

当Agent成为AI最大公约数

在商汤科技、亚马逊云科技、盛大集团、昆仑万维等一众科技企业高管齐聚的会场，Agent落地已然成为当下AI产业发展的最大公约数。5月20日，由量子位主办的第四届中国AIGC产业峰会在北京启幕。回顾行将过半的2026年，AI产业发展势头迅猛，以龙虾（OpenClaw）为代表的多款爆款Agent产品相继走红，各类全新业态与应用玩法不断涌现。行业从业者与普通使用者都面临诸多实际困惑，如何找准AI实用路径、把握产业发展机遇，成为行业共同思考的问题。

从AI到AI Agent，是从对话理解到执行交付的进步。参会嘉宾共同聚焦toB、toC Agent落地实践，业内普遍达成共识，AI Agent已然从行业热议概念走入产业现实应用。有观点认为，目前行业已完成全链路技术布局，但其层级繁杂、耦合性强的技术架构，也意味着AI规模化落地无法依靠单点突破实现，是一场需要全产业协同发力的系统工程。

理清企业级Agent工程化路径

本届峰会上，多位行业大咖从不同领域深度剖析行业现状与发展方向。昆仑万维董事长兼CEO方汉围绕市场主体发展，探讨个人与企业该如何平稳应对Agent带来的行业变革；亚马逊云科技产品技术部技术总监王晓野聚焦产业实操，深度剖析企业级Agent落地过程中存在的工程化难点；盛大集团副总裁、EverMind CEO邓亚峰解读长期记忆驱动的自进化技术，梳理AI向长效数字生产力转型的发展逻辑；商汤科技执行副总裁、首席科学家林达华则立足技术前沿，探析多模态融合与空间智能的全新发展趋势。

在王晓野看来，如今市面上各类AI模型、应用工具层出不穷，但能够真正扎根企业生产场景、实现常态化规模化运转的Agent产品数量并不多。个人端火热流行的各类Agent应用使用门槛低、玩法灵活，与严苛的企业商用场景存在本质区别。

峰会现场，他向北京商报等媒体分析称，

个人使用偏向自由体验，而落地到企业办公与产业场景中，安全管控、使用成本、技术工程化适配、企业内部组织流程磨合等一系列问题，都成为难以跨越的落地阻碍。

结合行业调研数据来看，当前绝大多数企业已经完成AI基础布局，行业整体布局普及率持续上涨，但仅有小部分企业成功将AI应用落地投产，真正依靠Agent提升实际生产力，多数项目依旧停留在测试试用阶段，产业落地呈现出布局热、落地冷的现状。

王晓野认为，企业级Agent落地绝非仅依靠大模型能力就能实现，想要完成规模化普及，企业必须搭建起完备的配套能力体系。落地过程中不仅要侧重优化推理算力，兼顾不同业务场景下的算力性价比；也要秉持开放思路灵活选用适配模型，严守企业核心数据安全；同时做好私有业务数据梳理，做好长期记忆生命周期管理，剔除冗余无效信息，兼顾内部数据记忆共享与权限隔离，让Agent深度贴合企业真实业务逻辑运转。

除此之外，企业还需补齐系统打通、人员权限划分、全天候稳定运维等工程化能力，在



满足员工高效便捷使用需求的同时，筑牢企业合规管理底线。他预判，AI将持续深度重塑传统工作模式，未来会有大批企业引入数字员工承接日常事务与基础决策工作，行业整体市场规模将迎来大幅扩张，逐一打通落地全流程中的各类痛点难点，是行业抓住新一轮产业发展机遇的核心关键。

探寻AI长期进化方向

AI赛道热度起落速度加快，仅凭短期流量走红的智能应用，很难持续扎根市场，扎实的技术沉淀与实际落地价值，才是行业稳步前行的底气。林达华围绕行业长期发展路径，在峰会上向北京商报等媒体分享了他的观察与思考。

在他看来，如今AI在数字创作、办公辅助领域快速普及，各类智能工具层出不穷，但不少热门产品热度褪去后便归于沉寂，核心在于未能形成完整的问题解决链路，难以贴合实际业务需求。企业落地AI过程中，异构数据整合是主要痛点，图文、表格、音视频、行业知

识库等各类信息梳理难度大，这一环节占企业AI落地成本七成以上。

依托多模态融合技术思路，商汤旗下智能应用小浣熊打通多元数据处理壁垒，搭建起数据梳理、智能分析到成果输出的完整服务体系。截至2025年，小浣熊已服务合计超1500万个人用户与企业客户，当年月活用户增长7倍，2026年1至5月再度实现5倍增长，即便一众大众AI工具热度退潮，它依旧保持稳健增长势头。

林达华坦言，当下主流多模态模型处理数字信息已十分成熟，但面对真实物理世界仍存在明显短板，难以理解现实场景中的空间逻辑与客观规律，极大限制了实体智能产品落地。同时单纯扩充训练数据无法破解这一困境，反而容易固化固有认知，影响实际判断效果。他认为，行业亟须革新研发与训练范式，整合视觉感知、逻辑推理、内容生成多项能力，搭建全新模型架构，摒弃盲目堆砌参数的模式，以精简高效的技术架构挖掘智能潜力。

放眼行业趋势，AI发展正从单一数字场景延伸至物理世界，空间智能成为产业新的

探索重心。未来成熟的智能体，需要兼顾数字信息研判与真实场景实操能力，实现虚实空间深度融合。他也呼吁行业摒弃浮躁风气，坚守长期主义深耕底层技术，稳步推动AI实现更深层次的产业落地。

从个人工具到组织生产力跃迁

在企业级落地路径与长期技术探索之外，行业机构也在为AI Agent的发展格局提供坐标系。

QuestMobile分析指出，当前AI Agent赛道的竞争，其本质是围绕操作系统和应用商店架构展开的生态博弈：手机厂商与互联网企业分别在底层系统与技能层占据优势，行业竞争已从用户时长争夺转向场景调用权卡位。目前国内已完成从底层核心能力到上层交互场景的全链路Claw布局，但多层次耦合的技术栈，决定了AI落地是一场横跨各环节的系统性战役，应用、平台、基建层各有专属竞争逻辑。主流产品虽已形成清晰的生态映射，但仍处于规模扩张阶段，尚未进入价值兑现的深水区。

增长黑盒发布的《2026中国OpenClaw生态现状：用户及企业应用调研报告》，则从组织视角拆解了落地痛点。报告显示，尽管Gartner预测到2028年将有15%的企业日常决策由AI Agent自主完成，但MIT数据显示，95%的企业AI试点止步于试验阶段。核心矛盾在于，多数企业仅停留在为员工配备个人工具，未重构组织流程与分工模式，AI节省的时间最终被旧协作链条消耗，个人提效无法转化为组织整体效率提升。

无论是生态博弈还是组织提效，最终都指向同一个行业共识：AI Agent的规模化落地，需要的是架构、生态、组织模式的系统性协调配合，打通技术、生态与组织的全链路，才能真正跨越AI Agent的落地鸿沟。

北京商报记者 陶凤 王天逸/文并摄

Market focus

阿里云的千亿美元新故事

一周前，阿里宣布2026年一季度阿里云AI收入占外部收入比重首次突破30%，一周后召开的2026阿里云峰会上，阿里发布基于平头哥新一代AI芯片真武M890的128卡超节点服务器、新一代千问旗舰模型Qwen3.7-Max以及面向Agent（智能体）的AI产品官网“千问云”，拉通了阿里云提出的“芯—云—模型—推理”技术体系。

峰会上，阿里云资深副总裁刘伟光表示，Agent时代，云在AI时代的作用从支持模型训练与Token生产，转变成承载海量Agent应用，被Agent调用。阿里云的同行也发现了这种转换，百度智能云在上周宣布，将全面升级为面向大规模Agent应用的全新栈AI云。腾讯云则将在下月围绕Agent上线新品。回到真金白银，阿里云预计，未来一年AI相关产品收入在外部商业化收入中的占比将超50%，再往远看，阿里集团CEO吴泳铭已提出目标：五年内阿里云和AI商业化年收入突破1000亿美元。

大云“生”小云

临近5月20日中午12点，刘伟光抛出最后一个亮点：千问云，“为Agent而生的全新AI产品”呼应了他开场的观点：我们正身处Agentic（智能化）时代。

据介绍，千问云提供Qwen、GLM、Kimi、DeepSeek、Wan、HappyHorse等150多款主流模型API（应用程序编程接口），并将模型服务的核心能力封装为Skills（技能）和CLI（命令行界面）工具，可让Agent工具高效地用模型和开发AI应用。

千问云网站在模型选择和模型调用环节，可为用户提供不同模型在参数、能力、价格、上下文长度等维度的对比信息，

对比后即可直接进入模型体验页面，用真实Prompt或任务验证输出效果。

千问云Skills将模型选型、模型调用、认证配置、用量查询等完整链路能力封装，用户无需编写任何代码，Agent即可动态路由不同模型，自动用视觉模型处理图片，用生图模型生成图片，用视频生成模型创作视频等；千问云CLI则可覆盖登录认证、模型查询、模型调用、用量查询和环境诊断等能力，可直接通过脚本或命令行自动化完成所有模型服务的工作流。

可以这么理解，千问云相当于开了一家面向开发者和Agent的大模型超市，提供千问及外部大模型的API供开发者接入，也针对Agent使用做了改进，让Agent更方便调用。

作为基础设施服务商，无间芯穹联合

创始人兼CEO夏立雪多次强调，基础设施需要AI原生重构。2026中关村论坛年会期间，夏立雪表示，“当前云计算时代的基础设施是为服务人类工程师设计的，而不是为AI设计的。现在的模式是基础设施上有一个接口，这个接口为人类工程师而设，往上加一层再去接Agent，这种方式限制了Agent的能力”。

芯片快跑

千问云是Agent的“前台”，挑模型、调接口、管账单。但Agent真正跑起来，靠的是“后台”的算力。

当天的峰会上，算力新产品被安排在第一个发布，那是基于平头哥新一代AI芯片真武M890的128卡超节点服务器磐久AL，搭载互联芯片ICN Switch 1.0，通信时延低至百纳秒级，可让128张AI芯片组成一台计算机。

真武M890采用自研并行计算架构，内置144GB显存，性能是真武810E的3倍，片间互联带宽800GB/s，芯片原生支持FP32到FP4等多种数据精度，可应用于高精度训练、低精度和超低精度推理的全场景。

平头哥互联芯片ICN Switch 1.0支持自研ICN互联总线协议和IPCCL通信库，吞吐量达25.6Tbps，可支持真武M890的高效互联，提升AI集群的算力效率。真武M890搭配ICN Switch 1.0芯片，可实现64卡全带宽互联。

当天，平头哥还首次公布真武系列芯片的规划。按照平头哥半导体副总裁高慧提供的时间表：真武810E已于2024年

二季度发布，2026年5月20日发布的是真武M890。2027年三季度，平头哥将推出真武V900，具有216GB显存，片间互联带宽1200GB/s，2028年三季度将推出真武J900。

同行也在行动。目前昆仑芯P800已完成规模化验证，2025年至今已交付多个万卡集群。基于昆仑芯的天池256卡超节点已经在上个月点亮，将于6月正式上市，吞吐性能较上一代提升25%，并完成包含文心、DeepSeek、GLM、MiniMax等主流模型的适配，推理效率提升50%，网络架构升级至HPN5.0，端到端时延优化50%，支持按需搭建数十万卡乃至百万卡超大集群。

以阿里云的说法，其技术体系为“芯—云—模型—推理”。站在百度智能云的角度，芯云模体是彼此衔接的全栈AI能力，在各种升级之间有效适配，达到整体提升。

“AI能做到毫秒到秒级别思考和发起任务，但人类是分钟级别的，这中间差了数量级。”在夏立雪看来，因为这个原因，我们需要打造更智慧的基础设施。换一个角度看，这正是Agentic时代基础设施要解决的推理需求问题。

下一关50%

谈到推理和训练需求的区别，CHIP中国实验室主任罗国昭向北京商报记者打了个比方：“有点像遇到难题，是要经验满满的老师随时随答，还是需要顶级专家立项、排期再攻坚。前者更像推理，后者则是训练。推理需要更低的响应延迟，在现

有的知识库中确认答案；训练需要顶尖的能力，在已知条件中探索未知的方式。”

会场内的焦点是千问云、Qwen3.7-Max、真武M890，会场外资本市场关心的是实实在在的财务数据。

根据5月13日发布的阿里最新财报，2026年一季度阿里云营收同比增长38%至416.26亿元，剔除阿里并表业务后的外部商业化收入增长40%，阿里云AI相关产品营收89.71亿元，占外部收入比重首次突破30%。

财报电话会上，阿里表示预计未来一年，AI相关产品收入在外部商业化收入中的占比将突破50%。百度也在秀肌肉，2026年一季度百度智能云AI业务收入136亿元，占百度一般性业务收入的52%。

回到阿里云，各种声音都有。文溯智库创始人王超认为，“阿里在探索商业化方面玩法挺多的，比如千问云，但并没有找到大模型真正应该发力的地方”。

在他看来，“AI厂商最核心的竞争力，也就是利润率最高的是模型，是自己的API售卖情况。阿里云做千问云的逻辑是跟其他模型企业更深度地绑定，通过给它们提供更适合Agent使用的云等服务，让模型厂商留在或绑在阿里云上，这样阿里云的营收会跟着其他模型厂商收益走，但是在为他人作嫁衣”。

一个关键指标即阿里云的外部商业收入。一季度阿里云外部商业化收入增速40%，高出整个阿里云增幅，但并没有披露千问模型的API收入等。而王超认为，“千问API收入才是阿里云真正核心的收入，诸多信息再次证明阿里变成了一个AI投资公司”。北京商报记者 魏蔚