

市场研究部

2025 年 11 月 27 日

谷歌 Gemini 3 与 NanoBanana Pro 发布

看好

事件

2025 年 11 月，谷歌推出新一代旗舰大模型 Gemini3 系列、专业图像生成模型 NanoBananaPro 及开发平台 Antigraity，标志着其“芯片-云-模型-应用”全栈布局的深化。

核心观点

Gemini3 依托自研第七代 TPU 芯片 Ironwood 构建的算力集群，实现单芯片 FP8 算力 4614TFLOPS、HBM 容量 192GB，集群规模达 9216 颗芯片，通过光电路交换网络达成 9.6Tb/s 互联带宽，为 AI 推理提供高性能、低能耗支撑。

Gemini3 支持百万级输入与 64K 输出，多模态基准测试全面领先：在 Humanity's Last Exam 学术推理测试中达 37.5%，Math Arena Apex 数学竞赛测试得分 23.4%，ScreenUnderstanding 能力以 72.7% 准确率居首。其 DeepThink 模式通过多步推理与验证机制，在 ARC-AGI 评测中取得 45.1% 成绩，解决复杂问题的能力显著提升。NanoBananaPro 则融合物理逻辑渲染与文本纠错技术，原生支持 2K 分辨率输出，解决图像生成领域的中英文文本渲染错误痛点，满足专业视觉创作需求。

谷歌通过 TPU 芯片、云计算基础设施与模型能力的深度协同，构建成本优势：TPUv7 每瓦性能较前代提升两倍，单位算力成本仅为英伟达高端 GPU 的五分之一。2025 年谷歌云业务订单积压达 1550 亿美元，同比增长 82%，显示市场需求强劲。Antigraity 平台整合智能体管理与多模态开发工具，支持开发者通过自然语言交互实现代码转换与系统操作，推动 AI 从工具向合作伙伴升级。

投资线索

算力基础设施：关注光模块、液冷技术及高带宽互联设备领域，受益于 AI 集群规模化部署带来的硬件升级需求。

应用场景落地：挖掘企业服务、营销科技等垂直领域机会，重点关注能整合多模态 AI 能力、优化生产效率与用户体验的场景化解决方案。

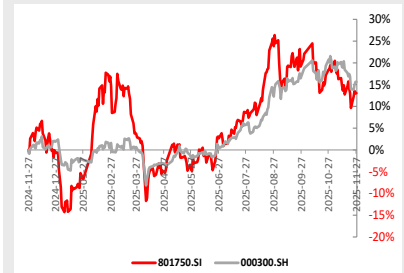
生态协同创新：紧跟头部企业资本开支节奏，布局算力硬件核心供应商，同时把握 AI Agent 技术驱动的软件开发与工具迭代机遇。

风险提示

1、技术发展不及预期；2、市场竞争加剧；3、地缘政治影响。

市场表现截至

2025.11.26



数据来源：Wind，国新证券整理

分析师：钟哲元

登记编码：S1490523030001

邮箱：zhongzheyuan@crsec.com.cn

证券研究报告

目录

一、谷歌大模型技术进阶	3
二、GEMINI 3 的技术架构与性能表现.....	3
三、NANOBANANA PRO 的图像生成与专业应用	3
四、全栈能力：从芯片到模型的协同效应	4
五、商业化进展与生态整合	4
六、投资线索	5
七、风险提示	5

一、谷歌大模型技术进阶

2025 年 11 月，谷歌在人工智能领域掀起技术革新浪潮，接连推出新一代旗舰大模型 Gemini 3 系列、专业图像生成模型 NanoBanana Pro（官方命名为 Gemini 3 Pro Image），并同步发布创新性开发平台 Antigravity。这一系列发布不仅彰显了谷歌在多模态理解、推理与 Agent 技术领域的全面突破，更凸显其从底层算力基础设施到上层应用生态的全栈布局战略，标志着 AI 技术正式迈入“自主推理+多模态执行”的新阶段。尤为关键的是，Gemini 3 的研发与训练全程依托谷歌自研第七代 TPU 芯片 Ironwood (TPUv7) 构建的大规模集群，该芯片专为推理场景优化，单芯片 FP8 峰值算力高达 4614 TFLOPS，HBM 容量达 192GB，芯片互连带宽 1.2Tbps，集群规模可扩展至 9216 颗，通过光电路交换（OCS）网络实现 9.6Tb/s 的超高互联带宽，既为模型提供了坚实的底层算力支撑，也印证了谷歌“芯片-云-模型-应用”全链条协同的技术路线优势。

二、Gemini 3 的技术架构与性能表现

作为本次发布的核心旗舰模型，Gemini 3 在技术架构与性能表现上实现了全方位提升，支持最高约 100 万 tokens 的输入上下文窗口与最大 64K tokens 的输出容量，能够高效处理文本、图像、视频、音频乃至代码等多种模态数据，在几乎所有主流 AI 基准测试中均显著超越前代产品及行业竞品。在学术推理领域，其无工具辅助状态下在 Humanity's Last Exam 测试中得分 37.5%，GPQA Diamond 测试中更是达到 91.9% 的高分；在 Math Arena Apex 测试中以 23.4% 的成绩拿下最新 SOTA，较前代实现翻倍提升。尤为突出的是其 ScreenUnderstanding 能力，在 ScreenSpot-Pro 评测基准中获得 72.7% 的高分，大幅领先 Claude Sonnet 4.5 (36.2%) 和 GPT 5.1 (3.5%)，能够精准解读结构化与非结构化文档，为 AI 直接操作图形界面（GUI）奠定了基础。此外，Gemini 3 创新性引入 DeepThink 模式，通过 Thought Signatures（思维签名）与 Thinking Levels（思考等级）的产品化封装，让模型在输出结果前能够进行多步推理、反思与验证，开发者可根据任务复杂度灵活配置思考时长，在 ARC-AGI 评测中更是取得 45.1% 的优异成绩，展现出解决新颖复杂挑战的强大能力。其智能体能力通过 Antigravity 开发平台深度集成，实现了代码编辑器与终端的直接访问，支持复杂软件任务的自主规划、多步骤工具调用与高效执行，在 LMArena 大模型竞技场上以 1501 Elo 的突破性评分位列公开模型首位。

三、NanoBanana Pro 的图像生成与专业应用

NanoBanana Pro 作为谷歌面向专业视觉内容创作的重磅产品，融合了 Gemini 3 的先进推理架构与世界模型技术，在图像生成领域实现了物理逻辑与文本

渲染的双重突破。该模型超越传统扩散模型的局限，在渲染场景前会进行专业的场景规划，能够生成具备物理精确光照效果的高质量图像，原生支持 2K 分辨率输出，更完美解决了长期困扰图像生成领域的文本渲染错误痛点，支持中英文等多语言文本的清晰呈现，用户可自定义字体类型或模拟手写风格，为海报制作、复杂图表设计、精细产品模型展示等专业场景提供了可靠支撑。依托 Gemini 3 对现实世界的深刻理解与强大推理能力，NanoBanana Pro 能够生成精准、细致且富含专业知识的图像内容，不仅可以为图片添加专业注释、将复杂数据转化为直观信息图表，还能实现手写笔记到标准化图表的转换，在人物一致性控制上最多可保持 5 人特征稳定，同时支持局部编辑与 4K 分辨率拓展输出，为广告营销、科研可视化、教育科普等多个领域提供工作室级的视觉生成能力。

四、全栈能力：从芯片到模型的协同效应

谷歌此次技术发布的核心竞争力，源于其从芯片到应用的全栈协同效应，这种协同体系构建了难以复制的技术护城河。在芯片层面，TPU 系列芯片经过多代迭代已形成成熟的训推一体 ASIC 体系，TPUv7 采用 HBM3e 显存，带宽达 7.4TB/s，每瓦性能较前代提升两倍，在实现相近算力水平时，成本仅为英伟达高端 GPU 的五分之一，有效降低了模型训练与推理的整体拥有成本（TCO）。在云端部署方面，谷歌持续加大 AI 基础设施投入，将 2025 年资本支出上修至 910-930 亿美元，并预计 2026 年资本支出将“显著增加”，强劲的投入支撑起庞大的云业务规模，2025 年三季报披露其云业务订单积压已达 1550 亿美元，环比增长 46%，同比增长 82%，展现出旺盛的市场需求。模型层面，Gemini 3 通过与 TPU 芯片的深度协同优化，实现了算力效率与模型性能的同步提升，而 Antigravity 开发平台则作为连接模型与开发者的关键枢纽，提供了 AI 驱动的 IDE 编辑器视图与多智能体管理界面，让智能体拥有专属工作空间，支持“人+Agent”的协同开发模式，将 AI 辅助从工具升级为积极的合作伙伴。这种全栈协同模式，使得谷歌能够在技术迭代速度、成本控制与应用落地效率上形成复利优势，持续巩固行业领先地位。

五、商业化进展与生态整合

在商业化进展与生态整合方面，谷歌已构建起覆盖 C 端用户、企业客户与开发者群体的完整生态闭环。面向 C 端市场，Gemini App 作为独立 AI 助手已积累 6.5 亿全球月活跃用户，其 AI Overview 功能深度集成谷歌搜索，直接覆盖数十亿次搜索请求，通过生成式 UI 技术实现“搜索即应用”的全新体验，同时 AI 能力已全面融入 Google Workspace（Gmail/Docs/Sheets）等核心办公产品，未来将依托订阅服务与广告变现模式实现规模化商业价值转化。面向企业客户与开发者，谷歌开放了 Gemini 3 的 API 接口，截至 2025 年 Q3，模型每分钟处理 Token 数量已达 70 亿，Android Studio、Firebase、Gemini CLI 等开发工具已全部接入该模型，开

发者可通过自然语言交互方式快速实现 Shell 命令转换与系统级操作，大幅降低 AI 应用开发门槛。商业合作层面，谷歌与 AI 公司 Anthropic 签署了 100 万颗定制 TPU 芯片的采购协议，标志着 TPU 芯片正式进入商业级服务市场，进一步拓展了其算力生态的商业化边界。这种多维度的生态整合，使得 Gemini 3 的技术能力能够快速渗透到消费、办公、开发、企业服务等多个场景，加速了 AI 技术的商业化落地进程。

六、投资线索

谷歌凭借“芯片-模型-云”全栈闭环构建了独特竞争优势，而阿里则通过开源生态与企业级市场渗透形成差异化路径。投资布局应把握两大主线：一是紧跟巨头资本开支周期，布局算力硬件核心供应商；二是挖掘垂直场景中具备落地能力的 AI 应用企业。谷歌生态核心参与者及 AI Agent 领域技术领先的公司，将在这一轮产业变革中占据有利地位，建议重点关注光模块、液冷、OCS 等硬件环节，同时把握企业服务、营销科技等软件应用领域的价值重估机遇。从硬件基础设施视角看，谷歌的算力扩张为上游供应链带来明确增长动能，其大规模采用 OCS 光交换架构，直接拉动高速光模块需求（中际旭创、新易盛、天孚通信）；液冷技术作为算力集群稳定运行的关键保障，使相关供应商（英维克）迎来发展机遇；OCS 交换机供应链（德科立、腾景科技）也将受益于谷歌 TPU 集群的持续部署，随着谷歌 2026 年 TPU 出货量有望大幅突破，这些硬件环节将进入高速增长周期。在软件与应用层面，投资机会集中于能够有效整合先进 AI 能力的场景化企业，企业服务领域（光云科技、税友股份）可通过接入 Gemini 3 API 接口提升产品智能化水平；营销科技公司（因赛集团）则能利用 NanoBanana Pro 的图像生成能力优化内容创作效率。

七、风险提示

- 1、技术发展不及预期；
- 2、市场竞争加剧；
- 3、地缘政治影响。

投资评级定义

公司评级		行业评级	
强烈推荐	预期未来 6 个月内股价相对市场基准指数升幅在 15%以上	看好	预期未来 6 个月内行业指数优于市场指数 5%以上
推荐	预期未来 6 个月内股价相对市场基准指数升幅在 5%到 15%	中性	预期未来 6 个月内行业指数相对市场指数持平
中性	预期未来 6 个月内股价相对市场基准指数变动在-5%到 5%内	看淡	预期未来 6 个月内行业指数弱于市场指数 5%以上
卖出	预期未来 6 个月内股价相对市场基准指数跌幅在 15%以上		

免责声明

钟哲元，在此声明，本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。

本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿等。国新证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，以下简称本公司）已在知晓范围内按照相关法律规定履行披露义务。本公司的资产管理和证券自营部门以及其他投资业务部门可能独立做出与本报告中的意见和建议不一致的投资决策。本报告仅提供给本公司客户有偿使用。

本公司不会因接收人收到本报告而视其为客户。本公司会授权相关媒体刊登研究报告，但相关媒体客户并不视为本公司客户。本报告版权归本公司所有。未获得本公司书面授权，任何人不得对本报告进行任何形式的发布、复制、传播，不得以任何形式侵害该报告版权及所有相关权利。

本报告中的信息、建议等仅供本公司客户参考之用，不构成所述证券买卖的出价或征价。本报告并未考虑到客户的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时可就研究报告相关问题咨询本公司的投资顾问。本公司市场研究部及其分析师认为本报告所载资料来源可靠，但本公司对这些信息的准确性和完整性均不作任何保证，也不承担任何投资者因使用本报告而产生的任何责任。本公司及其关联方可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务，敬请投资者注意可能存在的利益冲突及由此造成的对本报告客观性的影响。

国新证券股份有限公司市场研究部

地址：北京市朝阳区朝阳门北大街 18 号中国人保寿险大厦 11 层（100020）

传真：010-85556155 网址：www.crsec.com.cn