

开源闭源模式选择与中美大模型产业竞争

渠慎宁

(中国社会科学院工业经济研究所, 北京 100006)

摘要: 开源与闭源的模式选择, 是大模型产业发展中最具战略意义的问题之一, 也是理解中美人工智能竞争的关键。本文在厘清大模型语境下“开放权重”与传统开源软件本质差异的基础上, 系统比较两种模式的优势与局限, 从技术位势、利润池位置、可占有性与制度环境四个维度阐释模式选择的内在逻辑, 进而考察中美大模型产业的发展态势与模式分布, 重点分析模式选择对中美大模型产业竞争、上下游相关产业竞争、中国追赶进程以及智能化扩散与治理格局的影响。研究表明: 美国形成前沿闭源为主导的金字塔格局, 中国形成开放权重模式主导的雁阵格局。开源使我国实现了以扩散换时间、以生态换位势的非对称追赶, 但也面临前沿模型差距再度拉大、价值向算力环节集中而算力受制于人、生态优势具备可逆性等约束。立足“十五五”规划纲要部署, 应将人工智能作为培育未来产业和新质生产力的主攻方向, 健全多元化科技投入机制, 统筹算力底座建设与国产化替代, 巩固开源生态优势, 以智能化赋能实体经济和中小企业, 深化国际开放合作, 在中美人工智能竞争和整体科技产业竞争中赢得战略主动。

关键词: 大模型; 开源闭源; 模式选择; 中美竞争; 未来产业

中图分类号: F124

文献标识码: A

文章编号: 1005-0566(2026)09-0029-12

Open versus closed: Model strategy choices and the China-U. S. competition in large models

QU Shenning

(Institute of Industrial Economics, Chinese Academy of Social Sciences, Beijing 100006, China)

Abstract: The choice between open and closed model strategies is one of the most strategically significant issues in the development of the large model industry, and a key factor in understanding the China-U. S. competition in artificial intelligence. Building on a clarification of the fundamental differences between “open weights” in the large model context and traditional open-source software, this paper systematically compares the advantages and limitations of the two strategies, and explains the underlying logic of strategy choice along four dimensions: technological position, profit-pool location, appropriability, and institutional environment. It then examines the development trends and strategy distribution of the Chinese and U. S. large model industries, with a focus on how strategy choice affects competition in the large model industry itself, competition in related upstream and downstream industries, China’s catch-up process, and the pattern of intelligent-technology diffusion and governance in both countries. The study finds that the United States has formed a pyramid structure dominated by frontier closed-source models, while China has formed a flying-geese

基金项目: 研究阐释党的二十届四中全会精神国家社会科学基金重大专项“未来产业投入增长与风险分担机制的政策体系研究”(26ZDA011); 中国社会科学院学科建设“登峰战略”优势学科(产业经济学)(DF2023YS24)。

作者简介: 渠慎宁(1986—), 男, 江苏南京人, 中国社会科学院工业经济研究所研究员、中国社会科学院大学教授, 博士, 研究方向为新兴产业和未来产业。

structure dominated by open-weight models. Open-sourcing has enabled China to achieve asymmetric catch-up by trading diffusion for time and ecosystem for position, yet it also faces constraints such as a renewed widening of the frontier-model gap, the concentration of value in the compute segment where China remains dependent on others, and the reversibility of its ecosystem advantage. In line with the deployment of the 15th Five-Year Plan, China should make artificial intelligence a primary focus for cultivating future industries and new quality productive forces: maintaining diversified mechanisms for technological investment, coordinating the build-out of compute infrastructure with domestic substitution, consolidating its open-source ecosystem advantage, empowering the real economy and small and medium-sized enterprises through intelligent technologies, and deepening international cooperation and openness, so as to gain strategic initiative in both the China-U. S. AI competition and the broader science and technology industrial competition.

Key words: large models; open versus closed source; strategy choice; China-U. S. competition; future industries

以大语言模型为代表的人工智能大模型,正在成为智能时代的通用技术^[1],其影响沿着“技术突破、产业渗透、范式重构”的路径持续展开。与历史上的蒸汽机、电力和信息技术相比,大模型的特殊性在于,其能力的形成高度依赖巨额且不可逆的训练投入,能力的扩散却可以通过权重文件的复制在边际成本趋零的条件下完成。这种“生产高度集中、扩散近乎免费”的内在张力,使“是否开放、向谁开放、开放到什么程度”成为大模型产业中最具战略意义的决策因素,开源与闭源的模式分化由此构成观察大模型产业组织与国际竞争的重要切入点。模式分化在中美之间表现得尤为鲜明。2025 年 1 月,DeepSeek 以 MIT 许可证开放 R1 推理模型权重,性能逼近同期最强闭源模型而训练与使用成本低一个数量级,引发全球资本市场与产业界的剧烈反应。此后,阿里通义千问、智谱、月之暗面、MiniMax 等持续高频开源,百度将文心 4.5 系列转为开源,中国形成了系统性的开源格局。美国方面,OpenAI、Anthropic 等头部企业坚持旗舰闭源,但 OpenAI 于 2025 年 8 月发布 gpt-oss 系列开放权重模型,是其自 2019 年 GPT-2 以来的首次开放。同年 7 月美国政府发布的《赢得竞赛:美国人工智能行动计划》明确把发展和推广美国开源模型列为国家战略事项。开源闭源的选择,已经从企业层面的商业决策上升为中美人工智能产业竞争的关键变量。

党的二十届四中全会和 2026 年 3 月十四届全国人大四次会议审议通过的《中华人民共和国国民经济和社会发展第十五个五年规划纲要》,

把加快高水平科技自立自强、引领发展新质生产力摆在突出位置,专门部署深入推进数字中国建设,提出深化拓展“人工智能+”行动、抢占人工智能产业应用制高点。习近平总书记高度重视人工智能发展,在 2025 年 4 月中央政治局第二十次集体学习时强调,要充分发挥新型举国体制优势,坚持自立自强,突出应用导向,推动我国人工智能朝着有益、安全、公平方向健康有序发展,并指出要正视差距、加倍努力,集中力量攻克高端芯片、基础软件等核心技术,构建自主可控、协同运行的人工智能基础软硬件系统,同时深刻阐述人工智能可以是造福人类的国际公共产品。“十五五”时期是基本实现社会主义现代化夯实基础、全面发力的关键时期,也是中美科技产业竞争最为胶着的窗口期。大模型作为本轮人工智能革命的底座技术,其产业竞争的成败直接关系到未来产业建设的进度、新质生产力培育的成色和国家竞争新优势的塑造,而开源闭源的模式选择,正是这场竞争中牵一发而动全身的关键变量。本文围绕这一主线展开:第一部分厘清大模型开源闭源的技术范式与主要特征;第二部分比较两种模式的优缺点并阐释模式选择的内在逻辑;第三部分考察当前中美大模型产业的发展现状与模式选择;第四部分分析模式选择对中美大模型产业竞争、上下游产业竞争与中国追赶进程的影响;第五部分立足我国现实条件提出应对策略。

一、AI 大模型开源闭源的技术范式与主要特征

准确把握开源与闭源的内涵差异和技术特

征,是分析模式选择问题的逻辑起点。大模型语境下的“开源”与传统软件产业中的开源并不等同,开源与闭源也不是非此即彼的二元对立,而是呈现多层次、多维度的策略组合。这就需要界定清楚开放的层次,说明大模型语境中的“开源”究竟开放了什么,再分析两种范式各自的运行方式及其背后的技术经济根源。

(一)从源代码开放到权重开放的概念演变

开闭源之争并非大模型所特有。凡是具有平台属性、需要外部开发者共同投入的技术领域,历史上几乎都出现过类似情况。在操作系统领域,微软以 Windows 的专有授权建立起个人计算机时代的事实标准, Linux 则凭借 GPL 许可证与全球开发者协作,在服务器与超算市场取得了压倒性份额。移动端 Android 以源代码开放换取设备厂商与应用生态的快速集聚, iOS 则以软硬一体的封闭体系维持高溢价。在数据库领域, Oracle 长期依靠专有授权获取高额许可收入, MySQL、PostgreSQL 等开源产品则通过降低使用门槛成为互联网应用的默认选择。在上一轮人工智能浪潮中, TensorFlow 与 PyTorch 围绕深度学习框架的开源竞争,直接决定了开发者的归属。这些案例表明,开放并不等于放弃商业利益,而是把收益获取的位置从技术本体转移到互补环节,封闭也不必然意味着落后,其成败取决于能否以持续的性能领先支撑溢价。大模型的开源闭源选择是在新技术条件下的延续,但开放的对象与开放的含义都已经发生了实质变化。

在传统软件产业中,开源具有明确的制度边界:源代码公开,允许自由使用、修改与再分发,并由开放源代码促进会(OSI)认证的许可证体系加以约束。开发者拿到源代码,就拿到了软件的全部“生产过程”,可以完整复现、审计乃至改造软件本身。大模型语境中的“开源”则与之存在较大差异,当前业界所称的开源大模型,绝大多数属于开放权重(open weights)模式,即企业公开训练完成的模型参数文件与推理代码,允许下载、部署与微调,但训练数据、数据处理流程、完整训练代码与超参数配置通常并不公开。严格

满足数据、代码、权重三重开放的完全开源模型只占极小比例。两者相比,源代码公开意味着生产过程可复现、可审计,权重公开只意味着生产结果可复用。形象地说,前者开放的是“配方”,后者开放的是“成品”。

开放程度的差异还体现在许可证设计上。例如, DeepSeek 与通义千问的多数模型采用 MIT、Apache 2.0 等宽松许可证,允许无差别商用与再分发; Llama 系列则采用附加限制条款的社区许可证,对超大规模商用等用途设限。闭源阵营内部同样存在差别,多数旗舰模型仅以接口形式对外提供服务,少数模型甚至完全不对外开放。综合开放对象与使用约束两个维度,大模型的开放程度可划分为完全开源、宽松许可开放权重、限制许可开放权重、接口开放与完全封闭五个层次(见表1)。本文所讨论的“开源闭源模式选择”,正是企业在这连续区间上的定位选择,既包括开放层次的选择,也包括开放时点(同步开源或滞后开源)与开放结构(旗舰闭源加轻量开源的分层安排)的选择。为避免概念混用,本文将表1中第二、第三层次统称为“开放权重模式”,下文出现的“开源”若无特别说明,均指开放权重意义上的开源,与数据、代码、权重三重开放的“完全开源”相区别。

表1 大模型开放程度的层次划分

开放层次	开放内容	典型代表	使用约束
完全开源	权重、训练代码与训练数据全部公开	艾伦人工智能研究所 OLMo 系列	宽松许可证,可复现、可审计
开放权重(宽松许可)	公开权重与推理代码,训练数据与全流程不公开	DeepSeek 系列、通义千问系列、gpt-oss	MIT、Apache 2.0 等,可自由商用与再分发
开放权重(限制许可)	公开权重,附加使用限制条款	Llama 系列	社区许可证,对超大规模商用等设限
接口开放(闭源)	权重不公开,仅通过接口提供服务	GPT 旗舰系列、Claude 系列、Gemini 旗舰	按调用付费,受服务条款约束
完全封闭	仅供内部使用,不对外提供	部分企业内部模型	不对外开放

资料来源:作者根据各企业模型发布公告与许可证文本整理,信息截至 2026 年 6 月。

(二)两种技术范式的运行特征

闭源大模型以“模型即服务”为基本形态。企业将模型部署在自有或合作的云端,通过应用程序接口和订阅产品对外输出能力,用户按调用量或时长付费。其运行特征是“集中化”,模型版本、

安全护栏与内容策略由企业统一管理,可以持续开展对齐优化与滥用监测,问题响应与责任承担的主体清晰。代价则是过程不透明,用户无法审计模型的内部机制,数据需要出域调用,存在“黑箱”决策与供应商锁定的隐忧。

开源大模型则以“权重即产品”为基本形态。企业把训练完成的权重文件发布到开源社区平台,任何机构与个人均可下载部署、微调改造、蒸馏压缩,并在许可证允许的范围内再分发。其运行特征是“分布式”,能力的改进与场景适配由全球开发者社区共同完成,衍生模型围绕基座家族不断繁衍,工具链、教程与最佳实践随之沉淀。代价则是“发布即失控”,权重一经流出便不可召回,安全护栏可以被低成本移除,企业对下游的使用方式不再具有技术层面的约束能力。

两种范式还对应着两种不同的创新组织方式。闭源大模型下,创新集中于企业内部实验室,依靠高密度的人才、算力与数据闭环快速试错,优势在于方向聚焦、保密性强、工程整合度高。开源大模型下,创新分散于全球社区,基座企业提供公共底座,海量开发者在各自的场景中并行探索,改进成果通过衍生模型与工具链被生态即时吸收,优势在于试错面广、适配速度快、长尾需求覆盖充分。前者类似集中攻关,后者类似群体演化,二者各有所长,也各自需要付出相应的组织与协调成本。

(三) 开源闭源分化背后的经济技术根源

开源与闭源之所以在大模型领域形成两条不同路径,根源在于大模型独特的成本结构。大模型的生产呈现“重资产训练、零成本复制、有成本服务”的三段式结构:在训练端,前沿模型单次训练的算力成本达数千万乃至上亿美元且完全沉没;在复制端,权重文件的复制与分发边际成本趋近于零;在服务端,每一次推理都消耗真实算力。这一结构同时定义了两种模式的商业基础。对闭源而言,“训练贵”意味着必须依靠接口收费回收巨额投入,对权重的排他控制是其定价权的生命线。对开源而言,“复制零成本、推理有成本”意味着公开权重并不等于让渡全部价值,企业仍可通

过云上托管推理、企业级服务与行业解决方案实现变现。传统软件“开源即放弃收入”的逻辑由此在大模型时代被改写,这正是开放权重能够成为头部企业主动战略选择的经济前提。

可占有性与网络效应的此消彼长,进一步塑造了两种模式的力量对比。一方面,大模型的技术可占有性天然偏弱,蒸馏技术使闭源模型的输出本身成为可供学习的“教材”,架构与算法创新随论文发表和人才流动快速扩散,前沿性能优势的“半衰期”以月计,闭源护城河持续变浅、租金加速折旧。另一方面,开放权重却能构筑另一种壁垒,围绕基座模型家族形成的开发者、衍生模型、工具链与社区,呈现报酬递增与路径依赖的网络效应^[2-4],用户越多则生态越厚,生态越厚则越难被替代。技术领先难以持久,生态优势却不易撼动,这一对比从根本上提升了开源策略的相对吸引力。再考虑到大模型价值的实现高度依赖算力、数据与应用场景等互补资产^[5],以及权重一经发布即不可召回的高度不可逆性,开源与闭源的选择就成为企业在租金获取、生态构建、抢占互补资产、控制风险之间的系统性权衡,而绝非单纯的技术路线偏好。

二、开源闭源模式的优缺点比较与选择逻辑

开源与闭源并无抽象意义上的绝对优劣,两种模式各有其收益、代价与适用条件,企业的模式选择是在特定竞争位势、资产结构与制度环境下反复权衡的结果。以下分别考察两种模式的收益与代价,并从技术位势、利润池位置、可占有性与制度环境四个维度,解释企业在什么条件下会倒向其中一端。

(一) 开源模式的优势与局限

开源模式的优势集中体现在扩散与生态两个层面。第一,零许可成本极大加速了模型的扩散与采用,开放权重可以被自由下载部署,迅速进入企业、高校与开发者的工作流,在短时间内积累庞大的用户与开发者基数。第二,开放本身是一种信任机制,模型可审计、可本地化部署,能够满足政企客户数据不出域的合规要求,显著降低获客的信任成本。第三,开源动员了分布式的群体创

新,社区围绕基座模型开展微调、蒸馏、量化与工程优化,形成衍生模型与工具链的繁荣生态,反过来巩固基座模型的事实标准地位^[6-8]。第四,开源为间接变现打开了通道,企业可以通过云上托管推理、企业级服务、行业解决方案等互补环节获取收益,使开放与商业化得以兼容。

开源模式的局限同样明显。一是直接收益薄弱且可持续性存疑,权重免费发布意味着放弃了对模型本体的定价权,第三方托管商可以利用开放权重获利而无须向开发者付费,存在突出的“搭便车”问题。持续训练与发布高水平开源模型又需要巨额投入,若互补环节的变现不足以覆盖成本,开放承诺便难以长期维持。Llama 系列在社区许可证中对超大规模商用附加限制条款,正是开源者对这一问题的制度性回应。二是能力外溢资敌,竞争对手可以通过研究权重、蒸馏输出获得技术洞察,开源者的研发投入部分转化为行业公共品。三是安全不可控,权重发布后不可召回,护栏可被移除,一旦被恶意利用,声誉与法律风险难以完全切割。

(二) 闭源模式的优势与局限

闭源模式的优势集中体现在收益占有与安全可靠两个层面。第一,接口收费与订阅服务使创新收益的占有直接而充分,前沿性能优势可以即时转化为定价能力与现金流,进而支撑下一轮高强度的前沿投入,形成“领先、收费、再投入”的正向循环。第二,集中化的部署便于持续对齐、内容安全管理与滥用监测,模型版本与服务体验保持一致,责任界面清晰,更适合高风险能力的审慎管理。第三,闭源保护了训练方法、数据配方等核心知识资产,延缓了技术外溢的速度,有助于维持代际领先。

闭源模式的局限在于扩散与信任。一是扩散受到定价约束,价格敏感的长尾用户与发展中国家市场难以充分覆盖。当性能接近的开放模型出现时,闭源服务的“价格伞”会被系统性挤压。二是黑箱与锁定顾虑,用户无法审计模型机制,数据出域调用引发主权与合规担忧,政企客户对单一供应商依赖心存疑虑。三是护城河折旧快,蒸馏

与人才流动使闭源的技术领先难以长期保持,维持优势要求持续的巨额军备投入,一旦迭代节奏放缓,租金便迅速流失。两种模式的系统比较见表2。

表2 开源与闭源模式的优劣比较

比较维度	开源(开放权重)模式	闭源模式
收益占有方式	间接变现:云服务、行业方案、生态溢出	直接变现:接口调用与订阅收费
扩散速度与生态构建	扩散快,易形成开发者生态与衍生创新	扩散受定价约束,生态依赖官方工具链
安全与可控性	权重不可召回,护栏可被移除,但透明可审计	集中管控、持续对齐,但存在黑箱与锁定
创新投入的激励	依赖互补资产变现,存在搭便车风险	前沿租金直接激励高强度投入
用户信任与合规	可本地部署,满足数据主权与定制需求	责任主体清晰,便于内容安全管理
典型适用主体	追赶者、平台型企业、互补资产持有者	领先者、模型即产品的专业公司

(三) 模式选择的内在逻辑

上述优劣并不会自动决定企业的选择,模式选择的关键在于:哪一种优势对特定企业更有价值,哪一种代价对它更可承受。综合来看,四个因素从根本上塑造了选择的方向。

首先,技术位势决定开闭的基本倾向。处于性能前沿的领先者,其领先优势可以直接转化为定价能力与市场份额,闭源收割前沿租金是占优选择。处于追赶位置的企业在性能维度难以正面取胜,开源则提供了另一条竞争路径,即以零许可成本加速扩散,扩大用户与开发者基数,侵蚀领先者的租金空间,并把竞争从性能单一维度转移到成本、生态与可控性维度。开源因此带有“追赶者武器”的色彩。产业史上不乏先例。1998年,网景公司在浏览器份额被微软IE大幅挤压之后公开了Navigator的源代码,由此催生的Mozilla项目虽未夺回原有市场,却把竞争从产品性能拉到了标准与生态层面。当性能维度难以正面取胜时,开放是改换竞争维度的手段。位势变化还会引起选择的动态调整,当追赶者逼近前沿后,其开源承诺可能弱化,而领先者的生态受到侵蚀后,也可能进行防御性的局部开放。

其次,利润池位置决定开放的商业可持续性。当企业的利润池位于模型的互补品层,如算力云服务、社交与广告终端、行业应用时,模型层的免

费供给会放大对互补资产的引致需求,开源相当于把模型“商品化”,把产业租金推向自己占优的环节,开放因此与利润最大化相容。当模型本身就是最终产品、模型服务收入是唯一利润来源时,闭源则是实现创新收益占有的必要条件。这一逻辑在信息产业中反复出现。2000 年前后,IBM 大举投入 Linux 开发,本身并不从操作系统收费,目标是带动服务器硬件与技术服务收入。2007 年,谷歌牵头成立开放手机联盟并宣布 Android 开源,次年公开源代码,同样不向手机厂商收取授权费,收益来自搜索、广告与应用分发。企业开放的从来不是利润所在的那一层。这一机制与信息产业史上“开放互补品、收获核心资产”的经典策略一致^[9],也与创新收益占有理论判断吻合,即当可占有性较弱而互补资产重要时,创新收益将主要流向互补资产的持有者^[5]。

再次,可占有性强弱决定闭源护城河的深浅。闭源策略的有效性取决于企业守住技术租金的能力,而大模型的可占有性天然偏弱。例如,蒸馏技术使闭源模型的输出本身成为可供学习的“教材”,能力可以低成本外溢,架构与算法创新随学术发表和核心人才流动快速扩散,性能差距的收敛周期以月计。数字经济条件下可占有性的整体弱化趋势在大模型领域表现得尤为突出^[10]。可占有性越弱,闭源租金的护城河越浅,且折旧越快,维持闭源优势就要求持续地高强度投入。当这种投入的边际回报递减时,开源或混合策略的相对吸引力随之上升。反向的力量则来自安全与合规,模型滥用的责任风险与持续对齐的运维要求,为旗舰模型的闭源可控发布提供了现实必要性。产业史上 Windows 与 Linux 的长期竞争也表明,面对开源竞争者,闭源领先者的理性反应是加速创新与策略性降价,而非退出市场^[11]。

最后,制度环境塑造模式分布的国别差异。创新融资结构、外部管制与产业政策可共同作用于企业的选择。例如,资本市场对“前沿叙事”给予高估值的环境,支撑模型公司以闭源服务收入持续融资投入。平台型企业占主导的产业结构,则使开源与既有商业体系天然兼容。尖端算力

获取受限等外部约束,会提高动员分布式创新的开放策略的相对收益。政府还可以把开源作为生态培育与国际影响力工具纳入国家战略,直接改变企业开放的成本收益结构。监管规则同样在改变天平。欧盟《人工智能法案》对以自由开源许可证发布的通用人工智能模型作了部分义务豁免,但对被认定具有系统性风险的模型不予豁免。这一安排降低了中小规模开放模型的合规成本,却未放松对前沿能力的约束。模式选择因此不仅是企业决策,也是国家创新体系特征的映射。

四个因素叠加,现实中的企业策略很少落在两端。混合形态更为常见:分层开放(旗舰闭源加轻量开源)、滞后开放(前沿能力保留一至两代后再开源)、许可证调控(以条款限制竞争性使用)等做法普遍存在,它们对应着开放接入与让渡控制等不同的开放维度^[12-13]。模式选择也随位势与环境变化而动态切换,追赶者登顶后可能收紧开放,领先者承压后可能局部开放。可以预期,大模型产业的均衡不是单一模式的胜出,而是开源与闭源在不同层次、不同环节上的长期共存与互动。

三、当前中美大模型产业的发展态势与模式选择

中美同属全球大模型产业的第一梯队,但两国在模式选择上走出了方向迥异的路径:美国形成前沿闭源为主导的金字塔格局,中国形成系统性开放权重的雁阵格局。两国的差异并非单纯企业偏好的加总,其背后是位势、资产结构与外部约束的系统性差别。

(一)美国以前沿闭源为主导的金字塔格局

美国大模型产业呈现“前沿闭源为主、开源为辅”的金字塔结构。OpenAI 与 Anthropic 以闭源服务为核心商业模式,旗舰模型长期占据前沿性能榜单,接口调用与订阅收入支撑其巨额研发投入。谷歌采取分层策略,旗舰 Gemini 系列闭源,同时开放轻量级 Gemma 系列作为开发者入口。Meta 是美国阵营最重要的开源力量,Llama 系列一度是全球开源生态的事实基座,但 2025 年 Llama 4 发布

后市场反响不及预期,叠加内部组织调整,其对旗舰模型的开源承诺出现明显摇摆。

闭源主导格局的背后是空前的资本投入强度。据美中经济与安全审查委员会 2026 年发布的报告估计^[14],微软、亚马逊、Meta 与谷歌四家公司 2025 年的人工智能资本开支超过 3 500 亿美元,2026 年预计超过 4 000 亿美元,远超中国主要云厂商的同期投入。与此同时,美国阵营也出现了向开放回摆的标志性动向,OpenAI 以 Apache 2.0 许可证发布的 gpt-oss 系列开放权重模型,被普遍解读为对中国开源攻势与生态份额流失的防御性回应。《赢得竞赛:美国人工智能行动计划》文件也明确提出发展具有全球影响力的美国开源模型、推广美国技术栈出口。

(二) 中国以开放权重模式为主导的雁阵格局

中国大模型产业形成了以开放权重模式为主导、头部力量呈梯次推进的“雁阵”形态。DeepSeek 以 MIT 许可证开放 V3 与 R1 系列,把前沿级推理能力带入开源世界。阿里通义千问坚持全尺寸、多模态的高频开源,衍生模型规模位居全球开源模型家族前列,成为全球开发者二次开发的主要基座之一。智谱、月之暗面、MiniMax 等持续开放旗舰级能力,月之暗面的 Kimi K2 为万亿参数级开放权重模型。百度于 2025 年年中将文心 4.5 系列转为开源,标志着原闭源阵营的策略转向。与此同时,国内并非单一模式,字节跳动豆包等产品仍以闭源低价服务参与竞争,国内同样存在双轨格局。

产业规模与应用生态同步扩张。据国家互联网信息办公室统计,截至 2025 年 12 月 31 日,全国累计已有 748 款生成式人工智能服务完成备案,435 款生成式人工智能应用或功能完成登记,仅 2025 年就新增备案 446 款,备案主体覆盖互联网企业、高校科研机构与央企国企等各类机构,服务形态从通用能力延伸到行业垂类与多模态场景。旺盛的模型供给与超大规模市场的应用需求相互促进,为开放权重模式提供了广阔的落地土壤。

中美代表性企业的模式选择见表 3。

表 3 中美代表性大模型企业的模式选择

国别	企业	代表模型	模式选择	主要商业模式
美国	OpenAI	GPT 系列、gpt-oss	旗舰闭源,辅以开放权重	接口调用与订阅服务
美国	Anthropic	Claude 系列	闭源	接口调用与企业服务
美国	谷歌	Gemini 系列、Gemma 系列	旗舰闭源加轻量开源	云服务与广告生态
美国	Meta	Llama 系列	开放权重(限制许可),趋于收紧	社交广告,模型为互补投入
中国	深度求索	DeepSeek-V3、R1 系列	开放权重(MIT 许可)	低价接口与技术影响力
中国	阿里巴巴	通义千问系列	系统性开放权重(以 Apache 2.0 为主)	云计算与企业服务
中国	智谱	GLM 系列	开放权重为主	企业服务与行业方案
中国	月之暗面	Kimi K2 系列	开放权重	接口调用与订阅
中国	百度	文心系列	由闭源转向开源	云与搜索生态
中国	字节跳动	豆包系列	闭源低价	应用生态与广告

资料来源:信息截至 2026 年 6 月。

(三) 性能差距波动与生态重心迁移的发展态势

在性能维度,中美差距呈反复波动之势,每当美国推出新一代前沿模型便再度拉开,中国大致滞后 3~12 个月后追平,如此循环往复。斯坦福大学《人工智能指数报告 2026》^[15]显示,到 2025 年中美顶尖模型在主要基准上的差距一度收敛至 3% 以内,全球顶尖模型贡献机构前三位为 OpenAI、谷歌与阿里巴巴。但同一报告也记录了反向变化,截至 2026 年 3 月,顶尖闭源模型对顶尖开源模型的领先幅度已从 2024 年 8 月的 0.5% 扩大到 3.3%,竞技场排行榜前十名中六款为闭源模型。尤其是 2025 年 11—12 月美国头部企业密集发布新一代旗舰模型形成集中的能力跃升后,中美之间最前沿模型的差距呈现重新拉大的态势。同时,中美模型的差距结构也值得注意。目前在代码、数学等标准化任务上中国模型已十分贴近美国领先模型,在性价比维度上中国模型甚至具有一些优势,但在长程自主任务、复杂推理任务、超长上下文的连贯任务等接近人类专家连续决策任务上的差距在变大。这表明,当前沿竞赛进入算力密集驱动的新阶段后,单纯依靠工程效率创新维持追赶速度的难度正在上升。

在生态维度,开放权重生态的重心明显向中国迁移。麻省理工学院与 Hugging Face 研究团队的测算^[16]显示,截至 2025 年 8 月的一年间,中国开放权重模型占全球模型下载量的比重达 17.1%,首次超过美国的 15.8%。海外开发者与初创企业依托中国开放权重模型进行二次开发渐成普遍现象。

此外,价格优势进一步强化了中国的生态吸引力。美中经济与安全审查委员会援引第三方评测指出,2026 年初中国头部开放模型在智能水平评分大体持平的条件下,价格仅为美国对标闭源模型的四分之一左右。概言之,美国守住了性能的“最高点”,中国则占据了生态的“基本面”。

从创新主体与迭代节奏看,中美竞争还呈现产业界主导、发布提速的特征。《人工智能指数报告 2026》统计显示,2025 年全球超过九成的顶尖模型由产业界产生,企业已取代学术机构成为前沿创新的绝对主力。中美头部企业的旗舰模型更新周期普遍缩短,重要版本的发布间隔由以年计压缩到以季度甚至月度计。高频迭代既是闭源阵营维持领先的手段,也是开源阵营保持生态热度的方式,这在客观上加快了全行业的技术折旧与成本下降,使任何一方都难以凭借单次突破获得持久优势。

(四) 开源闭源模式分化的深层逻辑

中美模式格局的差异,是竞争位势、资产结构、外部约束与政策导向四重因素叠加作用的结果。竞争位势决定开源闭源的基本激励,资产结构决定开放能否赚钱,外部约束改变策略的相对收益,政策导向放大并固化企业的选择,由此演化出两国内部能够自洽、却难以相互复制的产业范式。

从竞争位势看,开源的本质是追赶者发动的“竞争维度的转换”。在性能单一维度上,追赶者正面挑战领先者胜算有限,开放权重把竞争同时引入成本、部署灵活性与生态厚度三个新维度,在这些维度上领先者的性能优势无法直接兑现,追赶者的低价与开放反而成为长板。对领先者而言,跟进开源等于放弃旗舰收益,不跟进则坐视开发者流失,这种两难解释了美国阵营的折中反应,如 OpenAI 发布 GPT-OSS、谷歌开放 Gemma,开放的都是非旗舰能力,意在牺牲最小的租金换取开发者心智、阻滞生态全面倒向中国生态,是典型的“保旗舰、开周边”防御性开放。反过来,国内个别企业在细分能力登顶后,出现旗舰延迟开源的迹象,同样印证了位势与开放意愿之间的反向关系。

从资产结构看,模式分布的差异有清晰的财务逻辑。美国头部模型公司是“模型即全部”的专业公司,其收入与估值几乎完全建立在闭源接口与订阅之上,开源旗舰无异于动摇自身的商业根基,其背后的风险资本与公开市场,又以前沿领先和收入高增长作为给予高估值的前提,进一步锁定了闭源路径。中国头部开源者则多为平台型企业或生态型创业公司。例如,阿里巴巴的利润池在云计算与企业服务,权重免费做大了云上推理与行业方案的引致需求。DeepSeek 背靠量化投资母体不依赖模型收入生存,开源带来的技术声誉与人才集聚本身就是回报。模型在美国头部企业是利润中心,在中国头部企业更多是流量入口与生态投入,这一微观差异加总起来,便是两国宏观模式格局的差异。

从外部约束看,出口管制非但没有锁死中国的追赶路径,反而内生地催化了开源路线。管制压缩了单点算力供给,使“以算力规模换性能”的正面竞赛难以为继,中国企业的理性应对是改变创新的组织方式,通过把模型优化问题通过开放权重外包给全球开发者社区,以分布式试错替代集中式堆料,以混合专家架构、低精度训练、推理优化等工程效率创新替代算力规模扩张,用迭代速度对冲算力差距。开源同时也是对管制本身的反制,权重一经全球扩散,技术封锁的边际效果即递减,海外开发者转而构建在中国模型之上,反而加深了对中国生态的依赖。约束催生应对,应对又反过来作用于约束,这是理解中国系统性开源最关键的一环。

从政策导向看,两国都把开源升格为国家战略工具,但着眼点并不相同。美国《人工智能行动计划》把开源模型视为技术栈输出的载体^[17],意在通过模型、芯片、标准的整体出口维系全球生态主导权,将强化本国开源生态作为对冲中国技术扩散影响力的手段。我国《国务院关于深入实施“人工智能+”行动的意见》提出“推动人工智能技术开源可及”,把人工智能定位为造福人类的国际公共产品,开源既是对内动员创新资源的组织方式,也是对外服务全球南方、塑造开放形象的合作

姿态。党的二十届四中全会进一步对“十五五”人工智能发展作出顶层设计和体系化部署,为开放路线提供了制度保障。政策的双向加持,使开源闭源之争从企业策略层面固化为国家间的生态与标准竞争。

归结起来,美国“前沿闭源为主、防御性开源为辅”与中国“系统性开源为主、旗舰闭源探索为辅”的格局,是同一套选择逻辑在不同位势、资产结构、外部约束与政策条件下的两个解。也正因为决定因素具有动态的,模式格局并非定局,算力约束的松紧、前沿差距的张缩、商业变现的成败,都会推动两国在混合结构中重新校准开放的层次与节奏,这正是研判后续竞争走势时必须紧盯的重要因素。

四、模式选择对中美产业竞争的影响

模式选择的影响远不止于企业自身的经营方式。开源闭源的分化重塑了中美在模型层的竞争规则,沿着算力产业链向上传导及应用产业向下扩散,改变了两国创新企业的生存空间与产业链的利益分配,直接影响着两国产业综合竞争力。

(一)对中美大模型产业竞争的影响

在模型层,开源对闭源形成了持续的租金挤压,竞争的天平在不同维度上向不同方向倾斜。对美国闭源阵营而言,性能接近的中国开放模型经第三方托管后以接近边际成本的价格提供服务,系统性地压缩了接口服务的利润空间。闭源厂商被迫一面以更高频的旗舰迭代维持领先,一面以轻量低价产品防守长尾市场,研发与算力开支被推向极限,市场对其巨额资本开支可持续性的担忧随之上升。对中国开源阵营而言,生态份额的扩张并未同步带来利润,模型层收入微薄,竞争收益更多以云收入、生态位与战略影响力等间接形式体现,“赢了份额、薄了利润”是必须正视的另一面。

竞争重心也开始从模型本身向标准与协议层上移。随着智能体应用的兴起,模型上下文协议(MCP)、智能体间协议(A2A)等接口标准成为新的争夺焦点。这类标准之争的实质是生态主导权之争,谁的协议成为事实标准,谁就能在智能体经

济中占据类似操作系统的核心位置^[18-19]。目前主流的智能体协议多由美国企业发起,但中国开源模型在全球的广泛部署,为国内主体参与甚至共同制定标准积累了筹码。从总体趋势看,竞争焦点正由单纯的性能比拼转向成本、延迟、可靠性和工程化能力,这对以效率见长的中国企业相对有利。两国的模式选择也在趋同,逐步收敛到“旗舰闭源探索+开放权重生态”的混合结构,分歧不再是开源或闭源的非此即彼,而在于如何把握混合结构中的重心与节奏。

(二)对中美上下游相关产业竞争的影响

向产业链上游看,模式竞争重塑了AI算力环节的利益格局与竞争窗口。模型层越是开放与商品化,产业利润越是向AI算力环节集中,以英伟达为代表的美国芯片企业成为最大受益者,美国也因此更加倚重出口管制来守住算力这一咽喉环节。然而,管制的实际效果却被开源所削弱,中国开源模型与国产芯片协同适配,使国产算力获得了规模化的真实负载和持续迭代的应用场景,软硬件相互促进,反而为国产替代打开了窗口。中国头部开源模型面向国产芯片优化发布的做法,正在把模型生态优势转化为芯片生态的牵引力。这就意味着,开源在让美国芯片企业多分利润的同时,又削弱了美国用管制遏制中国的成效,对两国而言利弊并存。

向产业链下游与整体科技产业看,两国的智能化路径出现分化。中国依托开源形成了低成本、可私有化部署的智能化供给,制造、医疗、政务、金融等行业可以在数据不出域的条件下规模化用模,叠加完备的制造业体系与超大规模市场,智能化改造的广度与速度具备独特优势,数字技术通过深化服务分工、促进产业协同提升经济效率的机理在智能化阶段进一步强化^[20]。美国则以闭源接口加软件即服务的方式推进企业级应用,在高端知识工作场景的渗透更深。放眼整体科技产业竞争,大模型的模式之争还外溢到基础软件领域:中国开源人工智能生态的壮大,带动了操作系统、数据库、开发框架等全栈开源路线的信心与人才储备,对长期受制于人的基础软件体系形成

难得的牵引。美国则力图以模型、芯片、云的整体技术栈输出巩固全球主导地位。技术栈之争正在取代单点产品之争,成为中美科技产业竞争的主战场之一。

(三)对中国追赶进程的机遇与约束

对中国而言,开源最直接的意义是改写了追赶的可行路径,其机遇体现在四个方面。一是以扩散换时间。开放权重使中国模型在全球开发者中的采用规模迅速做大,生态位的占领先于性能的全面登顶,走出了一条后发者罕见的“先得生态、再攻高点”路径。二是以生态换位势。全球下载份额的反超意味着相当一部分国际应用创新建立在中国模型之上,技术依赖结构的天平开始向有利于中国的方向移动。三是以效率对冲封锁。开源动员的分布式优化使算力管制的实际约束被部分消解,中国在受限条件下保持了与前沿可对话的位置,这本身就是对“小院高墙”策略有效性的证伪。四是以普惠厚植基础。低成本智能能力加速向中小企业与中西部地区扩散,实证研究表明开源技术的使用与企业生产率提升显著相关^[21],这为全要素生产率改善与新质生产力培育提供了微观基础。同时也应看到,人工智能作为新一代通用技术,其渗透性、替代性、协同性与创造性在推动各部门增长的同时,会带来就业结构两极化和初次分配中劳动收入份额变化^[22-23],开放权重加速扩散,使收益与冲击的分布问题更需统筹应对。

需要指出的是,开源路线对中国追赶进程也存在一定约束与风险。一是顶尖模型差距重新拉大的压力。当前沿竞赛进入算力密集驱动阶段后,美国以数倍于中国的资本开支持续扩张,2025年年底以来中美顶尖模型差距呈再度拉大之势,若该态势延续,开源阵营存在被长期锁定在跟随位置的风险。二是价值捕获的错配。开源压缩了模型层租金,利润向算力环节集中,而算力恰是中国受制于人的环节,存在“中国做大蛋糕、美国多分蛋糕”的结构性隐忧,国产算力的替代速度因此具有全局意义。三是生态优势的可逆性。美国正以开放权重产品和技术栈输出战略反向争夺开源

阵地,若中国开源生态的商业回路迟迟不能闭环、头部企业的开放承诺出现松动,份额优势可能得而复失。四是开放的安全代价。权重扩散的不可逆性使滥用风险与责任压力上升,治理能力必须与开放程度同步提升。开源为中国赢得了时间与生态。但生态优势能否转化为前沿能力与产业利润,取决于算力底座、商业制度与治理体系能否协同跟进。

(四)对智能化扩散与治理格局的影响

扩散方式的差异,最终反映为两国模型企业所面对的应用市场结构不同。中国的开放权重供给把用模门槛压到很低,数量庞大的中小企业可以本地部署、按需微调,把模型嵌入既有的生产经营范围,由此形成的海量长尾场景又反过来成为国内模型企业验证能力、沉淀数据与获取云服务收入的来源。美国闭源阵营则以接口与订阅方式服务付费能力较强的企业客户,应用集中于高端知识工作场景,单客户价值高但覆盖面窄。宽而薄的用户基础与窄而厚的付费市场,对模型企业的变现路径、迭代节奏乃至融资逻辑都构成了不同约束。

在治理与国际层面,两种模式塑造了不同的风险结构与全球角色。美国以闭源集中管控加出口管制构筑“高墙”,可控性强,但加深了全球对单点依赖与技术霸权的疑虑。中国以备案管理加分级评估探索开放条件下的安全治理,并把开放模型作为面向全球南方的公共产品供给。开放权重的不可召回性决定了治理必须跟上开放的步伐,治理框架的设计基准应当是能力分级与场景风险,而非对模式本身作制度性褒贬。若地缘竞争持续深化,全球智能生态可能走向围绕中美两套技术栈并行的格局,发展中国家的选择将成为决定生态规模的关键变量。

五、我国更好应对大模型产业竞争的策略

中美开源闭源之争没有预设的终局,真正的较量在于谁能率先把模式优势转化为生态优势,再把生态优势转化为产业优势和发展优势。我国已经在开放权重生态上赢得先手,下一步的关键是坚持开放路线不动摇,补齐算力与制度短板,用

好超大规模市场纵深,把智能化转型同扩内需、稳就业、促升级紧密结合,在高水平开放与高水平安全的动态平衡中赢得战略主动。

(一)把人工智能作为培育新质生产力的主攻方向,健全与前沿竞赛相匹配的投入机制

落实党的二十届四中全会部署,将“人工智能+”行动与扩大内需战略协同推进,以智能终端、智能体普及带动新一轮信息消费和服务消费升级,将智能化改造纳入大规模设备更新与技术改造支持范围带动有效投资,使智能经济加快成长为对冲传统动能走弱的重要增长极。把人工智能纳入中央和地方“十五五”规划重大工程和重大生产力布局,统筹全国算力、数据与应用布局,形成跨周期的稳定政策预期,引导社会资本与创新资源持续投入。

投入强度是这一主攻方向能否落地的前提。美国头部企业的人工智能资本开支已达年均数千亿美元量级,单个中国企业难以独自承担前沿竞赛成本,应构建财政引导、金融协同、市场主体相结合的多元投入体系。发挥国家级产业基金和政府引导基金作用,吸引保险资金、社保基金等长期资金进入,培育敢于支持前沿模型探索的耐心资本。用好科技创新债券、并购重组等工具拓宽企业融资渠道。加大前沿模型研发费用加计扣除和算力使用支持力度,使我国前沿探索的投入强度与追赶目标相匹配。同时应避免以有限资源简单复制美国的资本开支竞赛,在模型架构、训练方法与数据质量上形成差异化投入,把工程效率优势延伸到前沿探索环节。

(二)统筹算力能源底座建设与国产化替代,综合整治“内卷式”竞争

以全国一体化算力网为抓手优化算力基础设施布局,强化跨区域调度和绿电协同,提高存量算力利用率,防止智算中心重复投资和低水平价格竞争,把综合整治“内卷式”竞争的要求落实到算力领域。坚持开源模型与国产芯片协同适配的主线,组织龙头企业与开源社区联合开展软硬件协同攻关,缩短国产算力在主流开源生态中的适配周期,把模型生态优势转化为芯片生态牵引力,加

快摆脱对受限算力的依赖。

算力的成本结构同样需要统筹。前沿训练与规模化推理的能耗强度高,应将智算中心布局同电力供给、绿电资源相匹配,推动“东数西算”通道与新能源基地协同建设,依托能源成本优势部分抵消芯片性能差距。在需求侧,可通过算力券、公共算力平台等方式把闲置算力导向中小企业和科研机构,避免出现建得多、用得少的局面。国产芯片的适配也不能停留在能够跑通层面,需在主流开源框架和推理引擎上形成稳定可用的工具链,这是规模化替代的前提。

(三)巩固开源生态优势,以智能化赋能实体经济和中小企业

坚持模式中性和双轨并行,既巩固开放权重生态的基本盘,也为旗舰闭源探索保留空间,防止生态优势因商业闭环缺失而得而复失。加快建设国际兼容的开源许可证体系和中立的开源基金会、模型托管与评测公共平台,把开源贡献纳入科研评价和人才评聘体系。围绕云服务、行业解决方案和政府采购完善开源的商业变现机制,明确开放权重的知识产权界定与责任分配规则,从制度上解决“开源者无回报”的可持续性难题。

生态优势最终要在实体经济的用模规模中兑现。面向制造业转型升级和中小企业纾困的现实需要,依托开源模型建设行业级智能化公共服务平台,降低中小企业用模门槛,推动人工智能同制造、医疗、教育、政务等领域深度融合,整体提升全要素生产率。同步实施大规模技能再培训和就业帮扶,完善对受冲击群体的社会保障托底,健全智能时代劳动者权益保护制度,使智能化转型成为稳就业、提收入、促共富的动能而非加剧分化的风险。

(四)深化国际开放合作与全球治理参与,拓展战略回旋空间

落实“推动人工智能技术开源可及”部署,把开放模型作为造福人类的国际公共产品,面向全球南方深化人工智能能力建设合作,以低成本、可本地部署的方案帮助发展中国家弥合智能鸿沟,在扩大“朋友圈”中对冲封锁打压。支持联合国在

人工智能全球治理中发挥主渠道作用,主动参与智能体协议、评测基准、安全规范等国际标准制定,办好世界人工智能大会等交流平台,把生态规模优势逐步转化为规则话语权。

合作要有具体抓手。可依托现有开源社区和公共平台,面向发展中国家提供多语言小尺寸模型、本地化部署方案与培训课程,把能力建设做实,而非停留在倡议层面。在治理侧,推动建立中立的国际开源模型评测与安全披露机制,主动公开模型卡、能力边界与风险评估,以透明度换取信任。这既是负责任开放的应有之义,也是在两套技术栈并行的格局中争取中间地带的现实办法。

参考文献:

- [1] BRESNAHAN T F, TRAJTENBERG M. General purpose technologies: engines of growth? [J]. *Journal of econometrics*, 1995, 65(1): 83-108.
- [2] KATZ M L, SHAPIRO C. Network externalities, competition, and compatibility [J]. *The American economic review*, 1985, 75(3): 424-440.
- [3] ARTHUR W B. Competing technologies, increasing returns, and lock-in by historical events [J]. *The economic journal*, 1989, 99(394): 116-131.
- [4] DAVID P A. Clio and the economics of QWERTY [J]. *The American economic review*, 1985, 75(2): 332-337.
- [5] TEECE D J. Profiting from technological innovation: implications for integration, collaboration, licensing and public policy [J]. *Research policy*, 1986, 15(6): 285-305.
- [6] LERNER J, TIROLE J. Some simple economics of open source [J]. *The journal of industrial economics*, 2002, 50(2): 197-234.
- [7] VON HIPPEL E, VON KROGH G. Open source software and the private-collective innovation model: issues for organization science [J]. *Organization science*, 2003, 14(2): 209-223.
- [8] CHESBROUGH H. Open innovation: the new imperative for creating and profiting from technology [M]. Boston: Harvard Business School Press, 2003.
- [9] SHAPIRO C, VARIAN H R. Information rules: a strategic guide to the network economy [M]. Boston: Harvard Business School Press, 1999.
- [10] TEECE D J. Profiting from innovation in the digital economy: enabling technologies, standards, and licensing models in the wireless world [J]. *Research policy*, 2018, 47(8): 1367-1387.
- [11] CASADESUS-MASANELL R, GHEMAWAT P. Dynamic mixed duopoly: a model motivated by Linux vs. Windows [J]. *Management science*, 2006, 52(7): 1072-1084.
- [12] WEST J. How open is open enough? melding proprietary and open source platform strategies [J]. *Research policy*, 2003, 32(7): 1259-1285.
- [13] BOUDREAU K. Open platform strategies and innovation: granting access vs. devolving control [J]. *Management science*, 2010, 56(10): 1849-1872.
- [14] LUONG N. Two loops: how China's open AI strategy reinforces its industrial dominance [R]. Washington, D. C.: U. S. -China Economic and Security Review Commission, 2026.
- [15] STANFORD INSTITUTE FOR HUMAN-CENTERED ARTIFICIAL INTELLIGENCE. The 2026 AI index report [R]. Stanford: Stanford University, 2026.
- [16] LONGPRE S, AKIKI C, LUND C, et al. Economies of open intelligence: tracing power and participation in the model ecosystem [EB/OL]. (2025-11-27) [2026-06-30]. <https://arxiv.org/abs/2512.03073>.
- [17] DANIELS O J, DOHMEN H. Open models, soft power, and the spectrum of U. S. -China artificial intelligence competition [R]. Santa Monica: RAND Corporation, 2026.
- [18] ROCHET J C, TIROLE J. Platform competition in two-sided markets [J]. *Journal of the European economic association*, 2003, 1(4): 990-1029.
- [19] JACOBIDES M G, CENNAMO C, GAWER A. Towards a theory of ecosystems [J]. *Strategic management journal*, 2018, 39(8): 2255-2276.
- [20] 江小涓, 靳景. 数字技术提升经济效率: 服务分工、产业协同和数实孪生 [J]. *管理世界*, 2022(12): 9-26.
- [21] NAGLE F. Open source software and firm productivity [J]. *Management science*, 2019, 65(3): 1191-1215.
- [22] 蔡跃洲, 陈楠. 新技术革命下人工智能与高质量增长、高质量就业 [J]. *数量经济技术经济研究*, 2019, 36(5): 3-22.
- [23] 渠慎宁. 未来产业的效益创造、风险识别与投入机制设计 [J]. *改革*, 2026(3): 65-74.