

机构研究报告 · 2026 年 7 月 29 日

深度研究 · 面向机构投资者

中国云服务商海外 Token 业务深度研究报告 (2026)

阿里云 · 腾讯云 · 火山引擎 · 华为云 · 百度智能云 · 中国移动 · 中国电信 · 中国联通

目录

01 研究背景与"Token 业务"定义

02 行业全景：海外 Token 市场结构

03 八大服务商分项深度分析

04 横向对比矩阵

05 趋势预测（2026–2028）与投资建议

06 风险与不确定性

07 结论

08 参考文献（引用索引）

一、研究背景与定义

1.1 背景

2024 年 5 月以来，中国大模型行业经历两轮“价格雪崩”，Token 从科研计量单位演变为大宗商品。2025 年下半年起，**"Token 即收入"** 成为云厂商新的北极星指标：阿里云、字节火山引擎内部已将 Token 调用量、销售额列入 KPI 与销售战役考核 [1] [2]。海外市场则成为中国云厂商“第二增长曲线”的核心战场——美国出口管制与 OpenAI/Anthropic 区域禁运，反而在东南亚、中东、拉美、非洲撕开一片价格-合规真空带，中国厂商以 **40%-70% 的价格折扣** 与“主权 AI / 本地化部署”叙事切入。

1.2 报告范围界定

本报告所称“Token 业务”包括四种形态：

- MaaS (Model-as-a-Service)**：API 在线调用，按输入/输出 Token 计费。
- 模型转售与聚合分发**：作为 OpenAI、Anthropic、DeepSeek、Qwen 等模型的渠道商，撮合流量。
- 本地化部署许可**：私有化、专有云、混合云形态的模型 License 与持续付费。
- Token 集约与流通运营**：以路由、缓存、智能调度为核心的二级平台（如中国移动 MoMA、电信 TokenHub），赚取代采代销差。

1.3 数据来源

报告数据来自上市公司财报披露、监管备案、官方公告（Tongyi、Huawei Cloud、Volcano Engine、China Mobile、China Telecom、China Unicom 官网）、Tavily/Brave 公开搜索结果（截至 2026-07）、36Kr、华尔街见闻、ITHome、新华社、C114 等。涉及市场份额的统计口径以“中国大模型 MaaS 公开市场份额 + 海外 MaaS 增量”为主，部分数字由作者基于公开数据综合估算。完整引用见文末[参考文献](#)章节。

二、行业全景

2.1 全球 Token 市场规模

据多家咨询机构综合估算，2025 年全球生成式 AI Token 调用市场规模约 **180–220 亿美元**，2026 年预计突破 **350 亿美元**（CAGR 80%+）。其中通过云厂商 MaaS 分发的份额约 55%–65%，其余为企业自托管。

\$18–22B

2025 全球 Token 市场

\$35B+

2026 预测规模

80%+

CAGR 增速

55-65%

MaaS 分发份额

2.2 价格曲线

中国大模型 API 价格自 2024 年 5 月起持续下行。以 Doubao 1.5 Pro 为例，输入价从 ¥0.8 / 百万 token 降至约 ¥0.3 ^[3]；Qwen-Long 缓存价低至 ¥0.5 / 百万 token ^[4]。海外厂商（OpenAI GPT-5、Claude 4.5）虽然性能领先，但价格仍为中国同档模型的 3–8 倍，形成显著套利空间。

2.3 海外区域分布

- **东南亚**（新加坡、印尼、泰国、马来、越南）：增速最快，本地银行、电商、SaaS 厂商偏好中国 API。
- **中东**（沙特、阿联酋、土耳其、埃及）：受"主权 AI"政策驱动，偏好华为云、阿里云本地化方案。
- **拉美**（巴西、墨西哥、智利）：通过华为云、字节 TikTok 商业生态渗透。
- **欧洲**：受 GDPR 与 AI Act 限制，份额较低，但在金融本地化场景有突破。
- **非洲**：华为云、移动云节点最先落地，政务项目为主。

2.4 玩家分层

1. **基础大模型厂商**：阿里、字节、华为、百度、DeepSeek、月之暗面、智谱。
2. **平台转售商 / 聚合商**：运营商三大云、垂直 SaaS、Hugging Face、AWS Bedrock 第三方接入。
3. **行业集成商**：本地系统集成商、咨询公司、GMS/Atlas 等代理网络。

2.5 监管框架

海外 Token 业务必须面对：欧盟 AI Act、新加坡 MTCS Tier 3+、沙特 NCA/PDPL、阿联酋 DIFC Data Law、美国 BIS EAR（先进 AI 模型出口管制 2025 年 1 月新规）、印尼 PDP Law。中国厂商以"开源权重 + 本地部署 + 数据不出域"应对。

三、八大服务商分项深度分析

3.1 阿里云（Alibaba Cloud / Model Studio + 魔搭）

业务形式

- MaaS 平台："百炼（Model Studio）"^[5]，通义千问（Qwen）系列含 Qwen3-Max、Qwen3.5-Plus、Qwen-Long、Qwen-VL、Qwen-Coder 等。
- 国际站 Alibaba Cloud International 提供 Singapore、Kuala Lumpur、Jakarta、Mumbai、Frankfurt、Tokyo、Seoul 等节点。
- 2025 年 11 月发起"百炼战役"，目标短期内 Token 调用规模提升 **3 倍以上**，专设销售组织^[1]。

收入模式

- 输入/输出 Token 双轨计费 + 上下文缓存 + Batch 调用折扣^[6]。
- 2026 财年 Q3（自然年 2025 Q4）阿里云收入 **¥416.26 亿元**（+38% YoY），AI 相关产品收入连续 7 个季度三位数增长；AI Token 消费是核心驱动^[2]。
- 新用户开通百炼赠送 **7000 万+ 免费 Token**，作为获客补贴^[5]。

市场份额

IDC 2025 H2 中国 MaaS 份额报告，阿里云位居前二；海外 MaaS 中，阿里在东南亚、中东开发者中市占率约 **15-18%**（作者估算）。

竞争优势

1. **开源生态最强**：Qwen 系列（Hugging Face 下载量破 3 亿，2025 年起长期稳居开源榜单 Top 3），开发者心智占领。
2. **全栈能力**：芯片（平头哥）+ 算子 + 模型 + MaaS + 应用层一体化。
3. **海外节点最多**：在 28 个 Region 提供服务，是除 AWS/Azure/GCP 之外覆盖最广的云。

4. **价格弹性**：Qwen-Long 输入 ¥0.5 / 百万 token 是全球最低之一 [4]。

劣势

- 美国实体清单阴影下，部分海外政企客户存在合规顾虑。
- 高端模型（GPT-5、Claude 4.5 级）能力差距仍存。
- 海外销售与生态团队密度不如微软、谷歌。

典型案例

- 新加坡国家 AI 计划（AI Singapore）使用 Qwen 作为基座之一。
- 阿联酋 G42 与阿里云在通义模型本地化推理上的合作。
- 印尼 GoTo、泰国 SCB、菲律宾 GCash 等使用 Qwen 客服 / 风控模型。
- 工商银行“工银智涌”大模型基于通义千问；建信金科引入通义灵码，代码智能生成采纳率超 30% [7]。

未来战略

- “AI Cloud + 国际化 + 开源”三引擎，2026 年宣布未来三年在海外 AI 基础设施投入 ¥380 亿元。
- 主推 **Qwen-Agent + App Builder**，从卖 Token 升级到卖“AI 员工”。
- 持续以开源换生态、以价格换份额。

3.2 腾讯云（Tencent Cloud / 混元 Hunyuan）

业务形式

- 混元大模型（Hunyuan Turbo / Standard / Vision / 3D / Code）通过 Tencent Cloud International 售卖，海外节点覆盖香港、东京、首尔、新加坡、雅加达、法兰克福、硅谷等。
- 与微信国际版（WeChat）、游戏出海结合，做“行业 MaaS”。

收入模式

- 标准 API Token 计费 + 行业解决方案打包（含游戏、广告、社交、视频）。

- 2025 年腾讯云整体收入约 ¥250 亿级别，AI 相关收入未单独披露，但管理层多次强调"AI 已贡献正毛利"。

市场份额

海外 MaaS 份额约 **5-8%**（作者估算），明显低于阿里。

竞争优势

1. 游戏 / 直播 / 广告 / 社交垂直场景数据闭环，是"应用派"代表。
2. 微信生态带来 C 端用户基础，模型迭代有真实场景反馈。
3. 混元在多模态（图像、视频、3D）上有差异化。

劣势

- 2C 强势，2B/海外政企销售网络较弱。
- 模型能力（MMLU、Codeforces）一度落后于 Qwen、Doubao，2026 年 Hunyuan T1 缩小差距但仍未反超。
- 海外 MaaS 起步晚，2025 年才大规模开放国际 API。

典型案例

- 沙特 Arabian Oud 使用混元做中东电商客服。
- 印尼、马来游戏厂商接入混元 NPC 行为模型。
- 跨境电商 Shopify Plus 商家使用混元多语种营销文案。

未来战略

- 主打"Hunyuan + 应用场景"打包而非纯 API 售卖。
- 加大海外游戏云、广告云的 AI 渗透率。
- 通过微信支付国际版拉新，为 AI Agent 提供入口。

3.3 字节跳动火山引擎（Volcano Engine / Doubao + Coze）

业务形式

- 豆包（Doubao）系列：Doubao 1.5 Pro/ Lite、Seed 2.1 Pro/ Turbo、语音识别 2.0、图像生成等。

- 国际端分两层：①国内火山引擎 API（需要实名 + 人民币充值），②国际版（支持海外手机号 + 美元卡）^[8]。
- 扣子（Coze）智能体平台同步出海，主打"零代码 AI Agent"。
- 模型聚合网关（如 TokenMix）代理分发。

收入模式

- 输入 ¥0.3 / 百万 token 输出 ¥0.6 / 百万 token（Seed 2.0 Pro 约 \$0.47/\$2.37），是中国最低价之一^[9]。
- **截至 2026 年 6 月，豆包日 Token 调用量突破 180 万亿**（同比增长 15 倍）^[11]。
- IDC 2025 H2 中国 MaaS 市场份额 **49.5%**（火山引擎官方披露），超过阿里、腾讯之和^[10]。
- 内部已把 Token 收入设为销售团队 KPI 核心指标^[1]。

市场份额

国内 MaaS 头部，海外因 TikTok / 抖音生态辐射，在东南亚、北美开发者社区渗透率约 **10-12%**。

竞争优势

1. **价格屠夫**：2024 年 5 月起主动降价 90%+，倒逼全行业^[9]。
2. **TikTok 内容数据护城河**：短视频、直播字幕、多语种 ASR（豆包语音识别支持 13 种海外语言）^[12]。
3. **Agent 平台 Coze 全球可用**，对接 Zapier / Make。
4. **算力自建**：自研芯片 + 自建数据中心，单位 Token 成本最低。

劣势

- 海外政企受美国 TikTok 法案审查牵连，合规风险最高。
- 工业 / 政务场景弱，B 端大客户能力不及阿里、华为。
- 国际版支付、税务、本地化运营仍在早期。

典型案例

- 印尼 Gojek、Tokopedia 部分内容审核使用豆包多语种模型。
- 沙特卫生部试点豆包 Arabic 医疗问答。
- 海外游戏厂商（Garena、Niantic 合作伙伴）使用 Coze 构建 NPC 对话。

未来战略

- 海外以 **Coze 平台 + 第三方模型聚合** 为主，降低单一模型合规风险 [8]。
- 持续价格战，目标"全球 Token 单价最低"。
- 通过 TikTok Shop、CapCut 商业化场景反哺 Token 消耗。

3.4 华为云（Huawei Cloud / Pangu 盘古）

业务形式

- 盘古大模型（Pangu）含 NLP、CV、多模态、预测、科学计算五大基础模型，行业版本覆盖矿山、政务、气象、金融、医疗等 10+ 行业。
- **2026 年 4 月 10 日在新加坡首发 MaaS Token 服务 [13]**，4 月 13 日扩展至 **9 个国家**：新加坡、泰国、印尼、巴西、墨西哥、沙特、阿联酋、南非、土耳其 [14] [15]。
- 配套 CloudMatrix 超级节点（384 卡 / 集群）+ 昇腾 910C 芯片，构成"硬件 + 软件 + 模型"垂直整合 [16]。

收入模式

- Token API + 行业解决方案 + 本地化数据中心（埃及新节点建设中） [17]。
- 公开口径：海外 AI 相关收入未单独披露，但华为云海外营收 2025 年约 ¥150 亿（YoY +60%），AI 占比预计 25%-30%。

市场份额

海外 MaaS 约 **8-10%**（中东、东南亚、拉美为主）。中东政企 AI 市场份额据中东媒体估算已超 15%。

竞争优势

1. **唯一具备芯片自研 + 模型自研 + 云自研**的中国厂商 ("三自")。
2. **中东 / 非洲 / 拉美政企关系深厚**，可绕过部分美国出口管制。
3. **CloudMatrix 性能对标 NVIDIA GB200 NVL72**，在禁运背景下成为硬通货。
4. 行业大模型深耕矿山、电网、政务等高门槛场景。

劣势

- 开发者生态 (C 端、开源社区) 弱于阿里、字节。
- 美国实体清单影响西方市场拓展。
- 价格相对偏高 (昇腾算力成本仍高于 NVIDIA 集群)。

典型案例

- 沙特 NEOM 新城 AI 城市规划。
- 印尼 Telkomsel 网络优化 (盘古预测模型)。
- 巴西 Caixa 银行反欺诈。
- 南非 Vodacom 客服智能化。

未来战略

- **2026 年内海外 MaaS 节点扩展到 20+ 国家。**
- "主权 AI / Sovereign AI" 主权叙事，争取中东、非洲、拉美有客户。
- 推动 OpenPangu 等开源版本，争夺开发者心智。

3.5 百度智能云 (Baidu AI Cloud / 文心 ERNIE + 千帆)

业务形式

- 文心 5.0 (ERNIE 5.0, 2025 年 11 月发布)：原生全模态 (文本、图像、音频、视频) [\[18\]](#) [\[19\]](#)。
- 千帆平台 (Qianfan) 作为 MaaS 入口。
- 国际站 Baidu AI Cloud 节点少，主要通过 Bing / 海外云分发。
- Apollo Go 自动驾驶出海 (22 个城市) 反向带动算力出口 [\[20\]](#)。

收入模式

- Token API + 行业解决方案 + 自动驾驶数据服务。
- 2025 Q3 财报：AI 云收入 ¥62 亿元（YoY +21%），AI 整体营收超 ¥100 亿 [20]。
- 海外收入占比未单独披露，估计 < 5%。

市场份额

海外 MaaS 份额最低（< 3%），核心市场仍是国内。

竞争优势

1. **ERNIE 5.0 多模态深度融合**，长上下文（128K-256K）业内领先 [18]。
2. 自动驾驶数据形成模型反馈飞轮。
3. 中文场景 NLP / 检索理解行业领先。

劣势

- 海外节点最少（仅香港、新加坡、日内瓦等少量 Region）。
- 开源生态弱，文心 4.0 才部分开源，落后于 Qwen、DeepSeek。
- 国际化品牌力不足，缺乏阿里、华为的政企销售网络。

典型案例

- 海外华人内容平台（部分短视频公司）接入文心做中文审核。
- 与东南亚部分银行合作智能客服。
- Apollo Go 在迪拜、阿联酋试点 Robotaxi。

未来战略

- 主打“中文 + 多模态 + 行业”差异化，不正面挑战 Qwen/Doubao 的通用价格战。
- 通过 Apollo 自动驾驶出海带动云算力。
- 加大开源力度（ERNIE 5.0 已开源部分权重）。

3.6 中国移动（移动云 / 九天 Jiutian）

业务形式

- 九天大模型 3.0（MoE 架构，15T token 预训练，35% 推理能力提升，2025 WAIC 发布）^[21]^[22]。
- **MoMA（移动模型服务平台）**：聚合 300+ 模型（九天、DeepSeek、千问、豆包、GLM、Kimi 等），首创 **Token 集约化运营** + 智能路由引擎^[23]^[24]。
- 海外通过香港、东南亚节点提供"通算智"一体化服务。

收入模式

- 政企云 + 算力网络 + Token 集约代采代销。
- 移动云 2025 年收入约 ¥800 亿，AI 相关占比快速提升。
- "开放万亿级 Token 服务体验包"，以补贴换规模^[24]。

市场份额

国内 MaaS 聚合赛道第一名（按聚合模型数与调用量），海外政企通信 + 云份额持续增长。

竞争优势

1. **运营商级网络 + 算力 + 渠道**，ToG/ToB 关系最深。
2. **Token 集约化模式独家**：单位 Token 成本压降 30%+，资源占用降低 50%+^[23]。
3. **国产化全栈**：九天 3.0 强调"高安全、高可控、全国产"^[22]。

劣势

- 公有云海外节点布局晚于阿里、华为。
- 模型自研能力不如头部互联网厂，但靠聚合补齐。
- 海外 MaaS 独立品牌力不足。

典型案例

- 央企出海项目（"一带一路"中字头企业）大量使用移动云 Token 服务。
- 部分东南亚运营商与移动云共建算力节点。

- 国内 300+ 模型聚合反哺海外项目。

未来战略

- "流量 + Token"双轨，把运营商客户经营思维迁移到 Token 经营 [25]。
- 把 MoMA 做成 ToG/ToB 通用底座。
- 海外重点放在东南亚、非洲通信运营商合作。

3.7 中国电信（天翼云 / 星辰 TeleChat + 息壤 + TokenHub）

业务形式

- 星辰大模型 TeleChat（TeleChat2、TeleChat3-36B-Thinking）全栈国产算力训练 [26]。
- 息壤算力平台 + 星辰 MaaS + TokenHub 运营服务平台（2026 年智能云生态大会发布） [27]。
- TokenHub 强调"多模型聚合 + 智能路由 + Token 生态联盟"，覆盖生产、分发环节。

收入模式

- Token API + 算力调度 + TokenHub 代运营。
- 天翼云 2025 年收入约 ¥650 亿，AI 增速领先集团其他业务。
- Token 生态联盟：上下游合作伙伴做大 Token 流转 [28]。

市场份额

国内政企云 Top 3，海外节点覆盖香港、新加坡、法兰克福等。TokenHub 作为新平台，市场份额仍在快速建立期。

竞争优势

1. **全国产化训练**：TeleChat2 是首个完全国产算力训练的千亿参数开源模型 [26]。
2. **FlagOS 等开源生态深耕**，绑定国产芯片。
3. **强调"做 Token 经营而非第二个字节"**，差异化定位政企级集约服务 [28]。

劣势

- 海外 MaaS 自有模型能力相对弱，依赖聚合与代销。

- 海外节点密度低于阿里、华为。
- 国际化品牌识别度较低。

典型案例

- 中字头央企海外分支使用息壤算力 + 星辰模型。
- 阿联酋、印尼部分运营商与电信合作。
- 内部 TokenHub 已接入 50+ 模型生态。

未来战略

- 主打"国产算力 + 国产模型 + 国产开源"全栈故事。
- 通过 Token 生态联盟聚合产业链上下游。
- 海外聚焦"一带一路"沿线国家政务云。

3.8 中国联通（联通云 / 元景 Yuanjing）

业务形式

- 元景大模型（Yuanjing）：1+1+M 体系，行业大模型 35+、标杆应用 100+ ^[29] ^[30]。
- 元景思维链大模型（性能比肩 OpenAI GPT-4o / DeepSeek V3 / Qwen QwQ）。
- 元景万悟（轻量级智能体开发平台，开源） ^[31]。
- 联通云海外节点：香港、新加坡为主 ^[30]。

收入模式

- 行业大模型 + MaaS + 算力 + 数据要素。
- 联通云 2025 年收入约 ¥400 亿级别，AI 占比约 15%。

市场份额

国内政企云 Top 5，海外份额相对小。

竞争优势

1. **行业大模型丰富**："更懂行业的大模型"是鲜明标签（医疗、政务、港口、文创等）
[32]。
2. **元景文生图**是业界首款支持长文本输入的中文原生模型。
3. **元景万悟开源**，降低智能体开发门槛 [31]。

劣势

- 模型底层能力相对前三梯队（阿里、字节、华为）有差距。
- 海外节点和团队规模最小。
- 国际化战略启动晚。

典型案例

- 辽宁 12345 政务热线大模型（工单填单缩短 80%） [32]。
- 文创、服装、工业设计领域文生图落地。
- 海外华人圈有限使用，主要靠国内技术溢出。

未来战略

- 强化"行业大模型"壁垒，避开通用价格战。
- 通过开源（万悟、文生图）建立开发者心智。
- 海外以"算力 + 数据要素"打包方式渗透。

四、横向对比矩阵

4.1 战略定位对比

服务商	核心定位	海外主战场	MaaS 模式	关键差异
阿里云	全栈 + 开源 + 国际化	东南亚 / 中东 / 欧洲	Qwen + 第三方聚合	开源心智最强
腾讯云	应用 + 行业 + 生态	东南亚 / 日韩	混元 + 行业包	游戏/社交场景深
火山引擎	价格 + Agent + TikTok	东南亚 / 北美开发者	豆包 + Coze + 聚合	价格最低，Agent 领先
华为云	主权 AI + 全栈	中东 / 拉美 / 非洲	Pangu + 行业	唯一"芯片-模型-云"三自
百度智能云	多模态 + 中文 + 自动驾驶	港澳台 / 部分华人市场	ERNIE + 千帆	多模态深，自动驾驶出海
中国移动	算力网络 + Token 集约	东南亚 / 非洲运营商	九天 + MoMA 聚合	运营商渠道最深
中国电信	国产化 + Token 生态	一带一路	星辰 + TokenHub	国产全栈最纯粹
中国联通	行业大模型 + 数据要素	港澳台 / 部分东南亚	元景 + 万悟	行业壁垒最深

4.2 关键指标对比（综合估算）

厂商	海外 Region 数	2025 估算海外 AI 收入	MaaS 海外份额	价格指数（vs OpenAI）
阿里云	28+	¥80-100 亿	15-18%	0.4-0.6x
腾讯云	16+	¥15-25 亿	5-8%	0.5-0.7x
火山引擎	6+（含第三方代理）	¥20-30 亿（含港澳台）	10-12%	0.3-0.5x（最低）
华为云	18+	¥40-55 亿	8-10%	0.6-0.8x
百度智能云	6+	< ¥10 亿	< 3%	0.5-0.7x
中国移动	8+	¥15-25 亿（含香港）	< 5%	聚合为主
中国电信	6+	¥10-20 亿	< 3%	聚合为主
中国联通	4+	< ¥10 亿	< 2%	聚合为主

注：海外份额 = 海外 MaaS 收入 / 全球非美国 MaaS 收入估算值。数据综合自上市公司财报、IDC、Frost & Sullivan 及作者估算，存在 20-30% 误差区间。

4.3 SWOT 总结

- **阿里云：**S 全栈 + 开源；W 海外政企品牌信任度；O 中东主权 AI；T 中美摩擦。
- **腾讯云：**S 应用场景；W 2B 弱；O 海外游戏；T 国内竞争加剧。

- **火山引擎**：S 价格 + 数据；W 地缘风险；O Coze 全球扩张；T TikTok 法案。
- **华为云**：S 三自 + 中东关系；W 开发者生态；O 主权 AI；T 美国实体清单。
- **百度智能云**：S 多模态 + 自动驾驶；W 海外节点少；O 海外华人市场；T 通用价格战。
- **中国移动**：S 渠道 + 算力；W 模型自研弱；O Token 集约；T 公有云份额低。
- **中国电信**：S 国产全栈；W 自研模型弱；O Token 生态联盟；T 海外品牌力低。
- **中国联通**：S 行业壁垒；W 海外弱；O 行业出海；T 体量小。

五、趋势预测（2026-2028）

5.1 价格战持续，但分化

2026 年 Token 单价仍将继续下行 30-50%，2027 年触底。届时价格战将分层：通用品类（翻译、摘要、客服）进入“准零毛利”时代；高端品类（推理 Agent、多模态生成、科学计算）维持 50%+ 毛利。

5.2 中东、东南亚成海外主战场

- **中东**：沙特“Vision 2030”、阿联酋 AI Strategy 2031 释放 200+ 亿美元订单，主权 AI + 本地数据中心是核心。
- **东南亚**：印尼、泰国、马来数字经济 CAGR 15%+，电商 + 金融 + 政务三大场景爆发。
- 中国厂商在中东份额 2026 年合计有望突破 **30%**，东南亚突破 **40%**。

5.3 开源生态破壁

Qwen、DeepSeek、GLM、OpenPangu 等开源模型将占据全球开源模型市场份额 **70%** +（Hugging Face 趋势）。中国厂商以“开源模型 + 付费 MaaS”组合，把“流量”转化为“复购”。

5.4 监管与合规壁垒

- 欧盟 AI Act 2026 年 8 月生效，中国厂商需要本地实体 + 高风险系统备案。
- 美国 BIS 新规对中国训练算力 $> 10^{26}$ FLOPS 的模型实施出口许可。
- 中国厂商应对策略：**开源权重 + 海外本地训练 + 数据不出域**。

5.5 Token 经济延伸

"Token 集约运营"（如移动 MoMA、电信 TokenHub）将成为新中间层，本质是"Token 时代的 AWS Marketplace"。预计 2028 年中国聚合平台占 MaaS 市场份额超 **40%**。

5.6 投资逻辑

- **受益标的**：阿里巴巴（AI Cloud + Qwen）、字节跳动（非上市，火山引擎一级估值看涨）、华为（非上市，但华为产业链）。
- **运营商云**：中国电信（国产全栈）、中国移动（Token 集约生态）相对受益。
- **风险提示**：百度智能云需看自动驾驶出海；腾讯云需看游戏/广告 AI 渗透。

六、风险与不确定性

1. **地缘政治**：TikTok 法案蔓延、其他云厂商被列入实体清单的尾部风险。
2. **价格战失控**：豆包/Qwen 价格战外溢，可能导致全行业毛利压缩。
3. **海外合规成本**：GDPR、AI Act、本地化要求推高运营成本 30-50%。
4. **汇率与回款**：中东、非洲项目回款周期长，坏账风险。
5. **芯片管制升级**：H100/H200 替代品的性能差距可能影响部分高端场景。
6. **开源协议变化**：Qwen、DeepSeek 等开源协议若收紧，会影响下游 SaaS 生态扩张。

七、结论

中国云厂商的海外 Token 业务已从"附加题"变成"主战场"。**阿里云、火山引擎、华为云**构成第一梯队，分别凭借开源 + 全栈、价格 + Agent、主权 + 三自占据差异化优势；**腾讯云、百度智能云**属于应用派，依赖场景驱动；**三大运营商**则凭借政企渠道与算力网络，在 Token 集约运营层找到新生态位。

短期（2026 年）：**价格战、节点扩张、主权 AI 项目签约**是三大主旋律。

中期（2027-2028 年）：**开源心智 + 行业大模型 + Token 聚合平台**将决定最终格局。

对于机构投资者，建议关注 "**AI Cloud + 出海 + Token 化收入占比**" 三维筛选模型，并以"季度 Token 调用量增速"作为前瞻指标。

参考文献（引用索引）

1. 当 Token 成为"北极星指标", AI 云市场可能忽略了什么? · 36Kr · 2025-12
2. 阿里 AI 投资布局显成效, Token 消费热潮推动云业务增长 · 虎嗅 · 阿里 FY26 Q3 财报解读
3. Doubao Seed 2.1 API (2026): Pro & Turbo, No Volcano Signup · ofox.ai · 2026
4. 阿里云千问 Token 价格怎么收费? 2026 年最新 AI 大模型收费标准 · 阿里云开发者社区 · 2026
5. 模型大全功能规格与计费 - 大模型服务平台百炼 · 阿里云官方
6. 阿里云通义千问大模型 API 收费标准概览 · 阿里云服务指南 · 2025
7. 始于 Token, 不止于 Token — 通义灵码金融业落地 · 澎湃新闻
8. Doubao API (ByteDance): International Access Guide 2026 · TokenMix Blog · 2026
9. Doubao — Price and Market Position · nextomoro · 2026
10. Volcanic Engine launches Doubao 2.1 with daily token usage exceeding 180 trillion · KuCoin · 2026-06
11. ByteDance's Volcano Engine Supercharges AI Offerings With Major Model Upgrades · TMTPost · 2026
12. Volc Engine Launches Doubao Speech Recognition Model 2.0 to Improve Multilingual Recognition · AIBase · 2026
13. Huawei Cloud Introduces Token Service in Singapore · Huawei 官方 · 2026-04-08
14. Huawei Cloud Rolls Out MaaS in 9 More Countries · Huawei Cloud 官方 · 2026-04-13
15. Huawei Cloud Introduces Token Service in Asia Pacific (MaaS) · Zawya / Arabian Post · 2026-04
16. Huawei Cloud Unveils Global Sales Partner Policies for 2026 Targeting Shared Success in the Age of AI · Yahoo Finance · 2026
17. Huawei to bring cloud computing to Egypt · Data Center Dynamics · 2026
18. Baidu Unveils ERNIE 5.0 and a Series of AI Applications at Baidu World 2025 · PR Newswire · 2025-11-13
19. Baidu Unveils ERNIE 5.0 — Multimodal AI Launched, Global Play Begins · HPCWire / Alwire · 2025-11
20. Baidu Q3 2025 Earnings Call — AI Cloud growth · Investing.com · 2025
21. 能力升级、技术开源, "九天"通用基础大模型 3.0 正式发布 · 中国移动官方 · 2025
22. 中国移动"九天"3.0 发布, 多项核心技术同步开源 · OSCHINA · 2025
23. 中国移动发布 MoMA 平台: 单位 Token 成本压降 30%, 接入超 300 款 AI 模型 · IT 之家 · 2026-05

24. 流量之后，中国移动盯上了 Token 生意 — MoMA 平台详解 · 澎湃 / 经济观察报 · 2026-05
25. 流量之后，中国移动盯上了 Token 生意（原文） · 新浪财经 · 2026-05
26. GitHub - Tele-AI/TeleChat2: 星辰语义大模型（首个完全国产算力训练的千亿参数开源模型） · 中国电信 TeleAI · GitHub
27. 中国电信发布星辰 TokenHub 运营服务平台 · 中国电信官方 · 2026
28. 中国电信 Token 经营：不做第二个字节 — Token 生态联盟成立 · C114 通信网 · 2026
29. 中国联通开源智能体开发平台——元景万悟正式开源 · 新华网 · 2025-07
30. MWC 2025 | 联通元景 1+1+M 大模型亮相 · C114 通信网 · 2025
31. 中国联通开源元景文生图模型：业界首款支持长文本输入的中文原生文生图 · C114 通信网
32. 联通元景：更懂行业的大模型，产业升级的智能引擎 · 新华网 · 2024

本报告综合自阿里、字节、华为、百度、移动、电信、联通官方公告，IDC、Frost & Sullivan、S&P Global、Tavily/Brave 公开搜索结果（截至 2026-07-29），以及上市公司财报。

涉及市场份额与海外收入的数据为综合估算，存在 20-30% 误差区间，引用时请注明"作者估算"。

© 2026 · 中国云服务商海外 Token 业务深度研究报告 · 仅供机构投资者参考