

中国主要大模型公司海外业务与产品布局研究

报告日期：2026年8月 | **研究对象：**月之暗面（Moonshot AI）、DeepSeek、智谱（Zhipu AI / Z.ai）、MiniMax、阶跃星辰（StepFun）、零一万物（01.AI）

受众：关注中国 AI 产业的投资人及出海企业决策者

方法论：基于官方披露（招股书、年报、官方文档）、主流媒体报道及第三方数据平台（Sensor Tower、HuggingFace、LMArena/Arena.ai、OpenRouter 等）的公开信息研究，数据截至 2026 年 8 月上旬。未公开数据一律标注“未披露”，估算数据注明估算依据。**本报告不构成投资建议。**

执行摘要

中国大模型公司出海已从“零散尝试”进入“系统性布局”阶段。六家头部公司分化出四条差异化路径：**开源技术派**（DeepSeek、月之暗面，以开源权重+MaaS API 获取全球开发者）；**开源+政企双轮**（智谱，海外品牌 Z.ai + 东南亚/中东本地化部署）；**C 端消费出海**（MiniMax，Talkie 陪伴社交 + Hailuo 视频生成，海外收入占比超 70%）；**主权 AI 与企业方案**（零一万物，聚焦“一带一路”市场）；**端侧+B 端出海**（阶跃星辰，依托手机/汽车终端出海）。

关键数字：MiniMax FY2025 海外收入占比 **>70%**；零一万物海外收入占比约 **50%**；月之暗面海外收入已超国内（具体比例未披露）；DeepSeek 通过 Vercel 网关处理海外约 **23%** 的企业级 token；智谱 2025H1 海外占比约 **11.6%**（主要来自东南亚）。融资与资本化方面，智谱、MiniMax 已于 2026 年 1 月先后港股上市，DeepSeek 完成中国史上最大单笔 AI 融资（约 \$7.4B，估值超 \$500 亿），月之暗面估值达约 \$350 亿并筹备港股 IPO，阶跃星辰 Pre-IPO 估值 40–60 亿美元、目标 2026 年内递表，零一万物目标 2027 年赴港。

主要风险集中于地缘政治与数据合规：DeepSeek 被 10+ 个国家和地区限制、智谱被列入美国实体清单、MiniMax 深陷好莱坞版权诉讼与美区监管、月之暗面遭“蒸馏指控”与美政府关注。这些风险与“低价高性能+开源生态”带来的竞争力并存，构成中国 AI 出海的核心投资变量。

第一章 研究背景与行业全景

1.1 为什么出海：四大推手

- 国内价格战与变现困境：**2024 年 5 月 DeepSeek-V2 将 API 价格打到 GPT-4 Turbo 的约 1%，引发全行业降价，部分轻量模型免费，国内推理毛利一度为负。纯国内 API 变现无望，美元计价海外市场提供数倍毛利（如 Z.ai 国际价约为国内价的 2 倍）。
- 算力供给被卡：**美国对华芯片管制在 2023–2026 年间反复“禁售—恢复—拒买”：H20 于 2025-04 被实质禁售、2025-07 恢复；H200 于 2026-01 名义放开逐案审查但中国拒绝进口。中国企业对 H200 的订单超 200 万颗、远大于供给；2026 年 H100/H200 租赁价上涨约 40%。算力缺口倒逼企业境外建算力（如字节在马来西亚部署约 3.6 万颗 B200）并借出海换取美元收入。
- 估值与资本化诉求：**智谱、MiniMax 已上市，其余四家均在筹备 IPO，“全球 AI 平台公司”叙事是估值的核心支撑。
- 国内监管与数据环境：**模型备案、内容合规要求叠加中国《数据出境安全评估办法》，出海既是市场选择也是结构性对冲。

1.2 全球竞争格局与对标基准

截至 2026 年 8 月，海外头部厂商旗规定价（美元/百万 token，输入/输出）：OpenAI GPT-5.5 约 \$5/\$30；Anthropic Claude Sonnet 5 约 \$3/\$15、Claude Opus 4.8 约 \$5/\$25；Google Gemini 2.5 Pro 约 \$1.25/\$10；Meta Llama 5（开源）托管约 \$0.8/\$2.4。中国旗舰模型以低一个数量级的定价参与竞争（详见第四章）。

性能层面，中国开源模型已进入全球第一梯队：Arena.ai（原 LMArena）2026 年 2 月 Code Arena 中，智谱 GLM-5 居全球开源第一，MiniMax M2.5、月之暗面 Kimi K2.5 Thinking 并列第二；Vellum 开源榜（2026 年中）显示 Kimi K3、GLM-5.2、DeepSeek V4 系列、MiniMax M3 均位列推理/编程榜单前列。开源生态上，阿里 Qwen 家族累计下载超 10 亿次（2026-03），DeepSeek V4-Pro 单模型超 500 万下载，中国模型已占据 HuggingFace 头部下载的半壁江山（中国模型占全球头部开源下载约 45%+，2025 年新增口径）。

1.3 四种出海模式

模式	代表	核心逻辑
开源生态出圈	DeepSeek、月之暗面	开源权重绕过 App 商店/API 封锁，全球自托管
API 国际版直连	DeepSeek、智谱 Z.ai、阶跃	美元计价 API，毛利数倍于国内
消费级 App 出海	MiniMax（Talkie/Hailuo）、DeepSeek App	C 端订阅+广告，Sensor Tower 可见的高增长
企业级本地化部署	智谱、零一万物	主权 AI、东南亚/中东政企私有化

第二章 六家公司概况与经营基本面

2.1 横向对比表

维度	月之暗面	DeepSeek	智谱（Z.ai）	MiniMax	阶跃星辰	零一万物
总部	北京	杭州	北京	上海（开曼上市主体）	上海	北京
成立时间	2023-03	2023-07	2019	2021-12	2023-04	2023-05
创始人/CEO	杨植麟（清华/CMU，Transformer-XL、RoPE 作者）	梁文锋（幻方量化背景）	CEO 张鹏；创始人唐杰、李涓子（清华教授）	闫俊杰（前商汤副总裁）	CEO 姜大昕（前微软全球副总裁）；董事长印奇（旷视联创）	李开复
员工规模	约 300（2026，Wikipedia）	<200（估算，未披露）	883（2025H1 招股书）	415（2025 招股书）	400+（研发占比 80%+）	约 240（2026-07）
累计融资	累计约 \$40 亿+（含 2026 年多轮）	首轮外部融资约 \$74 亿（2026-06）	累计约 \$15 亿（13 轮）	IPO 前约 \$15.5 亿	B+ 轮超 50 亿元，Pre-IPO 近 25 亿美元	约 \$3-5 亿（第三方估算）
最新估值/市值	约 \$350 亿 （2026-07）	超 \$500 亿 （2026-06；7 月传 \$71-74B）	发行估值约 \$6.7B；市值峰值 HK\$1.27 万亿（2026-06）	发行估值约 \$6.5B；首日市值 \$11.5B+	Pre-IPO 40-60 亿美元；IPO 目标 100 亿美元	未披露（2024-08 约 104 亿元）
上市/IPO		计划 2027 年科创板 IPO	2026-01-08 港股上市（02513.HK）	2026-01-09 港股上市（0100.HK）	2026 年内递表港股	目标 2027 年赴港 IPO

维度	月之暗面	DeepSeek	智谱（Z.ai）	MiniMax	阶跃星辰	零一万物
	港股 IPO 筹备中 (pre-IPO 目标 \$500 亿)					
最新 营收	ARR 约 \$3 亿 (2026-06)	年化约 \$4-5 亿 (The Information/ Bloomberg)	2025 年 7.24 亿元 (+131.9%)	FY2025 \$79.0M (+158.9%)	2025 年近 5 亿元 ；2026E 约 12 亿元	2025 年 2.5 亿 元；订单 5 亿 元
海外 收入 占比	海外已超国内 (比例未披露)	未披露	约 11.6% (2025H1， 主要东南亚)	>70% (FY2025 官方)	未披露	约 50%

2.2 逐家要点

- **月之暗面**：收入结构以 API 许可为主（占比超 70%），ARR 从 2026 年 3 月的 \$1 亿增至 6 月 \$3 亿；K2.5 发布后 20 天营收超 2025 全年；媒体报道海外收入已反超国内，海外 API 收入自 2025-11 起翻 4 倍。风险面：2026-02 被 Anthropic 指控对 Claude 蒸馏，2026-07 遭白宫科技顾问指控，美方酝酿对华模型限制。
- **DeepSeek**：不融资、不商业化的实验室形象在 2026 年 6 月被打破——首轮外部融资约 \$7.4B（国家大基金领投、腾讯与宁德时代参与），投后估值超 \$500 亿，为史上最大单笔私人 AI 融资；The Information 报道其年化收入 \$4-5 亿。Vercel 披露 DeepSeek 占其企业级 AI 网关处理 token 约 23%，仅次于 Anthropic，是海外渗透的最强指标。
- **智谱**：2026 年 1 月成为"中国大模型第一股"，2026-06 市值一度达 HK\$1.27 万亿。商业模式高度依赖境内本地化部署（2025 年占收入 73.7%），海外收入以东南亚为主（占比约 11.6%），仍处早期。被列入美国实体清单（2025-01）是海外机构客户采用的重大障碍。
- **MiniMax**：六家中海外收入占比最高（>70%）。FY2025 收入 \$79.04M，其中 C 端 AI 原生产品 \$53.1M（占 67.2%）、开放平台/企业服务 \$26.0M（+197.8%）。2026-01 港股上市首日大涨 109%。深陷好莱坞版权诉讼与美区 AI 陪伴监管。
- **阶跃星辰**：走"AI+终端"路线（手机装机超 4200 万台、吉利银河 M9 智能座舱随车出海），B+ 轮超 50 亿元刷新行业纪录；2025 年收入近 5 亿元，2026E 约 12 亿元。海外收入占比未披露，国际开放平台设于新加坡、服务器位于美国。
- **零一万物**：已完成从基座模型到"企业 AI 决策"的转型（2025-01 放弃万亿参数预训练、与阿里云成立联合实验室，All in ToB），海外聚焦主权 AI 与"一带一路"市场（哈萨克斯坦 Q.AI 合资、泰国正大万蜂智能等），海外收入占比约 50%。2026 年 5 月锁定订单超 15 亿元（近半为可重复 ARR）。

第三章 产品矩阵与开源策略

3.1 开源模型与许可证

公司	代表开源模型	许可证	备注
DeepSeek	V3 / R1 / V3.1 / V3.2 / V4-Pro (1.6T MoE) / V4-Flash	MIT（全系）	1M 上下文；V4 为纯文本模型
月之暗面		K2 系列 Modified MIT ；K3 为自定义 "Kimi K3 License" (open-weight)	

公司	代表开源模型	许可证	备注
	Kimi K2 / K2 Thinking / K2.5 / K2.6 (1T MoE) ; K3 (2.8T MoE)		K2 要求月活>1 亿或月收>\$2000 万标注 Kimi; K3 新增 MaaS 连续 12 月收入 >\$2000 万须签约条款
智谱	GLM-4.5 / 4.6 / 4.7 / GLM-5 (744B) / GLM-5.2 (753B)	MIT (2025-07 起全系)	GLM-5.2 1M 上下文; AutoGLM 2.0 于 2026-07 全量开源
MiniMax	MiniMax-M2 / M2.5 (230B) / M3 (427B) ; H3 (视频)	M2/M2.5 MIT ; M2.7/H3 为 限制性社区许可	H3 商用限年收入<\$20M 机构, 且 排除美国、欧盟、英国、韩国部署
阶跃星辰	Step 3 (321B) / 3.5 Flash / 3.7 Flash; Step-Video-T2V; Step-Audio 系列	多为 Apache 2.0 ; Step-Video-T2V 为 MIT	开源节奏激进, 训练框架 Steptron 亦开源
零一万物	Yi / Yi-1.5 / Yi-Coder / Yi-VL	Apache 2.0	社区影响力已随战略转型减弱; "Yi-X1/ Yi-2"核实无发布记录

许可证趋势判断：行业正从"完全开放" (MIT/Apache 2.0) 走向"开放权重+商业授权"双轨。月之暗面 K3 首创 MaaS 营收签约门槛, MiniMax M2.7/H3 采用排除欧美英韩的地域条款, 均是面对版权诉讼、数据合规与商业化平衡的折中。对出海企业而言, **使用中国开源模型时须逐条审阅许可证的商业与地域限制条款。**

3.2 闭源模型与旗舰产品形态

- **DeepSeek:** API 模型 V4-Pro / V4-Flash; App 与网页免费开放; 无独立 C 端付费订阅。
- **月之暗面:** Kimi 助手 (Web/App/桌面/浏览器扩展)、Kimi Code (约 \$19/月, 兼容 Claude Code/Cursor)、Kimi Work、Kimi Claw (云端 Agent)、Agent Swarm (K2.6 可编排 300 子 Agent) ; 订阅梯度 \$19-199/月。
- **智谱:** GLM-5.2 (旗舰)、GLM-5V-Turbo (多模态)、AutoGLM 2.0 (云端智能体, 操作 50+ App)、GLM Coding Plan (海外开发者编码订阅, \$3/月起促销) ; 大陆"智谱清言"+海外 chat.z.ai。
- **MiniMax:** Talkie (陪伴社交 App, 海外) / 星野 (国内)、Hailuo AI (视频生成, 累计生成超 6 亿条)、H3 (视频生成旗舰)、MiniMax Agent、MiniMax Code; Token Plan 订阅 \$20-120/月。
- **阶跃星辰:** Step-2 (万亿参数)、Step-1o 系列 (端到端文本/视觉/语音三模态)、Step Edge (端侧基座)、Step AOS/Amoo (智能体操作系统)、STEPX Neo (全球首款大模型原生智能体手机) ; Step Plan 订阅 \$6.99-99/月。
- **零一万物:** PopAi (海外生产力工具, 用户近千万)、万智企业大模型平台、万策决策中枢 (老板 AI/销冠 AI/投资官 AI 三款"一号位 AI") ; Yi-Lightning 为最后一代自研基座, 之后转向依托开源模型。

3.3 社区运营与开源商业模式

- **社区运营:** DeepSeek-V3 GitHub 超 17 万 stars、V4-Pro HuggingFace 超 500 万下载; 月之暗面 K3 HF 超 112 万下载 (发布约两周) ; 智谱 GLM 家族官方称全球累计下载超 4000 万次; 阶跃 Step 3.5 Flash 上线一月 HF 超 30 万下载、在 OpenRouter 全球调用量登顶 (Step 3.7 Flash 累计处理超 1.91 万亿 tokens) ; MiniMax HF 组织 7.9K followers; 零一万物 Yi 系列 GitHub 约 7.8K stars, 下载量级偏低。
- **开源商业模式:** ① API 按 token 收费 (DeepSeek、月之暗面、智谱、MiniMax、阶跃均有国际定价) ; ② 云伙伴托管/生态分成 (AWS Bedrock、Azure、NVIDIA NIM、OpenRouter、Fireworks、Together、Groq 等) ; ③ 企业私有化部署与模型定制 (智谱本地化部署占收入 73.7%、阶跃端侧 License) ; ④ 开源权重"政策免疫"——实体清单与封禁主要限制托管 API, 开源权重可自托管绕过。

第四章 海外 API 端点与部署方案

4.1 海外 API 端点与运营实体

公司	海外 API 端点	海外运营实体/数据位置	中国端点
DeepSeek	api.deepseek.com（全球直连，无海外区域节点）	服务器位于杭州；依赖第三方托管规避数据主权	同一端点（国际/国内计费可切换）
月之暗面	api.moonshot.ai/v1（OpenAI/Anthropic 兼容）	新加坡实体 Moonshot AI PTE	api.moonshot.cn/v1
智谱	api.z.ai（OpenAI 兼容，zai SDK）	新加坡实体 JINGSHENG HENGXING；承诺"零数据保留"	open.bigmodel.cn（人民币计价，约为国际价 50%）
MiniMax	api.minimax.io/v1+ / anthropic	新加坡 Nanonoble Pte. Ltd.（美国-facing 运营方）	api.minimaxi.com
阶跃星辰	api-intl.stepfun.ai/v1	新加坡 SPARKLING AI Pte. Ltd.， 服务器位于美国	api.stepfun.ai/v1
零一万物	无自有海外端点	依赖 Replicate / OpenRouter / SiliconFlow 第三方托管	platform.lingyiwanwu.com

4.2 海外本地部署与云合作

- **本地/私有化部署**：六家均因开源权重支持自托管。DeepSeek（MIT，vLLM/Ollama 生态最成熟）、月之暗面（K3 需 ≥64 加速卡、约 1.5TB 内存，部署门槛高）、智谱（MIT+FP8 版本齐全，端侧 GLM-Edge 支持 ARM/RISC-V）、阶跃（Apache 2.0，GELab-Zero 提供即插即用本地部署）、MiniMax（受 H3 地域条款限制）、零一万物（一体机+AMD 合作）。
- **海外云合作**：DeepSeek 覆盖面最广（AWS Bedrock 首个全托管、Azure AI Foundry 欧盟区域、Groq、Cerebras、SambaNova、Together、Fireworks）；月之暗面（AWS Bedrock 法兰克福、Azure、NVIDIA NIM、Cloudflare Workers AI）；智谱（Google Cloud Gemini Enterprise 托管、OpenRouter/Together/DeepInfra，及高通骁龙端侧合作）；MiniMax（AWS Bedrock 上架 M2 系列、Azure 目录需自托管）；阶跃（NVIDIA NIM 全球部署、SiliconFlow）；零一万物（阿里云、AMD）。

4.3 海外旗舰模型 API 定价（美元/百万 token）

公司	模型	输入（缓存未命中）	输出
DeepSeek	V4-Pro	\$0.435	\$0.87
DeepSeek	V4-Flash	\$0.14	\$0.28
月之暗面	Kimi K3	\$3.00	\$15.00
月之暗面	Kimi K2.6	\$0.95	\$4.00
智谱	GLM-5.2	\$1.40	\$4.40
智谱	GLM-5	\$1.00	\$3.20
MiniMax	M3	\$0.30	\$1.20
阶跃星辰	Step 3.7 Flash	\$0.20	\$1.15

公司	模型	输入（缓存未命中）	输出
阶跃星辰	Step 3.5 Flash	\$0.10	\$0.30
对标：OpenAI	GPT-5.5	\$5.00	\$30.00
对标：Anthropic	Claude Sonnet 5	\$3.00	\$15.00
对标：Meta	Llama 5（托管）	\$0.80	\$2.40

定价观察：中国旗舰模型输出定价约为 Anthropic/OpenAI 同级模型的 **1/10–1/30**（如 DeepSeek V4-Pro 输出 \$0.87 vs GPT-5.5 \$30）。但推理模型（如 Kimi K3 始终开启推理）实际任务成本需按有效输出测算，名义低价≠实际低价。零一万物 Yi-Lightning（0.99 元/百万 token，约 \$0.14）为历史低价标杆，但该 API 业务已实质收缩（2026 年 LLM-Price 等汇总站点已不再收录）。

4.4 海外推理基础设施明细：自建还是公有云？

针对"是否建有海外本地推理、建在哪里、自建或与哪家公有云合作"逐一核实如下（截至 2026-08）：

公司	官方海外 API 端点	海外推理建设情况	服务器/数据位置	形态	主要公有云/平台合作
DeepSeek	api.deepseek.com（全球直连， 无海外区域节点 ）	未建设海外推理 ；官方数据全部存于中国境内（隐私政策明示），无法选择美/欧数据中心	中国大陆（杭州总部等）；社区观测部分经华为云 GeoDNS 分流	境内自建为主	AWS Bedrock（含 EU 区域 V4）、Azure AI Foundry（含欧盟区域）、Google Vertex；Groq / Cerebras / SambaNova / Together / Fireworks / DeepInfra / OpenRouter / 腾讯云 TokenHub
月之暗面	api.moonshot.ai/v1（OpenAI/Anthropic 兼容）	官方未披露自建海外数据中心；海外运营实体为新加坡 Moonshot AI PTE，底层推理位置未披露	未披露（存疑）	第三方公有云合作	AWS Bedrock（K2 Thinking/K2.5，法兰克福 EU 区域）、Azure AI Foundry、NVIDIA NIM、Cloudflare Workers AI、OpenRouter / Fireworks / Nebius / Modal / Vercel
智谱（Z.ai）	api.z.ai（OpenAI 兼容）	服务由新加坡提供 （官方隐私政策明示"generally provide the Services from Singapore"），海外运营实体 JINGSHENG HENGXING PTE. LTD.（新加坡）；有"零数据保留"承诺（媒体）	新加坡（官方）；⚠️ 个别第三方称"默认中国基础设施"，与官方口径相左，标注存疑	未披露具体云厂商	Google Cloud（Gemini Enterprise 托管 GLM-4.7/ GLM-5）、OpenRouter / Together AI / DeepInfra；端侧高通/ARM
MiniMax	api.minimax.io/v1+ / anthropic；海外运营主体新加坡 Nanonoble Pte. Ltd.	海外推理跑在 AWS 上 ：M2/M2.1/M2.5 在 AWS Bedrock 全托管，覆盖 14	AWS 全球区域	第三方公有云（AWS） ；	AWS（Bedrock，明确合作） ；Azure 目录可查但需自托管 VM

公司	官方海外 API 端点	海外推理建设情况	服务器/数据位置	形态	主要公有云/平台合作
		个 AWS 区域（美东/美西、法兰克福/伦敦/斯德哥尔摩/米兰/都柏林、东京/孟买/悉尼/雅加达/墨尔本、圣保罗），推理完全由 AWS 基础设施承载		国内侧用阿里云数据平台	
阶跃星辰	api-intl.stepfun.ai/v1（OpenAI 兼容）；运营主体新加坡 SPARKLING AI Pte. Ltd.	隐私政策明示服务器位于美国；具体云厂商未披露	美国（官方隐私政策）	未披露自建/云；以第三方平台承接	NVIDIA NIM（on-prem/cloud/hybrid）、OpenRouter、SiliconFlow、Modal、DeepInfra、Fireworks AI
零一万物	无自有海外端点；platform.lingyiwanwu.com 主要为境内服务	未建设海外推理	—	依赖第三方平台	OpenRouter（9 款模型）、Replicate、SiliconFlow；阿里云（境内联合实验室）

结论要点：

- 六家中仅 MiniMax 有明确的海外推理公有云落点（AWS Bedrock 14 区域），其余多为"境内自建 + 第三方平台转发"或运营实体/法律文本层面的海外定位。
- DeepSeek、零一万物未建海外推理；DeepSeek 官方 API 数据强制存中国境内，是海外企业采用的最大阻力，实践中通过 AWS/Azure/美国第三方托管"曲线落地"。
- 智谱、阶跃通过新加坡/美国运营实体+服务器承诺解决数据主权话术，但具体云厂商未披露（智谱官方口径新加坡托管，但第三方存在"默认中国基础设施"的相反观测，需尽调确认）。
- 月之暗面海外推理位置未披露，其国际 API 注册需中国手机号、支付以支付宝/微信为主，海外开发者多走第三方聚合平台。
- 对出海企业决策者：海外本地推理的"法律文本落点"≠"物理服务器落点"，采购时须在合同中约定数据驻留区域与子处理商清单（subprocessor list），并验证第三方托管商的真实区域。

第五章 海外市场表现与竞争格局

5.1 海外产品表现

- DeepSeek：App 2025 年 1 月登顶 140 个市场 App Store 榜首，19 天下载 2300 万；2025 全年累计下载约 1.73 亿、MAU 峰值 1.3 亿（绝大部分在中国）；美国用户占比仅约 5%。受多国封禁叠加话题退潮影响，海外 App 季度下载从 2025Q1 的 8300 万降至 2026Q1 的 1300 万。但开源生态与第三方托管持续扩张，Vercel 网关 token 占比 23%。
- 月之暗面：K3 发布周全球下载 93 万（+200%）、美国 8.6 万（+387%）；K2.5 后海外收入超国内；K2.6 曾为 OpenRouter 第二大常用模型。美国市场渗透存在地缘阻力（美企内部追踪到约 70 万次 Kimi 交互，合规风险被点名）。
- 智谱：海外约 10 万月活 API 用户、300 万免费聊天用户（SCMP 2025）；GLM-4.6 发布后两个月海外付费用户增长 10 倍；GLM-5 曾居开源模型 LMArena 第一。海外收入仍以东南亚为主。

- **MiniMax**: Talkie 长期位列全球陪伴类 AI App 下载前三（2024 年全球下载约 1700 万、MAU 峰值超 2900 万）；Hailuo 累计生成视频超 6 亿条，a16z 多次列入全球 Top 100 GenAI Apps；FY2025 海外收入占比超 70%。
- **阶跃星辰**: 国际站（美国服务器）提供全球 API；Step 3.7 Flash 在 OpenRouter 累计处理 1.91 万亿 tokens（全球调用前列）；终端随吉利银河 M9 等车型出海。海外 API 客户数未披露。
- **零一万物**: PopAi 高峰 ARR 近千万美元（早期主要现金流）；主权 AI 落地哈萨克斯坦（AlemLLM 哈语大模型、Q.AI 合资、500 所学校试点）与泰国（正大万峰智能），并推进沙特、卡塔尔项目。

5.2 与全球竞争对手的差距与定位

- **性能**: 中国开源模型已在推理、编码、Agent 场景进入全球第一梯队，但综合"多模态+Agent 稳定性"仍弱于 OpenAI/Anthropic 闭源旗舰（如 GLM-5.2 SWE-bench 自报 77.8% vs Claude Opus 4.6 约 80-81%）。
- **生态**: DeepSeek 的第三方托管覆盖、Qwen 的衍生模型规模（超 18 万）证明开源生态影响力；但海外开发者对中国模型的信任受地缘与数据合规议题拖累，西方企业倾向通过 AWS/Azure/第三方美国托管使用，而非直连官方 API。
- **差异化**: 中国公司的核心武器是"性价比 + 开源 + 快速迭代"，短板是"数据主权信任 + 生态成熟度 + 地缘政治敞口"。

第六章 风险与合规挑战

6.1 地缘政治与出口管制

- **芯片管制**: H100/Blackwell 始终禁售；H20 在 2025 年"禁售-恢复"反复，H200 名义放开但中国拒买。中国算力缺口巨大，2026 年 H100/H200 租赁价上涨约 40%。
- **模型层管制升级**: 美国已开始将"模型权重/模型服务"纳入管制视野（如对 Anthropic 前沿模型向非美国人限流、中国拟对顶尖模型出境设限），2026-07 美方酝酿重启对华 AI 模型限制，这是出海模式最大的政策变量。
- **个案**: 智谱被列入实体清单（2025-01，推定拒绝）；DeepSeek 被 10+ 个国家和地区政府/机构限制（意大利禁令持续、美国联邦立法未过但机构与州级禁令广泛、韩国整改后解禁、捷克 2025-07 全面禁用）；MiniMax 遭美区苹果商店下架（2024-12）与好莱坞三大片厂版权诉讼（2025-09，索赔 7500 万美元，2026-05 进入取证）；月之暗面被 Anthropic（2026-02）与白宫科技顾问（2026-07）指控蒸馏与获取受限芯片。

6.2 数据合规

- **欧盟**: AI Act（Regulation 2024/1689）对在欧提供服务的非欧盟企业同样适用，GPAI 需满足透明义务；GDPR 要求数据出境有充分性认定或 SCC——中国无充分性认定，柏林 DPA 已认定 DeepSeek 向中国传输数据不合法（2025-06）。
- **中东**: 数据本地化与算法双重认证（中国 DSG + 中东 GCC）是硬约束，曾有中企因数据未本地化失去沙特市场准入的案例，合规成本可占项目预算 25-28%。
- **中国侧**: 《数据出境安全评估办法》对训练数据出境设限，"进不来、出不去"的双向约束需要双轨合规。

6.3 开源许可证与知识产权

- 中国开源模型许可证从 MIT/Apache 2.0 转向"开放权重+商业授权"（K3 的 MaaS 签约门槛、MiniMax H3 的地域排除），下游使用方须审慎评估商用与地域条款。
- "蒸馏指控"（Anthropic 2026-02 指控 DeepSeek、月之暗面、MiniMax）与版权诉讼可能进一步收紧海外对中国模型的信任与采用。

第七章 结论与投资观察

7.1 竞争力分层

维度	海外成熟度	收入证据	地缘敞口
MiniMax	★★★★（C 端出海最深）	海外 >70% 收入	高（版权诉讼、美区监管）
DeepSeek	★★★★（开源+托管生态最广）	年化 \$4-5 亿（全球）	高（多国封禁）
月之暗面	★★★☆☆	海外收入超国内，API 占 70%+	高（美方指控、限制酝酿）
零一万物	★★★★（主权 AI 细分）	海外约 50%	中（避美，聚焦一带一路）
阶跃星辰	★★★☆☆（终端出海为主）	未披露	中（国际站美国服务器）
智谱	★★★☆☆（东南亚早期）	海外约 11.6%	高（实体清单）

7.2 关键投资观察

- 海外收入质量>规模**：MiniMax 与零一万物披露了海外收入占比，DeepSeek/月之暗面虽规模可观但披露不足，估值时须对"未披露项"打折。
- 开源是"政策免疫"资产**：开源权重可绕过实体清单/API 封锁，DeepSeek、月之暗面、智谱均受益；但开源红利正在被许可证商业化条款稀释。
- 算力即地缘**：芯片管制的反复（H20/H200）决定中国模型公司的训练与推理供给，境外算力布局与国产芯片适配（华为昇腾、寒武纪）是两条对冲路径。
- 监管是最大尾部风险**：美国模型层管制若落地，将直接冲击官方 API 直连模式；开源权重+第三方托管+主权 AI 合资是相对抗风险的形态。
- 跟踪指标**：海外收入占比及披露透明度、AWS/Azure 等第三方托管上架进度、OpenRouter/Vercel 等代理平台 token 份额、LMArena/Arena.ai 排名、合规认证（GDPR/SOC 2）进展、以及中美监管动态。

附录：数据来源与免责声明

主要信源：各公司官网与官方文档（platform.kimi.ai / api.moonshot.ai、api-docs.deepseek.com、docs.z.ai、platform.minimax.io、platform.stepfun.ai、lingyiwanwu.com）；招股书与年报（智谱 02513.HK、MiniMax 0100.HK）；Reuters、Bloomberg、The Information、FT、SCMP、TechCrunch、CNBC、AP；36氪、财新、21世纪经济报道、人民网、证券时报、科创板日报；Sensor Tower/AppMagic（经第三方转引）；HuggingFace、GitHub、OpenRouter、LMArena/Arena.ai、Vellum、Artificial Analysis。

数据质量声明：本报告数据基于上述公开来源交叉验证，标注三类来源：官方披露、媒体报道（注明媒体/日期）、估算（注明依据）。以下关键项存在不确定性，请谨慎引用：DeepSeek 营收（\$4-5 亿为 2026-07 The Information 单信源）、月之暗面估值（\$350 亿为 Bloomberg 报道）、阶跃星辰中东合作（经核实为传闻/未证实）、"Yi-X1/Yi-2"与"MiniMax Spike"（经核实无发布记录）、智谱"2024 年 36 亿美元估值"（无权威来源，主流口径为 \$2.7-3B）。**本报告不构成投资建议，数据以官方最新披露为准。**