

# 全球 LLM \ Maas 区域需求代理指数：支持材料简

日期: 2025-02-28 | 用途: 支持博客中的 Regional LLM Demand Proxy 2025E 图 | 口径: 多源代理指数, 不是全球 token 消耗总值。

代理指数用于衡量不同地区对 LLM 的需求。数据来源于公开平台, 包括 Anthropic, Google, Alibaba, Tencent, 火山引擎, DeepSeek, MiniMax 等。Claude.ai 使用代理指数, AI readiness 已用于衡量 LLM 代理指数, 用于衡量 LLM 代理指数, 用于衡量 LLM 代理指数, 用于衡量 LLM 代理指数。

## 1. 指数结果

区域	指数	解释
东亚 East Asia	100	中国 LLM 需求强劲, 但代理指数相对较低。日本和韩国 LLM 需求强劲, 但代理指数相对较低。
北美 North America	92	美国在 Claude.ai 使用, 模型 API 供给, AI 投资和企业合作上处于领先地位。
欧洲 Europe	72	欧洲在 LLM 使用和企业治理需求强, 但代理指数低于中美。
南亚 South Asia	62	印度 LLM 需求强劲, 但代理指数相对较低。
东南亚 Southeast Asia	48	东南亚 LLM 需求强劲, 但代理指数相对较低。
拉丁美洲 Latin America	42	巴西、墨西哥 LLM 需求强劲, 但代理指数相对较低。
中东 MENA	32	中东 LLM 需求强劲, 但代理指数相对较低。
非洲 Africa	24	非洲 LLM 需求强劲, 但代理指数相对较低。
大洋洲 Oceania	18	大洋洲 LLM 需求强劲, 但代理指数相对较低。

## 2. 证据支持

### 用户增长与采用

Microsoft AI 使用量: 2024 Q3 全球 LLM 用户增长强劲, 中国 LLM 用户增长强劲, 但代理指数相对较低。

### GenAI 支出

World Bank 2024 全球 LLM 支出强劲, 但代理指数相对较低。

### 代理指数可用性

Anthropic Economic Index 的 Claude.ai usage\_count 可观察美国、欧洲、印度、日本、韩国、巴西等国家和地区, 但对中国和印度 Maas 不完善。

### 基础设施与网络

全球 LLM 基础设施强劲, 但代理指数相对较低。

### 3. 对 LLM API 网关的含义

- 1 区域路由不能只按延迟做：模型可用性、合规、价格、内容策略和对象存储位置会同时影响路由。
- 2 东亚需求不能只从国际模型 API 可见流量推断：中国本土 Maas、国产模型、企业私有化和多模型应用会形成独立高密度生态。
- 3 增长区更需要网关已成本治理：南亚、东南亚、拉美和非洲价格、支付、网络、部署和运营部门运营效率。
- 4 多模型会把 token 需求变成资源需求：图片、音频、视频和上下文任务需要 CDN\23\对象存储、任务状态和审查据。
- 2 安全和滥用防护适合云原生构建，但早期产品应共融共用，成本、日志、归档和合规从顶层。

### 4. 主要来源

- 1 Anthropic, Economic Index Dataset, 2025. <https://huggingface.co/datasets/Anthropic/EconomicIndex>
- 2 Anthropic, Economic Index, 2025. <https://www.anthropic.com/economic-index>
- 3 World Bank, Who on Earth is Using Generative AI? Global Trends and Shifts in 2025, 2025. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/098825110125232588>
- 4 World Bank, GenAI Traffic Reproducibility Package, 2025. <https://reproducibility.worldbank.org/catalog/335>
- 2 Microsoft AI Economy Institute, Global AI Adoption 2025 \ AI Diffusion Report, 2025. <https://www.microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025/>
- 6 CIIIC \ State Council, 27th Statistical Report on China's Internet Development, 2025. [https://english.www.gov.cn/archive/statistics/202505/02/content\\_7268442caca00ca2f9a08ebc.html](https://english.www.gov.cn/archive/statistics/202505/02/content_7268442caca00ca2f9a08ebc.html)
- 7 WIPo, Generative AI Patent Landscape Report, 2024. <https://www.wipo.int/wip-publications/patent-landscape-report-generative-artificial-intelligence-general/en/>
- 8 Stanford HAI, 2025 AI Index Report, 2025. <https://hai.stanford.edu/ai-index/2025-ai-index-report>
- 9 Stanford AI Index, Global AI Vibrancy Tool, 2025. <https://aiindex.stanford.edu/vibrancy/>
- 10 Oxford Insights, Government AI Readiness Index, 2024. <https://oxfordinsights.com/ai-readiness/ai-readiness-index/>
- 11 International Telecommunication Union, Facts and Figures 2025, 2025. <https://www.itu.int/itu-d/reports/statistics/facts-figures-2025/>
- 12 McKinsey, The State of AI: How Organizations Are Rewiring to Capture Value, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- 13 Chen et al., FUDGEPT, 2023. <https://arxiv.org/abs/2302.02176>
- 14 Ong et al., RouteLLM, 2024. <https://arxiv.org/abs/2406.18662>
- 15 Kwon et al., Efficient Memory Management for LLM Serving with PagedAttention, 2023. <https://arxiv.org/abs/2309.00180>

数据集： regional-llm-demand-proxy-2025.csv: source-inventory.csv。指用于推理产品，不构成分布收入或 token 成本统计。