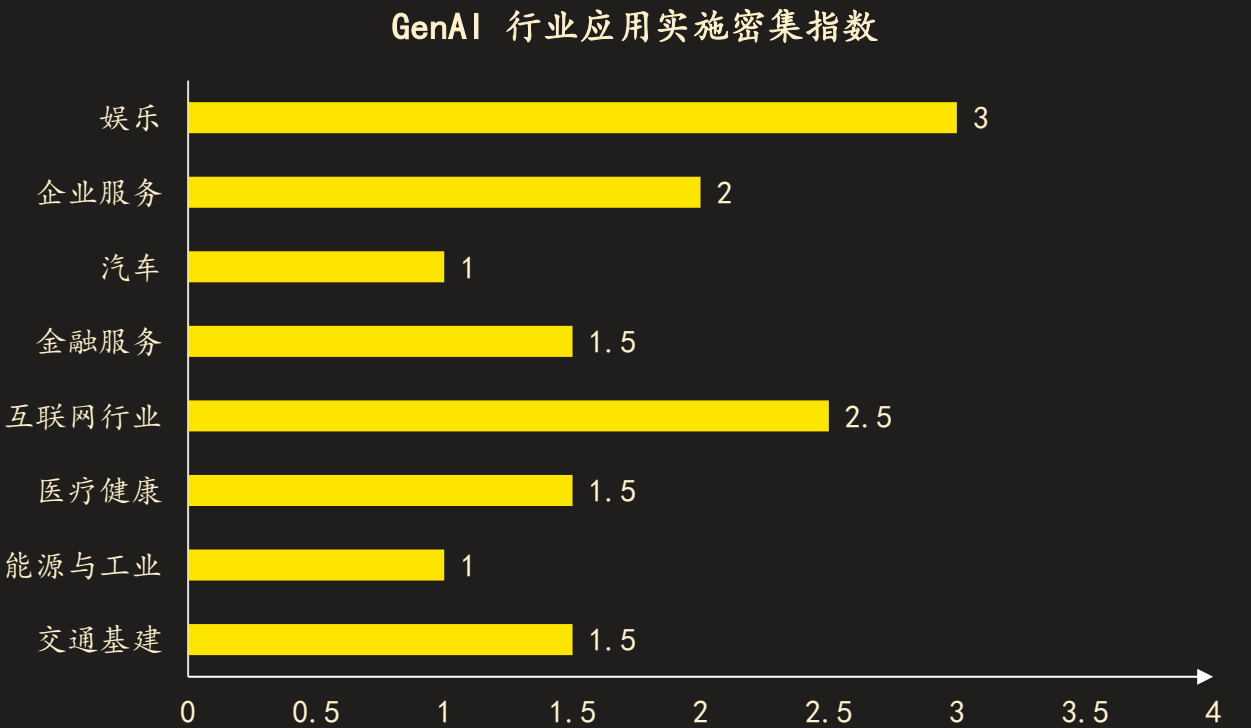


章节四 生成式AI最佳实践案例汇编

- 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。



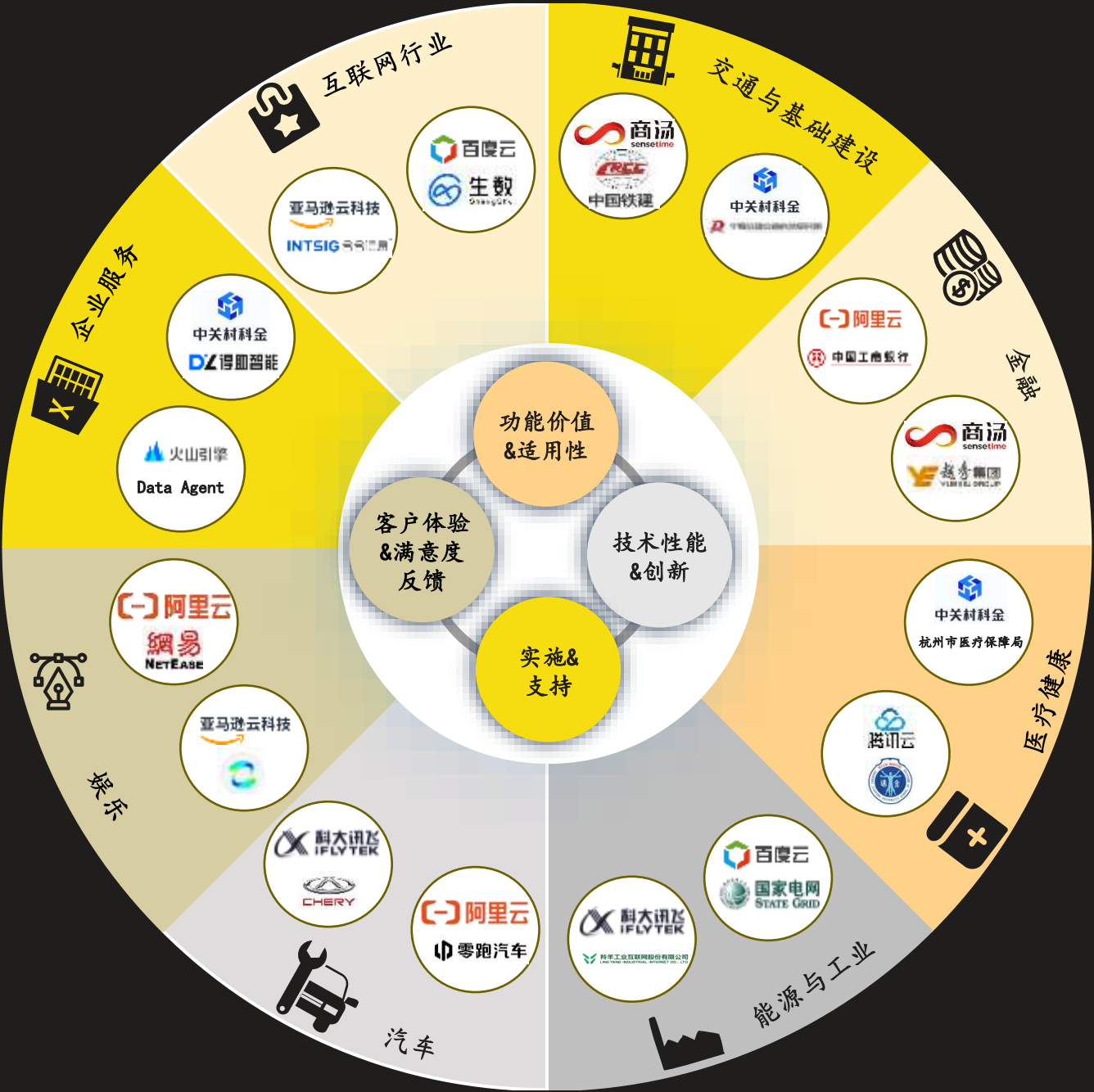
行业	关键信号
娱乐	生成式AI应用迅速崛起；文生图、文生视频；工作流程发生变化
互联网行业	AI 原生内容兴起；广告模式变化；传统巨头仍在增长
企业服务	效率提升；客户服务方式改变
医疗健康	进行文档/转录类试点；临床模型基本未变
金融服务	自动化客户支持；自动化文档处理
汽车	后台流程自动化；设计自动化
交通基建	维护类试点为主；工作流试点
能源与工业	新型系统建设；全产业链服务

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



交通基建行业的发展挑战与最佳应用实践

- 交通基建行业具备多个核心痛点与挑战：重复性工作繁重，大量时间消耗在编写技术文档、核算报表、编制投标文件等重复性、流程化的工作上。知识难以系统化，资深工程师的经验和知识难以系统化沉淀和传承给新人等。
- 该行业存在高质量、专业性强的数据获取与标注需求，与现有工作软件深度集成的需求，对可靠性与准确性要求极高，因此多采用垂类大模型+行业智能体模式，克服挑战，将AI应用深度融入企业。
- 在某些直接从事交通基础设施（如公路、铁路、桥梁、隧道等）规划、设计、施工和管理的工程企业的应用场景。生成式AI应用更侧重于内部生产效率提升、工程知识管理、设计施工流程自动化以及项目管理智能化。
- 生成式AI逐步成为“数字员工”或“智能助手”，深度融入其核心业务流程，实现“AI读文件、懂规范、用软件、控设备”的全链条赋能。

中国交通基建行业最佳应用实践

灵筑智工——中关村科金×宁夏交建

灵筑智工



客户痛点

缺乏统一的领域化训练和开发平台

文件格式多样且管理复杂

收集、输出的过程中会耗费大量时间且出现纰漏

核心评价关键词

交通基建专属大模型

智能规章识别与风险防控

全流程设计自动审核

效率与合规性双提升

功能价值与适用性维度

需求适配性：聚焦交通基建垂直领域，精准匹配施工、核算、投标等关键业务场景，显著提升目标达成效率与企业战略协同效益。

核心功能完整性：自然语言理解、文档生成、数据分析与设备控制等核心能力的深度融合，支撑从写文档到设备操作的完整业务闭环。

技术性能与创新力

多模态融合能力：借助得助大模型平台3.0，通过结合灵活配置的模块和组件，可快速搭建面向不同行业的音视频解决方案，助力企业高效连接客户。

生成内容的合规安全：拥有五重安全防护体系，既满足全栈国产化要求，也适配国际化本地部署，全方位保障数据安全合规。

实施与支持

智能体应用：其大模型平台拥有丰富的应用，能够三位一体管理，自由选择算力、模型、应用。其专业知识回答准确率较通用大模型提升40%。它的价值不仅是工具升级，更是行业DNA的重塑。

培训与支持：提供完善的培训机制、案例库与实时技术服务，显著降低用户使用门槛。

客户体验与满意度反馈

体验和定制化：通过模块化操作界面和流程引导机制，用户可轻松发起“文档编制”等功能，提高交互效率。

满意度追踪情况：平台界面清晰，智能体调用简单，配套培训与技术支持体系健全，用户满意度较高，认为“能用、好用、敢用”。

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

灵筑智工——中关村科金×宁夏交建

中关村科金与宁夏交建联合推出全国首个交通基建垂类大模型“灵筑智工”，通过上万份行业规范与工程技术文档训练实现模型专业性提升，并构建覆盖知识问答、文档写作、报表生成、数据分析、智能投标等智能体平台，满足交通基建行业对专业智能理解、高效文档处理与智能决策支持的迫切需求。

灵筑智工大模型应用平台：服务交通基建领域智能体快速搭建



4个解决实际痛点的“智能体”



中国交通基建行业最佳应用实践

商汤科技 × 中铁第一勘察设计院集团有限公司

医保小智



中铁第一勘察设计院集团有限公司

中国铁建 CHINA RAILWAY FIRST SURVEY AND DESIGN INSTITUTE GROUP CO.,LTD.

五大功能

知识检索

智能问答

文本生成

文本审核

多模态对话

核心评价关键词

铁路行业专属大模型

智能规范与风险识别

全流程设计赋能

效率与精度双提升

功能价值与适用性维度

专属数据与场景适配：基于商汤自研大模型，构建420GB多模态知识库，覆盖专业术语、图纸和规范，保障知识检索与应用的准确性。

全流程能力覆盖：支持方案设计、图纸查找、公式推理、报告生成与审核等关键环节，提升设计效率与成果质量。

技术性能与创新力

多模态融合与表达创新：实现文本、图纸、表格、公式等深度数据融合，构建行业领先的知识表达与推理体系。

智能风险识别与合规辅助：融合行业规范，自动识别潜在设计风险并提供修正建议，提升安全性与审查效率。

实施与支持

部署灵活高效：支持私有化快速部署，涵盖模型训练、上线与持续迭代。

平台化工具保障维护：借助LazyLLM与ModelStudio，简化定制与维护流程，降低使用门槛与长期成本。

客户体验与满意度反馈

一线使用体验佳：支持语义检索和智能问答，显著提升设计师工作效率，内测问答准确率达90%以上。

落地效果明确：已在铁一院实际业务中完成内测并取得成效，是业内首个应用于铁路设计的大模型。

技术性能与创新力

实施与支持

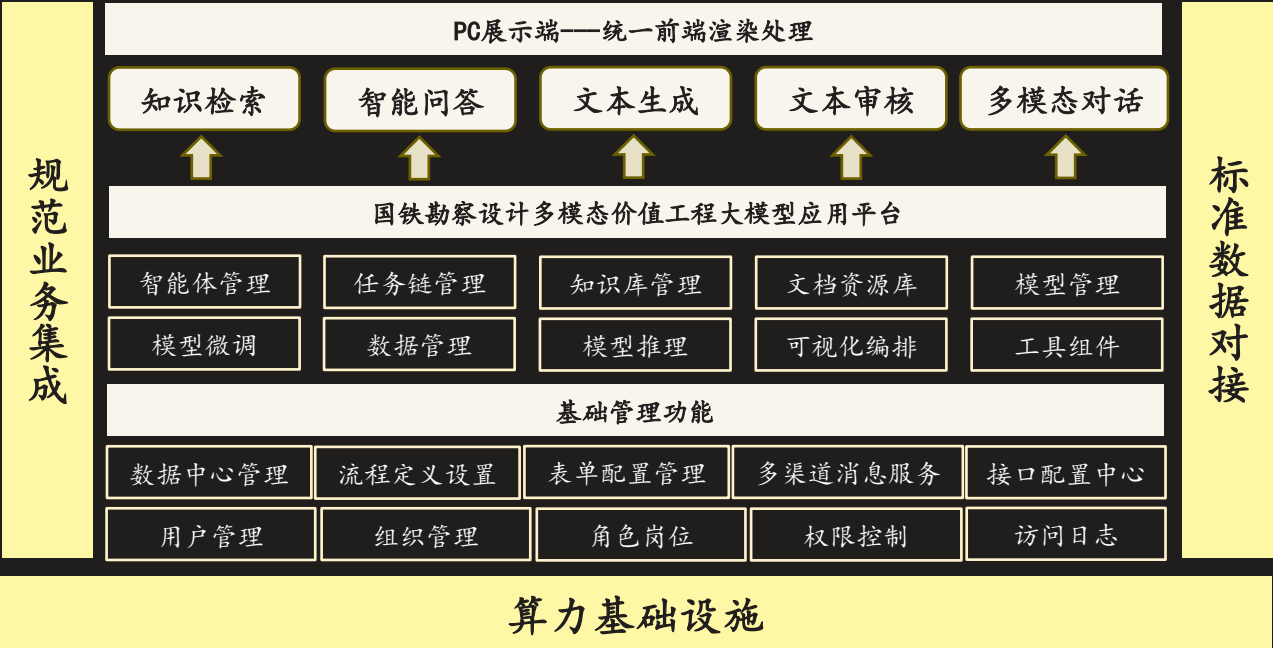
功能价值与适用性

客户体验与满意度反馈

商汤科技× 中铁第一勘察设计院集团有限公司

针对铁路工程设计行业文档类型多、跨模态整合难、专家经验传承效率低等痛点，商汤科技联合中铁第一勘察设计院集团，基于自研大模型构建覆盖全流程的铁路工程设计多模态智能平台，满足铁路设计对安全性、合规性与高效交付的综合需求。

产品架构及核心优势



解决方案成效

客户痛点与需求

解决方案

- 检索效率低：铁路工程设计领域文档类型多、数量大（超过20万份），传统检索依赖人工查找，效率低、准确率不稳定。
- 知识传承难：专家经验难以系统化沉淀，新人设计师学习曲线长，且人工审核耗时高。
- 多模态信息割裂：设计任务涉及图纸、公式、规范、报告等多模态数据，人工跨模态整合耗时耗力，易出错。
- 多模态深度融合：实现文本、图纸、公式、表格等多类型数据的统一表达与智能推理，支持语义检索与智能问答。
- 风险识别与合规辅助：融合行业规范，自动识别潜在设计风险并提供修正建议，提高设计安全性与审核效率。

实施成效

- 显著提升效率：在铁一院实际业务中，语义检索与智能问答准确率达90%以上，显著缩短资料查找与报告生成时间。
- 降低人工成本：减少跨模态信息整合与人工审核的人力投入，提升项目交付速度。



金融行业的发展挑战与最佳应用实践

- 金融行业特有的高密集数据属性以及经济环境的动态更变，为其在数据分析预测、数据治理层面带来主要挑战，数据分析预测的难点也进而使得客户服务建议合理化受到限制。
- 生成式AI的创新生成能力协助金融行业实现高效、低成本的功能拓展与数据信息处理。
- 生成式AI增强了金融行业的数据连接，协助机构及分析人员缩短了海量数据处理分析的时长，提升了数据处理决策的效率和准确性，同时赋能机构识别欺诈、异常数据等风险。
- 生成式AI赋能金融机构的数据处理高效性背后仍存在敏感数据泄露、模型偏见与过拟合等带来的交易决策失误等潜在风险，因此，金融机构及相关部门需建立完善的数据安全管理体系，并对重要信息进行二次验证核实。

中国金融行业最佳应用实践

大模型智能员工助手—商汤科技 × 越秀集团

大模型智能员工助手



客户痛点

多业态、多系统导致数据分散

成本与收益的不平衡

跨组织、跨部门信息共享与协同困难

核心评价关键词

企业级大模型基座

智能模型接入与算力优化

多模态解析与知识管理

功能价值与效能提升

功能价值与适用性维度

需求适配性：聚焦企业知识服务与智能化转型，支持多模型灵活接入（如商汤日日新、DeepSeek、Qwen等），适配集团级复杂业务场景。

核心功能完整性：涵盖模型接入、多模态解析、知识检索、应用评测及系统调用等核心模块，支撑企业从知识管理到业务应用的全流程闭环。

技术性能与创新力

多模态融合能力：具备多模态非结构化文档解析与结构化提取能力，支持文本、图像等多类型数据融合处理与知识生成。

生成内容的合规安全：支持权限隔离与用户级知识库访问控制，构建知识闭环机制，保障数据安全与模型持续优化。

实施与支持

智能体应用：通过智能体平台连接企业内部系统与业务流程，支持组件拓展与协议开放（MCP/A2A），实现业务链路一体化管理。

培训与支持：提供系统化培训与案例实践支持，助力企业高效接入与应用，推动AI能力与业务深度融合。

客户体验与满意度反馈

体验和定制化：提供清晰易用的平台界面与模块化操作流程，支持用户快速构建定制化知识服务与应用，显著降低使用门槛。

满意度追踪情况：通过完善的培训机制、案例库与实时技术支持，提升用户接受度与满意度，实现“易用、好用、放心用”。

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

大模型智能员工助手——商汤科技×越秀集团

商汤科技联合打造企业级AI中台。该项目将基于日日新大模型，通过对越秀集团内部海量多模态业务数据与知识资产进行深度挖掘与应用。旨在实现智能知识管理、业务流程优化等核心功能，以提升企业在业务场景中的认知与判断能力，满足大型集团企业对知识融合、流程敏捷与智能创新的核心需求。

AI中台大模型：赋能千行百业快速构建和部署多模态AI智能体的一站式企业级平台。

基础模型	视觉大模型 NLP大模型 代码大模型 AIGC大模型 多模态大模型 三维大模型 AI4S大模型							AI专家服务
大模型生产平台	数据管理平台 AI DataHub			人工智能开发平台AI Studio				
	数据采集标注服务 AI Annotator			模型推理 AI云开发机 AI Notebook 模型平台 自动化部署				
数据服务	AI算力池 (训练)ACP	AI容器实例CCI	AI裸金属实例BMS	AI对象存储AOSS	AI高速缓存	RDMA高速网络	专有网络VPC	
	异构AI算力池		AI存储池	软件定义AI网络		AI云安全		
	商汤人工智能计算中心（AIDC）							

知识闭环	企业本地知识库 多模态解析	结构化解析	核心功能
应用监控 精准问答 模型效能 场景评测 应用效能	多模型接入 日日新 Qwen Deepseek	权限隔离 用户隔离 访问控制 用户管理 数据权限	

已上线场景	资本制度助手	越秀审计知识问答	田园智能客服	法务案件追踪
	发展市场运营助手		越秀档案智能精灵	

中国金融行业最佳应用实践

商户智能审核助手——中国工商银行×阿里云

商户智能审核助手



客户痛点

效率低下：商户准入审核高度依赖人工处理

OCR 技术局限：泛化能力受限，信息提取复杂，鲁棒性不足，语义理解缺失

人工局限：在审核繁杂图文信息，存在效率低、易出错、成本高等痛点

核心评价关键词

全能多模态专家

极致性能标杆

精细交互交锋

专业应用导向

功能价值与适用性维度

复杂的图文推理能力：它不仅能够简单地描述图像内容，更能进行深层次的逻辑分析、因果关系推断和情感解读。

卓越的跨模态生态与创作：模型能够根据图像信息生成高质量的文字内容，实现了视觉与语言能力的深度融合。

技术性能与创新力

多模态融合能力：针对百万像素级的高清图像、极端宽高比的文档（如财务报表、学术论文）以及内嵌复杂文本的图片，它都能保持极高的信息提取和识别精度，解决了实际应用中的诸多痛点。

实施与支持

智能体应用：能够经学理解用户以坐标、边框或自然语言指出的指代请求，并能对图像的特定区域进行描述、分析和回答，实现了像素级信息的交互。同时能够比较不同图片之间的相似性与差异性，例如分析产品设计的迭代变化。

客户体验与满意度反馈

体验和定制化：特别优化了对扫描件、PDF、表格、流程图等专业文档的解析能力，能准确提取关键信息并回答相关问题，使其在办公、金融、科研等垂直领域具有巨大的应用潜力。其能力设计并非停留在技术演示，而是直接面向代码生成、视觉回答、内容审核、智能教育等需要高度可靠性和深度的商业与科研场景，是一款为落地应用打造的强大工具。

极致性能标杆

精细交互交锋

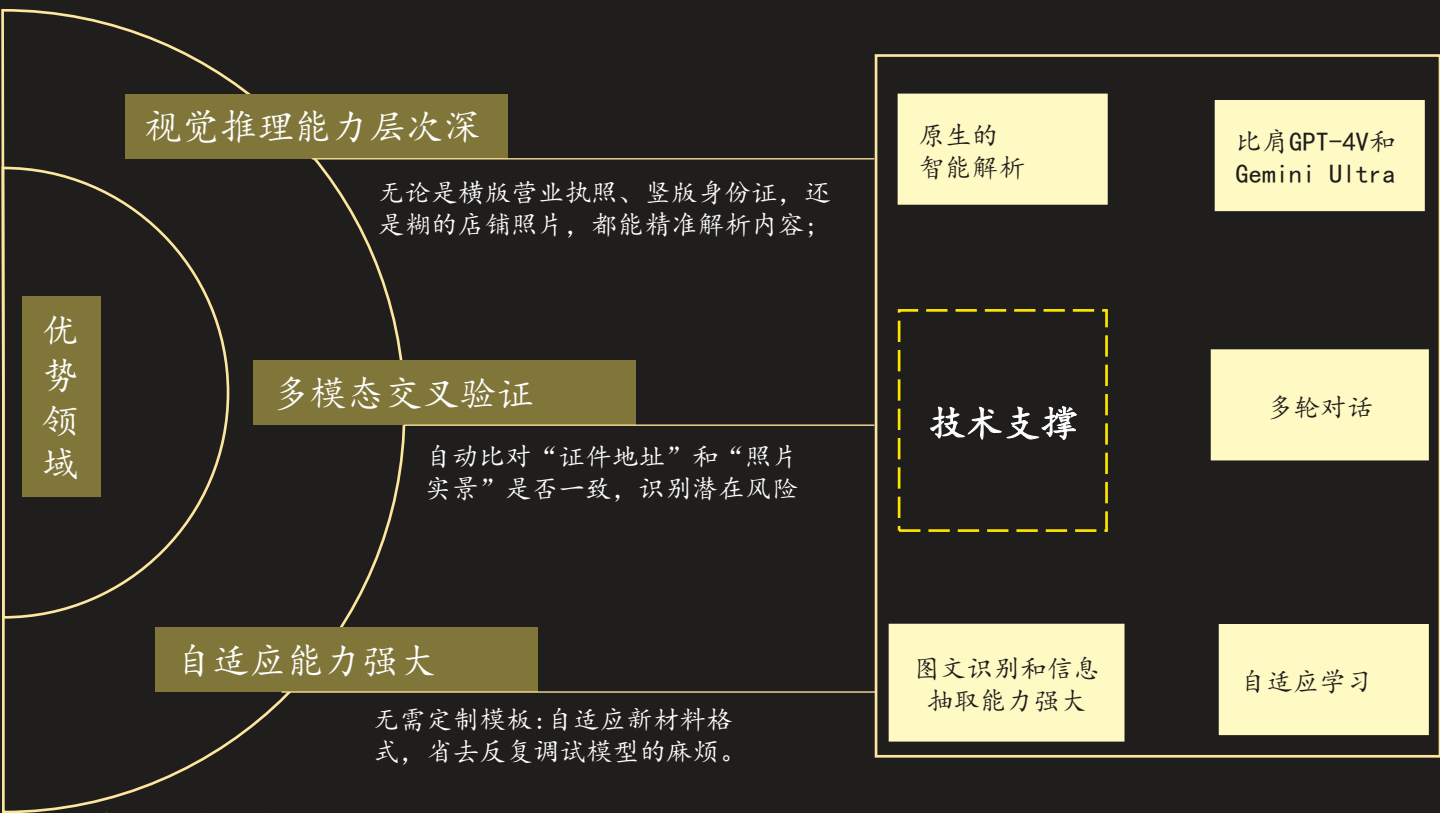
全能多模态专家

专业应用导向

■ 商户智能审核助手——中国工商银行×阿里云

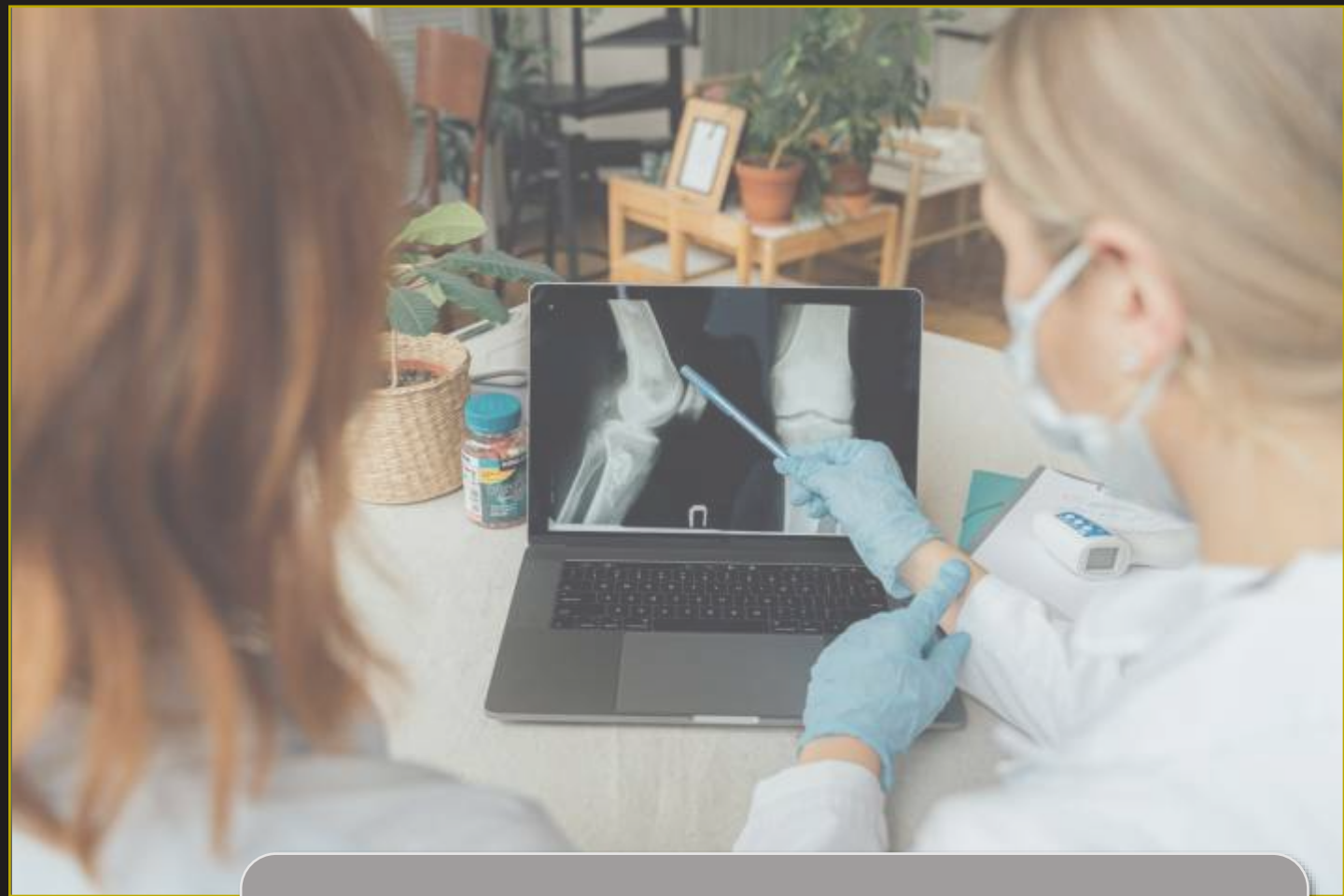
Qwen—VL—Max是阿里云通义千问团队推出的多模态大模型，属于Qwen-VL（Qwen Vision Language）系列中的顶级版本。具备强大的视觉与语言能力，能够处理图像、文本等多种信息输入，并在多个权威测试中表现优异。作为金融行业的领军者，中国工商银行敏锐洞察到技术趋势，率先选择将通义千问Qwen—VL—Max大模型应用在商户准入审核场景中，成功上线“商户智能审核助手”。

商户智能审核助手核心优势和技术支持



4个解决实际痛点的能力





医疗健康行业的发展挑战与最佳应用实践

- 生成式AI的数据信息生成能力协助药企突破了惯有的新药研发上市困难等难题，个性化的就诊建议助推行业由被动就医向精准化的主动就诊模式转变；此外，生成式AI通过合成数据解决了医患信息数据的隐私性难题。
- 生成式AI为医疗健康行业带来医药研发创新、管理效率优化、医学研究与服务支持以及定制营销等一系列的价值效益，其中医药创新与医学服务优化是目前生成式AI布局的重点领域。
- 生成式AI在助推医疗行业转型升级的同时，也带来了医疗内容偏误、责任事故归属、患者隐私安全保护、公平透明等风险挑战，以及融入了生成式AI的新模式是否有效仍需经过大量临床试验及多方验证证实。

中国医疗健康行业最佳应用实践

医保小智——中关村科金×杭州市医疗保障局

医保小智



杭州市医疗保障局

群众痛点

线上事务复杂难办

线下多次跑腿排队

沟通渠道单一低效

政府难题

线上服务有限体验差

系统孤立分析能力弱

业务繁杂协同效率低

核心评价关键词

医保垂类大模型

政策待遇识别与合规风控

经办流程自动审核与核验

服务效率与合规双提升

功能价值与适用性维度

覆盖核心医保场景：面向政策解读、待遇经办、投诉处理与远程交互等高频需求，支持“一点接入、全网通办”，把“难找、难懂、难办”转化为可指引、可办理。
支撑全流程闭环：结合用户身份与上下文提供个性化引导，配合智能预审、信息核验与知识推理，打通咨询—经办—反馈链路，减少线下多次跑。

技术性能与创新力

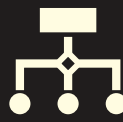
多模态与平台协同：RTC 实时音视频 + AI 算法（OCR/ASR/TTS/NLP/多模态）+ RPA 自动化联动，支撑视频办理、远程核验、流程编排等复杂场景。
持续进化与风控：建立闭环反馈，基于全量交互数据持续调优，形成热点识别、风险预警与智能监控能力

实施与支持

可运营落地：在经办机构与基层服务点快速部署，提供知识问答、经办指引、远程辅助、智能预审与流程提效等“即用型”能力；模型服务可按需选用。
赋能与保障：通过人机协同与远程交互降低上手成本，赋能基层经办人员并提升一致性；以持续运维与模型迭代保证稳定成效。

客户体验与满意度反馈

体验可感提升：线上渠道覆盖偏远与行动不便人群，流程引导更清晰，等待时间缩短、投诉率显著下降。
普惠与可信：服务触达更广、办理更便捷，群众感知从“脚尖奔波”转向“指尖智办”，形成正向口碑与满意度提升。



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

医保小智——中关村科金×杭州医疗保障局

中关村科金与杭州市医疗保障局联合推出面向医保服务的垂类大模型“医保小智”，基于政策法规、经办规程、历史工单与多模态业务数据的大规模训练，显著提升对医保术语与办理规则的专业理解与合规性，并构建覆盖政策解读与知识问答、经办流程引导与远程视频办理、流程预审与信息核验、智能质检与异常预警、报表生成与归因分析与趋势预测等能力的智能体平台，支持一点接入、全网通办，满足医保领域对精准政策理解、便民高效经办与智能监管决策的迫切需求。

医保小智大模型应用平台：技术框架及规范



落地开花：垂类大模型赋能医保服务的多维价值跃升

- 通过线上服务渠道，有效覆盖偏远地区、行动不便人群及工作时间冲突的群体，填补服务洼地。
- 人工智能分流海量简单咨询和预审工作，释放有限人力专注处理复杂、高价值业务。
- 利用多模态技术和大数据分析，实现医保基金智能监管，快速识别药店违规行为，提升监管效率。
- 为基层经办人员配备智能助手，大幅降低学习成本，提升服务能力与一致性。
- 大模型基于政策要点自动生成宣传文案，提升政策传播覆盖面与接受度。

中国医疗健康行业最佳应用实践

瑞金医疗大模型——腾讯云X瑞金医院

瑞金医疗大模型



案例核心评价关键词

医疗数据隐私程度较高

模型训练困难

电子病历费时

核心评价关键词

多模态交互

丰富的行业数据积累

多维医学表现解析

医患体验提升

功能价值与适用性维度

腾讯云通过开发和优化腾讯混元多模态大模型，以支持对医学文档的深度理解和处理。并整合文本、语音、图像等多种数据类型，实现对患者信息的全面分析和理解。

通过训练拥有海量医学数据和诊疗数据的医疗大模型，将医生从繁重的文书工作中解脱出来。

技术性能与创新力

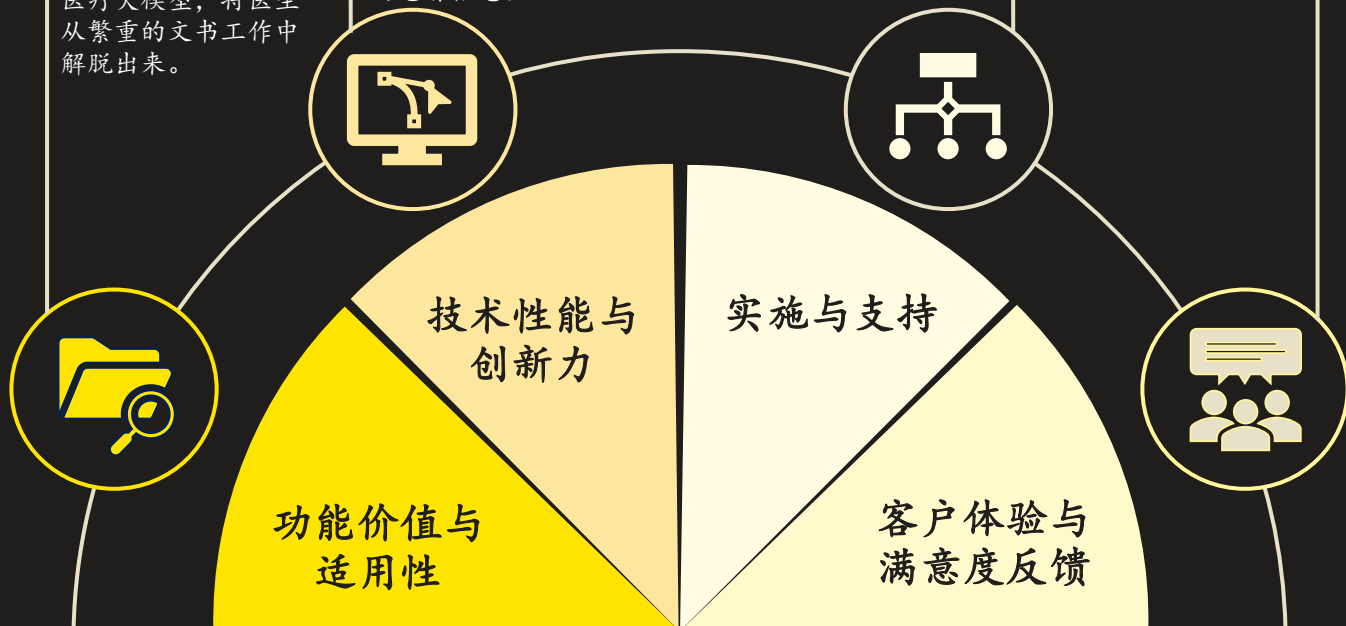
基于多模态交互的医学咨询和辅助决策技术研发服务的建设内容包括医学文档多模态大模型研发服务和基于多模态交互的医学咨询和辅助决策场景应用等。构建多模态交互平台，模拟真实的医患互动过程，支持医生在诊断和治疗过程中获取多维度的患者信息。

实施与支持

目前，电子病历生成、体检总检报告生成、健康咨询对话等AI应用已落地医院的多种业务场景。以体检报告生成为例，平均每5秒即可自动生成一份总检报告，每日自动生成超过500份，报告采纳率达到96%以上。

客户体验与满意度反馈

患者可基于大模型实现精准预约。利用OCR智能识别、智能语言分析等技术，智能评估肿瘤患者入院条件，显著提高了住院收治率，降低患者等候时间和经济成本，并有效解决了医院床位精准预约的问题。

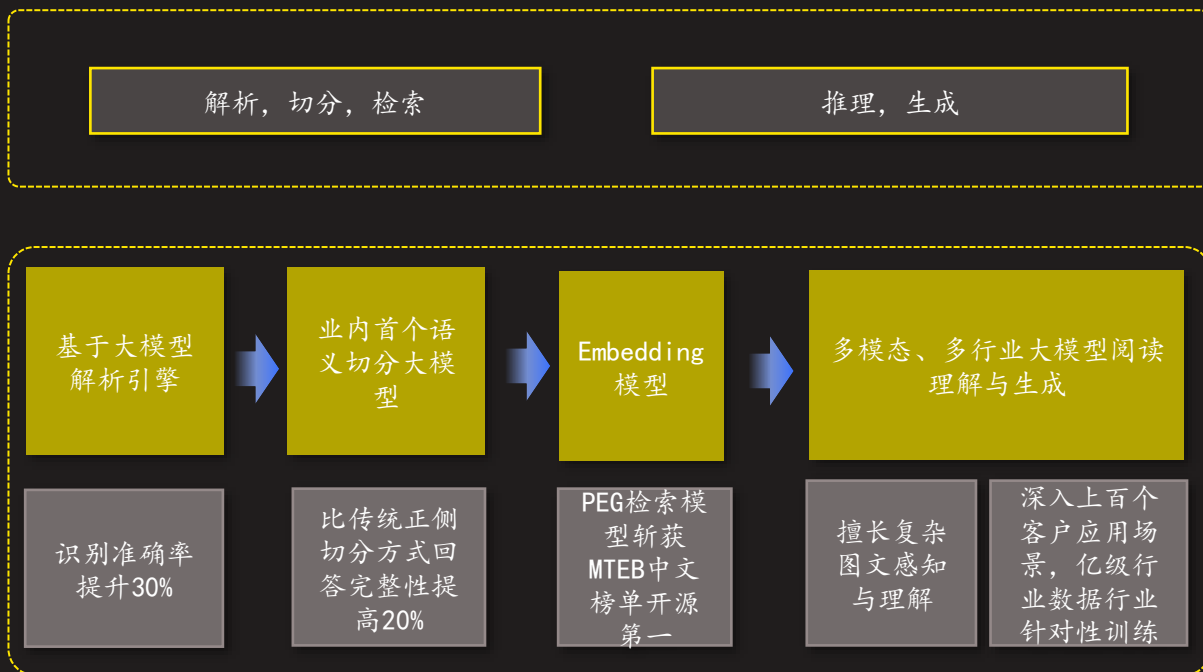


瑞金医疗大模型——腾讯云X瑞金医院

基于医疗健康行业共存的诊断报告书写任务繁重、患者预约等候时间较长等痛点，腾讯云通过开发和优化腾讯混元多模态大模型，以支持瑞金医院医生对医学文档的深度理解和处理。并整合文本、语音、图像等多种数据类型，实现对患者信息的全面分析和理解。

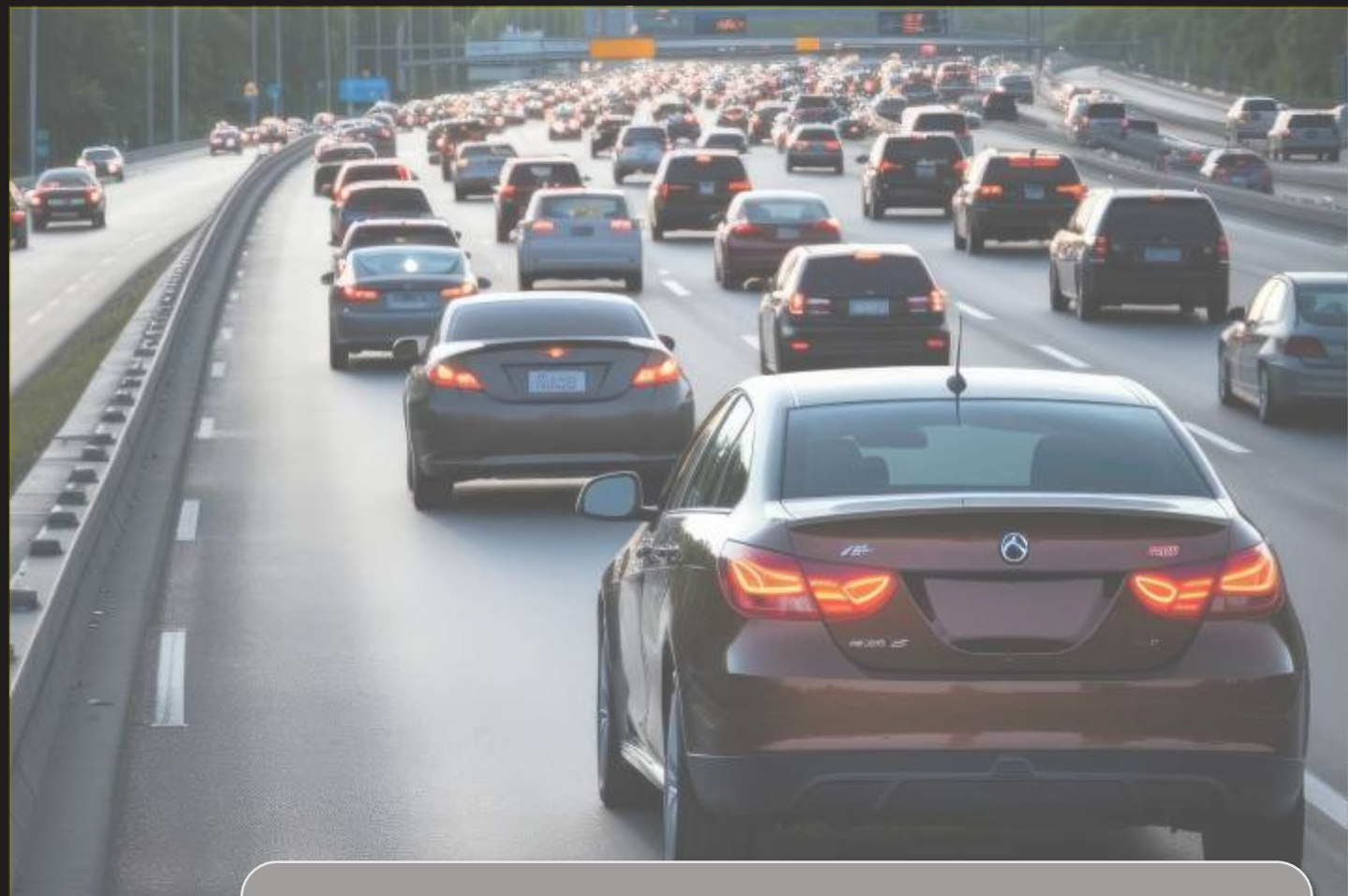
【瑞金医疗大模型】应用架构/核心优势示例

全链路解决复杂文档解析，切分，检索，推理，生成



优势分析

- 腾讯云构建了大模型知识引擎、图像创作引擎、视频创作引擎三大PaaS工具，将大模型技术封装，让各行各业的用户能直接上手使用。
- 通过大模型，医生可以深度理解患者的多模态信息（如文本、语音、图像），获取患者多维度的医学表现，模拟真实的医患交互过程。
- 大模型提供智能化的辅助决策支持系统，帮助医生在复杂病例中做出更精准的诊断和治疗决策，提升医疗服务质量。



汽车行业的发展挑战与最佳应用实践

- 生成式AI的引用重新定义了人机交互的形式，驱动汽车传统的指令式交互向主动服务模式转变，优化驾乘体验；此外，新技术的融入打通了汽车完整链路的数据，协助车企完善业务生态，打造“端到端”的感知决策一体化。
- 生成式AI正重塑汽车行业的创新格局，通过智能化设计、精准化营销和个性化服务，为汽车制造商带来前所未有的效率提升和用户体验优化。在这一变革中，智能座舱、路况检测和安全监控等应用成为AI技术应用的前沿领域。
- 生成式AI在汽车行业的应用虽然展现出巨大潜力，但同时也伴随着不可忽视的潜在风险。尽管企业普遍有能力应对生成式AI的战略和技术基础问题，但在科技伦理、数据治理及风险问题上信心不足，这些成为人工智能应用的重大障碍。

中国汽车行业最佳应用实践

阿里云X零跑

阿里云X零跑



客户痛点

智能座舱需求迫切

智能对话仅限于传统问答

缺乏开放统一的开发平台

核心评价关键词

全面性

前瞻性

可靠性

满意度

功能价值与适用性维度

智能驾驶核心能力：AI 驱动的感知、决策与控制技术已成为智能驾驶的核心，L2-L3 级别量产落地，L4-L5 正在研发。阿里云大模型与数据闭环技术支持端到端集成优化，保障复杂场景下的稳定性与可扩展性。

智能座舱与客服应用：通过语音交互、手势控制、情感识别与知识图谱，实现车内个性化体验与全生命周期智能客服，满足用户对安全、高效和个性化出行的需求。

技术性能与创新力

大模型引领智能驾驶：阿里云全栈 AI 云与灵骏智算集群提供超强算力支持，推动大模型在感知、决策、控制环节的深度融合，提升车辆自主性与安全性。

数据驱动创新：构建端到端数据闭环，打通研发、制造与用户使用环节，支撑动态保险定价、个性化订阅等新商业模式，推动产业链协同创新。

实施与支持

智能化应用落地：支持智能排程、自动化装配、质量检测与预测性维护，AI 视觉识别与工业机器人提升制造效率，降低生产成本。

培训与支持体系：阿里云生态为车企提供模型微调、应用开发与场景化落地的全链路服务，帮助研发团队快速上手并持续优化，缩短应用部署周期。

客户体验与满意度反馈

多维价值反馈：通过大数据与 AI 分析用户行为，实现车型推荐、故障预警与紧急救援等服务，增强用户安全感与信任感。

持续满意度提升：结合客户反馈与行业评价数据，阿里云 AI 不断优化智能驾驶与座舱体验，助力企业在激烈市场竞争中保持用户粘性与品牌价值。

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

阿里云X零跑

阿里云产品与平台

通义大模型

阿里云自主开发的大语言模型家族，自 2023 年推出以来经历了多个迭代版本，包括 Qwen2、Qwen2.5、QwQ-32B、Qwen2.5-Omni、Qwen3 等

百炼平台

阿里云提供的企业级大模型开发平台，支持模型训练、微调、部署、评测、RAG（检索增强生成）集成，以及零代码创建智能体（Agent）应用。该平台支持多模态模型调用、算力资源灵活配置，增强企业落地能力

技术能力与生态

强大的模型能力

- Qwen 系列模型支持超长上下文（如 QwQ-32B 支持 32K token），Qwen3 更支持 128K token，上线 reasoning 模式、coder、VL、audio 等多样能力。

广泛应用场景

- 包括企业客服、智能助手（如 Quark 浏览器、智能眼镜）、RAG 应用、内容生成、编码辅助等。LVMH 等国际品牌通过阿里云平台接入 Qwen，实现客服与供应链优化

开放与投资战略

- 阿里云通过开源 Qwen 系列权重（部分高级模型仍保持闭源），吸引开发者与企业用户，提高生态覆盖度。公司近期宣布将在未来三年投资人民币 3800 亿元，用于 AI 和云基础设施建设

阿里云的解决方案

阿里云通过百炼平台实现大模型落地零跑座舱，为用户的信息查询、车辆控制、娱乐与媒体服务、秒极生图等互动提供支持，提升用户的座舱体验。

□ 接入通义大模型实现开放式语音交互

- 阿里云通义大模型帮助零跑汽车改变了传统的固定形式的问答模式，支持用户与零跑智能座舱进行开放式语音交互（闲聊场景），进行自然、连贯的多轮对话，可秒级响应，同时结合企业知识库和互联网知识库，满足用户多元化的需求。

□ 基于语音调用通义万相实现秒级作画

- 零跑采用语音助手调用云端通义系列大模型，帮助用户通用语音调用通义万相实现文生图换壁纸，实现秒级作画，提升娱乐互动；支持语音查找如何使用汽车功能、规划路径等功能，丰富用户操作体验；知识库内容覆盖了零跑全系汽车知识和其他汽车品牌开放领域的信息。

□ 基于百炼构建大模型应用架构

- 基于百炼平台，零跑汽车构建了开放、可扩展的大模型应用架构，基于统一的大模型底座，实现了零跑座舱大模型应用场景的快速扩展与迭代，降低大模型应用的创新门槛与成本。

中国汽车行业最佳应用实践

科大讯飞X奇瑞汽车

科大讯飞X奇瑞汽车



客户痛点

汽车智能化水平继续提高

全场景 AI 交互需求明确

音频质感打造缺乏

核心评价关键词

智舱智驾融合

深度语义理解

声场重建优化

情感交互

功能价值与适用性维度

基于讯飞星火认知大模型，科大讯飞与奇瑞联合打造了Lion AI大模型平台。基于各自优势资源，在汽车智能座舱、智能音效、智能销服、国际多语种、智能驾驶、工业智能及企业数字化等七大领域全面深度合作

伯牙之音音响系统实现7.1.4全景声与225个音频焦点精准调控，独创头枕私享、虚拟现场等智能模式，打造移动音乐厅级的沉浸体验。

技术性能与创新力

Lion AI大模型加持AI语音助手，融合智舱与智驾能力，并包括营销服务场景、企业数字化升级、边缘计算和智能分析。

此外还为奇瑞汽车提供高质量的软硬件一体化汽车声学解决方案，涵盖智能音效系统、音频管理平台以及用户体验创新等多个领域。

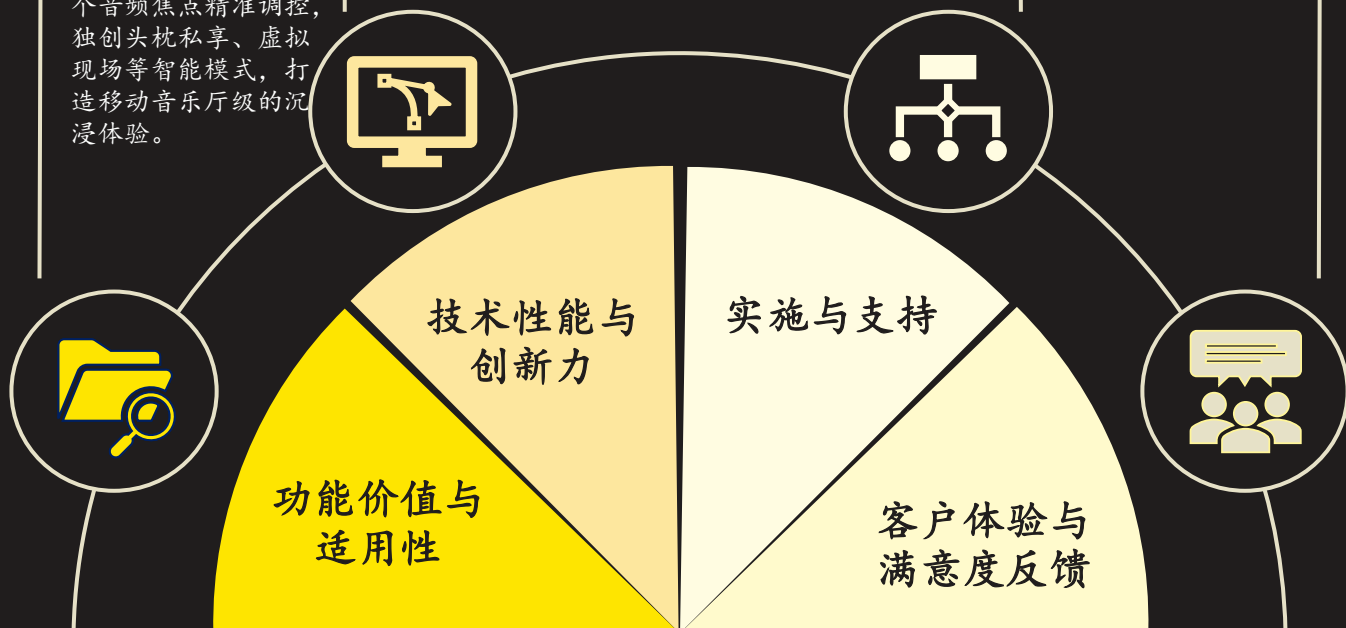
实施与支持

首搭星纪元ES，实现创新引领、能源管理、个性化服务、企业数字化、数据价值释放。

科大讯飞旗下安徽讯飞云创科技有限公司与奇瑞汽车旗下芜湖雄狮汽车科技有限公司成立合资公司——安徽祺声科技有限公司

客户体验与满意度反馈

基于大模型，奇瑞发力营销服务场景、企业数字化升级、边缘计算和智能分析三大方向，以供应链智能化创新、应用场景体验进阶和高效智能化决策为用户带来极致智能体验，进而奠定智能汽车大模型赛道新优势。

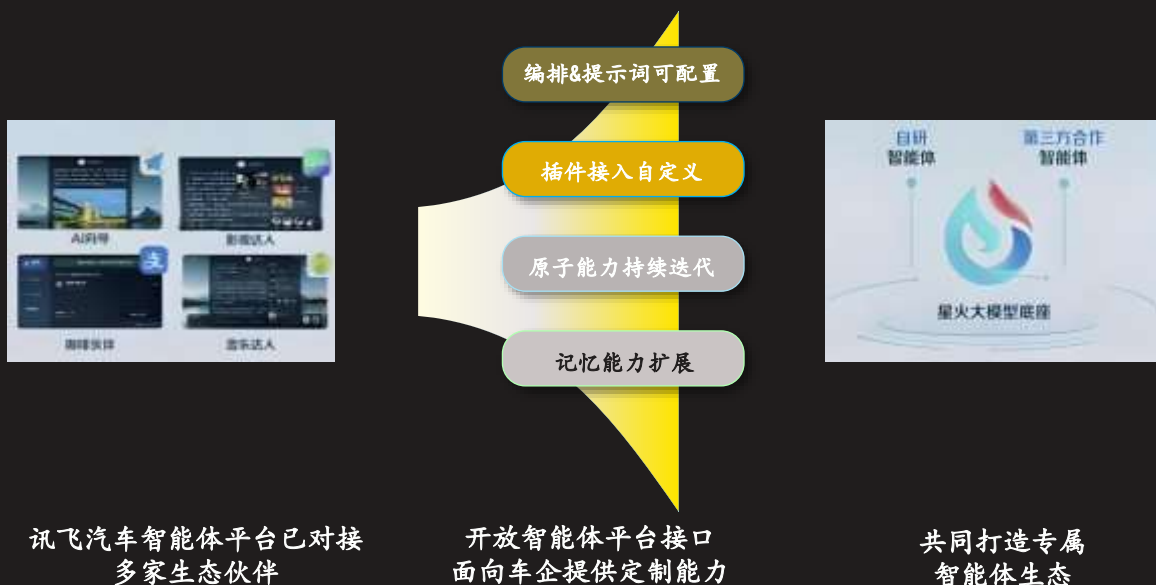


科大讯飞X奇瑞汽车

科大讯飞与奇瑞联合打造了Lion AI大模型平台。在汽车智能座舱、智能音效、智能销服、国际多语种、智能驾驶、工业智能及企业数字化等七大领域全面深化合作，通过单模到多模的生成式人工智能，推动智能驾驶与智能座舱融合。此外依托科大讯飞在人工智能大模型和声学技术方面的领先地位，以及奇瑞汽车在整车研发、量产及全球化市场拓展方面的深厚积累，联合打造了风云A9L的「伯牙之音」AI 音响，在车内带给用户百万豪车级的听享体验，共同打造高质量的软硬件一体化汽车声学解决方案。

应用架构/核心优势示例

科大讯飞与开发伙伴一起为车企构建丰富智能体生态



优势分析

- 在大模型的加持下，在人机交互表现中，任务与聊天将能实现无缝切换，同时，语音助手可实现深度语义理解，实现口语化表达理解，多意图分离，并能实现情感交互，实现风格切换、情感预测等功能。
- 大模型支持 **1800+** 功能点语音控制，覆盖全车 **90%** 的操作，满足全车人员娱乐、车控需求。
- 通过适配算法和精密调音，大模型协助奇瑞克服了车内复杂的声场环境，重建声场，突破车内空间限制，通过全音域优化覆盖，使得高音悠远、中音清晰、低音有力，还原器乐人声本真，让声音充满生命力。



能源与工业行业的发展挑战与最佳应用实践

- 在AI大模型在与能源场景深度融合中，对能源规划、建设、运营和管理带来了巨大变革。从能源行业的生产、分配与传输到消费环节，实现全流程的智能化升级与数字化转型。
- AI大模型在能源行业的应用不仅提高能源的生产效率和安全性，也为能源的可持续发展提供了强有力的技术支持，助推能源行业实现清洁、低碳、安全与高效。
- AI大模型目前已在能源各环节应用，在生产环节，大模型通过精准的数据分析与预测，实现能源生成过程的优化。此外，大模型对于设备的实时监控和维护，减少了故障停机的次数，大幅提高发电效率。在运输与运营环节，大模型技术的应用提高了电网的智能化水平，在稳定性与安全性层面全面提升，为电力分配和能源传输提供了强有力的技术支持。最后在消费环节，大模型可通过智能分析和推荐系统，优化能源的消费模式。此外，个性化的用户服务与智能问答系统，为消费者提供更好的消费体验。

中国能源与工业行业最佳应用实践

光明电力大模型——百度智能云X国家电网

光明电力大模型



客户痛点

海量数据中难提炼有效信息

波动性、随机性、间歇性

电力系统日趋复杂

核心评价关键词

新型电力系统建设

千亿级多模态模型

专业电力知识理解

全产业链服务

功能价值与适用性维度

基于国家电网的需求，百度智能云为国家电网提供了通用基础大模型与工具链平台、相关底层技术等重要支持，并在电网规划、电网运维、电网运行、客户服务等专业领域探索大模型应用，助力国家电网光明电力大模型的应用成果效果验证与落地推广。

技术性能与创新力

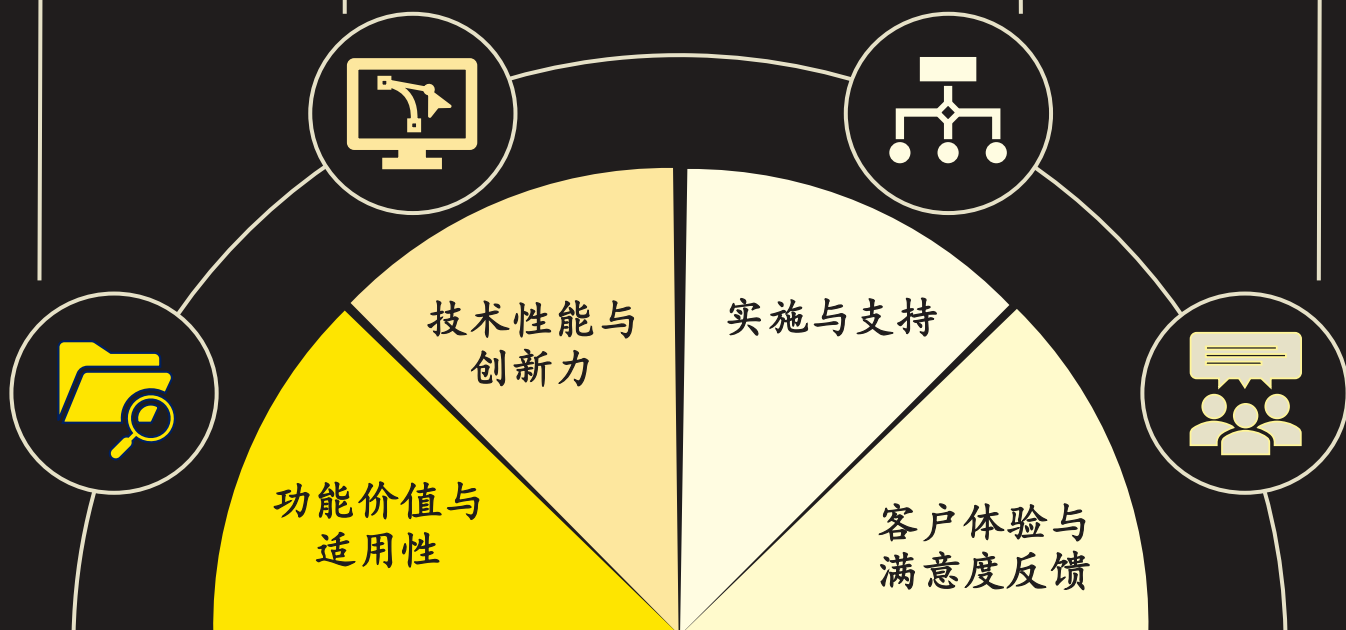
模型参数可达千亿级别，提升了模型的推理能力。实现了文字、图片、视频等多形态数据的融合分析，提升了模型的分析能力。此外通过光明电力大模型强大的智能交互和推理决策等能力，可快速生成满足调度运行高实时性、强可解释性需要的负荷转供策略。

实施与支持

具有波动性、随机性、间歇性的新能源大量并网，分布式光伏、新型储能、虚拟电厂等海量新型电力经营主体接入电力系统。面向复杂的业务形式与庞大的业务规模，提升电力提专决业数据和知识服务水平、加速模型应用于专业场景的效率，展行业级的人工智能平台工程化与体系化建设。

客户体验与满意度反馈

电网运维方面：光明电力大模型自动读取8类154个量测点实时和历史数据，逐一排除、综合分析，给出评估报告及建议。目前大模型已学习了2000多项故障案例，具备13项常见‘疾病’的诊断能力。



光明电力大模型——百度智能云X国家电网

为深入贯彻国家“人工智能+”行动，加快推进以大模型为代表的新一代人工智能技术融合应用，百度智能云支持国家电网打造了国内首个千亿级多模态电力行业大模型-光明电力大模型，共同推动电力行业数字化智能化转型升级，助力新型电力系统和新型能源体系建设。

应用架构/核心优势示例



优势分析

- 能源行业代表性大模型：光明电力大模型是能源行业大模型产业落地的典型代表，体现出了人工智能应用发展的场景优势。双方的深度合作是智能和能源的双向奔赴，助力国家“人工智能+”行动计划加速推进。
- 全周期的支持服务：百度智能云为国家电网提供了全周期的支持服务与完善的全流程解决方案，其中千帆ModelBuilder为国家电网提供电力大模型全流程模型开发工具链，并且实现大小模型一体化的开发、纳管和应用。此外，百度智能云提供的千帆AppBuilder支持高效开发电力行业企业级大模型应用，打通应用落地的最后一公里。

中国能源与工业行业最佳应用实践

羚羊能源大模型——科大讯飞X羚羊工业

羚羊能源大模型



客户痛点

提升能源效率

电力交易内容繁复

新能源功率稳定度低

核心评价关键词

产业链协同

能源6大行业场景覆盖

开箱即用的底座能力

低碳清洁

功能价值与适用性维度

广泛应用在能源安全生产监测等场景，解决传统机器学习方法样本需求量大、模型强定制、交付周期长、难复用等痛点问题。以升级版的讯飞星火大模型V4.0通用能力为核心技术底座，羚羊能源大模型2.0对能源视觉及时序大模型核心能力进行了重磅升级，能源数据及知识持续增强，工具链及落地实践持续完善，同时构建了大小模型统一纳管的AI平台，深度赋能能源发展。

技术性能与创新力

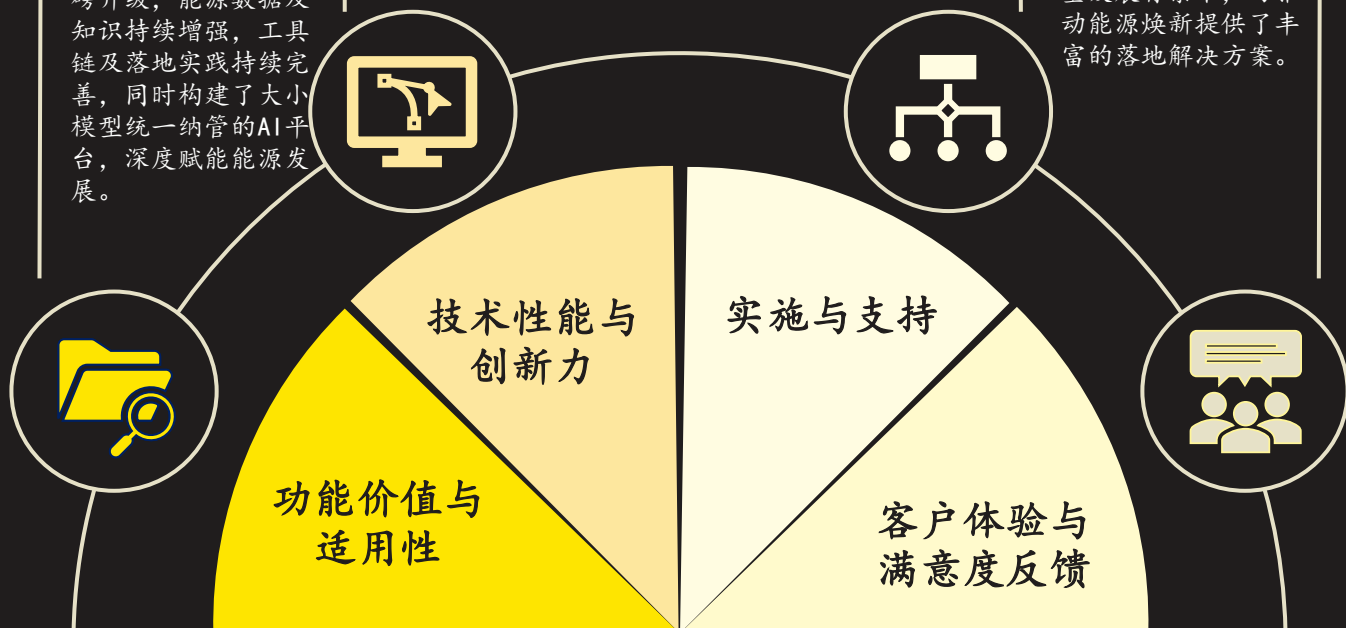
视觉大模型、时序大模型能力在能源绿色低碳背景下，为推动能源焕新提供丰富的解决方案。针对标注数据少、定制训练成本高、泛化性能弱等一系列小模型的痛点问题，打造了具备了低成本快速定义新功能、跨场景能力自适应迁移、深度语义视觉理解能力的视觉大模型。

实施与支持

智能体作为连接大模型与现实场景的载体，具备工具链调用、开箱即用、一站式开发、场景落地等重要作用，可在设备智能识别、设备配件匹配、销售数据分析等多个业务场景运用。

客户体验与满意度反馈

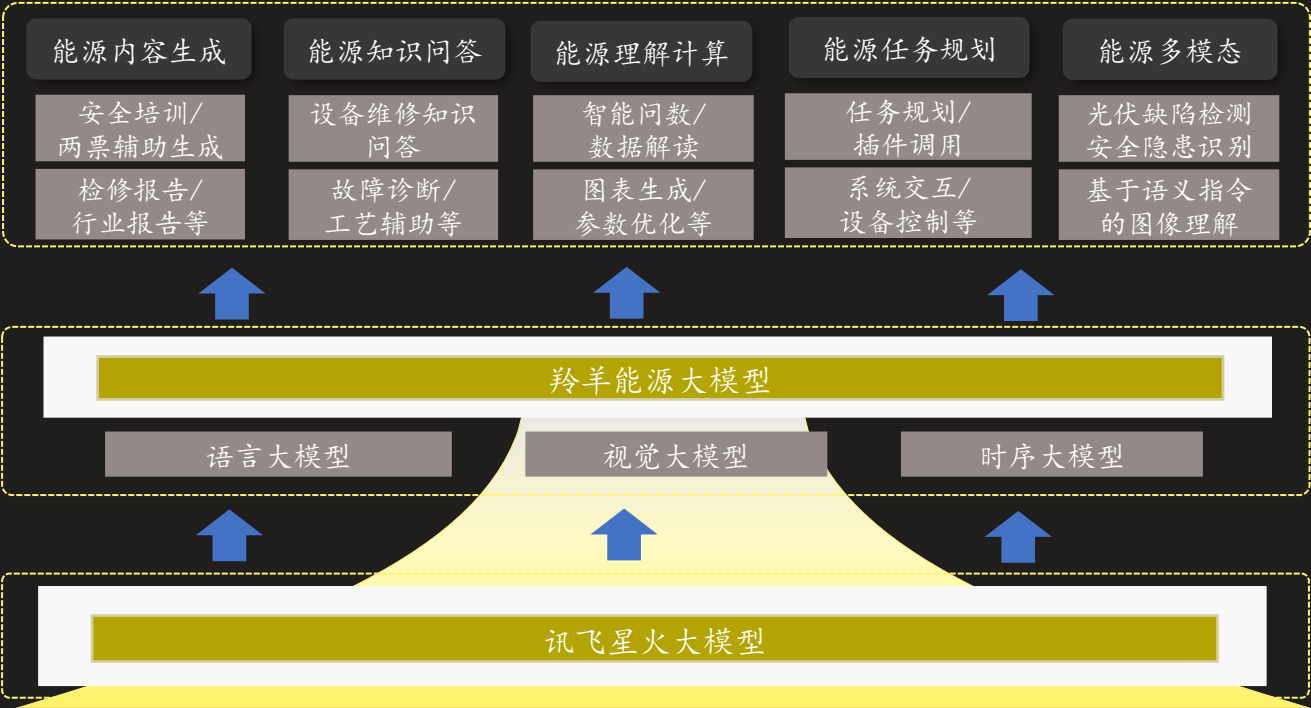
结合能源行业场景实际需求打造，具有能源内容生成、能源知识问答、能源理解计算、能源任务规划、能源多模态5大核心能力，拥有云边端协同、自主可控、数据与模型安全3大特性，涵盖风、光、水、火、核、储6大行业场景，在能源绿色低碳高质量发展背景下，为推动能源焕新提供了丰富的落地解决方案。



■ 羚羊能源大模型——科大讯飞X羚羊工业

羚羊能源大模型以讯飞星火大模型的通用能力作为核心技术底座，结合能源行业场景的实际需求打造了覆盖能源6大行业场景的专业大模型，在能源绿色低碳高质量发展的背景下，为推动能源焕新提供了丰富的落地解决方案。

应用架构/核心优势示例



优势分析

- 羚羊能源大模型五大核心能力：以讯飞星火大模型的通用能力为核心技术底座，结合能源行业场景实际需求打造，具有能源内容生成、能源知识问答、能源理解计算、能源任务规划、能源多模态5大核心能力。
- 3大模型特性：羚羊能源大模型拥有云边端协同、自主可控、数据与模型安全3大特性。
- 丰富的场景覆盖：羚羊能源大模型涵盖风、光、水、火、核、储6大行业场景，在能源绿色低碳高质量发展背景下，为推动能源焕新提供了丰富的落地解决方案。



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



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

提示词扩写与智能优化

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

依托快手自研多模态大模型“可灵”“可图”，平台具备高质量的视频与图像生成能力，并支持多语言输入满足全球需求。通过接入 Amazon Bedrock 与 DeepSeek，实现提示词扩写与生成优化，显著提升效率与可控性；再结合 AWS 的 EC2、S3、CloudFront 等架构，确保性能稳定与技术领先。

实施与支持

可灵AI已在 AWS Marketplace 上架，企业可便捷接入并快速落地。凭借 AWS 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。

技术性能与创新力

实施与支持

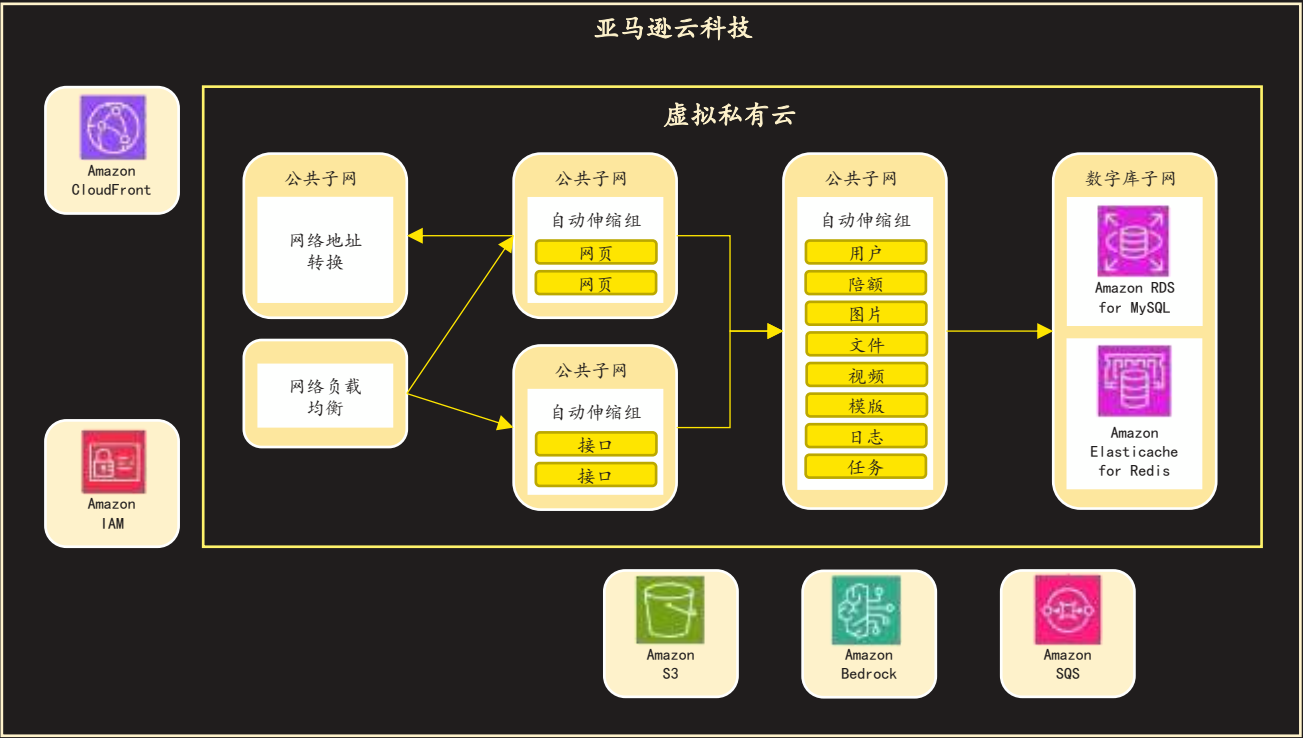
功能价值与适用性

客户体验与满意度反馈

■ 亚马逊云科技 × 可灵AI

快手旗下可灵AI与亚马逊云科技联合推出面向视频与图像创作的多模态生成平台“可灵”，基于文本、图像、视频等多模态数据与云端架构训练，显著提升对创作需求与生成质量的理解与可控性，构建涵盖文字转视频、提示词优化、模板与多语言支持、任务调度与存储分发等能力的智能平台，依托 AWS 的高并发与高可用，支持便捷接入与跨区域应用，满足广告、电商、游戏、娱乐等领域对高效低成本创作与规模化落地的需求。

可灵多模态视频生成大模型应用平台：技术架构图



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。

中国娱乐行业最佳应用实践

阿里云X永劫无间

阿里云X永劫无间

阿里云



永劫无间

客户痛点

延迟要求高

游戏稳定性要求高

游戏角色的自动托管

核心评价关键词

超低延迟

高并发场景

首发AI队友功能

多区域部署

功能价值与适用性维度

阿里云通过第八代 Intel 实例和 ACK One 方案，为《永劫无间》解决了低延迟、高并发和 AI 游戏功能的挑战，为游戏用户更稳定、更流畅的用户体验和更及时的数据分析能力。

技术性能与创新力

《永劫无间》手游的 Copilot 通过先进的算法和深度学习技术，能够分析战局中的每一个细节。从敌人的动向到队友的状态，再到地形和资源的分布，统统纳入考量范围。因此，Copilot 能够在瞬息万变的游戏环境中做出最优决策，为玩家提供精准的支持。

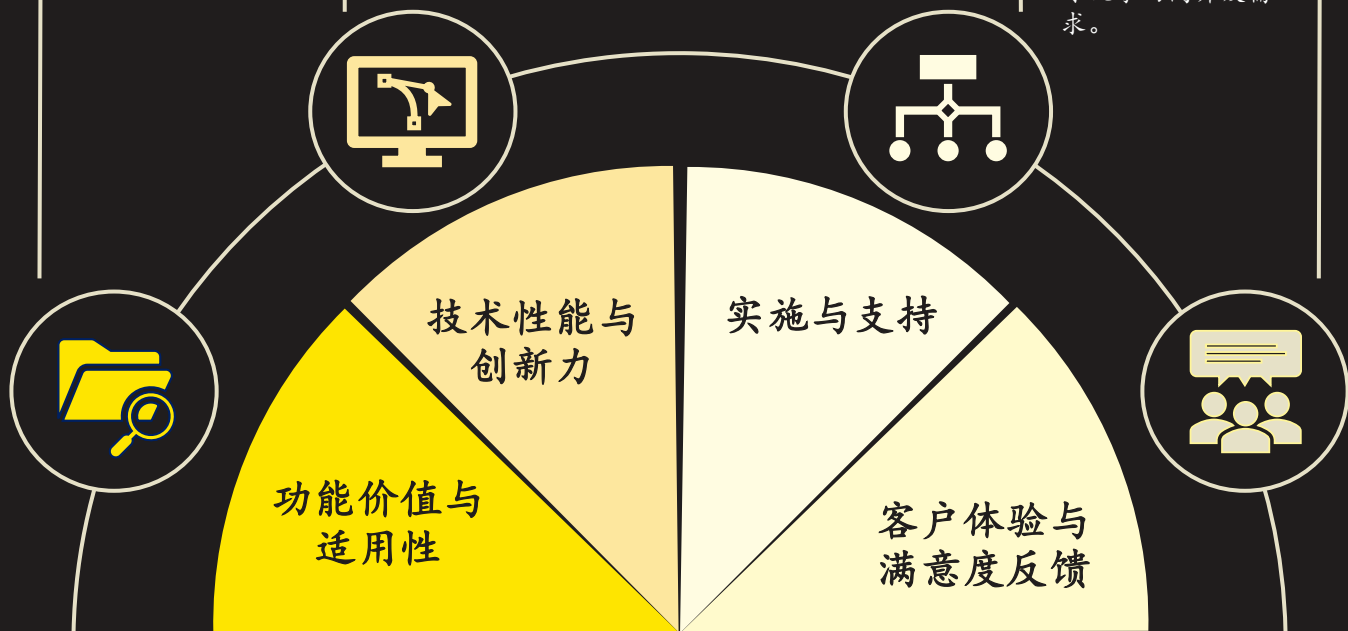
实施与支持

提供一致界面和API，对集群、计算、网络、存储、安全、监控、日志、作业、应用、流量等提供企业级运维管控能力。

面向云突发的业务波峰，业务自动或定时触发弹性扩容和缩容云上资源，降低综合业务成本。

客户体验与满意度反馈

多地域多集群智能资源调度能够跨地域跨集群寻找资源，最大限度避免缺货，提升资源利用率，提升业务效率。为《永劫无间》解决了低延迟、高并发和 AI 游戏功能的挑战，显著提升了游戏体验和运营效率，高效满足了全球玩家的高并发需求。



阿里云X永劫无间

网易公司与广大游戏热爱者一同成长，20 余年来推出数十款精品游戏，网易已跻身全球七大游戏公司之一。《永劫无间》是网易成功打造的又一个游戏大 IP。2021 年端游正式上线 Steam，《永劫无间》已与阿里云深度合作。2024 年 7 月《永劫无间》手游上线，游戏同时在线人数再次刷新记录。阿里云为《永劫无间》提供全面的云原生服务，以提升游戏体验和运营效率。

ACK One 应用架构/核心优势示例



优势分析

- 低延迟：《永劫无间》是一款动作竞技游戏，每一次精彩对战都需要后端服务同步交互，需游戏服务端提供更强的计算能力。高并发：全球预约超 4000 万，同时在线玩家百万级，周期峰谷特点，对后端服务器并发计算处理、云服务器的弹性能力有极高需求。AI 游戏：全球首发游戏 Copilot 队友功能，在听指挥、会配合、能闲聊的 AI 队友背后需要满足其实时性、模型高效迭代的 AI 基础设施。
- 多 Region 部署、玩家公平匹配就近调度，更好的战斗体验。
- ACK One 容器集群支持游戏 AI 服务构建大规模推理业务，一套模版发布多个虚拟集群，提升 AI 模型的部署与迭代效率。



企业服务行业的发展挑战与最佳应用实践

- 尽管生成式AI技术发展迅速，但在主流企业应用中仍处于早期阶段。部署生成式AI的企业需要自行完成大量繁重的工作，如筛选可用选项以及将其整合到业务流程中。此外数据隐私与机密也是企业关心的重点，如何保障在生成式AI工具使用过程中保护企业数据的机密性是企业成功的关键。
- 生成式AI协助企业突破了传统人工运维管理复杂的难点，驱动行业向流程优化、更快的问题洞察、任务准确率提升、办公效率提高等方向发展。
- 此外在企业服务的生成式AI产品中，AI Agent 愈发重要，它将AI从简单的命令生成器提升为真正的智能助手。Agent设计的核心理念是"目标导向"而非"命令导向"，用户只需描述想要达成的目标，Agent会自主规划和执行必要的步骤。



中国企业服务行业最佳应用实践

智能陪练解决方案——中关村科金X得助智能



客户痛点

传统课件制作成本高，且缺乏灵活性。

现有方案对能力评价不精准，建议质量低。

陪练效果缺乏真实感，且对练场景单一，无法触及长尾业务场景。

核心评价关键词

端到端解决挑战

大模型驱动性能卓越

灵活部署持续优化

效果量化客户认可

功能价值与适用性维度

该产品提供一站式企业培训解决方案。其核心功能在于将企业的隐性知识转化为显性学习资产，构建完整的培训闭环。它能快速应对新产品周期加速、合规要求严格等挑战。产品不仅适用于销售话术、客服对练，更覆盖培训全流程，有效提升企业整体效率。

技术性能与创新力

该产品以“大模型驱动”为核心，展现卓越性能。核心亮点是DeepSeek思维链技术，确保陪练对话的真实和逻辑。内容生成（AIGC）效率极高，15秒生成一份考卷，剧本搭建提效76%。它构建了多维度精准评估体系，准确率高达92.1%，提供比人工更客观的分析。

实施与支持

产品支持灵活部署和多样使用方式。它支持PC端、手机H5、飞书、钉钉等多种渠道接入，学员随时随地进行学习。在架构层面，支持私有化部署，满足数据安全问题。产品还能通过对海量会话数据分析，实现陪练模型自我优化与迭代，确保培训内容持续有效。

客户体验与满意度反馈

该产品为客户带来可量化业务成效，极大地提升了用户体验。在具体实践中，客服试岗时长缩短50%，首月话术违规率降低24%。学员获得“千人千面”的个性化指导，客户高度认可其效果，认为“已经全面超越了培训老师”，体现了产品的实用价值。

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

智能陪练解决方案———中关村科金X得助智能

中关村科金的解决方案实践

大模型升级

在课件生成与多轮对话推理中探索，结合培训各流程业务场景切入大模型的优势能力，大模型的深度学习能力使课件内容更加丰富多元生成时效更快，同时，在多轮对话中，大模型凭借精准推理，提供流畅自然的交互体验，大幅提升实战陪练中的培训效果。

客户挑战

行业趋势：金融产品创新周期加速，证券行业新产品从季度级更新提速至月级，传统培训体系难以跟进

业务挑战：很多理财经理在客户咨询新产品时无法提供专业解答，影响客户信任与转化

合规压力：监管要求持续提升，传统合规培训方式下违规事件仍高发，每年造成显著合规风险与声誉损失

合规风险识别与
预警

实时专业知识辅
助

千人千面的推荐
训练

实践
展示

知识库提取

选择客户/场景

个性化指导

实际练习



新一代得助智能陪练产品架构



中国企业服务行业最佳应用实践

火山引擎Data Agent

火山引擎DataAgent



客户痛点

数据价值释放难

分析门槛及成本高

决策依赖经验

核心评价关键词

企业级大模型底座

智能模型接入与算力优化

多模态解析与知识管理

功能价值与效能提升

功能价值与适用性维度

核心功能完整性：具备自然语言交互、多模态数据融合分析、自动化策略生成（如“一客一策”）、闭环任务执行等核心功能的成熟度与协同性，以满足实际业务闭环。

场景功能泛化性：灵活适配电商、金融、零售、SaaS等多业务场景，并支持跨场景的快速扩展与复用，保障技术的长期价值与投资回报。

技术性能与创新力

多模态融合能力：对文本、图像、音视频及结构化/非结构化数据的高效集成与跨模态分析，这是实现复杂业务智能化的技术基础。

生成内容的合规安全：数据隐私、模型合规性及生成内容的伦理风险控制，包括自动化监测、应急响应制度及安全审计流程。

实施与支持

使用成本优化：其云原生架构与轻量化模型设计的资源效率高，全生命周期部署、运维及迭代成本，且提供高性价比的规模化落地方案。

智能体应用：在客服、运营、营销等场景中自动化决策与执行，准确性、响应效率，以及对业务流程的实际优化贡献高。

客户体验与满意度反馈

体验和定制化：具备交互设计的直观性（如自然语言查询易用性）、API集成效率，以及参数调整、模块扩展等定制化能力，满足企业增长需求。

满意度追踪情况：持续追踪多角色用户满意度，识别共性痛点并驱动产品优化，形成“需求-反馈-迭代”的正向闭环。

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

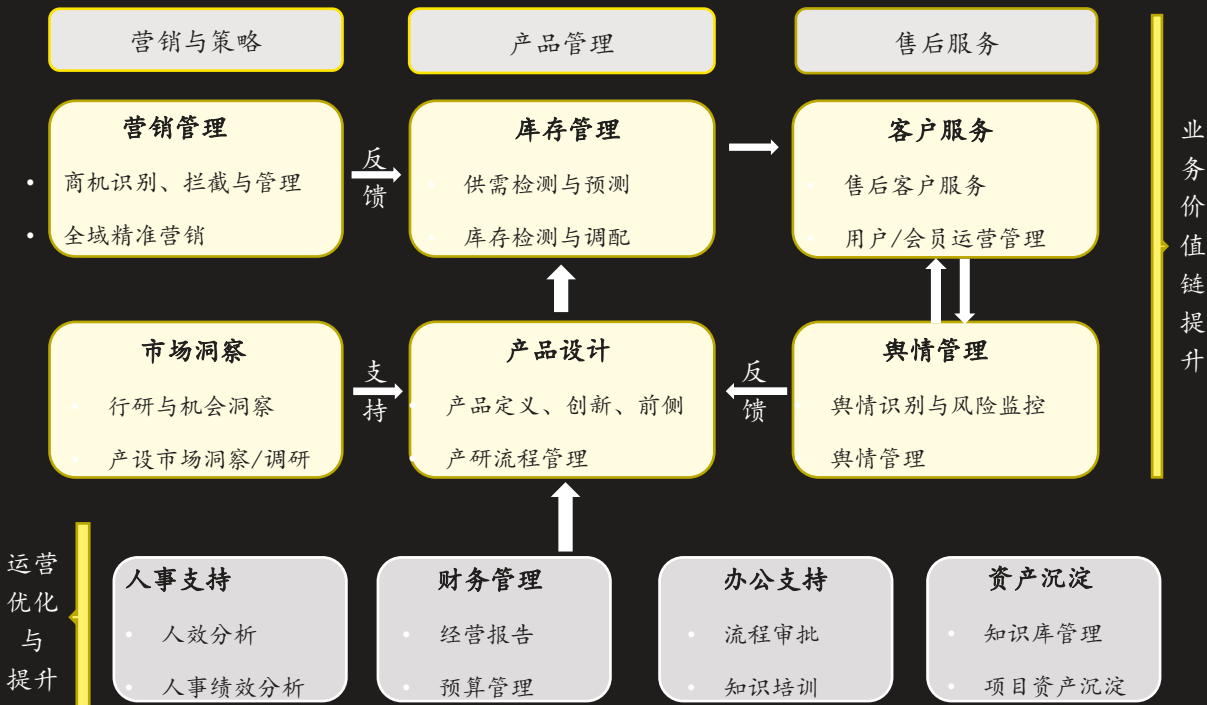
火山引擎DataAgent

DataAgent是火山引擎打造的企业级AI数据智能体，定位为企业的“第一位AI数据专家”。基于大模型与多模态理解技术，通过对企业内外部结构化与非结构化数据进行深度挖掘与分析，实现智能洞察、自动策略生成与“一客一策”精准营销，旨在降低数据使用门槛、提升决策效率与营销效果，助力企业实现数据驱动的智能增长。

火山引擎DataAgent：“数据+知识”的双向融合，赋能企业实现数据驱动的业务增长。



业务价值链条





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。



中国企业服务行业最佳应用实践

Chaterm——合合信息 × AWS

Chaterm



INTSIG 合合

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

客户痛点

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——合合信息 × AWS

Chaterm 是合合信息联合 AWS 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 AWS 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



主要特点



中国娱乐行业最佳应用实践

Vidu视频大模型——百度智能云X生数科技

Vidu视频大模型



客户痛点

视频生产一致性问题

底层算力基础设施需求大

高效运维与资源管理

核心评价关键词

全链路保障

领先的算力能效

多主体一致性

高动态高逼真高一致

功能价值与适用性维度

借助百度智能云的高性能引擎，Vidu也实现 AI特效等创新互动玩法的开发与应用，通过开放 API 支撑起大批泛互娱乐、图视频编辑、社交直播等应用型产品的流量增长。

产业协同发展方面，Vidu成为了首个接入百度智能云千帆大模型平台的视频模型。

技术性能与创新力

生数科技通过百度智能云的弹性队列、弹性伸缩、潮汐混部等能力实现整体资源分配率达到96%以上，资源利用率95%以上；通过训练容错能力实现秒级故障感知，分钟级定位恢复，保障有效训练时长达到99.5%以上。

实施与支持

百度百舸平台把异构算力进行统一管理，通过性能优化、自动选型等经验技术，大幅提升智能算力效能。提供了丰富的运维和可观测工具以及容错保障能力，提升了大规模集群长期运行的稳定性，降低因为故障导致训练任务异常终止的概率，减少业务损失。

客户体验与满意度反馈

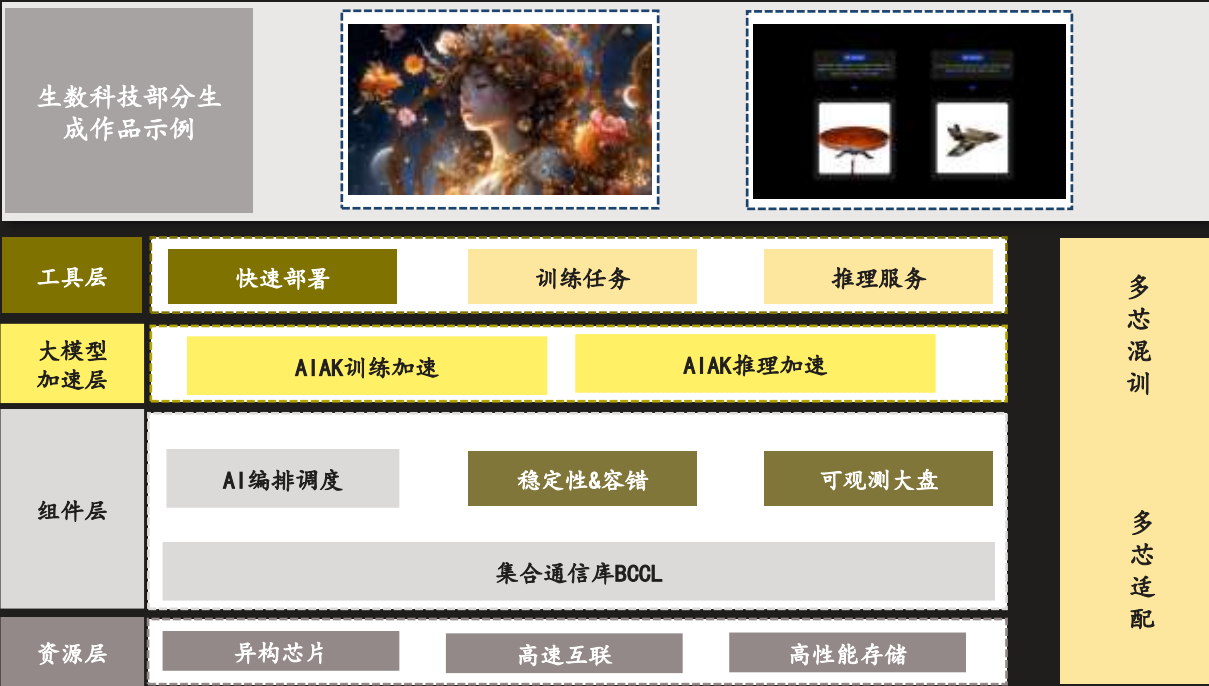
Vidu 简化了创作流程，提高了效率，让创作者得以专注于内容创意本身。产品中封装的任务自动容错机制确保用户的训练任务能够保持连续性和稳定性，不会因底层硬件故障而中断，从而避免资源浪费。



Vidu视频大模型——百度智能云X生数科技

基于百度百舸·AI 计算平台，百度智能云为生数科技提供了「快稳省」的大模型训推一体化基础设施，支持了国内第一个类 Sora 大模型 Vidu 的发布，已受到全球用户的广泛使用。同时，百度百舸支持了 Vidu 的持续迭代，Vidu在7月份全球上线之初，就率先突破了人物面部一致性难题，在9月份突破单主体一致性难题，并在11月再度实现多主体一致性能力。

应用架构/核心优势示例



优势分析

- 行业突破性技术：Vidu突破了视频模型从未被解决的‘一致性’难题，新增多图参考功能。仅需几张设定图，即可高效生成连续多个视频内容片段。
- 业界最快推理：Vidu实现了业界最快的推理速度，不到30秒就可以生成一段视频内容，与业界平均3-5分钟的生成速度比，实现了量级性的提升。
- 协助创作者提效：Vidu 简化了创作流程，提高了效率，让创作者得以专注于内容创意本身。
- 资源浪费减少：产品中封装的任务自动容错机制确保用户的训练任务能够保持连续性和稳定性，不会因底层硬件故障而中断，从而避免资源浪费。

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

江林烨、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

