

科技互联网周报 35 周 2026. 08. 24-08. 30

英伟达Q2营收962.2亿美元
纳德拉提出 “Token资本”

AI商业版

媒介360

本周热点索引 | 01

01

开源入口

英伟达拟129亿美元收购 Hugging Face，开源分发入口被重估

02

算力合约

Anthropic签下450亿美元算力合同，提前锁定Vera Rubin时代

03

资本市场

Anthropic最早9月底IPO，AGI商业模式迎压力测试

04

AGI定义

黄仁勋重写AGI标准：Token能盈利，才是经营意义上的智能

05

财报信号

英伟达二季度营收962.2亿美元，AI算力需求继续爆发

06

HBM供应

SK海力士美国HBM4E工厂计划2029年量产，内存成为AI战略资源

07

内存墙

美光警告“内存墙”：带宽瓶颈正在卡住大模型命脉

08

企业计费

ChatGPT Business开放Premium席位，企业AI进入用量分层时代

09

模型财务

DeepSeek前7个月营收4.75亿元，API毛利率达82.9%

10

AI出海

MiniMax 8月ARR突破8亿美元，ToB与海外收入成为增长主轴

本周热点索引 | 02

11

办公AI

千问办公上线**Qwen3.8-Flash**，95%日常任务走向低成本快模型

12

自研芯片

OpenAI公布**Jalapeño**芯片测试，模型公司加速软硬一体

13

端侧算力

苹果发布**M6**与**M5 Ultra**，**Mac**产品线全面押注本地AI

14

消费Agent

Meta计划推出**Hatch**，消费级AI智能体开始订阅化

15

办公入口

字节整合**TRAE**与扣子，统一办公AI品牌“豆包工作”

16

视频生成

阿里**Wan3.0**上线：30秒视频生成与文档输入进入生产流程

17

Robotaxi

小鹏**Robotaxi**获广州全无人测试资质，量产车路线加速

18

具身智能

小鹏机器人业务完成超9亿美元融资，车企造人路线被资本押注

19

机器人模型

蚂蚁灵波**Zero-WAM**发布，机器人开始“看视频学任务”

20

Token资本

纳德拉提出“**Token资本**”：企业不能把思考外包给单一模型

媒介360深洞察

AI进入“系统竞争周”：算力、开源、端侧、Agent与企业控制权同步重构

01 | 算力锁仓

英伟达财报、Anthropic天价合同、SK海力士HBM4E与美光“内存墙”说明，AI竞争底座正从GPU扩展到内存、封装和长期算力合约。

02 | 入口卡位

英伟达收购Hugging Face传闻、字节整合豆包工作、Meta Hatch与千问Flash显示，模型分发、办公入口和消费Agent都在抢用户高频场景。

03 | 成本重写

OpenAI自研推理芯片、DeepSeek高毛利API、ChatGPT Business分层计费和MiniMax ARR增长，共同推动AI从技术热度进入财务模型。

04 | 端侧觉醒

苹果M6/M5 Ultra与本地AI路线说明，端侧算力正在成为隐私、体验和成本之间的新平衡点。

05 | 具身落地

小鹏Robotaxi、小鹏机器人融资与蚂蚁Zero-WAM表明，具身智能竞争正从演示转向量产、运营和数据学习。

06 | 企业主权

纳德拉的“Token资本”提醒品牌，真正的AI资产不是某个模型，而是企业自己的上下文、流程、数据和交互经验。

媒介360

开源入口

01 | 英伟达拟129亿美元收购 Hugging Face，开源分发入口被重估

事件速读

- 据报道英伟达同意以129亿美元收购Hugging Face，若完成将成其史上最大收购。
- Hugging Face托管模型、数据集与工具，年化收入约1.5亿美元，是全球开发者使用开源AI的重要入口。

媒介360洞察

- AI竞争正在从“芯片供给”延伸到“模型分发权”。
- 英伟达若掌控开源入口，将在算力、工具链与开发者生态之间形成更强闭环。
- 开源平台的安全、治理与商业独立性会成为交易后的核心观察点。

给品牌的行动指南

- 品牌需要关注开源模型生态变化，避免工具链过度依赖单一平台。
- 技术团队应建立多模型、多平台的备选路线。
- 内容与产品创新可更多利用开源生态降低试错成本。

媒介360

算力合约

02 | Anthropic签下450亿美元算力合同，提前锁定Vera Rubin时代

事件速读

- Anthropic与Nscale签署约450亿美元、6年期算力租赁合同。
- 算力将来自西弗吉尼亚州数据中心，并采用英伟达下一代Vera Rubin系统。

媒介360洞察

- 前沿模型公司正在用长期合同锁定未来算力供给。
- 算力不再只是技术资源，而是决定模型迭代速度和商业规模的战略资产。
- 大模型公司的财务压力会越来越像重资产基础设施公司。

给品牌的行动指南

- 品牌采购AI服务时，要评估供应商的长期算力保障能力。
- 企业应把AI预算从单次项目转为持续资源规划。
- 高频AI应用要提前测算推理成本、峰值需求和供应稳定性。

资本市场

03 | Anthropic最早9月底IPO，AGI商业模式迎压力测试

事件速读

- 据报道Anthropic计划在美国劳动节后公开IPO招股书，上市窗口指向9月下旬或10月初。
- 公司考虑允许现有股东出售部分股份，并可能设置更长锁定期。

媒介360洞察

- Anthropic IPO将成为AI基础模型公司商业化的标志性检验。
- 资本市场会同时审视收入增长、算力成本、客户留存和治理风险。
- AI公司的估值逻辑正在从故事驱动进入财务证明阶段。

给品牌的行动指南

- 品牌要关注AI供应商的财务稳定性，尤其是长期交付能力。
- 企业合作不只看模型能力，也要看商业持续性。
- AI投资与采购决策应加入上市、融资和成本结构变量。

AGI定义

04 | 黄仁勋重写AGI标准：Token能盈利，才是经营意义上的智能

事件速读

- 黄仁勋在财报电话会上表示，很多任务上可以说AGI已经实现。
- 他提出AI能完成生产性工作、推理Token能带来利润、更多算力能生成更多盈利Token。

媒介360洞察

- AGI正在从技术终点转化为企业经营指标。
- “智能”不再只看能力测试，而看能否创造可计量的生产力和收入。
- Token经济将成为AI商业化最核心的度量单位之一。

给品牌的行动指南

- 品牌使用AI要从“好不好玩”转向“能不能创造收入和效率”。
- 每个AI项目都应设定可衡量的Token产出与业务结果。
- 组织需要建立AI ROI看板，跟踪成本、效率与转化。

媒介360

财报信号

05 | 英伟达二季度营收962.2亿美元，AI算力需求继续爆发

事件速读

- 英伟达第二财季营收962.2亿美元，数据中心业务收入890亿美元。
- 第三财季营收指引1080亿美元，并计划向AWS增配200万块GPU。

媒介360洞察

- AI算力需求仍处于强劲扩张期，云巨头继续加码基础设施。
- 英伟达从芯片供应商进一步成为AI产业节奏的核心定价者。
- 数据中心增长正在重塑全球科技公司的资本开支结构。

给品牌的行动指南

- 品牌应把AI应用成本波动纳入年度预算。
- 大型企业可与云厂商协商更长期、更稳定的AI资源方案。
- 营销、客服、内容等高频AI场景要提前做成本压力测试。

媒介360

HBM供应

06 | SK海力士美国HBM4E工厂计划2029年量产，内存成为AI战略资源

事件速读

- SK海力士计划2029年第三季度在美国印第安纳州工厂量产HBM4E。
- 公司预计内存短缺将持续到2030年底，目前占全球HBM市场收入58%。

媒介360洞察

- HBM正在从周期性存储产品升级为AI时代的关键工业资源。
- 供应链布局周期拉长，意味着算力成本压力会长期存在。
- AI基础设施竞争已经进入芯片、内存、封装和产能协同阶段。

给品牌的行动指南

- 企业AI战略要关注底层硬件供需，而不是只看应用层价格。
- 关键AI项目应设计可降级、可替换的模型和算力方案。
- 品牌方要预判生成式内容成本可能随硬件周期波动。

媒介360

内存墙

07 | 美光警告“内存墙”：带宽瓶颈正在卡住大模型命脉

事件速读

- 美光指出AI计算能力每两年增长3倍，但HBM带宽增速不足2倍。
- Meta训练Llama 3时，高达17%的意外中断由HBM故障引发。

媒介360洞察

- AI瓶颈不只在GPU，也在内存带宽、可靠性和封装技术。
- 模型越大，系统稳定性越成为训练和推理成本的重要变量。
- 底层硬件故障会直接影响AI服务质量与商业连续性。

给品牌的行动指南

- 品牌部署AI服务时要设置容灾、降级和备份方案。
- 企业不要只比较模型效果，也要评估平台稳定性。
- 对实时客服、交易和营销系统，应建立AI服务SLA指标。

媒介360

企业计费

08 | ChatGPT Business开放Premium席位，企业AI进入用量分层时代

事件速读

- OpenAI开放ChatGPT Business Premium席位，按月每人125美元，按年折合每月100美元。
- Premium提供Standard的5倍用量，并允许同一工作区混合分配不同席位。

媒介360洞察

- 企业AI收费正在从单一订阅走向分层用量管理。
- AI服务越来越像企业软件与算力资源的混合商品。
- 管理员视角、支出限额和席位弹性会成为B端AI产品标配。

给品牌的行动指南

- 企业应按岗位价值划分AI席位，而不是平均分配。
- 高频使用者可以配置Premium，普通岗位采用Standard。
- 品牌团队要把AI账号管理纳入组织效率和预算管理。

媒介360

模型财务

09 | DeepSeek前7个月营收4.75亿元，API毛利率达82.9%

事件速读

- DeepSeek前7个月营收约4.75亿元，为去年全年10倍，同期净亏损约7.15亿元。
- API销售毛利率达82.9%，但AI基础设施投入约110亿元。

媒介360洞察

- 高毛利API与高额基建投入并存，揭示AI商业的双重结构。
- 模型公司要同时比推理效率、资本能力和生态规模。
- AI价格战背后，是基础设施投入与开发者增长之间的平衡。

给品牌的行动指南

- 品牌调用API时应关注价格之外的稳定性和长期成本。
- 技术团队要建立多模型路由，避免单一价格波动影响业务。
- AI项目应分清实验性调用和生产级调用，分别管理预算。

媒介360

AI出海

10 | MiniMax 8月ARR突破8亿美元，ToB与海外收入成为增长主轴

事件速读

- MiniMax披露7月Token消耗量达1月20倍，8月ARR超过8亿美元。
- ToB业务占比约80%，海外收入约占60%，开放平台和AI企业服务增长703.1%。

媒介360洞察

- 中国AI公司正在从流量增长转向企业服务与海外变现。
- Token消耗成为衡量AI业务活跃度与收入潜力的重要指标。
- 开源下载、开发者客户和企业服务共同构成增长飞轮。

给品牌的行动指南

- 品牌出海可优先关注成熟AI平台的多语言和企业服务能力。
- B端营销要从曝光转向用量、留存和复购指标。
- 企业可用Token消耗和任务完成量衡量AI项目价值。

媒介360

办公AI

11 | 千问办公上线Qwen3.8-Flash，95%日常任务走向低成本快模型

事件速读

- 千问办公上线Qwen3.8-Flash标准模式，单任务生成速度提升约100%。
- Token消耗平均减少75%，官方预计95%日常任务可由标准模式完成。

媒介360洞察

- 办公AI正在进入推理经济学阶段：够快、够便宜比最强更重要。
- 模型分层会成为企业AI产品的核心策略。
- 高频低复杂任务将率先被标准化AI workflow接管。

给品牌的行动指南

- 品牌团队应把日常写作、总结、表格和PPT流程标准化。
- 不同任务使用不同模型档位，避免为简单任务支付高成本。
- 用速度和成本双指标评估AI办公工具，而不只看效果。

媒介360

自研芯片

12 | OpenAI公布Jalapeño芯片测试，模型公司加速软硬一体

事件速读

- OpenAI公布首款自研推理芯片Jalapeño测试结果，每瓦AI产出提升1.5至1.9倍。
- 该芯片由OpenAI与博通开发，年底前将部署进自家算力基础设施。

媒介360洞察

- 前沿AI公司正在把竞争推进到芯片、模型、内存和产品协同设计。
- 自研芯片的关键不是单点性能，而是推理成本和产品体验控制权。
- AI供应链从外部采购走向深度自建，将改变云和芯片生态。

给品牌的行动指南

- 品牌选择AI平台时要关注其成本控制和基础设施路线。
- 高并发AI产品应重视低延迟和能效指标。
- 企业要判断供应商是否具备长期降本能力。

媒介360

端侧算力

13 | 苹果发布M6与M5 Ultra，Mac产品线全面押注本地AI

事件速读

- 苹果上架M6与M5 Ultra芯片及新款Mac，Mac mini AI性能最高提升4倍。
- M5 Ultra支持512GB统一内存，Mac Studio被定位为端侧AI桌面设备。

媒介360洞察

- 端侧AI正在成为个人计算设备的新卖点。
- 当云端推理成本高企，本地大模型能力会重塑创作和办公体验。
- 苹果继续用硬件、系统和隐私形成端侧AI护城河。

给品牌的行动指南

- 品牌内容团队可提前测试端侧AI创作工作流。
- 涉及隐私的数据处理可优先考虑本地AI方案。
- 硬件更新决策应把AI性能纳入采购标准。

媒介360

消费Agent

14 | Meta计划推出Hatch，消费级AI智能体开始订阅化

事件速读

- Meta计划数周内推出AI智能体平台Hatch，基于OpenClaw虚拟助理构建。
- 高级版月费最高199.99美元，可访问DoorDash、Etsy、Reddit等网站。

媒介360洞察

- 消费级AI正在从聊天玩具走向任务执行和代操作服务。
- 订阅定价显示平台希望把Agent包装为生产力工具。
- 网站访问与代操作能力将重塑电商、内容和服务入口。

给品牌的行动指南

- 品牌要思考产品和服务如何被Agent直接调用。
- 官网、客服和交易流程应降低机器访问与理解成本。
- 内容SEO要升级为面向AI代理的可执行信息结构。

办公入口

15 | 字节整合TRAE与扣子，统一办公AI品牌 “豆包工作”

事件速读

- 字节将TRAE、扣子团队并入豆包体系，整合办公工作场景能力。
- TRAE IDE及CLI继续作为编程产品线发展，豆包被明确为AI主干业务。

媒介360洞察

- 办公AI进入品牌整合期，分散工具正在收拢为统一入口。
- 平台需要用清晰品牌心智承接用户的工作任务。
- 从写作、编程到流程自动化，AI办公将形成更完整产品矩阵。

给品牌的行动指南

- 品牌团队应减少工具割裂，建立统一AI办公入口。
- 内部培训要围绕场景而非单一工具展开。
- 内容、运营、商务和研发应共享同一套AI工作标准。

媒介360

视频生成

16 | 阿里Wan3.0上线：30秒视频生成与文档输入进入生产流程

事件速读

- 阿里Wan3.0支持单次生成30秒视频，并首次接受doc、xls、ppt、pdf、md等文档输入。
- 已进入短剧、广告、文旅、音乐视频等流程，并开放API调用。

媒介360洞察

- AI视频工具正在从创意玩具进入生产力阶段。
- 文档输入让PPT、表格、脚本可以直接转化为视频素材。
- 内容生产门槛下降后，品牌竞争会回到策略、审美和分发。

给品牌的行动指南

- 品牌应把PPT、产品资料和活动方案整理成可视频化资产。
- 短视频团队可建立AI生成视频的审核和风格规范。
- 营销内容要从“单条创意”升级为批量化资产生产。

媒介360

Robotaxi

17 | 小鹏Robotaxi获广州全无人测试资质，量产车路线加速

事件速读

- 小鹏Robotaxi获得广州远程测试资质，可开展车内无安全员的全无人道路测试。
- 车辆搭载4颗自研图灵AI芯片，有效算力3000 TOPS，计划下半年试运营。

媒介360洞察

- Robotaxi正在从演示测试进入更接近商业运营的全无人阶段。
- 小鹏以量产车平台和统一软件架构切入，强调工程化效率。
- 车企自研芯片与算法能力正在成为智能驾驶竞争关键。

给品牌的行动指南

- 汽车品牌要把自动驾驶叙事从技术参数转向真实运营体验。
- 城市出行、零售和本地生活品牌可提前探索车内服务场景。
- Robotaxi合作要关注安全、路线、服务与数据闭环。

媒介360

具身智能

18 | 小鹏机器人业务完成超9亿美元融资，车企造人路线被资本押注

事件速读

- 小鹏机器人业务完成首轮超9亿美元融资，投后估值约430亿元。
- 腾讯、阿里等战略投资者支持，计划第四季度量产，2027年进入门店导购。

媒介360洞察

- 具身智能融资进入大额工程化阶段，资本押注制造与供应链能力。
- 车企的量产体系、传感器经验和控制算法可能迁移到机器人。
- 机器人商业化正在从实验室展示转向门店、服务和海外场景。

给品牌的行动指南

- 零售品牌可提前规划机器人导购、陈列和服务流程。
- 体验店要从“人力服务”转向“人机协同”设计。
- 品牌应评估机器人带来的互动数据、服务效率和话题价值。

媒介360

机器人模型

19 | 蚂蚁灵波Zero-WAM发布，机器人开始“看视频学任务”

事件速读

- Zero-WAM可将人类演示视频作为上下文提示，让机器人无参数更新执行新任务。
- 在RoboTwin 2.0未见任务上平均成功率46.95%，较基线提升29.5个百分点。

媒介360洞察

- 具身智能的Scaling Law可能从机器人数据转向人类视频学习。
- 如果机器人能“看会”技能，训练数据成本会大幅下降。
- 视频、语言和动作的统一理解将推动机器人泛化能力突破。

给品牌的行动指南

- 品牌可把操作流程、服务动作和培训视频沉淀为AI可学习资产。
- 线下服务企业应标准化关键动作，便于未来机器人迁移。
- 内容资产不只是传播素材，也可能成为自动化训练数据。

媒介360

Token资本

20 | 纳德拉提出“Token资本”：企业不能把思考外包给单一模型

事件速读

- 微软CEO纳德拉认为，企业使用AI会积累提示词、交互记录和使用方式构成的Token资本。
- 他主张应用层与模型分离，保留上下文和记忆，避免绑定单一服务商。

媒介360洞察

- 企业AI护城河不只在模型，而在独有场景、数据和交互经验。
- 上下文、记忆和元数据将成为AI时代的新型知识资产。
- 多模型架构会成为企业保持控制力的重要方式。

给品牌的行动指南

- 品牌应沉淀自己的提示词库、知识库和任务流程。
- 不要把核心上下文完全绑定在单一AI平台。
- 建立可迁移、可审计、可复用的企业AI资产管理机制。

媒介360

本周结论：AI下半场开始比“谁更能组织系统”

从本周热点可以看到，AI竞争已经不只是模型能力竞争，而是算力、芯片、入口、生态、成本、场景与治理的系统竞争。

算力

算力合约、HBM短缺、内存墙与GPU涨价共同说明，AI规模化的真正门槛正在上移。

生态

开源入口、办公品牌、消费Agent与视频平台正在争夺AI时代的新分发权。

成本

API毛利率、Business席位、自研芯片和Flash模型都指向同一件事：AI要进入可持续财务模型。

端侧

苹果端侧AI和本地大模型路线，让个人设备重新成为AI体验和隐私计算的重要入口。

具身

Robotaxi、机器人融资与视频学习模型，正在把AI从屏幕带入物理世界。

品牌

品牌下一步不是追逐每个AI热点，而是重建内容、服务、数据、组织和合规的AI经营系统。

加入会员 增长快人一步

欢迎成为我们媒介360前瞻会员及智库会员！

媒介360将成为每位用户的超级连接器、前瞻中心及智库

- >> 营销增长指南：行业风向标、营销最佳实践、增长罗盘等
- >> 会员专享报告：媒介投资报告、CMO报告、品牌报告等
- >> 还可以参与我们的线下共研会，会员年会、媒介360大会等



扫码关注
m360服务号



扫码进入
m360小程序

加入媒介360会员，进一步洞察全媒体全营销趋势风向
敏捷把握市场脉动，让营销决策更敏捷、更科学！

- >> 第一时间了解行业内重大事件和热点新闻
- >> 最具专业度与前瞻性的媒体市场分析情报
- >> 深入观察广告主行业动态，透过现象解读本质

PC端会员注册

1. 打开媒介360主站登录页面：<https://www.media360.info/login>;
2. 手机号、验证码注册登录
3. 打开媒介360会员开通页面：<https://www.media360.info/account/vip>
4. 选择会员类型，点击购买，付费即刻成为会员

移动端会员注册

1. 微信搜索“媒介三六零会员”进入服务号
2. 点击底部菜单栏“会员-成为媒介360会员”
3. 选择“会员类型”，点击“立刻开通”，付费即刻成为会员

媒介360

媒介360

科技互联网周报

专业制胜 敏捷增长