FROST & SULLIVAN
沙利文

头豹
LeadLeo

2025年
中国大模型年度评测

AI变革行业创新发展

2025年03月

头豹研究院
弗若斯特沙利文咨询（中国）

观点摘要——大语言篇

01

中国大模型与国际差距加速收敛

2025年大模型年度评测结果显示，中国头部大模型整体评分已接近国际均线，排名前八的中国大模型平均得分几乎与海外顶尖模型持平。中国大模型在核心能力上已进入全球领先梯队，技术差距正在快速缩小。

02

大模型已成为“知识百科专家”

本次评测结果显示，所有参评大模型在常识、科学等知识类问题上的表现几乎达到满分，覆盖从基础常识到高阶科学问题的各类测试。这表明当前大模型在知识掌握方面已无明显短板，能够胜任“知识百科专家”的角色。

03

深度推理与数学是模型实力的重要分水岭

评测数据表明，大模型之间在逻辑推理与数学能力上的表现差距最为显著，在0-100的评分体系下，最大分差高达50分。这一现象凸显了推理与数学能力成为了衡量大模型实力的重要分水岭。

04

中国大模型的性价比远超国际大模型

本次评测数据显示，中国第一梯队大模型在整体得分超越国际大模型的情况下，其推理与生成成本却远低于海外竞争对手。中国领先大模型每100万token的平均价格仅38.2元，而国际大模型均价高达158.3元，形成近5倍的成本优势，展现出中国大模型在效率与性价比上的显著竞争力。

观点摘要——多模态篇

01

多模态理解能力整体尚处于发展阶段，识别准确率低于80%

在多模态理解能力的评测中，所有参评模型在各类图片和类型的整体识别准确率均未超过77%，其中最优模型的表现也未达到85%，显示出当前多模态理解在实际应用中的识别精度仍有较大提升空间。

02

多模态理解的核心挑战是物体定位

在多模态理解的九大细分维度中，物体定位维度的识别准确率最低，平均正确率仅为44.3%，物体精确定位依然是当前多模态理解技术的关键瓶颈。

03

模型的艺术创作能力显著优于商业创作能力

根据本次多模态生成的评测结果，所有模型在艺术性创作方面的均分为74.3，商业型创作的均分则为69.5，表明模型在满足美感和创造性等需求时表现较好，但在准确度和商业应用场景的适配性方面仍需进一步优化。

04

多模态生成的核心短板是指令遵循与文字生成

当前多模态生成面临两大主要问题：首先，模型在遵循指令方面存在频繁偏差，生成的图片与需求之间有一定程度的不符；其次，大部分模型无法准确生成文字。这些问题显著限制了多模态技术在更广泛应用场景中的可行性和发展潜力。

报告说明

沙利文联合头豹研究院谨此发布中国人工智能系列报告之《2025年大模型年度评测》报告。本报告全面解析中国大模型在大语言能力与多模态理解方面的最新表现，系统梳理过去一年国内大模型的技术进展、核心突破、短板挑战及应用落地情况。通过详尽的数据分析与专业评测，本报告旨在为行业决策者、投资机构、技术研发团队等提供深度洞察，助力精准研判产业发展趋势，推动大模型技术在实际场景中的优化与创新。

沙利文及头豹研究院发布的《2025年中国大模型年度评测报告》旨在全方位评估大模型在语言与多模态能力上的技术实力与应用进展。报告在2024年大语言模型评测的基础上，新增了对多模态理解与生成能力的深入考量，聚焦大模型技术的前沿突破及其在各行业深度融合的广泛影响。通过深入分析技术发展、市场竞争及创新趋势，报告为行业提供客观、专业的战略指导，助力各方把握未来技术变革的核心机遇。

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。

研究框架

章节一
中国大模型行业发展综述

- 大模型从文本向着多模态的发展历经三阶段：初期聚焦于模态理解与关联，中期扩展至模态生成能力，高级阶段实现任意模态转换与智能融合，逐步接近人类多模态智能水平。
- *人工智能技术的应用有效提高了工作效率，优化了工作流程，尤其在处理重复性工作和高脑力思考任务方面表现突出。目前，96.3%的人认为人工智能提升了工作效率，其中43%的人认为效率提升在20%-40%之间，23.9%的人认为提升幅度在40%-60%，21.2%的人认为提升幅度在60%-80%，而有4.8%的人认为效率提升超过80%。*
- *AI技术在文本和图像生成及理解上虽取得进步，但在语言风格、创造力、连贯性、错误率、复杂场景处理及细节真实性等方面，与人工相比仍有差距，需进一步提升技术水平和加强伦理考量*
- *目前，在多模态理解上，文本理解的技术最为成熟，广泛应用于搜索引擎、对话系统和内容推荐，市场渗透率高。而图像理解紧随其后，在医疗影像、自动驾驶、安防等领域已取得显著成果，但在通用场景中的性能仍待提升。其次音频理解和视频理解技术正在快速发展，音频理解在语音助手和客服领域应用成熟，但视频理解因计算复杂度高，应用多集中在短视频推荐和监控分析等特定场景，整体市场渗透度相对较低。*

中国大模型行业发展综述——发展路径

关键发现

大模型从文本向着多模态的发展历经三阶段：初期聚焦于模态理解与关联，中期扩展至模态生成能力，高级阶段实现任意模态转换与智能融合，逐步接近人类多模态智能水平

大模型的发展历程

初始探索阶段

多模态大模型的发展始于对不同模态间数据转换和关联的理解。早期研究集中在图像-文本、视频-文本和音频-文本的理解上，如BLIP-2、LLaVA等模型，它们能够识别图像、视频或音频中的信息，并与文本描述相对应，为后续的多模态交互打下基础。

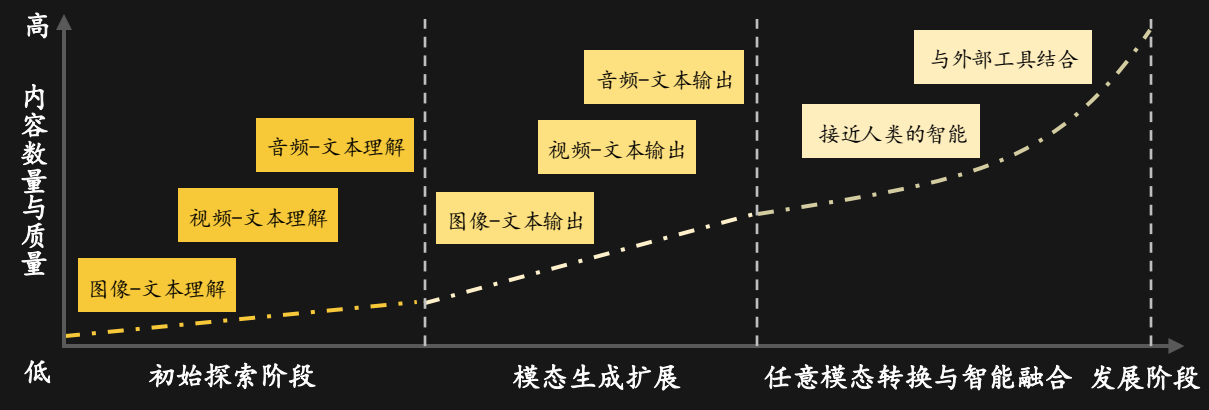
模态生成扩展

随着技术进步，多模态大模型开始支持特定模态的生成，如根据文本描述生成图像（GILL、Kosmos-2）或根据语音生成文本（SpeechGPT）。这一阶段标志着模型从理解向生成能力的拓展，进一步丰富了多模态交互的应用场景。

任意模态转换与智能融合

最新的发展聚焦于实现任意-任意模态的转换，研究者通过结合大型语言模型与外部工具（如搜索引擎、图像处理软件），使多模态大模型能够理解和生成来自不同模态的信息，如Visual-ChatGPT等，其智能水平逐渐接近人类，为更广泛的应用场景提供了可能。

➢ 多模态大模型性能的发展演变



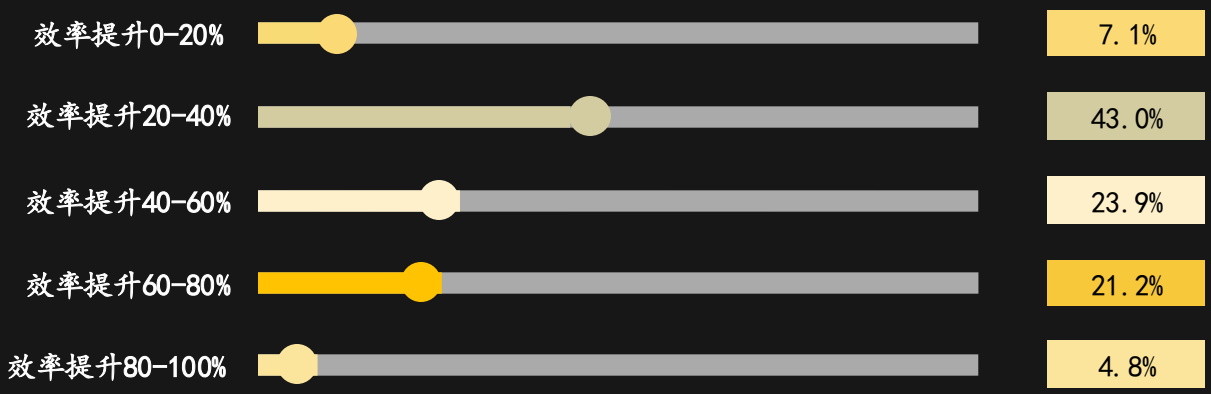
来源：沙利文、头豹研究院

中国大模型行业发展综述——价值效益

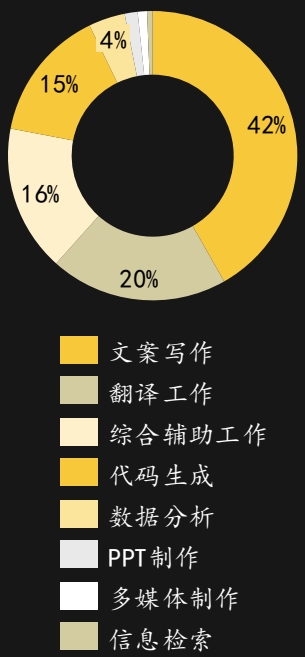
关键发现

人工智能技术在文案写作、绘画、视频及音频生成等方面显著提升效率、降低成本，优化工作流程，同时保留手工创作的独特价值于高端定制领域，为各行业带来革命性变化

AIGC提升工作效率区间，2023年



AIGC 辅助从事的具体工作内容，2023年



■ 人工智能技术应用广泛，显著提高工作效率，优化工作流程，正逐步改变各行业的工作方式，提升整体效率

人工智能技术的应用有效提高了工作效率，优化了工作流程，尤其在处理重复性工作和高脑力思考任务方面表现突出。目前，96.3%的人认为人工智能提升了工作效率，其中43%的人认为效率提升在20%-40%之间，23.9%的人认为提升幅度在40%-60%，21.2%的人认为提升幅度在60%-80%，而有4.8%的人认为效率提升超过80%。

人工智能的应用领域涵盖了多种任务，其中文案写作是最为广泛的应用方向，占比高达41.77%。通过自然语言处理技术，人工智能能够帮助用户快速生成高质量的文案内容，这不仅减轻了员工的负担，也大大提高了工作效率。特别是在内容创作密集的行业，如广告、营销、媒体和教育等，人工智能的应用让文案写作从传统的人工创作模式转变为高效、智能化的过程，从而节省了大量时间并提高了内容创作的质量。翻译工作的应用占比为19.85%，这表明人工智能在帮助组织打破语言壁垒，促进国际化交流和合作方面发挥了重要作用。其高效且准确的翻译能力，使得全球化合作变得更加顺畅。

此外，代码生成（14.8%）和综合辅助工作（16.39%）也是人工智能应用的重要领域。这些领域通常需要专业技能和大量时间投入，而人工智能化技术的应用使得这些工作变得更加轻松高效。尽管在数据分析（4.08%）、信息检索（0.6%）、PPT制作（1.45%）和多媒体制作（1.04%）等领域的应用占比较低，但这些领域的应用也证明了人工智能技术的广泛适用性。总之，人工智能的普及和应用正在改变工作方式，提高各行业的工作效率。

来源：沙利文、头豹研究院

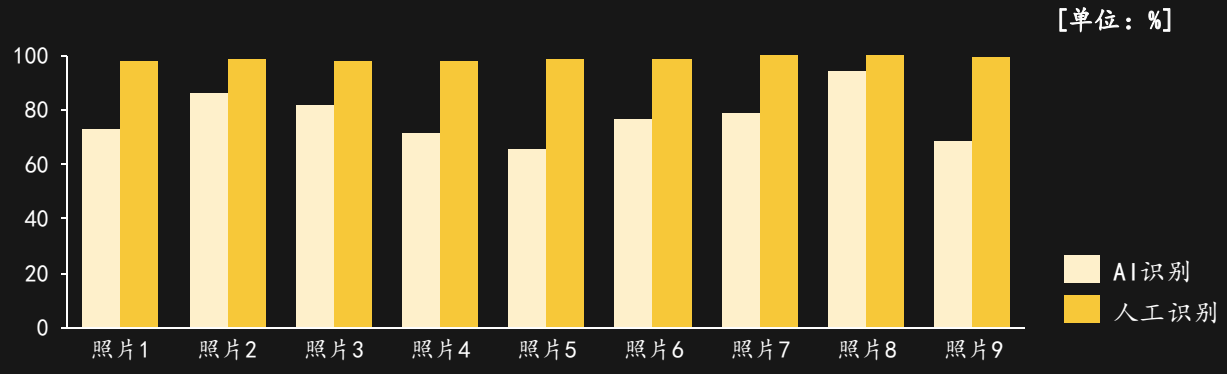
中国大模型行业发展综述——发展痛点

关键发现

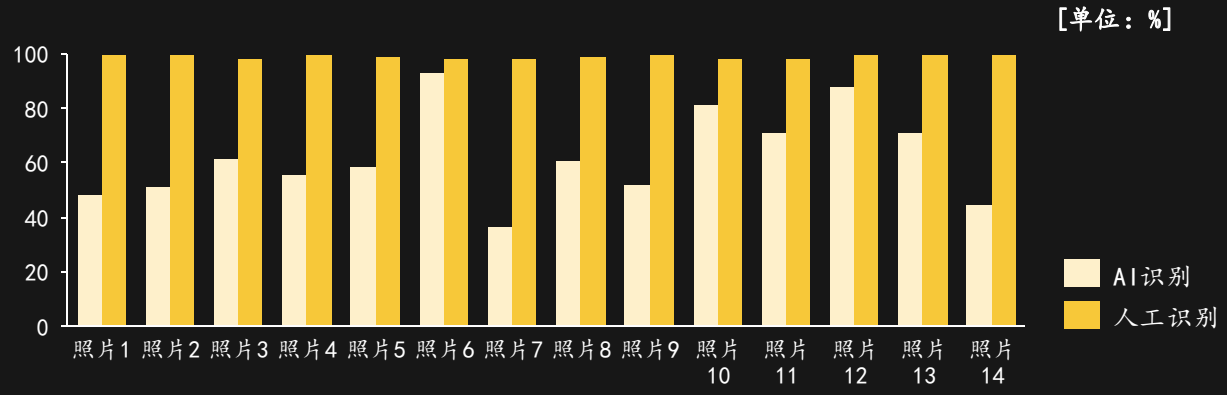
AI技术在文本和图像生成及理解上虽取得进步，但在语言风格、创造力、连贯性、错误率、复杂场景处理及细节真实性等方面，与人工相比仍有差距，需进一步提升技术水平和加强伦理考量

人工智能识别与人工识别的差异

➢ 人工智能识别与人工识别红外相机动物影像准确率比较（简单场景）



➢ 人工智能识别与人工识别红外相机动物影像准确率比较（复杂场景）



■ **人工智能识别虽取得进步，但在总体准确率和复杂场景应对上，目前仍无法超越人类的高水平识别能力**

人工智能识别总准确率为69.0%，均值为68.2%。人工识别总准确率为99.0%，均值为99.1%。人工识别准确率显著高于人工智能识别准确率。具体来看，在简单场景中，AI识别准确率为77.3%，人工识别准确率为98.9%。在复杂场景中， AI识别准确率为62.31%，人工识别准确率为99.1%。简单场景通常指的是那些背景信息较少、干扰因素较少、物体特征明显的情况，对于这类任务，AI已经能够达到一个相对较高的准确度，但仍难以匹敌人类几乎无误的表现。而在复杂场景中，AI识别的挑战进一步加大，其准确率下降至62.3%。复杂场景可能包含更多的变量，例如光照变化、遮挡、多角度视图、相似物体之间的区分等，这些都增加了识别的难度。然而，即便是在这样复杂的环境中，人工识别依然保持了极高的准确性，达到了99.1%。这反映了人类在处理不确定性和模糊信息方面的独特优势，以及在复杂环境下做出正确判断的能力。综上所述，尽管人工智能在特定领域的某些方面已经取得了令人瞩目的进步，但在总体识别准确率和应对复杂场景的能力上，目前还无法超越人类。

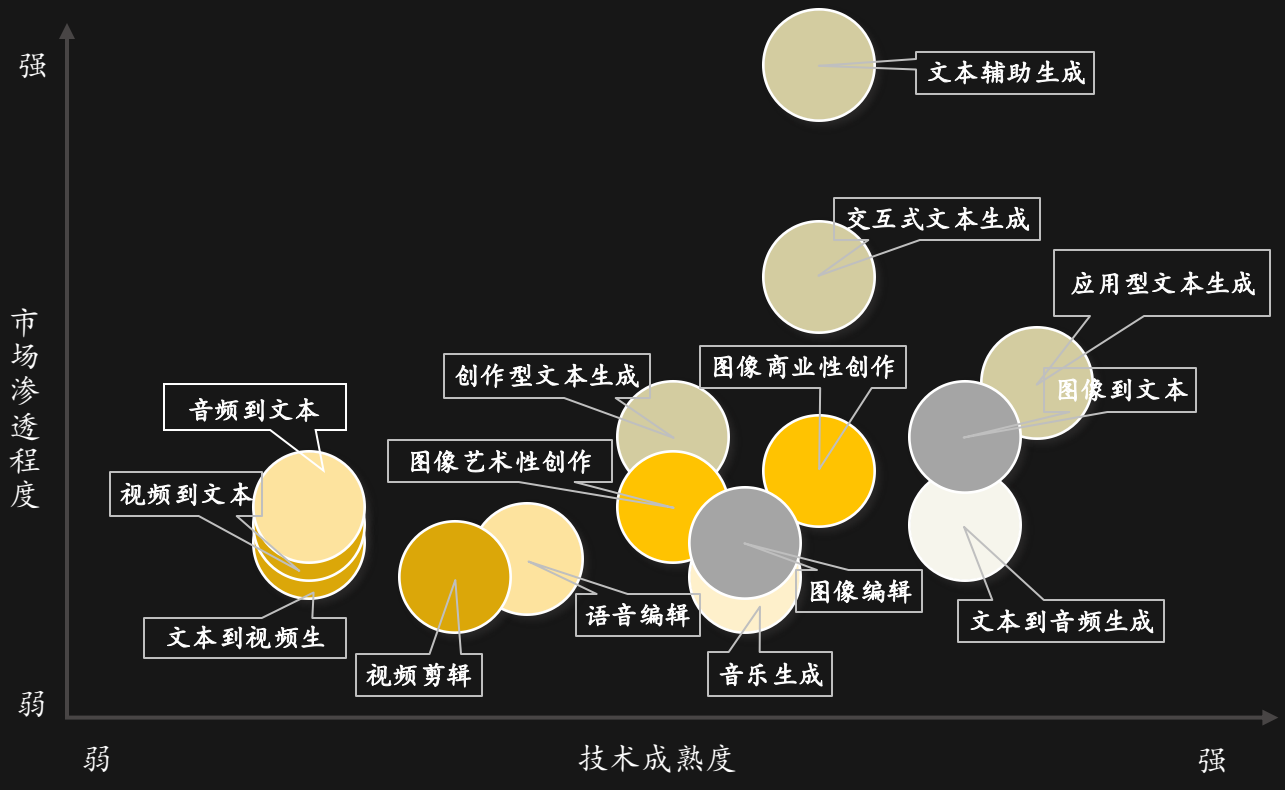
来源：沙利文、头豹研究院

中国大模型行业发展综述——技术成熟性

关键发现

多模态技术中，文本理解与生成技术成熟度高；图像、音频技术快速发展；而视频技术因计算复杂度高，尚需突破。整体而言，多模态生成技术潜力巨大，但广泛应用前需解决计算成本和质量等瓶颈

AIGC各模态技术成熟度和市场渗透程度分析



- **多模态技术中，文本理解与生成最为成熟，图像、音频技术快速发展，视频技术因复杂度高尚需突破**

目前，在多模态理解上，文本理解的技术最为成熟，广泛应用于搜索引擎、对话系统和内容推荐，市场渗透率高。而图像理解紧随其后，在医疗影像、自动驾驶、安防等领域已取得显著成果，但在通用场景中的性能仍待提升。其次音频理解和视频理解技术正在快速发展，音频理解在语音助手和客服领域应用成熟，但视频理解因计算复杂度高，应用多集中在短视频推荐和监控分析等特定场景，整体市场渗透度相对较低。

而在多模态生成上，文本生成技术最成熟，广泛应用于内容创作、对话系统和辅助写作，市场渗透率高。图像生成技术快速发展，得益于扩散模型和生成对抗网络，在设计、艺术创作和广告领域已有较高应用，但生成质量和通用性仍有提升空间。音频生成技术在音乐创作、语音合成等领域逐渐成熟，市场需求持续增长。视频生成技术因计算复杂度高、生成质量要求高，目前尚处于初步探索阶段，应用集中在短视频特效和动画生成，市场渗透率相对较低。总体而言，多模态生成技术仍在快速迭代中，未来潜力巨大，但在广泛应用之前仍需突破计算成本和内容质量等瓶颈。

来源：沙利文、头豹研究院

章节二
大模型年度评测背景介绍

- 大模型技术已进入全面发展阶段，竞争格局正从百花齐放逐步过渡到稳定发展。目前，中国在通用基础大模型领域的竞争者已减少至约20家，主要由互联网企业、云计算巨头和人工智能创业公司主导。在技术层面，多模态理解与生成能力取得了显著突破。头部大模型普遍具备对图像、文档、音频等多模态的理解能力，且多模态生成技术也在快速进步。与2023年相比，2024年大模型的生成能力大幅提升，尤其是在多模态生成领域的全面增强，极大拓宽了应用边界。这一进展不仅促使传统大模型厂商加大投入，还吸引了跨界垂直领域的图片和视频企业积极参与竞争，推动了市场竞争和技术创新的加速。
- *在此背景下，沙利文及头豹研究院发布2025年中国大模型年度评测报告。该报告在2024年大语言模型评测的基础上，新增了对多模态理解与生成能力的全面评估，旨在全方位衡量大模型在语言能力与多模态能力两个维度的技术实力与应用进展。沙利文将持续跟踪中国大模型领域的最新动态，为行业提供客观、专业的指导与参考。*

大模型年度评测背景介绍——评测背景与参与者

评测背景

大模型技术已进入全面发展阶段，竞争格局正从百花齐放逐步过渡到稳定发展。目前，中国在通用基础大模型领域的竞争者已减少至约20家，主要由互联网企业、云计算巨头和人工智能创业公司主导。

在技术层面，多模态理解与生成能力取得了显著突破。头部大模型普遍具备对图像、文档、音频等多模态的理解能力，且多模态生成技术也在快速进步。与2023年相比，2024年大模型的生成能力大幅提升，尤其是在多模态生成领域的全面增强，极大拓宽了应用边界。这一进展不仅促使传统大模型厂商加大投入，还吸引了跨界垂直领域的图片和视频企业积极参与竞争，推动了市场竞争和技术创新的加速。

从应用层面看，大模型的应用已不再局限于对话助手和简单的通用内容创作，逐步渗透至自动驾驶、医疗影像分析、3D角色生成等行业深度应用场景，展现出在多个行业领域的广泛潜力与商业价值。

在此背景下，沙利文及头豹研究院发布2025年中国大模型年度评测报告。该报告在2024年大语言模型评测的基础上，新增了对多模态理解与生成能力的全面评估，旨在全方位衡量大模型在语言能力与多模态能力两个维度的技术实力与应用进展。沙利文将持续跟踪中国大模型领域的最新动态，为行业提供客观、专业的指导与参考。

参评者概览

通用基础大模型	豆包大模型、讯飞星火、智谱·AI、通义千问、文心一言、腾讯混元、零一万物 01.AI、商汤日日新 sensenova、紫东太初、Moonshot AI、MINIMAX、阶跃星辰、百川智能 BAICHUAN AI、360智脑、deepseek、书生
多模态理解	通义千问、阶跃星辰、讯飞星火、智谱·AI、商汤日日新 sensenova、豆包大模型、零一万物 01.AI、MINIMAX、腾讯混元、deepseek、面壁智能 MODELBEST
多模态生成	通义万相、阶跃星辰、即梦AI、讯飞星火、智谱·AI、秒画、豆包大模型、天工、腾讯混元生图、可灵AI、360智绘

来源：沙利文、头豹研究院

大模型年度评测背景介绍——大语言评测参与者介绍

参评企业	模型系列	参评模型名称	模型发布时间	模型发布情况
商汤科技	商汤日日新	SenseNova-5.5-Pro	发布于2025年1月	此后版本无大更新，评测版本为评测周期对应的最新版本
阿里云	通义千问	Qwen2.5-max	发布于2025年1月	此后版本无大更新，评测版本为评测周期对应的最新版本
腾讯云	腾讯混元	Hunyuan-turbo-latest	发布于2024年10月	后续版本有持续更新，评测版本为评测周期对应的最新版本
零一万物	零一万物	Yi-Lightning	发布于2024年10月	此后版本无大更新，评测版本为评测周期对应的最新版本
智谱AI	智谱	Glm-4-Plus	发布于2024年6月	后续版本有持续更新，评测版本为评测周期对应的最新版本
360	360智脑	Zhinaoo2-o1	发布于2024年12月	此后版本无大更新，评测版本为评测周期对应的最新版本
字节跳动	豆包	Doubao-pro-32k	发布于2024年5月	后续版本有持续更新，评测版本为评测周期对应的最新版本
百度智能云	文心一言	EB 4.0 Turbo	发布于2024年6月	此后版本无大更新，评测版本为评测周期对应的最新版本
科大讯飞	讯飞星火	Spark Ultra 4.0	发布于2024年6月	后续版本有持续更新，评测版本为评测周期对应的最新版本
百川智能	百川智能	Baichuan4-Turbo	发布于2024年11月	此后版本无大更新，评测版本为评测周期对应的最新版本
月之暗面	Kimi.ai	Moonshot-v1-8k	发布于2024年2月	此后版本无大更新，评测版本为评测周期对应的最新版本
阶跃星辰	阶跃星辰	Step-2-16k	发布于2024年11月	此后版本无大更新，评测版本为评测周期对应的最新版本
上海人工智能实验室	书生	Internlm3-latest	发布于2025年1月	此后版本无大更新，评测版本为评测周期对应的最新版本
名之梦	Minimax	Minimax-Text-01	发布于2025年1月	此后版本无大更新，评测版本为评测周期对应的最新版本
深度求索	深度求索	Deepseek-V3	发布于2024年12月	此后版本无大更新，评测版本为评测周期对应的最新版本
中国科学院自动化研究所	紫东太初	Taichu-2.0	发布于2023年6月	后续版本有持续更新，评测版本为评测周期对应的最新版本

评测周期：25/01/20—25/01/24

来源：沙利文、头豹研究院

大模型年度评测背景介绍——多模态评测参与者介绍

■ 中国大模型多模态理解评测背景信息

参评企业	参评模型名称
商汤科技	SenseNova-5.5-Pro
阿里云	Qwen-vl-max-latest
腾讯云	Hunyuan-turbo-vision
阶跃星辰	Step-1v-32k
智谱AI	Glm-4v
科大讯飞	图片理解
字节跳动	Doubao-vision-pro-32k
面壁智能	MiniCPM-llama3-v-2.5
Minimax	海螺AI
零一万物	Yi-Vision-v2
深度求索	DeepSeek-VL-7b

均使用**模型调用API**的形式，模型调用参数与Prompt设置完全一致

评测周期：25/01/13-25/01/17

■ 中国大模型多模态生成评测背景信息

参评企业	参评模型名称
商汤科技	秒画
阿里云	通义万相
腾讯云	混元生图
阶跃星辰	Step-1X
智谱AI	CogView4
科大讯飞	讯飞星火
字节跳动	豆包·文生图
抖音	即梦AI
快手	可灵AI
360	360智绘
天工AI	AGI Sky-SaaS-Image

均使用**网页生成**的形式，模型调用参数与Prompt设置完全一致

评测周期：25/01/20-25/01/24

来源：沙利文、头豹研究院

章节三
大模型年度评测结果

- 大语言评测篇的综合结果显示，国际大模型整体表现优于中国大模型，通义千问、商汤日日新、腾讯混元以及智谱超越国际大模型均线，位居中国大模型的第一梯队。
- *在多模态评测中，阿里云、商汤科技及腾讯混元三家企业表现尤为卓越，凭借出色的多模态理解与生成能力，位居综合排名的前三甲，展现出其在多模态领域的前沿探索和技术优势。*

大模型年度评测评测结果

关键发现

大语言评测篇的综合结果显示，国际大模型整体表现优于中国大模型，通义千问、商汤日日新、腾讯混元以及智谱超越国际大模型均线，位居中国大模型的第一梯队。在多模态评测中，阿里云、商汤科技及腾讯混元三家企业表现尤为卓越，凭借出色的多模态理解与生成能力，位居综合排名的前三甲，展现出其在多模态领域的前沿探索和技术优势。

报告完整版登录www.leadleo.com 搜索《2025年中国大模型年度评测》

来源：沙利文、头豹研究院

方法论

- 头豹研究院布局中国市场，深入研究19大行业，532个垂直行业的市场变化，已经积累了近100万行业研究样本，完成近10,000多个独立的研究咨询项目。
- 研究院依托中国活跃的经济环境，从纵深防御、快速响应、轻量化部署等领域着手，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员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，以专业的视野解读行业的沿革。
- 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，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。
- 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。
- 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

法律声明

报告完整版登录www.leadleo.com
搜索《2025年中国大模型年度评测》

首席分析师

袁栩聪

15999806788

oliver.yuan@leadleo.com

研究总监

李庆

13149946576

livia.li@frostchina.com

www.frostchina.com ; www.leadleo.com

https://space.bilibili.com/647223552

https://weibo.com/u/7303360042

头豹 LeadLeo

FROST & SULLIVAN 沙利文