

豆包2.1 Pro 跨越“质变点”，火山引擎推动 AI 生产落地

本报记者 谷月

跨越“质变点” AI 从辅助走向交付

过去两年，大模型行业经历了从“百模大战”到落地困境的过山车。许多企业投入不菲，却始终停留在“尝鲜”阶段。谭待在大会上提出了破解这一困境的关键标尺——“质变点”：模型能力只有跨越特定阈值，才能从“辅助工具”进化为“生产级劳动力”。

全球范围内，视频生成领域的质变点是 Seedance 2.0，Coding 与 Agent 领域则是 Claude Opus 4.6。而豆包2.1 Pro 的使命，就是在 Coding 和 Agent 领域超越这一质变。

评测数据支撑了这一判断。在 SWE-Pro 软件工程基准测试中，豆包2.1 Pro 优于 Claude Opus 4.6；在 OSWorld 操作系统交互基准和 MMMU-Pro 多模态理解基准中，同样位居全球前列。

这些评测并非简单的“刷题”，而是模拟真实工程环境——Terminal Bench 2.1 考察命令行交互能力，SciCode 聚焦科学计算场景下的代码生成。豆包2.1 Pro 不仅能写“Hello World”，更能解决真实工程师面临的实际问题。

最能体现这一跨越的，是大会现场展示的芯片设计 RTL 测试案例。RTL（寄存器传输级）设计是芯片设计中最核心、最复杂的环节之一，对代码准确性、逻辑严谨性和工程完整性要求极高。豆包2.1 Pro 连续运行近18小时，经历9轮迭代，不仅完成了代码生成，还跑通了仿真、测试、综合检查等完整流程。这已不是“AI 辅助写代码”，而是“AI 独立完成工程交付”。

这种能力跃迁，直接回应了市场对“AI 幻觉”和“不可靠”的担忧。正如谭待所言，真正的生产级应用，是代码真正可以上线、跑通测试，是能够处理长程复杂任务。3D 虚拟城市场景演示中，500 余个智能 Agent 同步协作、完成上千轮工具调用的场景，不再是实验室里的 Demo，而是数字孪生和智慧城市的真实预演。

6月23日，在北京举行的2026夏季FORCE原动力大会上，火山引擎集中发布了豆包大模型2.1等重磅产品。当豆包大模型日均Token调用量突破180万亿、火山引擎以49.5%的份额领跑中国公有云MaaS市场时，一个关键问题浮出水面：大模型从“炫技”走向“落地”的关键转折点上，谁才是真正的领跑者？

“豆包2.1的Coding和Agent能力已经跨过生产级可用线。”火山引擎总裁谭待指着豆包2.1 Pro在Terminal Bench 2.1、SWE-Pro等代码评测中的成绩，语气笃定。这不仅是技术的宣言，更是产业落地的号角。



成本降至 1/5 价格策略推动 AI 生产落地

技术能力是“质变点”的一翼，成本则是另一翼。纵观大模型发展史，每一次大规模商业化落地都伴随着成本的指数级下降。豆包2.1 Pro 的定价策略，堪称对国际主流模型的一次“降维打击”。

每百万Tokens输入6元、输出30元，缓存命中仅1.2元，综合使用成本较 Claude Opus 4.6降低近80%。这意味着同样一笔预

算，企业可以调用5倍的Tokens。在AI应用规模快速扩张的今天，这不仅是价格优势，更是推动AI生产落地的有力支撑。面向高频调用场景的豆包2.1 Turbo，价格进一步降至2.1 Pro的一半，大幅降低了AI应用的边际成本。火山引擎还上线了 Doubao-Seed-Evolving 版本，以每月2至4次的频率快速迭代——这种“小步快跑”的节奏，契合中国互联网企业“快速试错、敏捷迭代”的工程文化。

价格优势背后，是火山引擎在算力调度、模型压缩、推理优化等方面的深厚积累。在MaaS（模型即服务）模式下，推理成本直接决定了客户的使用意愿和业务规模，

火山引擎已将AI推理的门槛压低到了“可规模化”的水平。

多模态商用闭环 拓展实体产业边界

如果说豆包2.1 Pro 代表了AI在逻辑与代码世界的深耕，那么 Seedance 2.5 的首次亮相，则标志着多模态生成技术在通往世界模型的道路上更进一步。

Seedance 2.5 实现了多项关键突破：30秒单段原生视频直出、支持50个全模态素

材联合生成、无损局部编辑，并首创3D白模分镜预览功能。长镜头叙事让AI视频承载完整剧情；多素材约束解决品牌视觉统一性难题；3D预演大幅降低试错成本。

更为深远的是，火山引擎正在将视频生成推向实体产业。从广告电商到工业制造，从具身智能到自动驾驶，Seedance 为数据合成、场景仿真、流程演示提供全新工具底座。企业可在数字孪生环境中模拟生产流程、预演设备故障，无需在物理产线上进行昂贵试错。

Seedance 并非孤立存在。它与豆包大模型、图像模型 Seedream 5.0 Pro、音频模型 Seed-Audio 1.0 共同构成多模态矩阵。Seedream 5.0 Pro 支持交互式精准编辑、多图层分离和原生多语种文字生成，可直接输出可编辑的分层设计图；Seed-Audio 1.0 具备零样本多模态参考能力，一次生成多角色对白、背景音乐和拟音特效。从文字到图像、视频到音频，火山引擎正构建完整的“内容生成闭环”。

Agent 云原生架构 破解规模化落地难题

大模型的下一个战场是 Agent（智能体）。大模型是“大脑”，Agent 是“手脚”——自主规划、调用工具、执行任务。豆包2.1 Pro 在 Agent 能力上的跃升，为这一趋势提供了底层支撑。

面向 Agent 时代，火山引擎升级了整套 AI 云原生架构。方舟 CLI 命令行工具让开发者通过一行指令将 Agent 接入火山方舟，大幅降低开发门槛。AgentKit 与 HiAgent 3.0 提供更丰富的工具链和开发环境。ArkClaw 企业版 Agent 工作台直接面向企业级场景，支持复杂业务流程编排与执行。AI Trust 产品体系则为大模型和 Agent 在金融、制造、医疗等强监管行业的合规使用提供保障。

这些基础设施正在改变火山方舟的生态格局。大会数据显示，目前已有超过110万企业和个人使用火山方舟大模型服务，年Token调用量超过1万亿的企业达200家，半年内增长一倍。企业级AI应用通常面临决策周期长、迁移成本高等问题，半年翻倍的增长曲线，说明市场对火山引擎的认可正在加速转化为市场份额。

知
晓
读
懂
用
好

法律
政策

护航中小企业健康、高质量发展

（全国中小企业法律政策宣传月）

