AI超级周期

智能体Infra发展趋势

周震刚

IDC中国研究副总裁

全球 AI 价值兑现提速，Agent 成为新增长驱动力

AI创造全球经济价值

22.5万亿美元

2031年前
AI累计创造的经济价值

远超此前任何一代通用技术

Agent 成为新增长驱动力

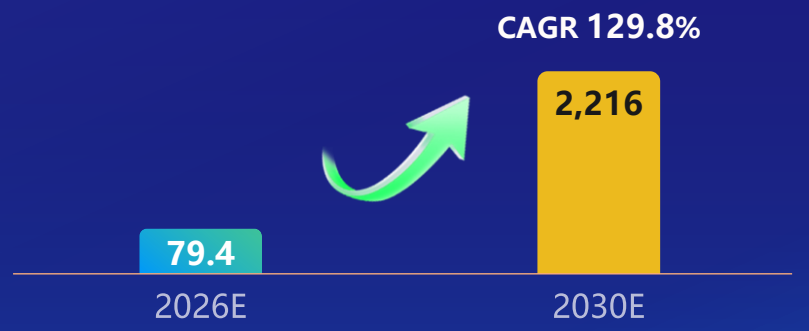
Agent成为核心驱动力

20% 智能体推动企业人工智能采购预算进一步增长

多Agent协同成主流

70% 2026年平均每个企业计划部署19.65个Agent；70%的组织将采用复合AI（Composite AI）

全球活跃Agent数量（百万个）



Token消耗量：0.026Peta（2026）→ 152,667Peta（2030）

AI已从“生成”迈入“执行”阶段：Token是成本核心，Agent是价值核心 —— 产出不再是一次性内容交付，而是持续的任务执行与结果反馈

看得见的 Agent 应用，看不见的算力基础设施

算力需求公式

Token 消耗 =



任务执行次数

- **2800亿次**（2026）→ **415万亿次**（2030）
- 中国：**178亿次**（2026）→ **39.41万亿次**（2030）

×



单任务Token消耗

- 数十Token → 数十万Token，增长数百倍

×



多智能体协同放大系数

- 一个用户任务 → 多个智能体子任务，推高任务执行次数
- 智能体间上下文传递与结果汇总延长任务链路，推高单任务Token消耗

Agent Infra：重构 AI 基础设施能力边界

Agent不是"一次调用即结束的请求"，而是**长期运行、持续交互的进程**

维度	过去	Agent Infra
需求特征	阶段性负载	**生产型负载：长期稳定、实时响应、支持高并发**
竞争焦点	芯片峰值算力	**系统有效算力：能、算、存、网低损耗协同**
运行重心	拥有资源	**高效使用资源：调度、数据供给、Token运营**
部署形态	集中式智算中心	**云—边—端协同：按时延、成本、数据安全配置**
价值实现	资源供给	**智能能力服务：买模型能力与Token产出**
行业配置	统一标准	**场景化组合：依据业务特征灵活匹配部署方案**

