

近期英伟达 CEO 黄仁勋在最新深度访谈中明确提出，Token 已成为全新商品，AI 计算机是生产 Token 的工厂，未来 AI 计算占 GDP 比重将实现百倍级跃升。2026 年以来，全球 AI 产业迈入规模化商用新阶段，Token 已从大模型的基础技术单元，升级为可计量、可定价、可交易的商业价值载体，Token 经济迎来爆发拐点。2026 年 3 月我国日均 Token 调用量突破 140 万亿，两年增长超千倍；全球 Token 周调用量年增速超 800%。Token 经济解决了 AI 产业商业化模糊的长期痛点，推动大模型厂商从“卖能力”的定制化模式转向“卖用量”的标准化变现模式，重构了 AI 全产业链商业逻辑。当前英伟达、阿里等全球科技巨头已完成核心战略卡位，在全球 Token 需求持续爆发的背景下，投资者应该关注国内市场哪些投资机会？

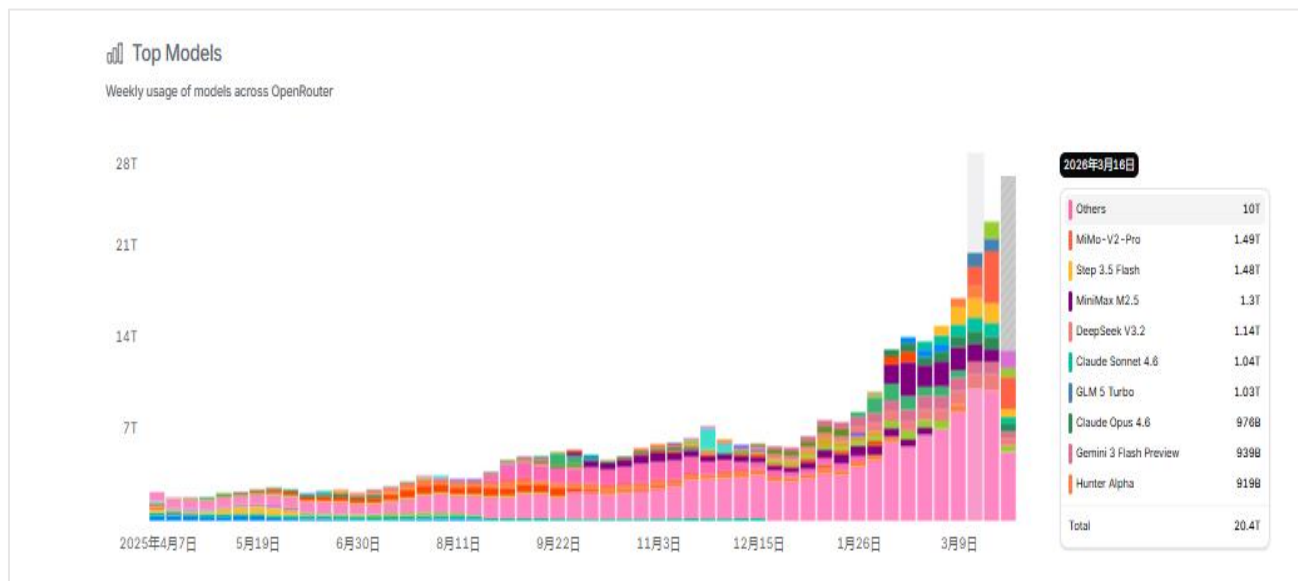
Token 经济崛起：AI 产业新周期，投资者如何把握时代红利？

2026 年以来，随着全球 AI 产业从技术研发迈入规模化商用阶段，Token（官方中文定名“词元”）已从 AI 大模型处理信息的最小技术单元，升级为可计量、可定价、可交易的全新商业价值载体，Token 经济正式迎来爆发拐点。国家数据局数据显示，截至 2026 年 3 月，我国日均 Token 调用量已突破 140 万亿，较 2024 年初实现两年超千倍增长；全球 Token 周调用量亦从 2025 年 3 月的 2.03 万亿飙升至 2026 年 3 月的 20.4 万亿，一年增长超 8 倍。

Token 经济的核心价值，在于彻底破解了 AI 产业长期面临的商业化模糊痛点，推动大模型企业从“卖能力”的定制化服务模式，转向“卖用量”的标准化、规模化变现模式，重构了 AI 产业全链条的商业逻辑。全球科技巨头已率先完成战略卡位：英伟达 CEO 黄仁勋明确提出“Token 是全新商品，AI 计算机已成为生产 Token 的工厂”，断言 AI 计算占 GDP 比重将翻百倍；阿里巴巴-SW(9988.HK)正式成立 ATH 事业群，以“创造 Token、输送 Token、应用 Token”为核心目标，完成全产业链布局。在全球 AI 行业 Tokens 需求快速提升的同时，会给国内市场的投资者带来怎样的投资机会？



图一：全球主要模型 Token 调用量增长情况



资料来源：OpenRouter 平台数据

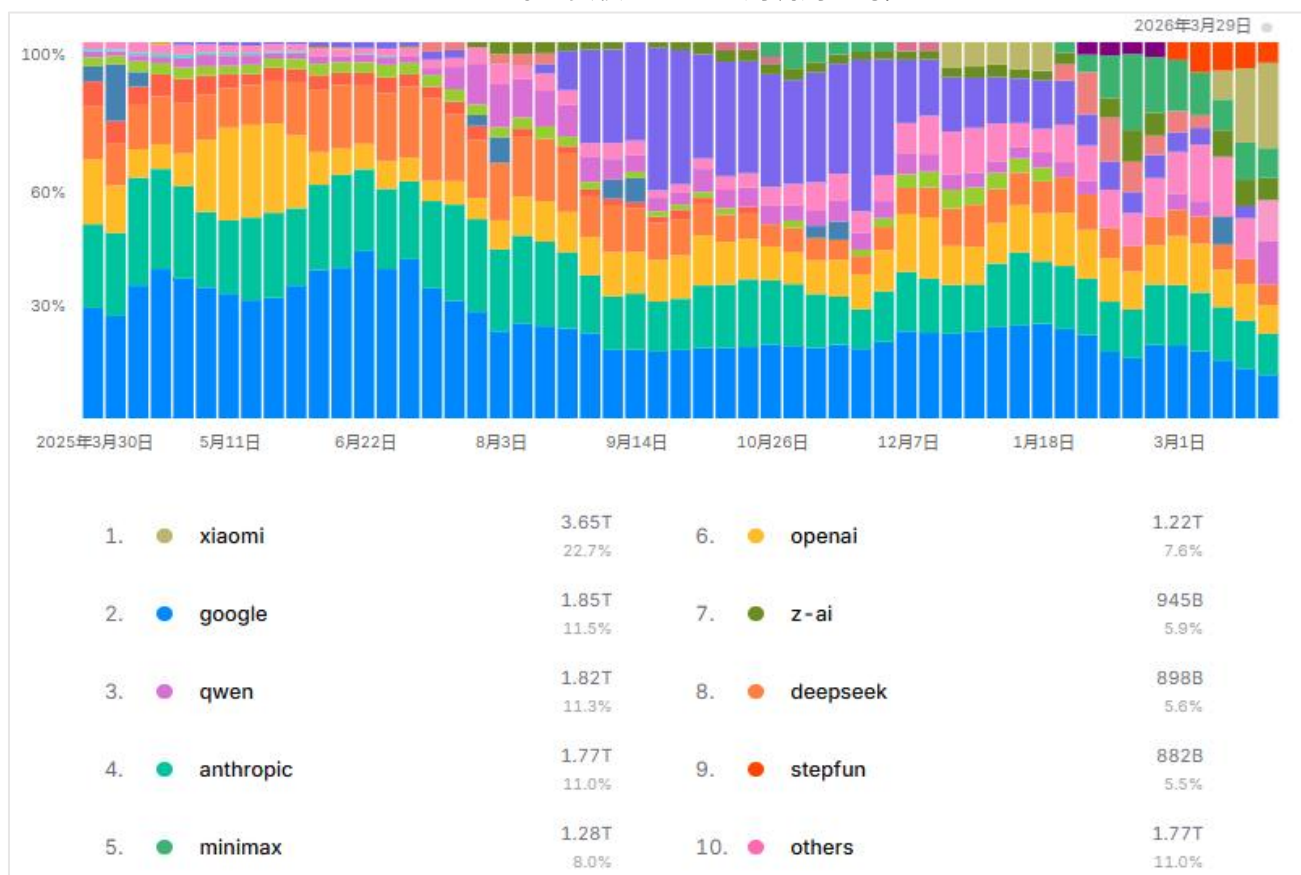
➤ AI 驱动 Token 经济兴起，未来增长潜力显著

2026 年 3 月 25 日，全国科学技术名词审定委员会正式发布公告，将人工智能领域 Token 的中文名定为“词元”，明确其是“智能设备中信息存储、处理和交换的具有一定语义的基本符号单元，是大模型处理和交换信息的最小单位”微博。国家数据局进一步明确，Token 具备可计量、可定价、可交易三大核心特征，是智能时代的价值锚点，也是连接技术供给与商业需求的核心结算单位。从产业视角看，Token 经济是以 Token 为价值标尺与流通媒介，围绕其“生产-调度-消耗-定价-交易”全生命周期形成的新型 AI 产业经济形态，其底层逻辑可概括为“Token=算力=能源=价值”，本质是将 AI 智能能力转化为可标准化交易的大宗商品。

Token 经济的崛起，核心驱动力来自 AI 推理需求的指数级爆发，这一增长动能集中体现在调用规模、应用场景、商业化模式三大核心维度。在调用规模上，国内 AI 日均 Token 调用量从 2024 年初的 1000 亿，跃升至 2025 年底的 100 万亿，2026 年 3 月进一步突破 140 万亿，两年实现超 1400 倍的增长；全球市场同样维持高景气度，OpenRouter 平台数据显示，2026 年 3 月中旬全球模型周调用量已达 20.4 万亿 Token，环比增速超 20%，其中中国模型周调用量连续多周超越美国，实现历史性赶超。



图二：全球主要模型 Token 周调用量对比



资料来源：OpenRouter 平台数据

在应用场景上，Token 的使用边界实现全维度渗透，AI Agent 规模化落地、企业级智能体 7×24 小时不间断运行、多模态内容生成集中爆发，推动单用户 Token 消耗量较传统对话场景提升 10-50 倍；除文本核心场景外，视觉、音频、视频等多模态 Token 需求快速扩容，1 张 1024×1024 高清图片约消耗 85 个视觉 Token，1 分钟高清视频便可消耗数千至数万个 Token，持续打开行业增长天花板。在商业化层面，Token 经济的商业模式已完成全面验证，按 Token 计费已成为全球 AI 服务的主流商业模式，彻底解决了 AI 能力难以量化的行业痛点，助力大模型企业实现规模化盈利，国内部分模型企业甚至出现“20 天收入超越 2025 年全年”的业绩爆发，充分印证了 Token 经济的商业可行性与长期增长潜力。



图三：AI agent 不同应用等级的特征

等级	tokens 特征	应用特征	典型场景
萌芽级	以输入 tokens 为主，输出 tokens 多为简单文本（如短回答、单轮对话）	仅在非核心场景零星试用，未接入企业系统，无标准化流程	中小企业市场部用 AI 写朋友圈文案 HR 用 AI 简历筛选（单轮提问）
普及级	输入 tokens 来源开始多样化（如对接部分业务系统 API），输出 tokens 涉及结构化内容（如报表摘要、多轮对话）	1-3 个非核心业务场景实现规模化应用，但未覆盖核心业务流程，tokens 消耗集中在单一部门	电商企业智能客服日均处理 10 万次咨询，制造业用智能体分析设备日志
融合级	输入 tokens 全面对接企业核心系统（如 CRM、ERP、生产系统），输出 tokens 直接影响业务决策多模态 token 占比提升	融入 2-5 个核心业务场景（如金融风控、供应链调度、精准营销），形成标准化的数据输入—AI 处理—业务执行“闭环”，tokens 消耗与业务指标强关联	银行数据分析智能体实时分析用户交易数据（日均 100 万条，每条转化为 100tokens），年 tokens 量约 3.6 亿
驱动级	tokens 来源覆盖企业全链路，输出 tokens 支撑端到端业务自动化，多模态 tokens 占比超 50%	AI 成为业务运营的核心引擎，80%以上业务场景依赖 AI 决策，tokens 消耗与企业营收、成本等核心指标直接挂钩，有专门的 AI 平台管理 tokens 分配与模型优化	大型互联网平台数据分析智能体分析用户行为数据（日 1 亿次交互，每天交互 100 tokens），年 tokens 量约 36 亿
原生级	tokens 成为企业“业务货币”——不仅是 AI 处理的单位，更是业务流程的核心载体（如用 tokens 衡量用户需求、产品功能），输出 tokens 可直接生成新业务（如 AI 自动设计产品、生成服务方案）	企业商业模式基于 AI 原生设计，tokens 消耗效率（如“每 100 万 tokens 的营收”）成为核心竞争力，AI 模型迭代与业务创新深度绑定	大型 AI 原生企业（如智能驾驶公司、生成式 AI 平台），日均处理亿+多模态交互，年 tokens 量超 360 亿

资料来源：英伟达 GTC 大会，Wind

Token 的价值内涵并非一成不变，其先后经历了三个核心演进阶段，最终完成了从底层技术单元到数字经济核心价值载体的本质跃迁。在最初的技术单元阶段，Token 是大模型完成文本分词、语义理解、推理生成的最小处理单位，是 AI 技术体系稳定运行的底层基础；随着行业商业化进程加快，Token 进入计费标准阶段，成为 AI 服务商业化的核心计量单元，按 Token 消耗量定价的“用多少付多少”标准化模式，类比水电煤的公用事业计费逻辑，大幅降低了 AI 服务的使用门槛，推动行业实现规模化普及；而在 AI 推理需求全面爆发的当下，Token 已进阶至价值载体阶段，逐步成为数字经济的核心价值流转工具，成功打通“电力-算力-智能-价值”的全链路转化，成长为 AI 时代可全球流通、可分层定价的全新大宗商品。

* TOKEN 经济如何重塑 AI 产业商业逻辑？

Token 经济的崛起，并非简单的计费模式创新，而是对 AI 产业全链条商业逻辑的底层重构，这种重构核心体现在商业模式迭代、产业竞争逻辑重塑、全链路价值分配体系重建三大维度。在 Token 经济成



熟之前，AI 大模型企业的商业化主要依赖定制化解决方案、私有化部署、固定额度订阅制三类模式，本质是“卖能力”，普遍面临技术能力难以量化、服务定价无统一标准、规模化复制难度高、盈利天花板明显的核心痛点。而 Token 经济推动 AI 产业商业模式实现了根本性转变，AI 企业的核心收入来源从一次性的能力交付，转向持续性的 Token 用量消耗，这一模式不仅以 Token 为统一计量标准消除了不同场景、不同客户的服务定价差异，实现了 AI 服务的工业化、规模化输出，更将项目制一次性收入转为用户持续使用带来的流水型收入，形成稳定的现金流闭环，大幅降低了企业经营波动。同时，该模式还将 AI 服务的覆盖边界从服务头部大客户的小众模式，拓展至覆盖中小微企业、个人开发者的普惠模式，全面打开了市场天花板，更让企业的算力投入、电力成本与 Token 产出直接对应，收入规模与 Token 消耗量直接绑定，最终形成了“投入-产出-盈利”的清晰商业闭环。

Token 经济同时彻底改变了 AI 产业的竞争逻辑，行业竞争的核心指标从比拼模型参数规模、榜单跑分，转向了优化单 Token 生产成本、提升每瓦电力 Token 产出率，谁能以更低的成本、更高的能效生产 Token，谁就掌握了行业定价权与核心竞争力。这一转变催生了两大清晰的产业趋势，其一，能源禀赋成为 AI 企业的核心竞争壁垒，电力成本占 Token 生产成本的 70%，电价直接决定了 Token 的底价，中国西部新能源电价低至 0.3 元/度，仅为欧美的 1/3 至 1/5，使得国产 Token 生产成本仅为海外价格的 1/6 至 1/10，形成了碾压级成本优势，也推动“Token 出海”成为中国 AI 产业的新增长曲线；其二，效率优化成为行业技术研发的核心方向，行业研发重心从“做大模型”转向“做高效模型”，MoE 稀疏架构、KV Cache 缓存技术、液冷散热、国产芯片能效优化等技术的迭代，核心目标均是提升 Token 生产效率，正如英伟达的 AI 系统在“每瓦 Token 数”指标上领先竞争对手一个数量级，直接决定了其在 Token 生产环节的核心话语权。

除此之外，Token 经济还为 AI 产业构建了一套全新的产业价值闭环，彻底重建了全链路的价值分配体系。在这套闭环中，电力是核心原料，AI 芯片是生产发动机，智算中心是 Token 工厂，大模型是 Token 炼化环节，云服务是流通渠道，AI 应用是最终消费场景，Token 成为全产业链价值分配的核心标尺，每个环节的价值贡献，都可通过其对 Token 生产、流通、增值的贡献来量化，打破了此前 AI 产业“算力厂商赚走大部分利润，应用厂商难以盈利”的失衡格局。与此同时，Token 经济还推动 AI 产业从“技术驱动”转向“需求驱动”，模型研发、算力布局不再盲目追求参数规模，而是围绕下游场景的 Token 需求进行精准匹配，实现了“需求-生产-供给”的正向循环，推动 AI 产业真正从技术探索阶段，迈入工业化、产业化的成熟阶段。

➤ AI TOKEN：算力与模型巨头的“新石油”战争

AI Token 已从大模型文本处理的最小语义与计算单元，升级为 AI 时代兼具生产资料、计价标尺、商业模式载体三重核心属性的“数字硬通货”，2026 年行业正式完成从“训练驱动”向“推理驱动”的历史性转折，推理算力占 AI 总算力投入的比例已升至 70%，全球竞争的核心已从模型参数规模竞赛，全面转向 Token 全产业链的生产力、定价权、生态控制权的系统性战争，Token 的生产效率、成本控



制与场景渗透能力，直接决定企业的核心竞争力与行业话语权。

底层算力层的竞争，是 AI Token 经济的根基之战，核心围绕单 Token 生产成本、能效比、吞吐量三大核心指标展开，本质是争夺 Token 生产的规则制定权与硬件壁垒。英伟达是这一赛道的绝对主导者，其在 2026 年 GTC 大会上正式提出“Token 工厂经济学”，首次将数据中心定义为“Token 工厂”，并建立了从免费层到 150 美元/百万 Token 的五级 Token 价值分层体系，重新定义了行业的价值评估标准。硬件层面，英伟达推出 Blackwell 与 Vera Rubin 两代核心架构，通过整合 200 亿美元收购的 Groq 公司 LPU 语言处理单元，形成“GPU 负责高吞吐、LPU 负责低延迟”的混合推理架构，宣称将单位 Token 成本降至上一代 Hopper 平台的 1/10，每瓦特推理性能提升 10 倍，同时依托 CUDA 生态垄断全球 90%以上的高端 AI 训练与推理算力，牢牢掌控 Token 生产的核心硬件标准，竞争对手上，AMD、英特尔、华为昇腾等厂商分别抢占中低端市场与国产替代赛道，各大云厂商也加速落地自研芯片，中国更凭借绿电产业链优势，从能源端构建了 Token 生产的系统性成本壁垒。

图四：AI 工厂 Token 成为标准化产出



资料来源：GTC 大会

模型层与定价权是产业链的核心利润之战，行业已形成清晰的分层竞争格局。美系头部厂商牢牢掌控高端 Token 市场的定价话语权，OpenAI 通过 GPT-4o、GPT-5 的模型能力分层，建立了行业通用的 Token 价格锚点，其高端模型 Token 定价最高达 150 美元/百万 Token，推动行业形成“为效果付费”的溢价逻辑；Anthropic 的 Claude 3 系列主打长上下文窗口优势，在专业长文本场景形成差异化定价能力；谷歌则通过 Turbo Quant 无损压缩技术，将 KV Cache 压缩至 3-3.5bit，实现 6 倍以上内存压缩与 8 倍注意力计算加速，在多模态场景实现 Token 成本的极致优化，与 OpenAI 形成正面抗衡。

中国厂商则实现了集群式崛起与定价体系的重构，2026 年 3 月全球 AI 模型 API 聚合平台 OpenRouter 数据显示，中国 AI 大模型周调用量达 7.36 万亿 Token，连续三周超越美国，全球前五名 AI 大模型中



中国占据四席，彻底打破了美国厂商的长期垄断。国产模型通过 MoE 混合专家架构、FP8 混合精度量化、KV 缓存优化等工程创新，实现单 Token 成本的断崖式下降，MiniMax M2.5 标准测试的 Token 消耗仅为 Claude Opus 的 36%，运行成本仅为海外头部模型的 1/40，主流国产模型 API 输入价格已全面进入 0.5 元/百万 Token 时代，仅为海外同性能模型的 1/10-1/20。值得注意的是，行业竞争已从单纯的价格战转向价值分层，头部国产厂商如智谱 AI 凭借 GLM-5 模型在编程等专业场景的性能突破，2026 年一季度两次上调 Token 价格，累计涨幅达 83%，而 API 调用量不降反升，印证了“模型智能上界决定定价权，Token 消耗规模决定价值体量”的行业核心逻辑，厂商纷纷通过涨价过滤非刚性测试需求，聚焦高价值企业客户，完成从“流量补贴”向“价值筛选”的商业化转型。

➤ 投资者应该关注哪些机会？

AI 时代，Token 的崛起推动计算的本质发生了根本性跃迁——从过去由人类预先录入、计算机负责检索匹配的“存储系统”，彻底升级为具备上下文感知与自主生成能力的“生成系统”，计算机在实体经济中的角色也随之完成颠覆性重构。英伟达创始人黄仁勋对此有精准的产业定义：他将过去以文件存储为核心功能的传统计算机，比作无法直接创造高额收益的“仓库”；当下的 AI 计算机，已转型为与企业收入创造直接挂钩的“Token 工厂”，AI 代工厂生产的 Token，已成为可精细化分层、标准化定价的全新数字商品。黄仁勋进一步指出，Token 对不同受众具备差异化的核心价值，目前已形成类似 iPhone 的免费、中端、高端产品分层体系，“市场愿意为每百万 Token 支付 1000 美元的应用场景已近在眼前，这只是时间问题，而非可能性问题”。基于“Token 工厂”的全新商业模式，计算设备彻底完成了从传统成本中心向利润中心的跨越，黄仁勋也对这一产业趋势给出了笃定的宏观推演：“生产力的跃升将推动全球 GDP 加速增长，我完全确信，未来计算在全球 GDP 中占比将达到此前的 100 倍。”在此背景下，Token 消费的爆发式增长，也催生了市场对其背后核心投资机会的深度挖掘。

投资者可优先关注 AI 算力产业链主线，这是 Token 经济最具确定性的“卖铲子”赛道，也是 AI 产业爆发的底层根基。Token 的本质是算力与能源的转化产物，算力供给直接决定 Token 产能，需求刚性极强。重点关注国产通用 AI 算力核心标的，直接决定 Token 生产的能效与成本，深度受益于推理算力需求爆发，例如可配置内存接口芯片龙头澜起科技(6809.HK)、光通信核心标的剑桥科技(6166.HK)与长飞光纤光缆(6869.HK)，以及晶圆制造龙头中芯国际(0981.HK)、算力基础设施运营商万国数据-SW(9698.HK)，这类标的直接绑定 Token 产能扩张的刚性需求，业绩兑现确定性强，是稳健配置的核心底仓。

中期成长聚焦 AI 软件与大模型主线，这是 Token 经济的定价核心与价值分配枢纽，马太效应极为显著。大模型是 Token 的核心“发行方”，头部厂商直接掌握行业 Token 定价标准，可凭借规模化生产形成“成本下降-份额提升”的正向飞轮。通用大模型领域重点关注 MiniMax-W(0100.HK)、商汤-W(0020.HK)，这类厂商已实现按 Token 计费的规模化 API 商业化，量价齐升逻辑明确；垂类 AI 赛道可布局金山软件(3888.HK)、金蝶国际(0268.HK)，这类企业深耕办公、企业 ERP、智能体等高价值



场景，深度绑定企业核心业务流程，能锁定长期稳定的 Token 消耗需求，成长弹性充足。

稳健配置优先选择全产业链布局的互联网大厂主线，这类标的兼具稳定现金流与 AI 成长确定性，是港股 AI 配置的压舱石。重点关注腾讯控股(0700.HK)、阿里巴巴-SW(9988.HK)，二者分别依托混元大模型+全场景生态、ATH 事业群 Token 全链路布局，完成了 Token “创造-输送-应用”的完整闭环，在 Token 定价、产能供给、场景落地全环节均具备不可替代的壁垒；同时可关注快手-W(1024.HK)、美团-W(3690.HK)等具备丰富落地场景的平台型企业，其 C 端与 B 端生态可持续释放规模化的 Token 消耗潜力。

免責聲明：本報內容所提供資料所述或與其相關的任何投資或潛在交易，均受限於閣下司法轄區適用的法律及監管規定，而閣下須單獨就遵守該等法律及監管規定負責。本報內容僅供參考，不構成任何投資建議。本公司對所提供的財經資訊已力求準確，但對其中全部或部分內容的準確性、完整性或有效性，不承擔任何責任或提供任何形式保證。如有錯失遺漏，本公司恕不負責。另請注意證券與虛擬資產價格可升可跌，尤其虛擬資產的風險極高，投資者應對有關產品保持審慎及自行承擔投資風險。

