

INDUSTRY RESEARCH • 2026

中国大模型厂商 海外拓展深度研究

布局、进入路径与商业模式

研究截止日：2026年7月28日

证据口径：公司公告、官方文档、监管文件与可靠第三方研究

云厂商

AI原生厂商

区域策略

商业模式

面向投资机构决策者与企业战略部门 • 公开资料研究

目录

执行摘要

一、研究范围与判断口径

二、海外布局全景：三种梯队同时形成

2.1 云基础设施型：阿里、华为、腾讯领先

2.2 产品级出海型：百度与BytePlus各有侧重

2.3 模型与应用驱动型：开发者影响力高，经营深度分化

表1：代表性厂商海外成熟度

三、进入海外市场的六条路径

3.1 开放权重与开发者生态

3.2 国际API和MaaS直接销售

3.3 云市场与第三方模型平台分销

3.4 本地运营商、政府与系统集成商合作

3.5 跟随中国企业出海

3.6 消费应用和行业解决方案

四、本地化：从语言适配到经营体系

五、商业模式与定价

5.1 四层收入结构

5.2 官方API价格快照

5.3 差异化竞争

六、区域优先级与进入策略

表2：区域优先级矩阵

七、典型模式案例

案例一：Qwen——开放生态与云平台协同

案例二：DeepSeek——低价API与第三方托管

案例三：华为云——主权AI和行业交付

案例四：01.AI——从国际API转向区域主权合作

八、风险与三种情景

九、投资判断与战略建议

十、结论

参考来源

免责声明

执行摘要

中国大模型厂商的海外竞争已经从“模型是否能被海外用户访问”，进入“能否在当地稳定部署、合规交付并持续收费”的阶段。研究显示，真正形成海外经营基础的厂商并不多：阿里云、华为云、腾讯云依托既有云区域、销售网络和系统集成能力，具备从算力、模型平台到行业方案的一体化交付能力；DeepSeek、Qwen、GLM、Kimi等模型则主要通过开放权重、国际API和第三方模型平台建立全球开发者影响力。两者分别代表“重资产基础设施出海”和“轻资产模型分发出海”，但最终都必须补足本地合规、企业服务与客户关系。

本报告形成六项核心判断：

- 1. 东南亚和中东是最具现实性的首轮商业化市场。** 两地既有较强的云与AI需求，也存在电信运营商、主权基金、政府数字化项目和中国企业客户等渠道。欧洲和北美仍具有技术影响力价值，但监管、采购审查、数据治理与地缘政治成本更高。
- 2. 云基础设施是企业级模型出海的关键护城河。** 截至2026年中，阿里云公开口径为32个地域、105个可用区，华为云为34个地域、102个可用区；腾讯云覆盖21个地域、64个以上可用区。相比之下，百度智能云和BytePlus虽有国际产品入口，却未公开同等完整的海外IaaS地域体系。¹²³
- 3. 开放权重是获客工具，不是收入本身。** Qwen、DeepSeek、GLM、Yi等模型可以在海外社区下载、由第三方平台托管或被企业自部署，但下载量、API流量与厂商实际海外收入属于三种不同口径。厂商能否控制计费入口、提供服务等级协议并获得续约，决定开源影响力能否转化为企业价值。
- 4. 中国厂商的价格优势真实存在，但不能简单等同于高毛利。** 例如DeepSeek V4 Flash官方输出价格为0.28美元/百万Token，Qwen3.7 Plus国际版为1.60美元，GLM-5为3.20美元，而OpenAI GPT-5.4、Anthropic Claude Opus 5的输出价格分别为15美元和25美元。价差有利于获取成本敏感型客户，但渠道分成、区域算力、客户支持、安全审计和销售费用仍会显著影响最终毛利。⁴⁵⁶⁷
- 5. 主权AI和私有化部署是海外企业市场的重要增量。** 在沙特、阿联酋、东南亚、非洲及部分欧洲市场，客户购买的不只是模型调用，还包括数据驻留、专属算力、本地语言、行业知识库、审计能力和本地服务团队。华为、阿里和腾讯在这一模式中更占优势；智谱、01.AI等模型公司则通过本地合作方、联合研发或合资公司进入。
- 6. 海外竞争的终局不是单模型性能，而是“可信赖的AI交付能力”。** 未来24—36个月，决定胜负的指标包括海外付费收入、企业客户续约、区域推理成本、本地数据处理、合规认证、渠道控制力以及行业方案的可复制程度，而非单一评测榜单。

一、研究范围与判断口径

本报告把“海外布局”分为五个成熟度层级：L1为模型在海外社区传播；L2为海外用户能够注册产品或调用API；L3为进入云市场、模型平台或建立渠道合作；L4为形成企业付费、私有化部署或行业项目；L5为设有本地团队、基础设施、合规体系和持续收入。只有达到L4以上，才能视为形成较稳定的海外商业化。

样本分为两类。第一类是拥有云基础设施或大型生态的厂商，包括阿里云/Qwen、华为云/盘古、腾讯云/混元、百度智能云/文心和BytePlus/Seed；第二类是AI原生厂商，包括DeepSeek、智谱/Z.ai、MiniMax、月之暗面/Kimi、01.AI和百川智能等。由于中国厂商普遍未单独披露海外模型收入，本报告不对海外收入规模进行无依据估算，而以官方产品可用性、地域部署、合作公告、价格和企业案例判断成熟度。

宏观背景支持海外需求继续增长。Stanford AI Index 2026显示，2025年全球企业AI投资为5,817亿美元，受访组织中88%已在至少一个业务环节使用AI，70%使用生成式AI；但OECD按10人以上企业普查加权得到的2025年企业AI采用率仅20.2%。二者差异说明，大企业自报调查与全体企业普查不能直接比较，也说明中小企业和新兴市场仍处在规模化早期。⁸⁹ Stanford还指出，达到GPT-3.5级别能力的模型推理成本从2022年11月约20美元/百万Token降至2024年10月约0.07美元，27个月下降约280倍，为低价API和行业嵌入创造了经济条件。¹⁰

二、海外布局全景：三种梯队同时形成

2.1 云基础设施型：阿里、华为、腾讯领先

阿里云/Qwen是目前布局最完整的样本。阿里云已形成亚洲、中东、欧洲、北美和拉美的地域网络；Model Studio明确在新加坡、美国弗吉尼亚、德国法兰克福、日本东京和中国香港提供独立工作空间、API Key和区域端点。¹¹ 其海外打法不是只输出Qwen，而是将云基础设施、模型平台、数据服务、安全、跨境电商和本地合作伙伴组合销售。在马来西亚、印尼、日本、中东和欧洲，阿里云可同时服务中国出海企业与当地客户。Qwen开放权重承担开发者获客，Model Studio按量计费和企业云合同承担变现。需要注意，阿里集团披露的AI相关云收入没有单列国际业务，不能据此反推Qwen海外收入。

华为云/盘古的优势是政企客户、通信网络、硬件、混合云和行业交付。其海外地域覆盖拉美、中东、非洲、欧洲和亚太，并通过运营商、政府IT部门、银行、零售商和本地集成商进入市场。盘古的海外可用性却不能与华为云地域数直接画等号：公开信息更多体现为面向拉美的西班牙语/葡萄牙语模型、土耳其本地AI算力和按需开通，而不是所有海外地域均可即时调用全部盘古模型。华为更接近“数字基础设施与行业能力出口”，项目金额可能较高，但交付周期、资本投入和地缘风险也更大。

腾讯云/混元依托游戏、音视频、社交、内容和小程序生态。腾讯云在东南亚、日本、韩国、欧洲、美国及沙特设有地域，并通过True IDC、GoTo、零售和金融科技伙伴拓展。混元3D已经提供国际API，适合游戏、广告、影视和3D资产生产；但通用混元模型在各海外地域的完整可用清单和海外收入未公开。腾讯披露的“海外客户增长”包含云、音视频、游戏和微信生态客户，不能等同于混元付费业务增长。

2.2 产品级出海型：百度与BytePlus各有侧重

百度智能云/文心已提供Qianfan国际产品入口和国际控制台，但未公开独立海外地域、可用区、国际定价和本地合作伙伴全表，因此更适合定义为“产品可用型”而非“海外区域经营型”。百度的国际优势更多来自Apollo Go、搜索和企业AI技术，而文心本身的海外商业化证据相对有限。

BytePlus/Seed的路径与传统云厂商不同。BytePlus ModelArk官方文档称可面向约160个国家和地区提供服务，但明确排除美国、中国、印度和俄罗斯；这表示产品可销售范围，不表示拥有160个自建数据中心。¹² 其核心竞争力是推荐、广告、视频、内容生成和消费应用运营经验。视频模型可较快形成按生成量计费的收入，但版权、人物肖像、内容标识和应用商店政策构成实质约束，特别是在美国和欧洲市场。

2.3 模型与应用驱动型：开发者影响力高，经营深度分化

DeepSeek是开放权重和成本效率路线的代表。其官方API提供缓存差异定价，并可由第三方云和模型平台托管。该模式让海外开发者能够快速测试或自部署，也降低了客户对单一中国云节点的依赖。但DeepSeek没有披露海外办公室、区

域销售、企业客户和海外收入，技术传播应与商业经营严格区分。其最可能的变现路径是直接API、企业私有化服务、生态授权或与云平台联合分发。

智谱/Z.ai已建立国际定价和英文开发者入口，并在东南亚、中东、非洲等地推进“主权大模型”与本地基础设施合作。公开报道显示，其路径通常与本地政府、运营商、数据中心或华为硬件结合。¹³ 这使智谱比纯API厂商更接近企业交付，但合资、联合方案与实际付费收入仍需分别核验。

月之暗面/Kimi已将国际API迁移至Kimi品牌平台。Kimi K3官方价格为输入3美元、缓存输入0.30美元、输出15美元/百万Token，上下文为1,048,576 Token。¹⁴ 其优势是长上下文、编程和Agent使用体验，进入方式以开发者API和第三方托管为主；公开资料尚不足以证明其已形成广泛本地团队或区域基础设施。

MiniMax代表“多模态模型+全球消费应用”路径：文本、语音、视频和角色互动能力可通过API或应用订阅变现。该路径比政企项目扩张更快，但用户获取成本、内容安全、未成年人保护和应用商店分成会侵蚀毛利，且消费用户规模不能直接代表企业API收入。

01.AI呈现明显战略转型。其历史Yi模型仍以Apache 2.0在Hugging Face和GitHub分发，PopAi继续由新加坡主体运营并提供订阅；但国际API站目前提示因业务调整暂停服务。公司2025—2026年转向企业AI和主权AI，官方时间线披露与哈萨克斯坦共建AlemLLM、成立Q.AI合资公司。¹⁵ 这是从“模型API全球化”转向“区域合作和企业方案”的典型案例。百川智能海外一手证据较少，本报告仅作为潜在行业模型供应商观察，不列入成熟案例。

表1：代表性厂商海外成熟度

厂商	主要出海资产	当前成熟度	核心路径	主要短板
阿里云/Qwen	全球云地域、Model Studio、开源生态	L5	云+MaaS+渠道+行业	海外AI收入未单列
华为云/盘古	云、硬件、运营商与政企交付	L4-L5	主权云、混合云、行业项目	盘古海外GA清单不透明
腾讯云/混元	云、游戏、音视频、3D生态	L4	云+内容/游戏方案	通用模型海外收入未披露
百度/文心	国际Qianfan入口、Apollo生态	L2-L3	产品入口、智能驾驶协同	海外地域与伙伴不足
BytePlus/Seed	全球应用运营、视频与推荐	L3-L4	SaaS/MaaS、内容工具	美国缺位、版权压力
DeepSeek	开放权重、低价API、第三方托管	L2-L3	开源获客+API	本地团队与收入不透明
智谱/Z.ai	国际API、主权AI合作	L3-L4	本地伙伴+私有化	合作公告多于财务披露
Kimi	国际API、长上下文、编程	L2-L3	API+开发者工具	地域经营和客户证据有限
01.AI	开源Yi、PopAi、哈萨克合资	L3	应用订阅+主权AI项目	国际API暂停、战略转型中

三、进入海外市场的六条路径

3.1 开放权重与开发者生态

开放权重是进入全球市场成本最低的方式。开发者可以绕过传统销售流程，在Hugging Face、GitHub或本地服务器直接部署模型；第三方推理平台再将模型封装为OpenAI兼容API。其作用是扩大技术影响力、形成评测反馈和生态适配，并为后续企业采购提供“已经在用”的基础。

局限也很明确：模型下载可能来自测试、镜像或重复拉取；第三方平台掌握计费和客户数据；企业自行部署后，模型厂商未必获得收入。因此应同时观察官方API调用、企业支持合同、云市场分成、私有化授权和续约，而不是单独使用下载量衡量商业化。

3.2 国际API和MaaS直接销售

Qwen Model Studio、DeepSeek、Z.ai和Kimi均形成国际或美元计费入口。该路径可通过免费额度、低价轻量模型、兼容SDK和在线文档降低开发门槛，再通过更强模型、批处理、缓存、微调、专属容量和企业SLA提升客单价。决定转化率的关键不是注册页，而是海外支付、低延迟端点、工单响应、日志审计、数据保留政策和合同责任。

3.3 云市场与第三方模型平台分销

进入大型云市场、开发者平台或模型聚合平台，可以借用其支付、合规、GPU和客户关系。对AI原生厂商而言，这是进入欧美和亚太企业采购体系的快捷方式；对客户而言，也便于在统一云账单中采购。代价是渠道分成、价格控制权下降、客户数据难以沉淀，并可能被多个同类模型放在同一页面直接比价。

3.4 本地运营商、政府与系统集成商合作

在中东、东南亚、拉美和非洲，本地电信运营商、主权基金、政府数字化机构和系统集成商往往掌握数据中心、牌照、客户和招投标渠道。中国厂商提供模型、云或一体机，合作方提供本地算力、销售和合规。收入可来自许可、算力、项目实施、运维和分成。这一路径适合华为、阿里、腾讯和智谱，但应区分谅解备忘录、试点、正式上线和持续计费四个阶段。

3.5 跟随中国企业出海

制造、汽车、电商、游戏、物流、金融科技和手机企业是中国云厂商最容易获得的首批海外客户。合同可以由中国总部统一采购，再扩展到海外分支，从而降低获客和信任成本。其风险是客户结构长期停留在“中国客户海外分支”，无法证明厂商已取得当地原生客户。

3.6 消费应用和行业解决方案

AI聊天、陪伴、视频、翻译、教育和生产力应用可通过订阅、广告、虚拟道具和按量生成直接变现，增长速度快；但平台抽成、内容审核和营销支出较高。企业行业方案则通过知识库、客服、代码、营销、制造和金融项目变现，销售周期更长，但客单价、续约和切换成本通常更高。最优组合往往是消费应用验证产品、开放模型获取开发者，再用企业版本获得利润。

四、本地化：从语言适配到经营体系

语言本地化只是第一层。东南亚需要处理英语与印尼语、泰语、越南语等混合输入；中东要求高质量阿拉伯语、方言与宗教文化适配；日本和韩国企业客户更关注专业术语、敬语、格式和本地知识。厂商应建立本地评测集与人工反馈，而非只用翻译语料微调。

第二层是数据与监管。欧盟AI Act于2024年8月生效，禁止性实践和AI素养义务自2025年2月适用，通用人工智能模型义务自2025年8月适用，多项透明度规则在2026年8月进入适用阶段。¹⁶ 韩国AI基本法于2026年1月生效，新加坡采用自愿治理框架与既有数据保护法规结合的方式。不同国家对跨境传输、自动化决策、内容标识和本地代表要求不同，因此全球统一API需要叠加区域化数据保留、模型版本和安全策略。

第三层是产品与商业运营，包括当地货币、税务、发票、支付、时区、客服、SLA、采购条款和责任保险。企业客户通常要求数据不用于训练、支持删除、提供日志和安全事件通报，并能接受第三方审计。具备本地法人和云区域的厂商更容易满足要求；轻资产模型厂商则需要可信的本地云和集成伙伴。

第四层是行业本地化。中东的政府、能源、金融和智能城市项目强调主权与本地部署；东南亚的高潜场景包括跨境电商、客服、金融科技和内容生产；日韩更适合制造、机器人、游戏与企业软件；欧洲侧重工业、汽车、金融和数据主权；拉美则更重视价格、客服、电商和本地支付。

五、商业模式与定价

5.1 四层收入结构

中国厂商的海外收入可以分为四层：第一层是API按Token、图片、音频或视频收费；第二层是开发者套餐、消费订阅和企业按席位SaaS；第三层是专属实例、微调、知识库、Agent平台和托管服务；第四层是私有化部署、模型授权、系统集成和长期运维。前两层标准化程度高、扩张快，后两层合同金额和客户黏性更高，但交付成本也更高。

云厂商还可以通过计算、存储、数据库、网络、安全和模型调用获得组合收入。即使模型API本身低价，客户的数据和应用迁移到同一云后，整体生命周期价值可能显著提高。因此阿里、华为、腾讯可以把模型视为云资源的需求入口；AI原生厂商则必须依靠模型能力、开发工具或应用场景直接获得收入。

5.2 官方API价格快照

下表选取各厂商具有代表性的公开模型。它只反映价格，不代表模型能力、延迟、上下文和服务等级完全可比。

厂商/模型	输入价	缓存输入	输出价	说明
DeepSeek V4 Flash	\$0.14	\$0.0028	\$0.28	1M上下文
DeepSeek V4 Pro	\$0.435	\$0.003625	\$0.87	1M上下文
Qwen3.7 Plus国际版，≤256K	\$0.40	依规则	\$1.60	256K以上分档加价
Qwen3.7 Max国际版	\$2.50	依规则	\$7.50	1M档位
GLM-5	\$1.00	\$0.20	\$3.20	Z.ai国际价

厂商/模型	输入价	缓存输入	输出价	说明
GLM-4.7	\$0.60	\$0.11	\$2.20	Z.ai国际价
Kimi K3	\$3.00	\$0.30	\$15.00	1,048,576上下文
Google Gemini 3.1 Pro, ≤200K	\$2.00	\$0.20	\$12.00	超200K为\$4/\$18
OpenAI GPT-5.4	\$2.50	\$0.25	\$15.00	Batch约五折
Anthropic Claude Sonnet 5	\$3.00	依缓存规则	\$15.00	2026-08-31前限时\$2/\$10
Anthropic Claude Opus 5	\$5.00	依缓存规则	\$25.00	1M上下文

单位：美元/百万Token。来源：各厂商官方定价页，研究截止日快照。[4567141718](#)

价格显示，中国厂商在中低价区间具备明显选择空间，尤其是DeepSeek、Qwen Flash/Plus和GLM。但“输出价格低5—20倍”不能直接推导同等能力下成本低5—20倍，也不能推导同等毛利。海外端点可能使用不同GPU和带宽，企业客户还会产生销售、合规、赔付和支持成本。更合理的投资指标是每单位有效任务成本，即完成一次合格客服、代码修改、文档审核或视频生成的总成本，而非单纯Token价格。

5.3 差异化竞争

与OpenAI、Anthropic、Google和美国云平台相比，中国厂商的优势主要是价格、中文及部分亚洲语言、开放权重、灵活私有化和制造、电商、游戏、物流等行业经验；不足是欧美品牌信任、全球企业采购覆盖、标准化合规披露、责任承担和本地服务网络。

与Mistral等欧洲厂商相比，中国开放模型在规模、更新速度和成本上有竞争力，但欧洲厂商更容易获得“本地主权AI”标签。中国厂商若直接强调来源和低价，可能触发审查；更有效的差异化是与当地云、运营商和软件商联合交付，让客户掌握数据、部署和审计权。

六、区域优先级与进入策略

表2：区域优先级矩阵

区域	市场机会	进入难度	推荐路径	综合判断
东南亚	高	中	新加坡区域枢纽+运营商/银行/电商合作	第一梯队
中东	高	中	主权云、政府/能源/金融项目、合资	第一梯队
拉美	中高	中	西葡语模型+电信/支付/中国企业渠道	第二梯队
日韩	高	中高	本地云、制造/游戏方案、严格语言适配	第二梯队
非洲	中长期	中	运营商、银行、教育、低成本私有化	第二梯队/长期
欧洲	高	高	开放模型、本地伙伴、主权部署	选择性进入

区域	市场机会	进入难度	推荐路径	综合判断
北美	高	很高	开源生态和第三方托管为主	技术影响力市场

东南亚适合以新加坡作为合同、研发和区域管理中心，在马来西亚、印尼、泰国和越南布局数据中心或合作节点。当地电商、金融科技、客服、教育和内容行业价格敏感，适合Qwen、DeepSeek和GLM等模型；但越南、印尼等地跨境数据要求更严格，需要按国家配置。

中东的机会来自政府主导的AI基础设施、能源资本和主权数据需求。最优模式是本地数据中心、运营商或主权基金合作，而非从中国端点跨境调用。沙特PIF于2025年设立HUMAIN，覆盖AI基础设施、云、模型和应用，反映国家级AI平台正在成为重要买方和合作方。¹⁹

日韩具有高付费能力和制造、游戏、机器人场景，但采购标准、语言质量和安全要求高。中国厂商更适合与当地云、运营商和软件公司合作，先从代码、客服、游戏资产和企业知识库切入，而非以通用聊天产品竞争。

欧洲的优势是工业和企业需求，障碍是GDPR、AI Act、数据主权与供应链审查。开放权重和私有部署可降低数据担忧，但仍需说明训练数据、版权、风险管理和模型更新责任。**北美**市场规模大，却面临出口管制、政府采购和品牌信任压力；对多数中国厂商而言，现阶段更现实的价值是开发者采用和第三方托管，而非大规模直销。

拉美和非洲适合长期布局。华为等厂商在运营商、网络、政府和银行已有基础，阿里云也在增加拉美资源。需求侧对价格、本地语言和部署灵活性更敏感，但支付、汇率、基础设施和售后网络会延长回收周期。

七、典型模式案例

案例一：Qwen——开放生态与云平台协同

Qwen通过开放权重形成开发者入口，再用Model Studio国际区域、企业云服务和当地伙伴承接付费。该模式的优势是模型、API、数据和计算可在同一平台组合，客户从试用到生产的迁移路径清晰。风险是开源用户可能在其他云上运行Qwen，阿里无法获得全部价值；因此其关键指标不是下载量，而是国际Model Studio活跃客户、海外Token调用、云资源联动收入和续约。

案例二：DeepSeek——低价API与第三方托管

DeepSeek凭借开放权重和低价迅速获得全球开发者关注，适合代码、Agent和成本敏感型推理。第三方平台分发降低了地缘与支付门槛，也使模型可以部署在客户选择的区域。但这一模式把部分客户关系交给平台，且缺少公开的本地企业服务网络。若要提高收入质量，需要增加企业支持、专属容量、合规版本和行业伙伴。

案例三：华为云——主权AI和行业交付

华为可将云、网络、服务器、数据库、模型平台和咨询服务打包，特别适合电信、政府、金融、能源和制造客户。与纯API相比，合同更大、切换成本更高；与美国云相比，华为在全球南方和中国企业供应链中具备差异化。主要风险是硬件供应、地缘审查和项目回款周期。

案例四：01.AI——从国际API转向区域主权合作

Yi开放模型维持开发者影响力，PopAi保留海外消费订阅，但官方国际API暂停；与此同时，公司通过哈萨克斯坦AlemLLM和Q.AI合资项目转向企业与主权AI。这一变化说明，纯模型API若难以形成持续规模，厂商可能转向本地政府、企业和合资项目。其收入更具项目属性，后续应观察合同执行、复用程度和海外回款。

八、风险与三种情景

监管与数据风险。 数据跨境、个人信息、自动化决策、生成内容标识和行业监管正在同时加强。模型厂商需要将合规成本视为产品成本，而非一次性法务支出。

地缘与供应链风险。 高端芯片出口、云服务限制、政府采购审查和制裁可能改变海外算力选择。重资产云区域面临投资锁定，轻资产模型则可能被应用商店、云市场或支付渠道限制。

商业风险。 低价API容易吸引试用，却可能带来低留存、高峰算力压力和低毛利；开放权重也会削弱厂商对客户控制。项目制虽然客单价高，但收入确认、回款和定制成本不稳定。

技术与品牌风险。 多语言幻觉、本地事实错误、提示注入、版权、模型安全和服务中断都会直接影响海外企业采购。对投资者而言，重大安全事件和监管处罚应进入估值折价。

未来三年可分为三种情景：

- **基准情景：区域化扩张。** 中国厂商在东南亚、中东和全球南方通过开放模型、本地云和行业合作扩大份额，欧美以开发者生态和第三方托管为主。
- **乐观情景：主权AI形成新市场。** 多国希望降低对美国模型与云的单一依赖，中国厂商通过本地部署、开放权重和成本优势成为第二供应体系，海外企业合同和续约明显增长。
- **保守情景：监管与供应链分割加剧。** 欧美及其盟友提高采购和算力限制，部分新兴市场也加强数据与模型审查，中国厂商只能通过白标、开源或中国企业海外分支间接经营，收入增长慢于技术影响力。

九、投资判断与战略建议

投资机构应重点跟踪六组指标：

1. **经营质量：**海外收入、付费企业客户、年度经常性收入、续约率、回款周期；
2. **单位经济性：**每百万Token毛利只是起点，还要观察缓存命中、GPU利用率、峰值容量和渠道分成；
3. **客户结构：**当地原生客户占比，以及是否过度依赖中国出海企业或单一政府项目；
4. **本地能力：**海外地域、数据驻留、本地法人、销售、法务、客服和系统集成网络；
5. **产品组合：**模型API能否带动云、数据、Agent、SaaS和行业解决方案收入；
6. **风险控制：**监管调查、版权争议、数据事件、芯片供应与重大客户集中度。

对云厂商，建议把区域Model Studio与本地云资源、数据库、安全和行业伙伴绑定，避免只靠模型降价。对AI原生厂商，建议优先选择2—3个区域和1—2个行业，把开放模型流量转成企业支持、专属部署和订阅收入，而不是同时铺开所有国家。对消费应用厂商，应将用户增长与付费率、获客成本、平台分成和内容治理共同考核。

在估值层面，应对“开源影响力高但收入不可见”的厂商给予技术期权价值，但不能按云收入倍数直接估值；对“项目合同大但复用性低”的厂商，应关注回款和人员交付成本；对同时拥有海外区域、开发者生态和企业续约的厂商，可给予更高确定性溢价。

十、结论

中国大模型厂商已经具备全球技术竞争力和显著价格弹性，但海外拓展仍处于从“可访问”向“可经营”过渡的阶段。阿里、华为和腾讯拥有基础设施与渠道优势，适合主权云和企业方案；DeepSeek、Qwen、GLM和Kimi通过开放权重、低价API和开发者工具扩大影响；BytePlus、MiniMax等厂商则从视频、内容和消费应用切入。不同路径并不存在唯一赢家，但商业价值最终都取决于是否能够持续收费、控制客户关系并承担本地责任。

因此，判断中国大模型出海不应只问“模型在多少国家可用”，而应连续追问：数据在哪里处理、谁签合同、谁负责交付、客户为何续费、渠道拿走多少收入、监管事件由谁承担。能够把技术能力转化为本地化、合规化和可复制交付体系的厂商，才可能在全球模型价格下降和竞争加剧的环境中形成长期价值。

参考来源

免责声明

本报告仅基于截至2026年7月28日可获得的公开资料，不构成证券研究、投资建议、法律意见或商业承诺。模型价格、区域可用性、监管要求和合作状态变化较快，实际采购前应以厂商控制台、合同、当地法律和监管机构最新文件为准。公司未披露的海外收入、客户数和利润率均未作确定性推算。

-
1. Alibaba Cloud, Global infrastructure and 2026 expansion announcements: https://www.alibabacloud.com/blog/alibaba-cloud-expands-global-ai-infrastructure-with-new-data-centers-in-france-japan-malaysia-and-mexico_603278 ↩
 2. Huawei Cloud, Global Infrastructure: <https://www.huaweicloud.com/intl/en-us/about/global-infrastructure.html> ↩
 3. Tencent Cloud, Global Infrastructure and Saudi region announcements: <https://www.tencentcloud.com/dynamic/news-details/100638> ↩
 4. DeepSeek API Docs, Models & Pricing: https://api-docs.deepseek.com/quick_start/pricing ↩↩
 5. Alibaba Cloud Model Studio, Model Pricing: <https://www.alibabacloud.com/help/en/model-studio/model-pricing> ↩↩
 6. Z.ai Docs, Pricing: <https://docs.z.ai/guides/overview/pricing> ↩↩

7. Anthropic, Claude API Pricing: <https://platform.claude.com/docs/en/pricing> ↔
8. Stanford HAI, AI Index 2026: <https://hai.stanford.edu/ai-index/2026-ai-index-report> ↔
9. OECD, AI use by individuals and firms, January 2026: <https://www.oecd.org/en/about/news/announcements/2026/01/ai-use-by-individuals-surges-across-the-oecd-as-adoption-by-firms-continues-to-expand.html> ↔
10. Stanford HAI, AI Index 2025—Research and Development: <https://hai.stanford.edu/ai-index/2025-ai-index-report/research-and-development> ↔
11. Alibaba Cloud Model Studio, Available Regions: <https://www.alibabacloud.com/help/en/model-studio/regions/> ↔
12. BytePlus ModelArk, Availability: <https://docs.byteplus.com/en/docs/ModelArk/availability> ↔
13. Reuters, Zhipu AI overseas expansion, 25 June 2025: <https://www.reuters.com/world/china/openai-says-chinas-zhipu-ai-gaining-ground-amid-beijings-global-ai-push-2025-06-25/> ↔
14. Kimi Platform, Kimi K3 Pricing: <https://platform.kimi.ai/docs/pricing/chat-k3> ↔
15. 01.AI, Company timeline and cases: <https://www.01.ai/about.html> ↔
16. European Commission, AI Act regulatory framework and application timeline: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> ↔
17. OpenAI, API Pricing: <https://developers.openai.com/api/docs/pricing> ↔
18. Google AI for Developers, Gemini API Pricing: <https://ai.google.dev/gemini-api/docs/pricing> ↔
19. Saudi Public Investment Fund, HUMAIN portfolio: <https://www.pif.gov.sa/en/our-investments/our-portfolio/humain/> ↔
20. World Bank, Digital Progress and Trends Report 2025—AI Foundations: <https://www.worldbank.org/en/publication/dptr2025-ai-foundations> ↔
21. UNCTAD, Technology and Innovation Report 2025: <https://unctad.org/publication/technology-and-innovation-report-2025> ↔
22. McKinsey, The State of AI 2025: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai> ↔