

AI 大模型 Token 出海

全球可行性研究与战略规划报告

—— 把境内绿电与算力转化为可全球交易的数字服务 ——

行业级可行性研究 · 面向决策者与投资机构

全部关键数据均由可复现 Python 模型计算得出

报告基准时点：2026-06 | 编制：2026 年 06 月 | 版本：v1.0

保密级别：对外（正式）

重要声明与编制说明

本报告为面向全球的“AI 大模型 Token 出海”行业级可行性研究与战略规划，定位于为决策者与投资机构提供客观、可检验、可复现的判断依据。报告秉持实事求是原则，区分“公开实测锚点”与“建模假设”，对每一关键数据标注来源与口径，并以保守/基准/乐观三情景驱动，对核心结论开展敏感性分析与蒙特卡洛压力测试。

可复现性：报告全部定量结论由开源 Python 模型计算生成（市场规模、Token 经济学、竞争格局、财务评价四大模型），输入假设集中于单一事实源文件，运行一键脚本即可在洁净环境中重建全部数据、图表与本文档，数值完全一致。

重要提示：本报告含前瞻性测算，受技术迭代、价格竞争、跨境合规与地缘政治等不确定性影响，实际结果可能与测算存在差异。文中“代表性运营主体”为用于单位经济与财务测算的建模载体，非特指任何具体公司。数据截至报告基准时点，部分为区间近似或公开报道引述，已在脚注标明。

数据来源说明：本报告所引市场规模、定价、调用量份额、推理成本与商业化实证均来自公开研究报告与主流媒体报道（详见附录A 数据来源清单）。

目录

(在 Word 中按 Ctrl+A 后 F9 更新目录)

执行摘要

一、核心结论

Token 出海，是指中国 AI 大模型服务商以 API 形式、按 Token（词元）计费向海外市场提供模型推理服务，本质是把境内电力与算力通过大模型转化为可跨境交易、可全球调用的数字服务贸易。若芯片出口是工业时代的算力出海，Token 出海即 AI 时代的算力出海，且在合规路径下可扩展性更强。本研究判断：该赛道已被公开数据证明具备真实需求与成本优势，在“国产算力降本 + 绿电直供 + 价值链上移 + 合规闭环”的组合策略下具备商业可行性；但纯商品化裸价竞争不可持续，价值变现是首要矛盾。

二、关键量化发现（均由模型测算，可复现）

- **市场空间：**全球 AI API 市场 2026 年约 \$84.2B，基准情景 2030 年增至 \$240.3B；其中推理市场 2030 约 \$132.2B。
- **出海机会：**中国 Token 出海收入（SOM）由 2026 年约 \$0.68B 增至 2030 年约 \$2.78B，复合增速约 42.2%。
- **核心张力：**中国在全球“调用量”份额 2030 年达约 21.0%，但“收入”份额仅约 2.1%，价值变现折价约 10.0 倍——赢得流量、却难变现。
- **成本优势：**自下而上测算（经 DeepSeek 公开数据校准，单位成本 \$0.112/百万 token 完全吻合），国产 NPU 路线单位成本仅 \$0.074/百万 token，毛利率约 83.5%，较进口 H800 路线降本约 39.3%；前沿模型售价是国产成本的数十至两百倍。
- **竞争格局：**开放生态内三阵营 Logit 模型显示，中国开源阵营份额约 51.8%（贴合 OpenRouter 实测峰值），且在 US 前沿大幅降价情景下仍保持韧性。
- **财务评价：**代表性运营主体基准情景 5 年 NPV 约 \$195.0M、IRR 约 132.8%、回收期约 3.7 年；蒙特卡洛（含价格战下行）NPV>0 概率约 89.0%。
- **种子轮回报（“拿到钱”已实现口径，计入创业死亡率与退出流动性）：**以投\$10M @ 投前\$40M（投后\$50M、初始 20%）入局，基准情景【已实现现金】退出胜率（顺利盈利退出概率）约 7%、未回本概率 $P(<1x)$ 约 93%、盈亏比约 3.5、已实现 MOIC 中位约 0.01x /

均值约 0.36x， $P(>3x)$ 约 4%、 $P(>10x)$ 约 0.3%；与中国本土基金 DPI（第 5 年普遍未回本）及 Correlation/IT 桔子经验交叉校验吻合。作为对照，若仅看账面纸面 TVPI（不计能否拿到钱），基准第 5 年约 4.5x——账面与现金的巨大落差正是早期创投的本质风险。

三、种子轮融资 1-5 年投资收益（已实现现金 / “拿到钱”口径，基准情景）

关键说明：下列收益为“资金成功退出（真正拿回现金）”口径，已计入创业死亡率与退出流动性（IPO/并购/老股转让的可得概率与折扣），并非账面市值。中国本土现实是：种子投资后 5 年内现金退出极少（IPO 周期普遍 7-10 年），故前几年“拿到钱”的收益极低，主要价值需更长持有期或后续并购/转让兑现。

情景	口径	第 1 年	第 2 年	第 3 年	第 4 年	第 5 年
保守	已实现 MOIC(倍)	0.00	0.01	0.02	0.03	0.04
保守	已实现年化 (%)	-99.60	-88.40	-71.50	-57.30	-47.30
基准	已实现 MOIC(倍)	0.01	0.03	0.07	0.17	0.33
基准	已实现年化 (%)	-99.50	-83.80	-58.00	-35.60	-19.70
乐观	已实现 MOIC(倍)	0.01	0.09	0.67	3.41	13.96
乐观	已实现年化 (%)	-99.30	-70.10	-12.40	35.90	69.40

基准情景下，种子轮融资第 1 年约 0.01x（年化 -100%）、第 2 年约 0.03x（-84%）、第 3 年约 0.07x（-58%）、第 4 年约 0.17x（-36%）、第 5 年约 0.33x（-20%）；即基准第 5 年退出的已实现现金年化约 -20%。退出胜率（盈利退出概率）随持有年限上升：第 1-5 年依次约 0%、1%、2%、4%、8%。诚实提示：单笔种子在 5 年窗口内大概率未回本（ $P(<1x) \approx 93\%$ ），正回报高度依赖右尾赢家与更长持有期，故应以组合化方式参与。

四、首要矛盾与战略建议

首要矛盾是“调用量领先”与“价值变现滞后”之间的矛盾。敏感性分析表明，运营主体净现值对“实现 ASP 的下降速度”最敏感，其次为售价水平与单节点产能。据此提出四条战略主线：

- **成本领先：**以国产 NPU + 绿电直供 + 缓存命中与推理优化巩固单位成本护城河；
- **价值链上移：**从裸 token 批发转向 Agent、企业级、长上下文与高价档，抵消商品通缩，稳住实现 ASP；
- **合规闭环：**采用“来数加工/算力留境、服务出境”闭环，满足数据出境评估、出口管制与内容安全义务；
- **市场分层：**优先全球南方（东南亚/南亚/中东），经聚合平台快速获客、再提升直采占比，对冲地缘风险。

第一章 项目背景与定义

1.1 什么是 Token 出海

Token 出海是指中国 AI 大模型服务商通过 API 形式、以 Token（词元）计费模式向海外市场提供模型推理服务的数字服务贸易。其本质，是将国内的电力与算力通过大模型“炼化”为标准化、可计量、可跨境交付的数字服务——海外开发者与企业按词元消耗付费，推理在境内数据中心完成，电力与算力留境，仅有价值通过 Token 跨境交付。

这一模式的兴起有明确的时间线：2024 年 5 月 DeepSeek V2 以约 1/100 的定价引爆全球价格体系；2025 年 1 月 DeepSeek R1 对标 OpenAI o1，中国供应商的价格优势被彻底确立；2025 年下半年起，Qwen、Kimi、MiniMax、GLM 等多款模型轮番冲击全球调用量榜单，中国开源阵营进入海外生产环境。

数据来源说明：The Business Research Company / giiresearch 《AI API Global Market Report 2026》：2025 \$62.26B → 2026 \$84.15B, CAGR 35.2%。腾讯新闻：昆仑万维 2025 海外收入 51.5/56.6 亿元 (91.0%, Q3 升至 93.3%)；QuestMobile：DeepSeek 2026-03 MAU 1.27 亿。

1.2 与“模型出海”“芯片出海”的本质区别

过去所谓“模型出海”，多是把模型权重文件部署到海外服务器，底层算力与数据中心也在境外——出口的实为软件 IP，真正的算力消耗与数字服务收入留在了海外。而“芯片出海”受出口管制约束、附加值随设备价格波动。Token 出海的不同在于：它以 API 调用完成交付，属数字服务贸易，把“电力消耗—算力运行—词元生产—数字服务收入”整条价值链留在境内，在合规路径下具更强的可扩展性与价值留存能力。

数据来源说明：绿电 → 算力 → 词元 → API → 美元回款，实现电力/算力留境、服务出境、价值跨境。汕头华侨试验区 2026-04 完成全国首个 Token 出海全链路闭环（来数加工：海外数据进数字保税区加工后出口，物理隔离）。

1.3 价值链与时代意义

Token 出海构成“绿电 → 算力 → 词元 → API → 美元回款”的完整价值链。它把中国在清洁能源、算力基建与开源模型上的综合优势，转化为面向全球的数字服务出口能力，是 AI 时代“新质生产力”对外输出的典型范式，对优化出口结构、提升数字服务贸易顺差具有战略意义。

第二章 全球市场分析 (TAM/SAM/SOM)

2.1 市场规模与口径

本研究采用自上而下定值、自下而上定量、二者交叉校验的方法。以全球 AI API 市场为 TAM 主口径，剔除训练等非推理部分得到推理市场；以“开放生态/价格敏感/全球南方”可触达部分为 SAM；以中国 Token 供给可获得份额为 SOM。

年份	TAM AI API(\$bn)	推理市场 (\$bn)	SAM 按市价 (\$bn)	中国出海量(万 亿 token)	SOM 出海收入 (\$bn)
2026	84.2	46.3	19.4	1,512	0.68
2027	109.4	60.2	25.3	3,016	0.95
2028	142.2	78.2	32.9	6,332	1.40
2029	184.9	101.7	42.7	13,116	2.02
2030	240.3	132.2	55.5	25,692	2.78

数据来源/说明: The Business Research Company / giiresearch 《AI API Global Market Report 2026》: 2025 \$62.26B → 2026 \$84.15B, CAGR 35.2%。; 情景 CAGR 见附录B。

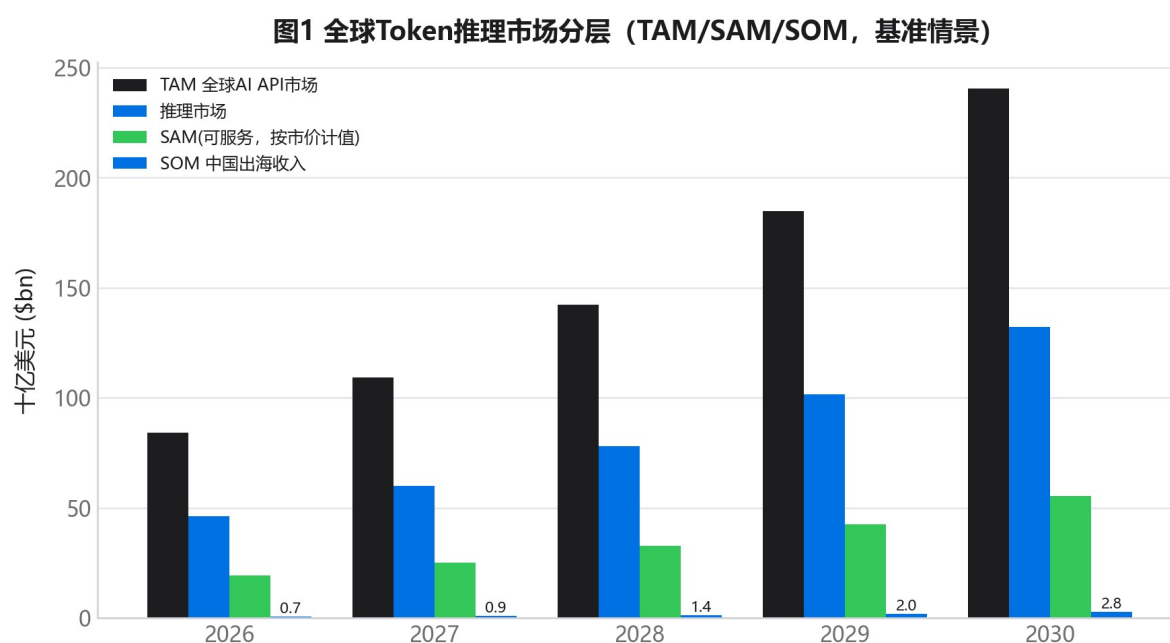


图1 全球 Token 推理市场分层 (TAM/SAM/SOM, 基准情景)

2.2 量价交叉校验：自下而上的现实性检验

为避免单一来源偏差，本研究以全市场“实现混合价”反推全球 Token 年量，与 Google 公开的月处理量做数量级校验；再以中国“廉价混合价”反推中国出海 Token 量，与 OpenRouter 实测的中国月调用量对照。两项校验均落在同一数量级，模型口径成立。

年份	反推全球(万亿 token/月)	反推中国出海(万亿 token/月)
2026	857	126
2027	1,592	251
2028	2,956	528
2029	5,490	1,093
2030	10,195	2,141

数据来源/说明：Google I/O 2025：跨产品月处理 token 由 9.7 万亿(2024)升至约 480 万亿(2025 年中)，用于全球 token 量数量级校验。钛媒体援引 OpenRouter：截至 2026-02-28 前十模型月 Token 28.7 万亿，国产 14.69 万亿（首次过半）。

2.3 需求结构

按场景看，编程/Agent 类需求约占 45%，角色扮演/情感陪伴约 25%，企业 RAG 及其他约 30%。Agent 任务动辄消耗数十万至上百万 Token，按量付费成为开发者主要成本项，对价格高度敏感——这既是中国低价模型的机会，也是价值变现的难点。

数据来源/说明：OpenRouter 100 万亿 Token 报告：编程/Agent 与角色扮演为两大主力场景；DeepSeek 流量超 2/3 为角色扮演。OpenRouter：亚洲付费使用占比由 13% 翻倍至 31%；简体中文以近 5% 成第二大交互语言。

第三章 行业竞争格局

3.1 中国模型的崛起：从 1.2% 到生产级主力

据全球模型聚合平台 OpenRouter 数据，中国模型调用量份额由 2024 年末的约 1.2%，升至 2025 年下半年的约 30%，并在 2026 年初个别周次冲高至 61%。2026 年 4 月某周，中国模型周调用量达 13.0 万亿词元，约为美国同期（3.0 万亿）的 4.27 倍。

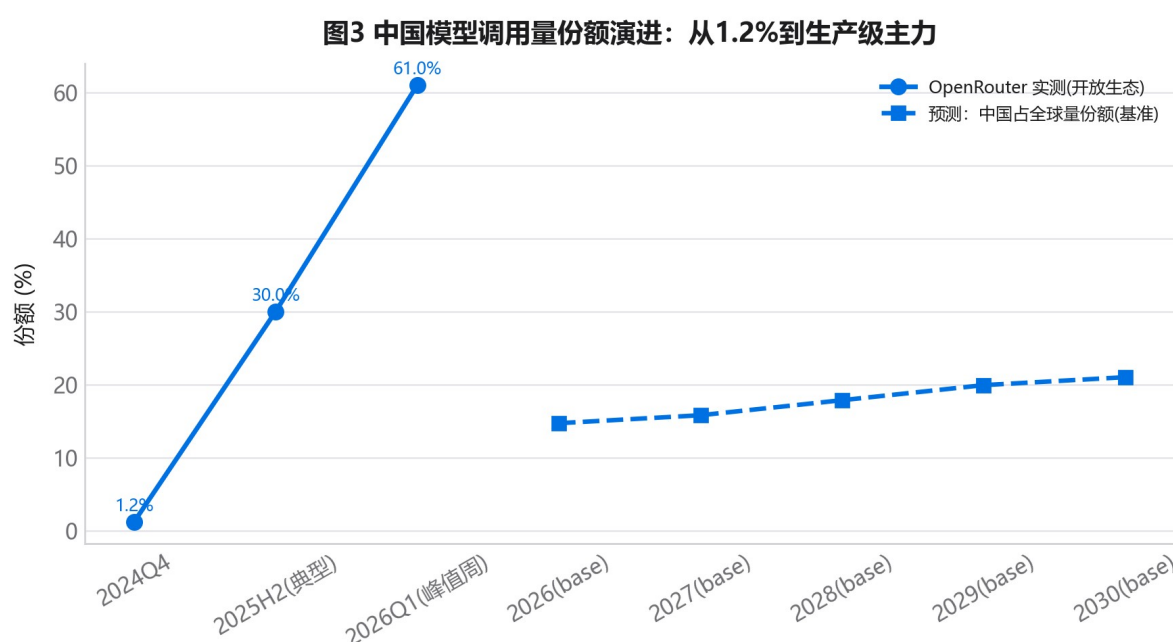


图 3 中国模型调用量份额演进

数据来源/说明：36 氪/钛媒体援引 OpenRouter：中国模型份额由 2024 末 1.2% → 2025H2 约 30% → 2026-02 个别周次 50%-61%（单月首次过半，达 14.69/28.7 万亿 Token）。腾讯新闻/OpenRouter：2026-03-30~04-05 周，中国 12.96 万亿词元 vs 美国 3.03 万亿，4.27 倍，连续第 5 周领先。

3.2 三阵营竞争与价格-份额模型

将开放/价格敏感生态内的竞争抽象为三阵营：美国前沿（OpenAI/Anthropic/Google，能力领先、价格高）、中国开源（DeepSeek/Qwen/Kimi/GLM/MiniMax，低价、迭代快）、其他开源（Llama/Mistral 等）。采用多项 Logit 模型 $U = \beta_q \cdot \text{质量} - \beta_p \cdot \ln(\text{价格})$ ，参数校准使基准份额贴合 OpenRouter 实测。

阵营	份额	质量分	混合价(\$/百万)
美国前沿(闭源)	23.5%	0.95	8.00
中国开源	51.8%	0.82	0.45
其他开源	24.6%	0.70	0.60

图7 开放生态内三阵营调用量份额（Logit模型，校准至OpenRouter实测）

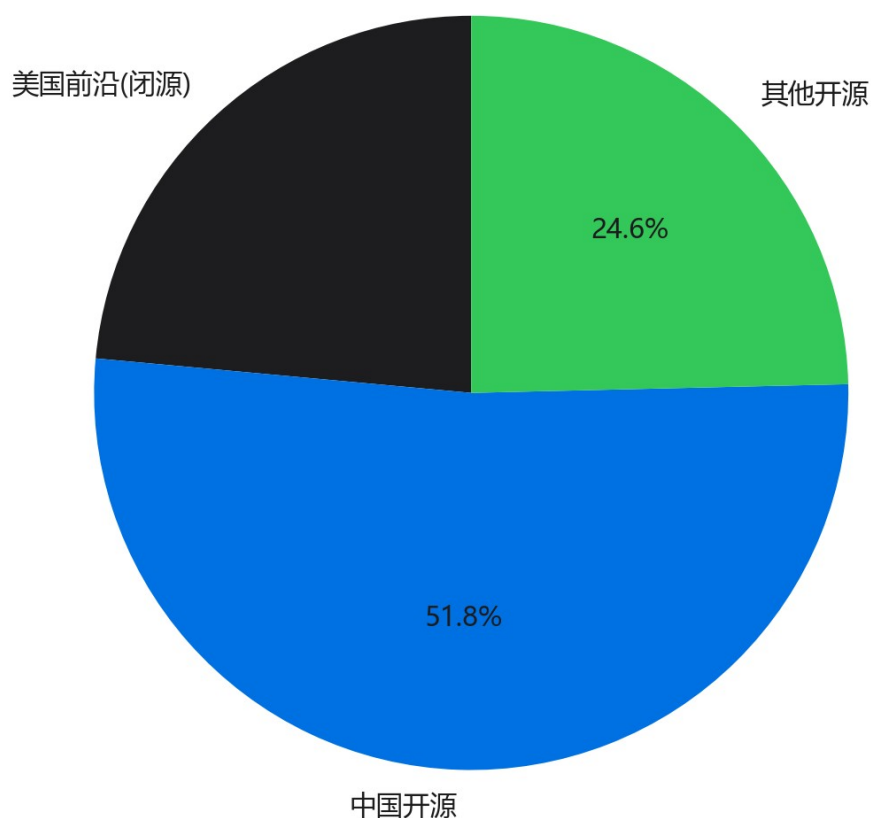


图 7 开放生态内三阵营调用量份额（Logit 模型）

数据来源/说明：多项 Logit；参数校准使基准下 $CN_open \approx 52\%$ （贴合 OpenRouter 实测峰值 50-61%）。 $\beta_{price} > 0$ 体现开发者价格敏感（Agent 任务动辄百万级 Token）。

3.3 价格弹性与价格战韧性

价格-份额扫描显示，中国份额对价格的弧弹性约 -0.25（0.30~0.60 区间），在常态价格区间相对稳健。若中国把质量分从 0.82 提升至 0.95（缩小能力差距），份额可升至约

67.3%；若美国前沿发动价格战、降价至 \$3/百万，对中国份额冲击仅约 6.7 个百分点，显示深度价格优势构筑的份额韧性。

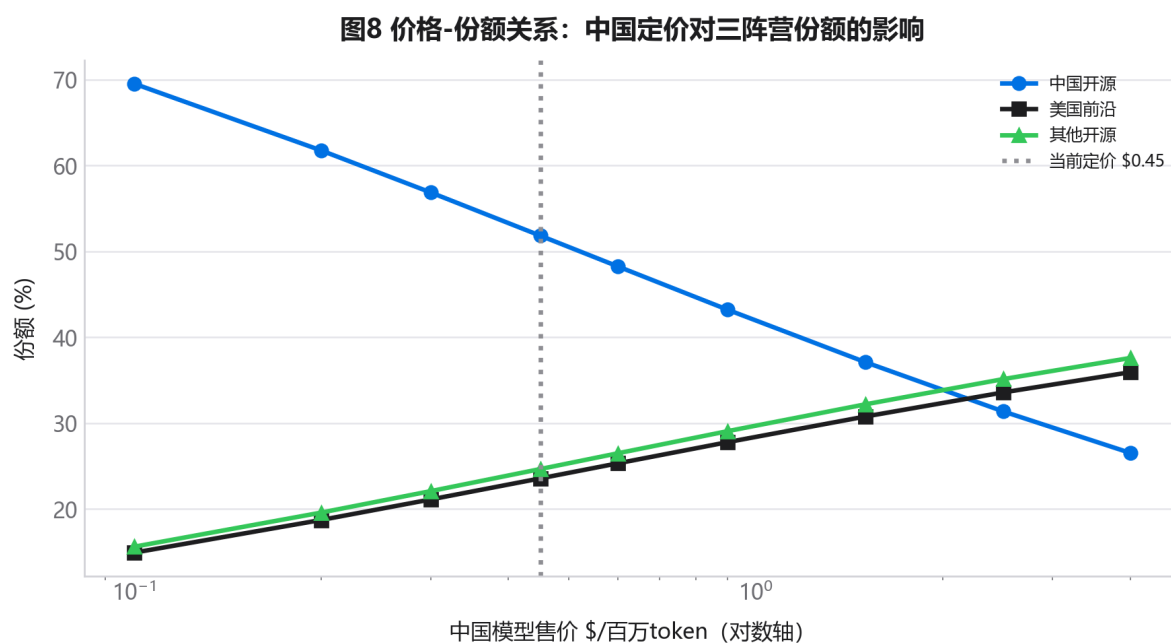


图 8 价格-份额关系

3.4 开源生态护城河

基于阿里通义千问、DeepSeek 等开源模型的全球衍生模型上传量已超过基于美国主流模型者。当东南亚等地的本土语言模型也开始依赖中国模型与工具链时，迁移成本将极高——生态一旦形成，便构成超越单点价格的长期护城河。这是“能力出海”相对“产品出海”的本质跃迁。

数据来源/说明：钛媒体/腾讯新闻：国产开源衍生模型上传量超越美国主流模型；东南亚本土模型依赖中国 Token。

第四章 技术可行性与 Token 经济学

4.1 自下而上的单位成本模型

以“节点=8 卡”为单位自下而上建模：单位成本 = 总成本/小时 ÷ (节点吞吐 × 利用率 × 3600 ÷ 1e6)。总成本含折旧（资本开支按寿命摊销）、电力（功率 × PUE × 电价）与运维加成。模型以 DeepSeek 官方披露的推理经济性校准。

校验项	数值
DeepSeek 反推吞吐 (tok/节点/s)	39,610.0000
DeepSeek 隐含单位成本 (\$/百万 token)	0.1122
模型复算单位成本 (\$/百万 token)	0.1122
日成本 (\$)	87,072.0000
日 token (十亿)	776.0000

校验结果：以 DeepSeek 披露的日成本 \$87,072、日 Token 776 十亿反推，节点吞吐约 39,610 tok/s，隐含单位成本 \$0.1122/百万 token，与模型复算值完全一致，验证方法可靠。

数据来源说明：DeepSeek 开源周 (2025-03-01) 官方披露的推理成本-利润测算 (H800 租金 \$2/h 假设)。

4.2 双技术路线与成本优势

对比进口 H800 路线与国产 NPU（昇腾类）路线。基准情景下，国产 NPU 单位成本显著更低，叠加绿电直供，构成中国 Token 的核心成本护城河。

路线	折旧	电力	运维	单位成本	售价	毛利率%
进口 H800 路线	0.0951	0.0025	0.0244	0.1221	0.45	72.9
国产 NPU 路线	0.0571	0.0022	0.0148	0.0741	0.45	83.5

图4 Token 单位成本构成与毛利空间（基准情景）

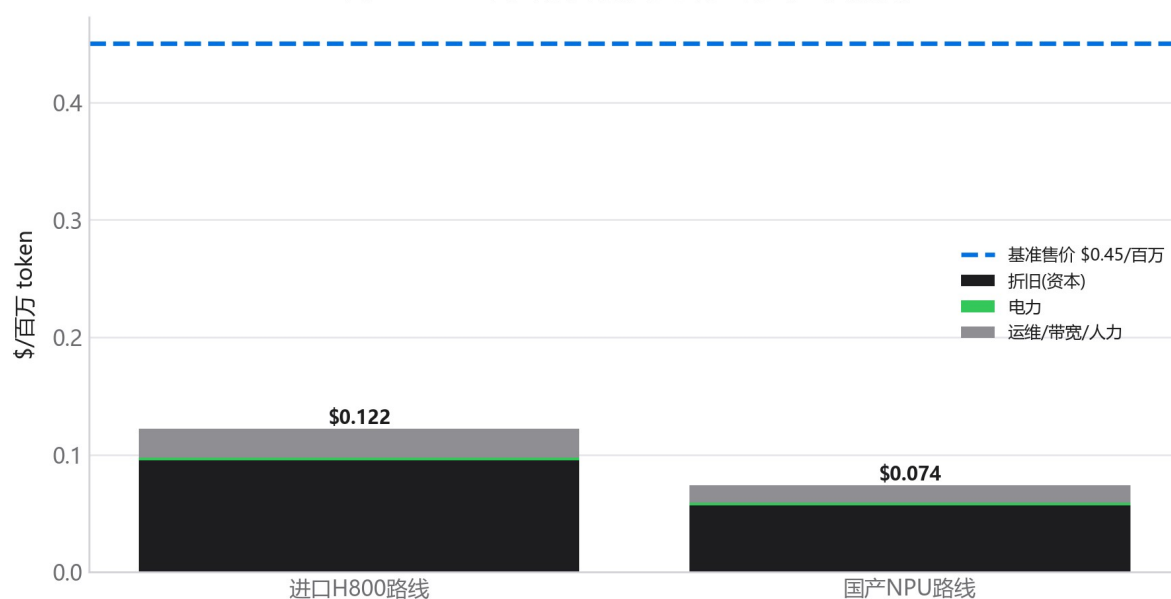


图 4 Token 单位成本构成与毛利空间（基准）

数据来源/说明：国产NPU(昇腾910B类)单卡资本开支与功耗低于进口卡；硅基流动等以国产算力降本30-70%。国内绿电/大工业电价约0.30-0.50元/kWh ($\div 7.1 \approx \$0.042-0.070$)，绿电直供更低。

4.3 成本 vs 主流售价：数量级优势

以国产NPU基准单位成本\$0.0741/百万token为基准，对照主流模型混合售价：前沿闭源模型售价高达国产成本的数十至两百倍，廉价档亦有数倍空间。这一数量级差异，正是中国把电力与算力“炼化”为高毛利数字服务的经济基础。

图5 主流模型售价 vs 国产NPU单位成本（条末标注为成本倍数）

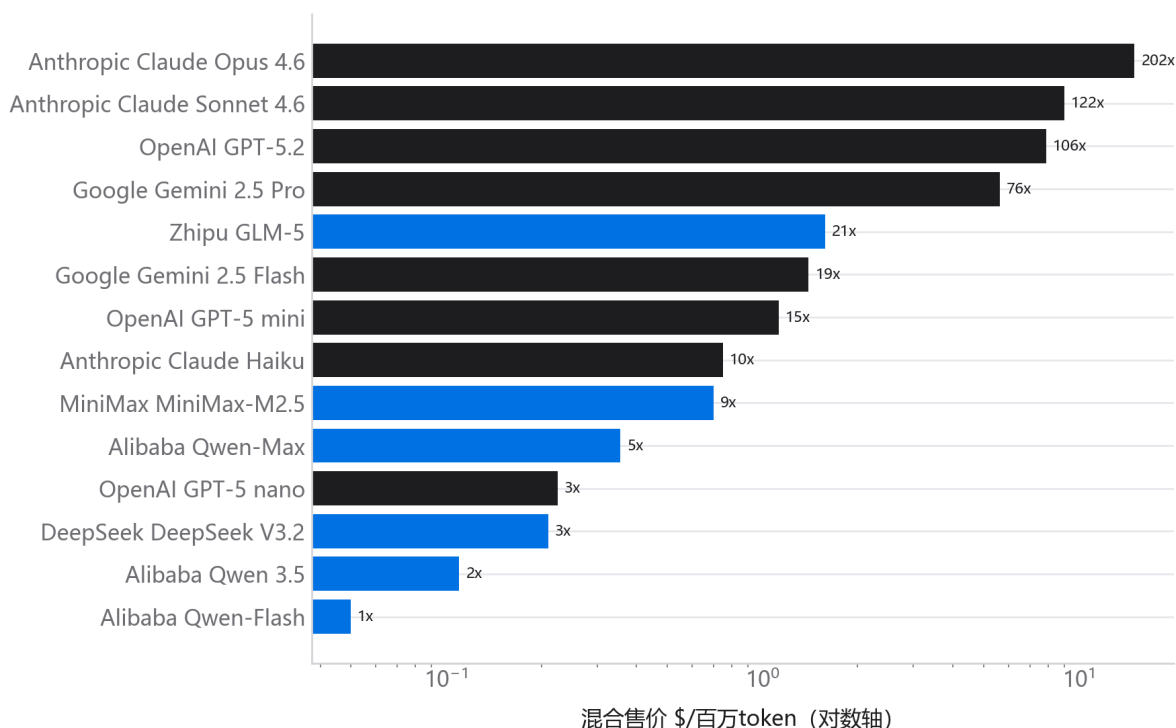


图 5 主流模型售价 vs 国产 NPU 单位成本（倍数）

4.4 成本敏感性：利用率与电价

单位成本对利用率高度敏感：利用率从 30% 提升至 70%，单位成本可下降逾半；电价则因绿电直供具备进一步下探空间。因此“提高利用率（调度/缓存/弹性）+ 绿电直供”是成本端的两大可执行抓手。

图6 国产NPU路线单位成本敏感性

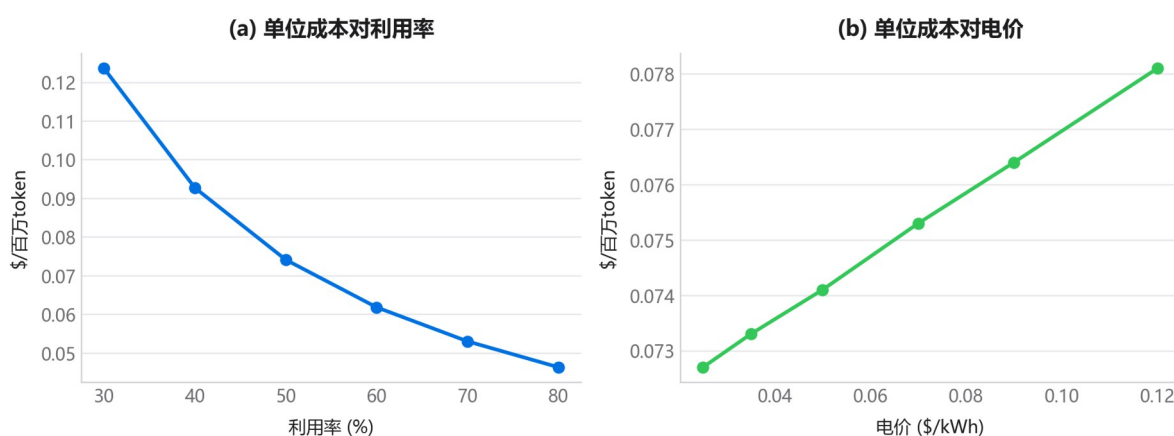


图 6 国产 NPU 路线单位成本敏感性

第五章 商业模式与价值链

5.1 三种交付模式

- **聚合平台分销**：经 OpenRouter 等统一 API 平台触达全球 400+ 模型的开发者，平台抽佣约 3-10%；获客快、起量快，是早期主渠道。
- **企业直采/SDK**：面向中大客户提供直连 API、专属 SDK 与 SLA，毛利更高、粘性更强，是中长期价值留存的关键。
- **来数加工闭环**：海外数据进入境内数字保税区加工，加工后数字服务再出口，全流程物理隔离、监管可控，兼顾合规与价值留境。

数据来源/说明：OpenRouter 统一 API 聚合 400+ 模型，平台抽佣约 3%-10%。汕头华侨试验区 2026-04 完成全国首个 Token 出海全链路闭环（来数加工：海外数据进数字保税区加工后出口，物理隔离）。

5.2 定价与价值链上移

纯裸 token 批发将完全暴露于商品通缩（市场商品价年降 20%-40%）。可持续的定价策略是产品结构上移：向 Agent 编排、企业级合规、长上下文、缓存命中与高价档迁移，使运营主体的“实现 ASP”下降速度显著慢于商品价，从而把价格战转化为结构升级。财务模型即以此为基准前提（实现 ASP 年降约 12%）。

数据来源/说明：运营主体实现 ASP 年降；区别于市场商品价(-20%~-40%)。基准 -12% 反映产品结构持续上移（Agent/企业/长上下文/缓存命中/高价档）对冲商品通缩，为可执行的战略前提。

5.3 商业化实证

商业化已被实证：昆仑万维 2025 年海外收入 5.15 亿元，占总收入 91.0%（Q3 升至 93.3%）；DeepSeek 2026 年 3 月月活达 1.27 亿，广泛分布于发展中国家。这些实证表明海外付费需求真实且具规模。

数据来源/说明：腾讯新闻：昆仑万维 2025 海外收入 51.5/56.6 亿元（91.0%，Q3 升至 93.3%）；QuestMobile：DeepSeek 2026-03 MAU 1.27 亿。

第六章 目标市场与出海战略

6.1 市场分层与进入次序

基于需求强度、价格敏感度、地缘中立性与本地化协同，提出四层市场与进入次序：

市场层级	机会逻辑	可检验 KPI
全球南方核心(东南亚/南亚)	DeepSeek MAU 大量分布于印度/印尼；跨境电商/玩具/纺织数字化升级需求；中文+本地语种。	首年东南亚开发者签约数、月活调用 Token、付费转化率。
中东(海湾)	主权 AI 投入大、对中美中立采购、绿能-算力协同。	主权基金/电信运营商合作 LOI 数、本地节点落地。
拉美/非洲	价格敏感、被前沿厂商覆盖不足、增长快。	拉美开发者占比、本地化语言模型调用占比。
发达市场长尾开发者	硅谷开发者已大量‘投喂’中国模型(角色扮演/编程);经聚合平台触达。	OpenRouter 上国产模型份额、欧美中小团队直采数。

数据来源说明：腾讯/钛媒体/澎湃：DeepSeek 用户分布发展中国家；汕头试点指向东南亚跨境电商等；硅谷开发者大量调用国产模型。

6.2 进入策略

- **先平台、后直采**：以聚合平台快速起量验证 PMF，再沉淀直采客户与自有结算，降低平台依赖；
- **先南方、后长尾**：优先全球南方核心市场，再以聚合平台触达发达市场长尾开发者；
- **本地化协同**：多语种与本地场景微调、就近接入与低时延专线，提升留存；
- **绿电-算力协同出海**：与中东等主权 AI、清洁能源富集地区开展算力-绿能协同，形成中立可控的供给节点。

第七章 合规、安全与地缘风险

7.1 三大强制合规义务

- **数据出境安全评估**：海外用户数据跨境需满足出境安全评估与最小化留存要求；
- **出口管制核验**：技术与服务出口需进行出口管制核验，避免受控用途与受限实体；
- **内容安全过滤**：推理输入输出需符合内容安全与当地法律的过滤与审计要求。

数据来源说明：数据出境安全评估、出口管制核验、内容安全过滤为三大强制合规义务。

7.2 “来数加工” 合规闭环

2026 年 4 月，汕头华侨试验区完成全国首个 Token 出海全链路闭环：海外数据进入数字保税区进行算力加工，加工后数字服务再出口，全流程物理隔离、监管可控，实现“数据从境外来、在专区内处理、结果再回到境外”的相对封闭跨境服务闭环；配套国际数据专线，至新加坡时延低至 32.7ms。该范式为 Token 出海提供了可复制的合规基础设施。

数据来源说明：汕头华侨试验区 2026-04 完成全国首个 Token 出海全链路闭环（来数加工：海外数据进数字保税区加工后出口，物理隔离）。广东移动 4.8T 国际数据专线，至新加坡时延低至 32.7ms。

7.3 地缘对冲

面对出口管制与地缘不确定性，应以“面向中立/全球南方市场 + 国产算力自主 + 合规闭环 + 多平台分发”组合对冲，降低对单一市场、单一平台与单一硬件来源的依赖。

第八章 财务分析与投资评价

8.1 代表性运营主体与建模口径

代表性运营主体：境内绿电直供的 Token 出海推理服务商，以国产 NPU 集群为主、进口卡为辅，经 OpenRouter 类平台 + 直采向全球供给。模型链条：单节点年化可售 Token = 吞吐×利用率×3600×8760；收入 = 年售 Token×实现 ASP（逐年）；净收入 = 收入扣减聚合平台抽佣；COGS = 现金算力成本加折旧；EBIT = 毛利减期间费用；自由现金流 = 净利加折旧减资本开支；投资评价含 NPV、IRR、回收期与退出 EV/EBIT 终值。

8.2 基准情景五年损益与现金流

年份	售出 Token(万 亿)	净收入(\$M)	毛利率%	EBIT(\$M)	资本开支 (\$M)	FCF(\$M)
2026	47.3	20.5	63.7	4.8	25.9	-15.3
2027	141.9	54.0	70.8	16.6	25.9	1.2
2028	258.6	86.7	68.1	24.3	38.0	5.1
2029	416.3	122.7	65.0	30.8	48.4	12.3
2030	615.0	159.6	65.2	40.2	60.5	16.9

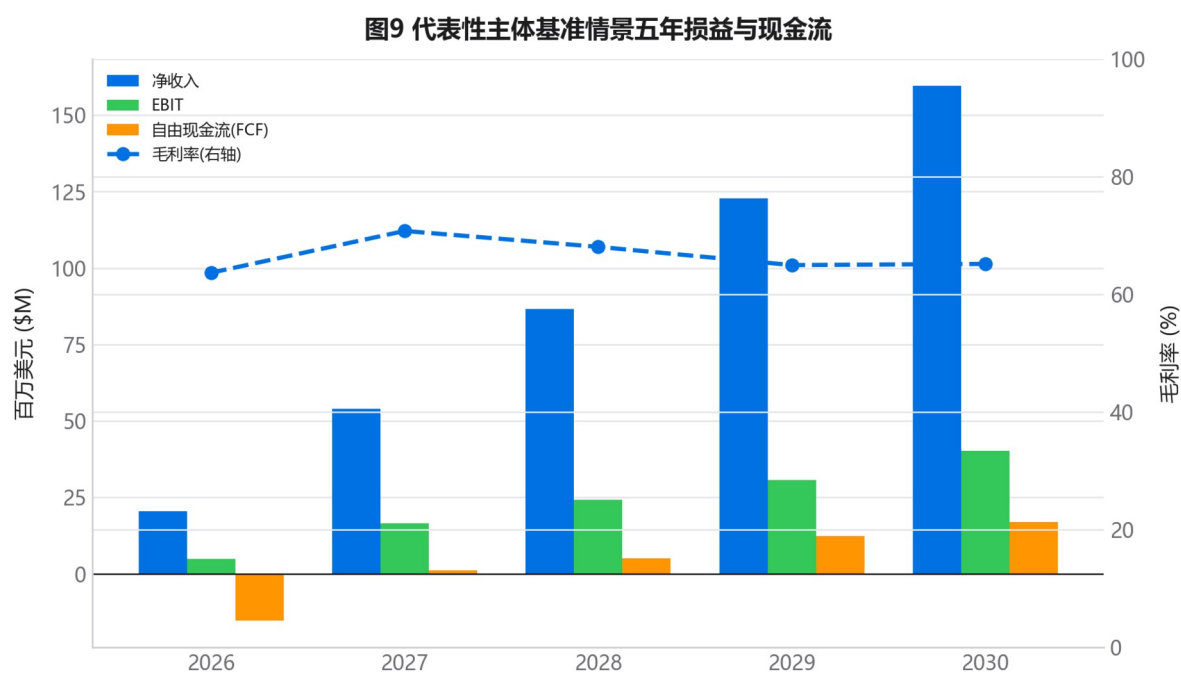


图 9 基准情景五年损益与现金流

8.3 三情景投资评价

情景	NPV(\$M)	NPV 不含 终值(\$M)	IRR%	回收期 (年)	累计 CapEx(\$ M)	2030 净收 入(\$M)	2030 EBIT 率%
保守	-53	-53	n.a.	n.a.	97	14	-143.9
基准	195	7	132.8	3.73	199	160	25.2
乐观	3,383	541	n.a.	0.00	276	843	51.6

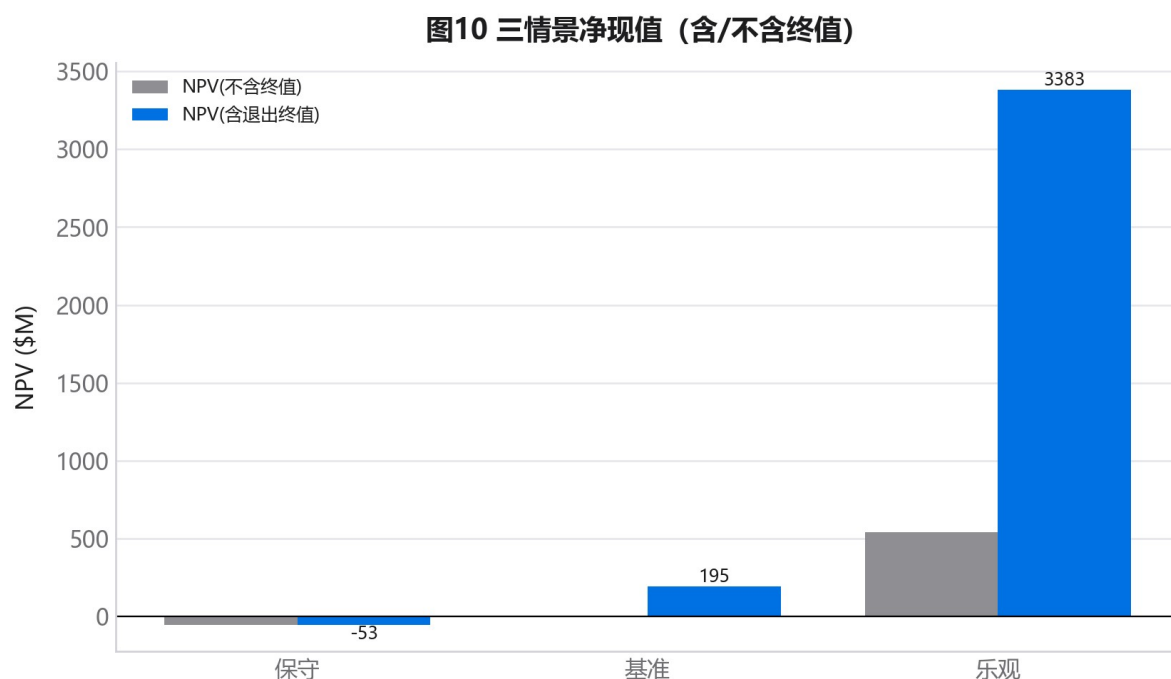


图 10 三情景净现值（含/不含终值）

口径说明（重要）：本节 NPV/IRR/回收期均为“条件于成功执行”口径，即假设主体成功完成扩张并存活；创业本身的失败概率不在此处计提，而在 8.6 节以“存活层”对种子轮投资人回报单独折算。基准情景含终值 NPV 约 \$195.0M、IRR 约 132.8%（条件于成功），但不含终值的 5 年 DCF 仅约 \$7.0M——即 5 年窗口内项目近似盈亏平衡，绝大部分价值来自持续经营特许权（终值）。这要求经营者必须在窗口期内建立可持续的份额与价值链地位，方能兑现终值。

数据来源说明：盈利型 AI 基础设施退出 EV/EBIT 约 6-12x（高增长高确定性取高值），分情景取值。折现率见附录 B。

8.4 敏感性分析（龙卷风）

单变量±扰动显示，NPV 对“实现 ASP 年降速度”最敏感，其次为售价水平与单节点产能（吞吐×利用率），再次为期间费用率与节点资本开支。这与战略判断一致：稳住价值变现（ASP）是第一要务。

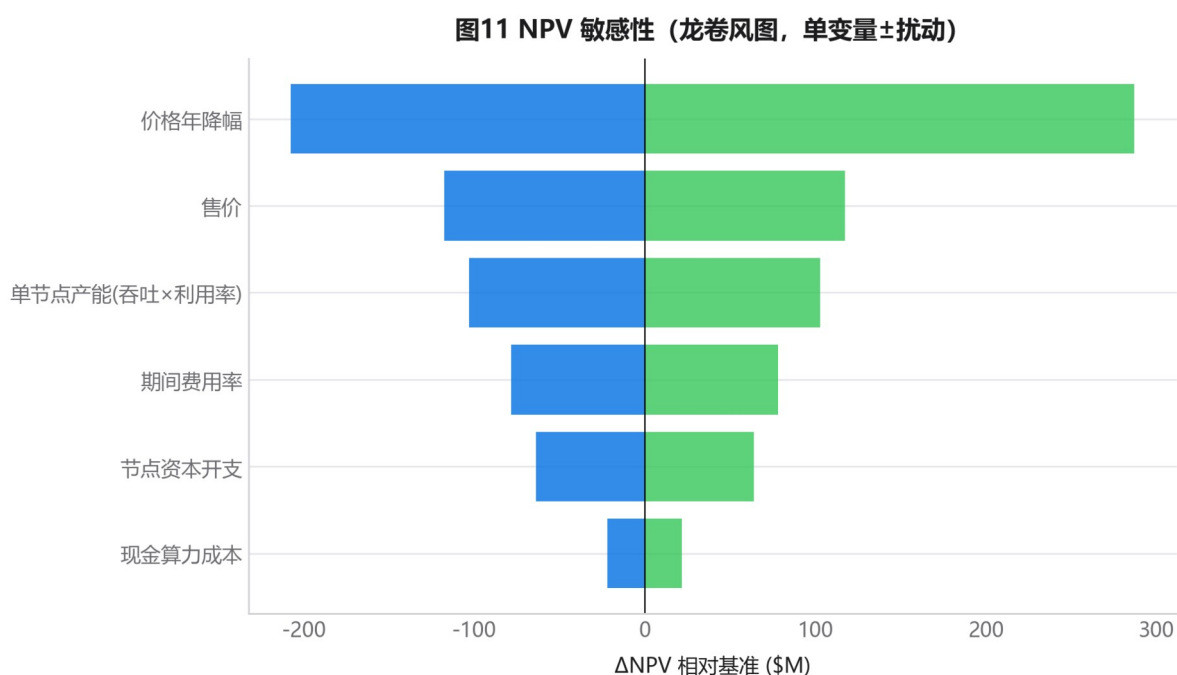


图 11 NPV 敏感性 (龙卷风图)

8.5 蒙特卡洛压力测试

对 6 个关键变量（售价、利用率、吞吐、电价、份额达成、ASP 年降，后者覆盖价格战 regime）做三角分布抽样，共 20,000 次模拟：NPV 中位数约 \$187.6M，P05 约 \$-43.1M，P95 约 \$557.3M，NPV>0 概率约 89.0%。结论：在合理假设区间内项目大概率创造正价值，但价格战下行情景下存在亏损可能，须以战略主线主动规避。

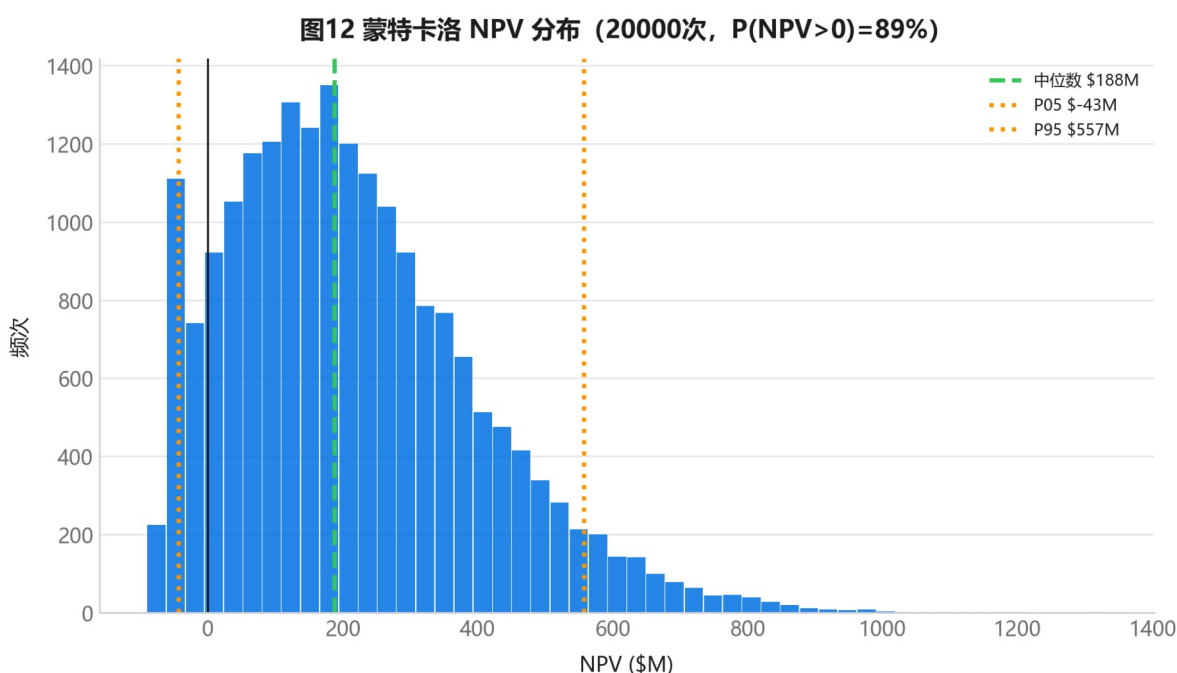


图 12 蒙特卡洛 NPV 分布

数据来源说明：蒙特卡洛对 6 大关键变量做三角分布抽样（端点=保守/乐观，众数=基准），含 ASP 年降以覆盖价格战下行 regime，评估 NPV 分布。

8.6 种子轮投资回报与风险（“拿到钱”已实现口径）

在上述代表性主体之上，从风险投资（VC）视角评价【种子轮】投资人的回报。已确认条款：种子轮投资 \$10M、投前估值 \$40M（投后 \$50M、初始持股 20%），其后经 Series A、Series B 两轮融资稀释；全部优先股按 1x 非参与、pari-passu 清算优先权设定。企业股权价值采用与第八章退出口径对齐的隐含 EV/Revenue 倍数法逐年市值标记（入场年市值 = 投后估值，自洽不跳涨；末年与退出 EV/EBIT 对齐）。

核心方法（本次重算的关键）：账面市值并不等于投资人真正能拿回的现金。本节以“资金成功退出（拿到钱）”为头条口径，对账面回报按【创业死亡率 + 退出流动性（IPO/并购/老股转让的可得概率与折扣）】折算，给出真正可兑现的现金回报；账面纸面 TVPI 仅作透明对照。全部存活与流动性参数锚定中国本土退出现实。

(1) Cap table 稀释路径（基准情景，as-converted 持股 %）

年份	企业股权价值	种子%	A 轮%	B 轮%	普通股%
----	--------	-----	------	------	------

	(\$M)				
2026	50	20.00	0.00	0.00	80.00
2027	130	16.24	18.80	0.00	64.96
2028	204	12.55	14.53	22.73	50.19
2029	284	12.55	14.53	22.73	50.19
2030	362	12.55	14.53	22.73	50.19

基准情景下，企业股权价值由投后 \$50M 起步（入场自洽、无瞬时跳涨），种子持股由初始 20% 经两轮稀释至退出时约 12.6%（据实计算）。保守情景因现金缺口大、融资额高，稀释显著更重；乐观情景因募资需求小，稀释温和。

(2) 三情景 年 1-5 种子轮收益：账面纸面 TVPI vs 已实现现金（拿到钱）

情景	口径	第 1 年	第 2 年	第 3 年	第 4 年	第 5 年
保守	账面 TVPI(倍)	1.00	1.00	0.26	0.16	0.09
保守	已实现现金 (倍)	0.00	0.01	0.02	0.03	0.04
保守	已实现年化 (%)	-99.60	-88.40	-71.50	-57.30	-47.30
基准	账面 TVPI(倍)	1.00	2.10	2.56	3.56	4.54
基准	已实现现金 (倍)	0.01	0.03	0.07	0.17	0.33
基准	已实现年化 (%)	-99.50	-83.80	-58.00	-35.60	-19.70
乐观	账面 TVPI(倍)	1.00	4.69	13.45	35.05	84.86
乐观	已实现现金 (倍)	0.01	0.09	0.67	3.41	13.96
乐观	已实现年化 (%)	-99.30	-70.10	-12.40	35.90	69.40

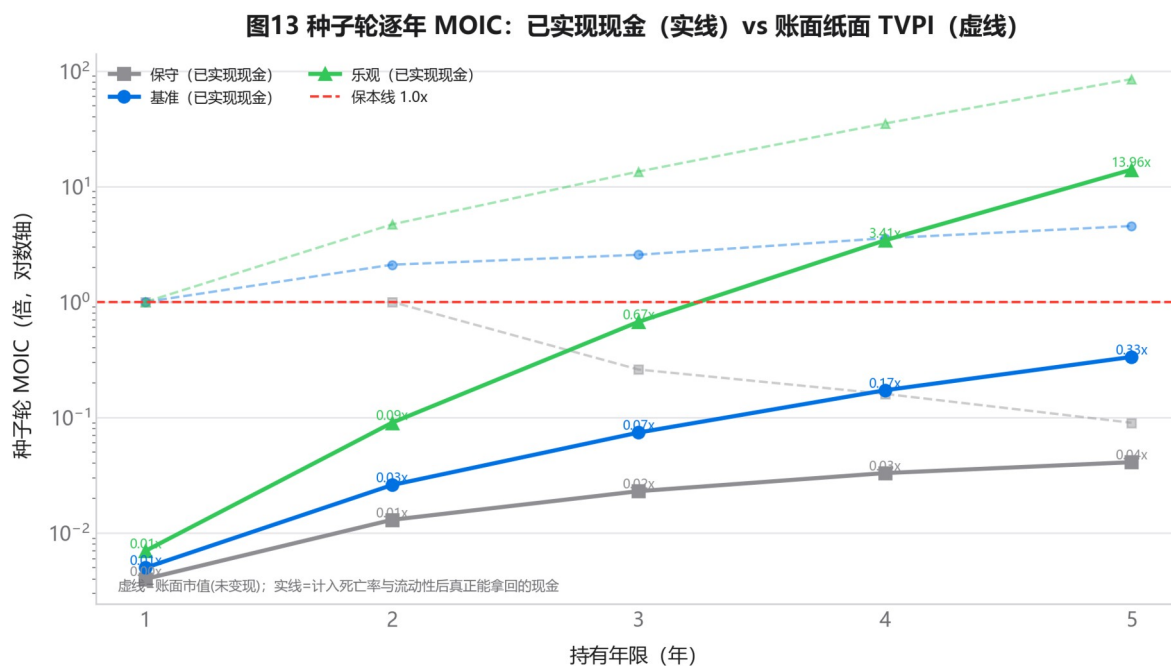


图 13 种子轮逐年 MOIC: 已实现现金 (实线) vs 账面纸面 TVPI (虚线)

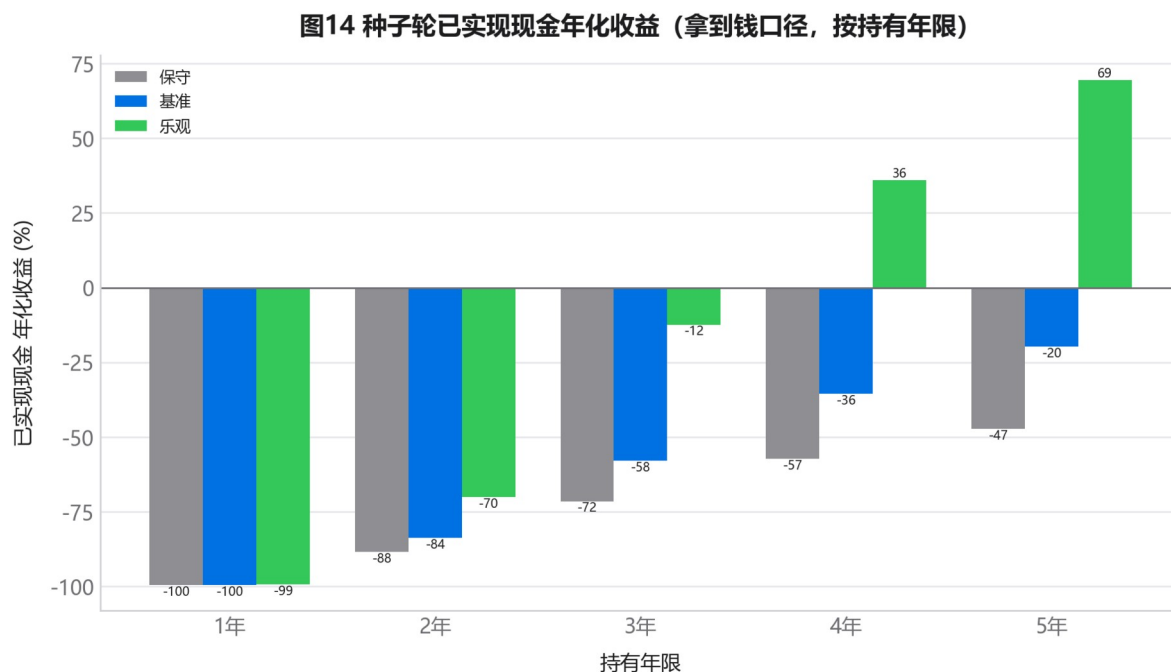


图 14 种子轮已实现现金年化收益 (拿到钱口径, 按持有年限)

数据来源/说明: 已确认: 种子轮 \$10M @ 投前 \$40M (投后 \$50M, 初始持股 20%)。口径A (用户已选): 逐年给出若该年主动退出可变现现金。胜率= $P(\text{已实现现金 MOIC} > 1)$, 即真金白银>投入方算赢, 账面存活不算赢; 年化= $\text{已实现 MOIC}^{1/(n-1)} - 1$; 盈亏比= $E[\text{MOIC}-1 | \text{盈利退出}] / E[1-\text{MOIC} | \text{亏损}]$ 。同时保留纸面市值 TVPI 作透明对照 (非头条)。

解读：账面 TVPI 在成长情景下逐年攀升（基准第 5 年约 4.5x），但这是不可直接变现的私募股权账面值；一旦计入“能否、何时、以何折扣拿到现金”，已实现回报大幅回落（基准第 5 年约 0.33x）。这与全球 VC 第 5 年 DPI 普遍仅 0.27-0.39x、中国本土基金第 5 年普遍未回本的现实一致。

(3) 已实现退出层：本土锚点与第 5 年现金分解

折算依据（全部公开数据）：IT 桔子《2022 中国天使投资行业报告》——获天使投资企业死亡率约 31%、进入下一轮约 45%、最终 IPO 仅约 2%；上市周期——成立 5 年内上市仅约 4%、最集中于 10-15 年，天使→独角兽平均 51 个月、独角兽→上市再约 2.75 年（即种子后 5 年内 IPO 现金退出≈0，须靠并购/老股转让兑现）；本土基金 DPI——政府引导基金均值约 0.7（未回本）、7 年 DPI=1 为优秀、9 年为常态；老股转让市场不成熟、折扣较深。据此设三结局（成长/僵尸/死亡）与逐年流动性可得概率 $L(t)$ 、折扣：

情景	成长%	僵尸%	死亡%	第 5 年可 变现%	流动性折 扣%	账面 TVPI(倍)	已实现现 金(倍)	已实现年 化%
保守	18	25	57	20	40	0.1	0.04	-47.3
基准	28	27	45	33	30	4.5	0.33	-19.7
乐观	40	28	32	50	18	84.9	13.96	69.4

图16 第5年退出：账面 TVPI vs 已实现现金 MOIC（计入死亡率/流动性/折扣的拿到钱折损）

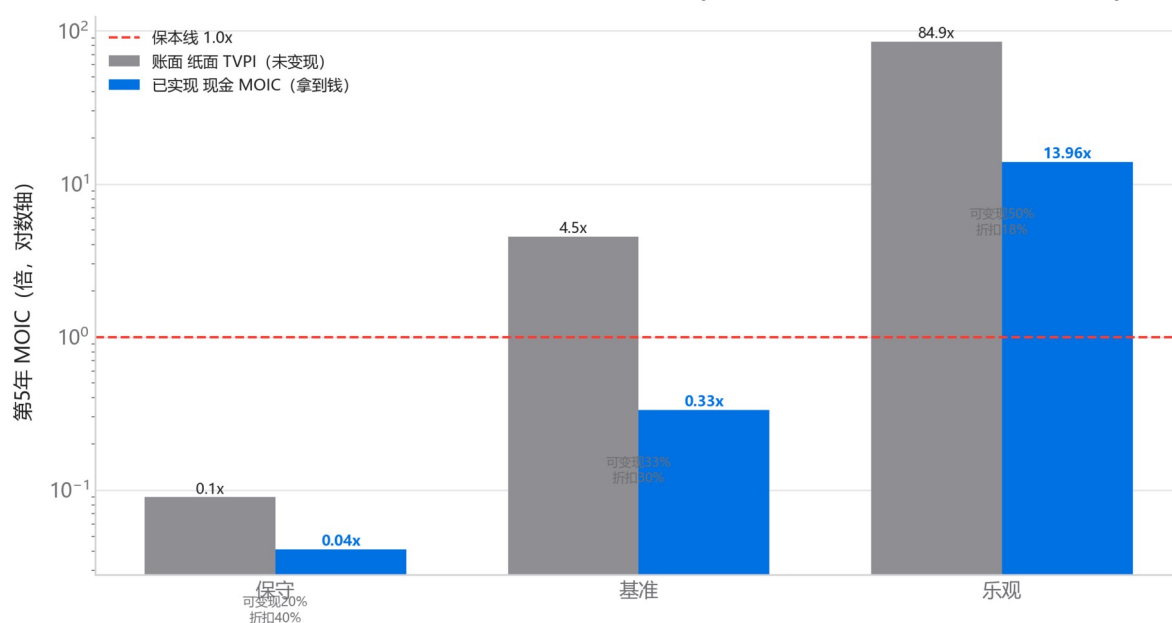


图 16 第 5 年退出：账面 TVPI vs 已实现现金 MOIC（拿到钱折损，按情景）

数据来源说明：36 氪援引上市公司数据：成立→上市最集中 10-15 年(占 28.9%)、成立 10 年以上上市占 75%、成立 5 年内上市仅约 3-4%(近年趋零，2021 约 3%)；36 氪/虎嗅：天使→独角兽平均 51 个月、独角兽→上市再平均 2.75 年→种子投资后 5 年内 IPO 现金退出≈0，须靠并购/老股转让兑现。界面/清科：政府引导基金均值 DPI≈0.7（未回本），7 年 DPI=1 为优秀标杆、9 年为常态，2013 年后成立基金多数 DPI<1、中位约十年回本；AVP/Carta/thevcfactory：全球 VC 平均第 8 年才到 1x DPI，第 5 年 DPI 常仅 0.27-0.39x（即便最终 8-11x 的优秀基金，year-5 DPI 亦低）。锚定本土现实：种子后 5 年内 IPO 现金退出≈0（成立 5 年内上市仅约 3-4%且种子在成立约 2 年后入场）；现金主要来自并购(占退出≈30%)与老股转让(市场不成熟，2024 同比-15.5%)；对照全球 VC 第 5 年 DPI 仅 0.27-0.39x、第 8 年才到 1x。故第 5 年成长企业可变现概率保守 0.20/基准 0.33/乐观 0.50，且早年极低 (liquidity_shape)。

(4) 已实现现金蒙特卡洛：胜率、盈亏比、年化与逐年退出概率

在第 8.5 节同一套 6 变量三角抽样与固定随机种子之上叠加已实现退出层，共 20,000 次模拟，得到第 5 年退出的【已实现现金】MOIC 分布。下表并列“账面纸面”（不计能否拿到钱）与“已实现现金”（拿到钱）两口径，差距即流动性与死亡率的代价：

指标	账面纸面(对照)	已实现现金(拿到钱·头条)
胜率 $P(\text{MOIC}>1)=\text{盈利退出概率}$	84%	7%
未回本 $P(\text{MOIC}<1x)$	15%	93%
盈亏比	13.9	3.5
MOIC 中位 / 均值	4.28x / 5.11x	0.01x / 0.36x
年化 中位 / 均值	34% / 31%	-57% / -62%
$P(>3x)$ / $P(>5x)$ / $P(>10x)$	64% / 43% / 12%	4.4% / 2.4% / 0.3%

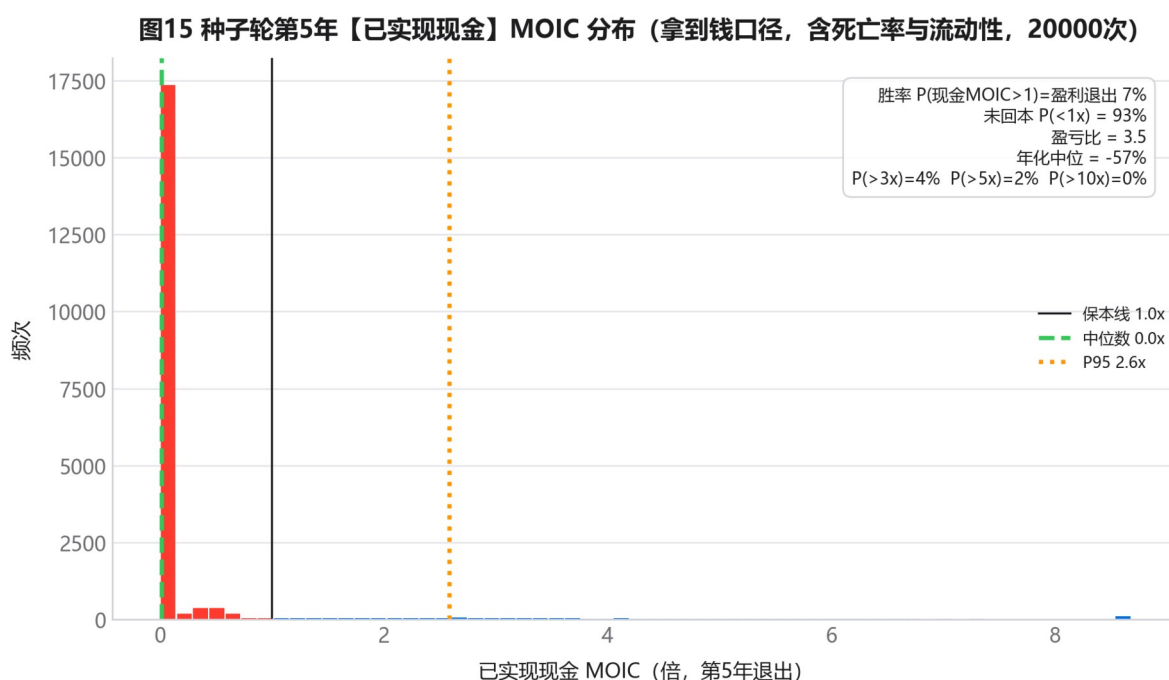


图 15 种子轮第 5 年【已实现现金】MOIC 分布（拿到钱口径）

逐年退出胜率（已实现现金 MOIC>1 的概率）随持有年限上升：第 1 年 0%、第 2 年 1%、第 3 年 2%、第 4 年 4%、第 5 年 8%——这正是私募股权 DPI 的 J 曲线，早年几乎无法变现，价值需更长持有期或后续并购/转让兑现。

交叉校验：将本模型已实现现金分布与中国本土基金 DPI 及 VC 经验分布对照，验证标定的公允性——

现金回报区间	本模型	经验区间	来源
P(已实现<1x) 未回本	93%	约 0.80-0.93	本土：政府引导基金均值 DPI≈0.7、2013 年后多数基金 DPI<1（本土处区间高位）；全球 VC 第 5 年 DPI 仅 0.27-0.39x → 多数未回本。
P(已实现≥1x) 现金回本	7%	约 0.08-0.20	5 年窗口内现金回本属少数（IPO≈0、靠并购/老股转让）；对照 IT 桔子天使 → IPO 仅 2%(终生)。
P(已实现≥3x) 重仓赢家	4%	约 0.04-0.08	Correlation：约 5%笔数

≥10x、20-30%回 1-5x；5
年窗口现金兑现进一步压
低。

图17 种子轮已实现现金分布 交叉校验：本模型 vs 本土DPI/VC经验（清科/Correlation/IT桔子）

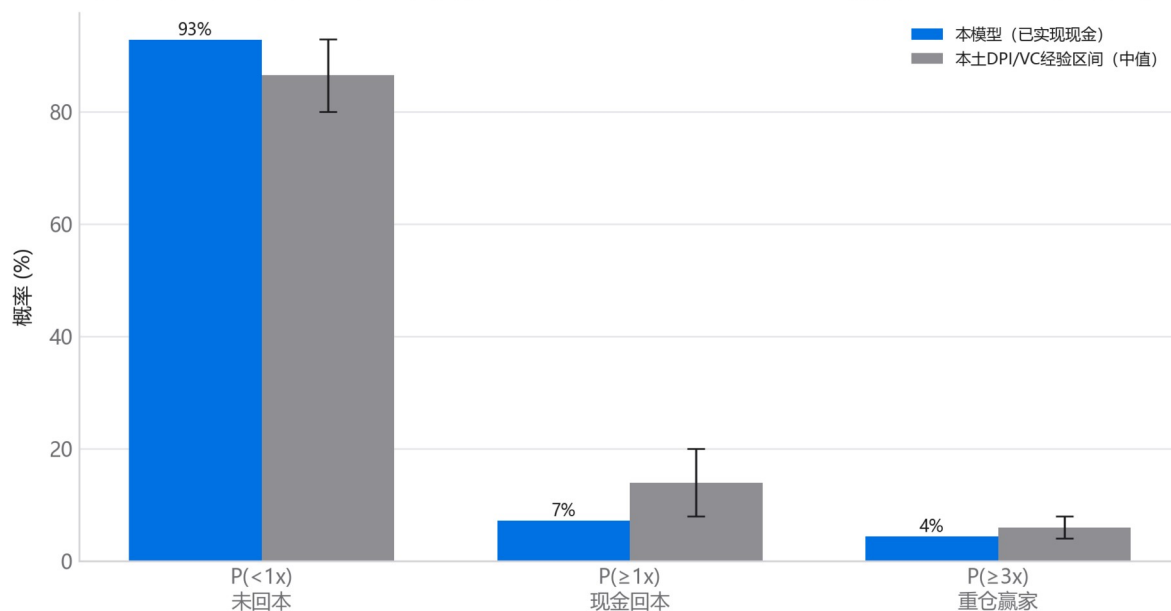


图 17 已实现现金分布交叉校验：本模型 vs 本土 DPI/VC 经验

数据来源/说明：界面/清科：政府引导基金均值 DPI≈0.7（未回本），7 年 DPI=1 为优秀标杆、9 年为常态，2013 年后成立基金多数 DPI<1、中位约十年回本；AVP/Carta/thevcfactory：全球 VC 平均第 8 年才到 1x DPI，第 5 年 DPI 常仅 0.27-0.39x（即便最终 8-11x 的优秀基金，year-5 DPI 亦低）。Correlation Ventures：约半数资本、约 2/3 笔交易回报<1x；<4% 资本达≥10x。典型组合：60-70% 回报≤1x、20-30% 回 1-5x、约 5% 回≥10x（幂律右偏）。

结论与诚实提示：(a) 以“拿到钱”口径计，种子轮第 5 年退出胜率（顺利盈利退出概率）约 7%、未回本概率约 93%、已实现 MOIC 中位约 0.01x、均值约 0.36x、年化中位约 -57%、盈亏比约 3.5；(b) 账面 TVPI 与已实现现金的巨大落差，根因是中国种子在 5 年窗口内现金退出极少（IPO 周期长、并购薄、老股转让不成熟）；(c) 模型分布落入本土 DPI 与 Correlation 经验区间，标定可信；(d) 1x 清算优先权仅对下行提供有限保护，且需有买方/退出窗口方能兑现；(e) 单笔种子大概率未回本，正回报高度依赖右尾赢家与更长持有期，应以组合化、并预留更长退出周期的方式参与早期创投。以上均为如实呈现。

第九章 风险评估与应对矩阵

风险	概率	影响	应对措施
地缘政治与出口管制	高	高	来数加工合规闭环、数据物理隔离、面向中立/全球南方市场对冲、合规出口核验。
数据跨境合规	高	高	数据出境安全评估、数字保税区、最小化数据留存、内容安全过滤。
价格战与毛利压缩	高	中	国产 NPU + 绿电降本、缓存命中优化、向高价值 Agent/企业场景升级。
前沿能力差距	中	中	开源生态绑定迁移成本、垂直/多语种微调、快迭代。
客户集中与平台依赖	中	中	提升直采占比、自建结算与 SDK、多平台分发。
汇率与回款	中	低	美元计价、跨境结算合规通道、汇率对冲。

总体判断：核心风险集中于地缘政治、数据合规与价格战。前两者以“来数加工合规闭环 + 面向中立市场”对冲，后者以“成本领先 + 价值链上移”化解。风险可控、可治本。

第十章 实施路线图与阶段目标

遵循“先定可分解的阶段目标，再用客观可检验标准衡量结果”的原则，制定三阶段路线图：

阶段	阶段目标	客观可检验标准
阶段一 验证（0-12 月）	跑通合规闭环+聚合平台起量	完成数据出境评估；接入≥2 聚合平台；月调用量达标；首批海外付费客户与美元回款。
阶段二 规模（12-30 月）	国产 NPU 集群扩张+直采占比提升	集群达基准节点计划；直采收入占比≥40%；毛利率≥60%；现金流转正。
阶段三 升级（30-60 月）	价值链上移+多市场布局	Agent/企业级收入占比显著提升；实现 ASP 年降≤12%；进入≥3 区域市场；达成基准 NPV/IRR 目标。

贯穿三阶段，坚持主动复盘纠错：建立月度指标看板（利用率、实现 ASP、毛利率、合规事件、客户集中度），小问题早处理，既纠错又治本。

第十一章 结论与建议

综合市场、技术、竞争、合规与财务多维论证，本研究结论如下：

- **机会真实且具规模：**中国 Token 出海收入 2030 年基准约 \$2.78B，复合增速约 42.2%。
- **成本优势确凿：**国产 NPU+绿电单位成本约 \$0.0741/百万 token，毛利率约 83.5%，经 DeepSeek 公开数据校准可靠。
- **商业可行但有前提：**基准 NPV 约 \$195.0M、IRR 约 132.8%（条件于成功执行），但价值高度依赖持续经营，须在窗口期建立份额与价值链地位。
- **种子轮回报以“拿到钱”口径如实呈现：**计入创业死亡率与退出流动性后，种子轮第 5 年【已实现现金】退出胜率约 7%、已实现 MOIC 中位约 0.01x、均值约 0.36x、年化中位约 -57%（账面纸面 TVPI 基准第 5 年约 4.5x）——中国种子 5 年窗口内现金退出极少，正回报依赖右尾赢家与更长持有期，宜组合化、留足退出周期参与。
- **首要矛盾明确：**赢“调用量”、难变现“收入”。稳住实现 ASP（价值链上移）是决定成败的第一杠杆。

总建议：以“成本领先 + 价值链上移 + 合规闭环 + 市场分层”四线并举，先在全球南方跑通闭环与正现金流，再以开源生态与直采沉淀护城河，把“流量领先”稳步转化为“价值领先”。该路径实事求是、可执行、经得起检验。

附录 A 数据来源清单

[A1] The Business Research Company / giiresearch 《AI API Global Market Report 2026》：2025 \$62.26B → 2026 \$84.15B, CAGR 35.2%。

[A2] giiresearch 《LLM Cloud Service Global Market Report 2026》：2025 \$12.28B → 2026 \$15.21B, CAGR 23.9%。

[A3] Technavio 《AI API Market 2026-2030》：2026-2030 增量 \$121.73B, CAGR 26.3%, 25-26 同比 20.9%。

[A4] 36 氪/钛媒体援引 OpenRouter：中国模型份额由 2024 末 1.2% → 2025H2 约 30% → 2026-02 个别周次 50%-61%（单月首次过半，达 14.69/28.7 万亿 Token）。

[A5] 腾讯新闻/OpenRouter：2026-03-30~04-05 周，中国 12.96 万亿词元 vs 美国 3.03 万亿，4.27 倍，连续第 5 周领先。

[A6] 钛媒体援引 OpenRouter：截至 2026-02-28 前十模型月 Token 28.7 万亿，国产 14.69 万亿（首次过半）。

[A7] Google I/O 2025：跨产品月处理 token 由 9.7 万亿(2024)升至约 480 万亿(2025 年中)，用于全球 token 量数量级校验。

[A8] analysis-atlas / 钛媒体 / 各厂商定价页（2026 年中）：2025 → 2026 token 价格约降 80%；MiniMax-M2.5 输出≈\$1.1/百万、GLM-5≈\$2.55/百万、Claude Opus4.6=\$25/百万；Qwen3.5≈0.8 元/百万。

[A9] DeepSeek 开源周（2025-03-01）官方披露的推理成本-利润测算（H800 租金 \$2/h 假设）。

[A10] 腾讯新闻：昆仑万维 2025 海外收入 51.5/56.6 亿元（91.0%，Q3 升至 93.3%）；QuestMobile：DeepSeek 2026-03 MAU 1.27 亿。

[A11] 汕头华侨试验区 2026-04 完成全国首个 Token 出海全链路闭环（来数加工：海外数据进数字保税区加工后出口，物理隔离）。

[A12] 广东移动 4.8T 国际数据专线，至新加坡时延低至 32.7ms。

[A13] IT 桔子《2022 中国天使投资行业报告》：获天使投资企业死亡率≈31%（破产/注销/宣布倒闭/产品官网下线，或 5 年未获新融资经人工核实）。

[A14] 财新/吴晓波援引国家企业数据：新设企业 1 年存活 92%、3 年 76%、10 年 50.3%。

- [A15] Carta：种子→A 轮 2 年晋级率 30.6%(2018Q1)→15.4%(2022Q1)；Crunchbase：近期队列(2023/2024) 16%-24%，历史最终晋级 45%-55%；AI 赛道估值与晋级高于非 AI。
- [A16] Correlation Ventures：约半数资本、约 2/3 笔交易回报<1x；<4%资本达≥10x。典型组合：60-70%回报 ≤1x、20-30%回 1-5x、约 5%回≥10x（幂律右偏）。
- [A17] 36 氪援引上市公司数据：成立→上市最集中 10-15 年(占 28.9%)、成立 10 年以上上市占 75%、成立 5 年内上市仅约 3-4%(近年趋零，2021 约 3%)；36 氪/虎嗅：天使→独角兽平均 51 个月、独角兽→上市再平均 2.75 年 → 种子投资后 5 年内 IPO 现金退出≈0，须靠并购/老股转让兑现。
- [A18] 界面/清科：政府引导基金均值 DPI≈0.7（未回本），7 年 DPI=1 为优秀标杆、9 年为常态，2013 年后 成立基金多数 DPI<1、中位约十年回本；AVP/Carta/thevcfactory：全球 VC 平均第 8 年才到 1x DPI，第 5 年 DPI 常仅 0.27-0.39x（即便最终 8-11x 的优秀基金，year-5 DPI 亦低）。
- [A19] 清科/投中/融中：2024 退出中 IPO≈36-48%、并购≈29-30%、老股转让市场不成熟且 2024 转让事件同 比降 15.5%；独角兽累计退出中仅约 52%体面上市、近半并购/关停/估值下调。
- [A20] Zanbato：直接 VC 老股二级折扣由 2023-12 的约 46%压缩至 2024 年末约 3%（优质资产）；中国老股 转让市场不成熟、缺体系化平台，折扣显著更深，建模取较深折扣并情景化。

附录 B 核心建模假设

假设项	取值	依据/来源
TAM CAGR(保/基/乐)	22% / 30% / 38%	介于 LLM 云服务 CAGR 23.9% 与 AI API 35.2% / 36.7% 之间，分情景取值。
推理占 API 比	45% / 55% / 65%	推理（含 Chat/Agent/RAG 在线调用）为 API 市场主体，建模假设区间。
SAM 量占比	30% / 42% / 55%	开放生态(OpenRouter 类) + 全球南方 + 中小开发者；参考亚洲付费占比升至 31%、开放模型占全球流量近 30% 等结构。
全市场实现价 2026	\$4.5/百万	全市场实现混合价(input/output、前沿+廉价加权)2026≈\$3-6/百万，取 \$4.5；由 \$84B AI API 市场与公开

		token 处理量数量级反推校准。
中国混合价 2026	\$0.45/百万	国产 token 混合售价 2026≈\$0.45/百万 (DeepSeek V3.2 \$0.14/\$0.28、 MiniMax/GLM 略高加权)； 2025 → 2026 价格约降 80%，预测期 年降 20%-40%。
国产 NPU 单卡资本/功耗	\$18k / 310W	国产 NPU(昇腾 910B 类)单卡资本开 支与功耗低于进口卡；硅基流动等以 国产算力降本 30-70%。
利用率(保/基/乐)	35% / 50% / 65%	推理集群常态利用率 35%-65% (峰谷 不均、弹性冗余)。
电价(保/基/乐)	\$0.070/0.050/0.035	国内绿电/大工业电价约 0.30-0.50 元/kWh ($\div 7.1 \approx \$0.042$ -0.070)，绿 电直供更低。
实现 ASP 年降(保/基/乐)	-25% / -12% / -5%	运营主体实现 ASP 年降；区别于市 场商品价(-20%~-40%)。基准 -12% 反 映产品结构持续上移 (Agent/企业/长 上下文/缓存命中/高价档) 对冲商品 通缩，为可执行的战略前提。
折现率(保/基/乐)	18% / 14% / 11%	早期算力/AI 基础设施 WACC 11%- 18% (高风险高资本)。
退出 EV/EBIT(保/基/乐)	6x / 9x / 11x	盈利型 AI 基础设施退出 EV/EBIT 约 6-12x (高增长高确定性取高值)， 分情景取值。
种子轮条款	投\$10M @ 投前\$40M(投后 \$50M,20%)	已确认：种子轮 \$10M @ 投前 \$40M (投后 \$50M，初始持股 20%)。
Series A 融资额(保/基/乐)	\$45M / \$30M / \$20M	A/B 轮金额按情景资助累计资本开支 (基准累计 CapEx≈\$199M，经营现 金流覆盖其余)；保守情景烧钱更 多、募资更大、稀释更重。投前估值 按当年市值法 V[t]。
Series B 融资额(保/基/乐)	\$90M / \$60M / \$35M	A/B 轮金额按情景资助累计资本开支 (基准累计 CapEx≈\$199M，经营现

		金流覆盖其余)；保守情景烧钱更多、募资更大、稀释更重。投前估值按当年市值法 $V[t]$ 。
种子估值法	隐含 EV/Revenue(入场自洽+退出对齐)	隐含 EV/Revenue 倍数法 (审计修正版)：年 0 市值=投后估值 (入场自洽、不凭空跳涨)，末年=退出 EV/EBIT×末年 EBIT (与 FINANCE 退出口径对齐)；倍数在两端之间对数线性过渡。修正了旧版恒定倍数下乐观情景'年 1 市值数倍于投后估值'的不合理跳涨。EBIT≤0 时用困境 EV/Revenue≈1x 兜底，反映仍具收入的资产残值。
清算优先权	1x 非参与, pari-passu	行业标准：1x 非参与、pari-passu 清算优先权；下行时优先股可取 1x 回收，退出价值低于优先额总和时按投入额 pro-rata、普通股归零。
存活层 成长/僵尸/死亡(保)	18% / 25% / 57%	三状态概率按公允锚点标定：成长概率锚定种子→A 轮 15-50% (AI 溢价取偏上) × A → B≈60%'链路，基准 0.28、保守 0.18、乐观 0.40；死亡概率参考 IT 桔子天使死亡率 31% (全样本均值，种子档更高) 与 Correlation'近半亏损'，基准 0.45、保守 0.57、乐观 0.32；其余为僵尸/停滞。死亡回收≈0 (1x 优先权但残值近零)、僵尸回收 0.15-0.70x (下行轮/并购，部分优先权保护)。目标：混合分布的 $P(<1x)/P(>10x)$ 落入 Correlation/IT 桔子经验区间 (见交叉校验)。
存活层 成长/僵尸/死亡(基)	28% / 27% / 45%	三状态概率按公允锚点标定：成长概率锚定种子→A 轮 15-50% (AI 溢价取偏上) × A → B≈60%'链路，基准

		0.28、保守 0.18、乐观 0.40；死亡概率参考 IT 桔子天使死亡率 31%（全样本均值，种子档更高）与 Correlation‘近半亏损’，基准 0.45、保守 0.57、乐观 0.32；其余为僵尸/停滞。死亡回收≈ 0（1x 优先权但残值近零）、僵尸回收 0.15-0.70x（下行轮/并购，部分优先权保护）。目标：混合分布的 $P(<1x)/P(>10x)$ 落入 Correlation/IT 桔子经验区间（见交叉校验）。
存活层 成长/僵尸/死亡(乐)	40% / 28% / 32%	三状态概率按公允锚点标定：成长概率锚定种子$\rightarrow$A 轮 15-50%（AI 溢价取偏上）$\times A \rightarrow B \approx 60\%$‘链路，基准 0.28、保守 0.18、乐观 0.40；死亡概率参考 IT 桔子天使死亡率 31%（全样本均值，种子档更高）与 Correlation‘近半亏损’，基准 0.45、保守 0.57、乐观 0.32；其余为僵尸/停滞。死亡回收≈ 0（1x 优先权但残值近零）、僵尸回收 0.15-0.70x（下行轮/并购，部分优先权保护）。目标：混合分布的 $P(<1x)/P(>10x)$ 落入 Correlation/IT 桔子经验区间（见交叉校验）。
死亡/僵尸回收 MOIC	≈ 0 / 0.15-0.70x	三状态概率按公允锚点标定：成长概率锚定种子$\rightarrow$A 轮 15-50%（AI 溢价取偏上）$\times A \rightarrow B \approx 60\%$‘链路，基准 0.28、保守 0.18、乐观 0.40；死亡概率参考 IT 桔子天使死亡率 31%（全样本均值，种子档更高）与 Correlation‘近半亏损’，基准 0.45、保守 0.57、乐观 0.32；其余为僵尸/停滞。死亡回收≈ 0（1x 优先权但残值

		近零)、僵尸回收 0.15-0.70x (下行轮/并购, 部分优先权保护)。目标: 混合分布的 $P(<1x)/P(>10x)$ 落入 Correlation/IT 桔子经验区间 (见交叉校验)。
第 5 年可变现概率 L5(保/基/乐)	20% / 33% / 50%	锚定本土现实: 种子后 5 年内 IPO 现金退出≈ 0 (成立 5 年内上市仅约 3-4%且种子在成立约 2 年后入场); 现金主要来自并购(占退出$\approx 30\%$)与老股转让(市场不成熟、2024 同比-15.5%); 对照全球 VC 第 5 年 DPI 仅 0.27-0.39x、第 8 年才到 1x。故第 5 年成长企业可变现概率保守 0.20/基准 0.33/乐观 0.50, 且早年极低 (liquidity_shape)。
逐年流动性形状(年 1-5)	2%/8%/25%/55%/100%	锚定本土现实: 种子后 5 年内 IPO 现金退出≈ 0 (成立 5 年内上市仅约 3-4%且种子在成立约 2 年后入场); 现金主要来自并购(占退出$\approx 30\%$)与老股转让(市场不成熟、2024 同比-15.5%); 对照全球 VC 第 5 年 DPI 仅 0.27-0.39x、第 8 年才到 1x。故第 5 年成长企业可变现概率保守 0.20/基准 0.33/乐观 0.50, 且早年极低 (liquidity_shape)。
流动性折扣(保/基/乐)	40% / 30% / 18%	Zanbato: 直接 VC 老股折扣 2023-12 约 46%$\rightarrow$2024 末约 3%(优质); 中国老股市场不成熟、缺体系化平台, 折扣更深, 建模取 保守 0.40/基准 0.30/乐观 0.18 并情景化。
死亡风险形状(年 1-5)	8%/30%/60%/85%/100%	财新/吴晓波: 新设企业 1 年存活 92%(8%死)、3 年 76%(24%死); IT 桔子: 种子天使为倒闭最高发轮次 (50.7%)。VC 支持的种子烧钱更快,

死亡集中于 2-4 年，年 5 收敛到情景
死亡率。

附录 C 可复现性说明

本报告全部数据与图表由以下开源脚本生成，输入假设集中于单一事实源 assumptions.py：

- assumptions.py：全部输入假设与数据来源（单一事实源）；
- model_market_sizing.py：市场规模 TAM/SAM/SOM 与量价交叉校验；
- model_token_economics.py：Token 单位成本与毛利（DeepSeek 校准）；
- model_competition.py：三阵营价格-份额 Logit 模型；
- model_financials.py：五年财务、NPV/IRR、龙卷风与蒙特卡洛；
- model_seed_returns.py：种子轮估值路径、cap table 稀释、1x 清算瀑布，以及【已实现现金（拿到钱）】退出层与 MOIC/胜率/盈亏比；
- make_figures.py / build_report.py：图表与本文档生成；
- run_all.py：一键复现全部产物（pip install -r requirements.txt && python run_all.py）。

在洁净环境运行 run_all.py 可重建全部 CSV、图表与本 docx，数值与本报告完全一致。