

新华三“豪赌”全栈路线

2026年, AI基础设施建设进入了高投入与高损耗并存的矛盾周期。

市场端, 中国Token(词元)调用量两年激增1400倍, 全球AI基础设施投入规模正向7万亿美元迈进; 但工程化落地端, 效率瓶颈却越发显著。英伟达数据显示, 分布式训练中的网络拥塞导致算力空转率普遍超过30%; Gartner则预测, 到2027年年底, 超过40%的企业级智能体项目或将因落地难而被迫取消。

应用层面的反馈更为直观: 据浪潮人工智能研究院测算, 我国智算中心平均算力使用率仅为30%。某千卡规模项目的实际运营情况也印证了这一现状——不仅上架率不足50%, 且已上架服务器的实际利用率也不到30%, 年运营成本却高达3000万元。

这暴露了当前行业的一个核心症结: AI基础设施的建设瓶颈, 早已不在单点算力芯片的性能指标, 而在于计算、网络、存储、云管、安全、运维等环节之间, 存在着一条尚未被系统性打通的“隐形断层”, 算力墙、存储墙、功耗墙等各种说法因此产生。

本报记者 张琪玮 实习生 王嘉琪

系统性的“木桶效应”

“如果AI是一辆超级赛车, 需要的不仅是强悍的发动机, 更需要配置完整、调校优异的整车。”今年5月, 在新华三集团的领航者峰会上, 总裁于英涛用一个形象的比喻来拆解行业迷思。他直言, 当前多数人的做法是不断更换更强的发动机: “但如果变速箱打滑、轮胎漏气、油路堵塞, 这辆车能跑多远?”

数据不会说谎。根据新华三公开的行业测试报告, 万卡集群中数据IO瓶颈导致GPU空等时间占比可达40%以上。这意味着昂贵的算力芯片有近一半时间在等待数据搬运。网络层面的拥塞和丢包, 则让跨节点通信效率大打折扣。而因安全事件、维护故障或调度失误导致的训练中断, 平均恢复时间以天为单位计算。对于千卡级以上的智算集群, 每一次意外停机都意味着数十万甚至上百万的直接损失。

“Token的性价比不是一张卡决定的, 也不是由某一个单一能力决定的。”于英涛强调。

这并非新华三的一家之言。上海智能算力科技有限公司副总经理辛帅在峰会上分享了从零搭建万卡集群的真实历程。他将挑战归结为四个维度: 规模之难——万卡级别意味着任何单点故障都可能演变为全局事故; 异构之痛——不同厂商芯片的指令集和通信库各自为政, 能否协同工作是个巨大问号; 运维之繁——故障定位须达毫秒级、链路级; 演进之感——今天购买的设备需适配未来3~5年的模型架构。“四点中任何一点考虑不到, 后面都可能推倒重来。”辛帅表示。

一场全栈的“押注”

一个越发清晰的事实是: 算力基础设施的系统协同能力, 已严重滞后于芯片性能的迭代速度。大量GPU不是“跑不动”, 而是“跑不满”。

面对这一困局, 业界分化出了不同的解题路径。有的企业选择“往前冲”, 以行业智



图为H3C UniPoD S80000超节点亮相NAVIGATE 2026

能体打包算力、模型与场景, 试图绕开底层协同的碎片化, 直接交付应用价值。有的企业试图“往深扎”, 将AI推理能力注入基站和接入网, 死磕连接层的智能化跃迁, 为上层应用扫清时延障碍。也有的云厂商选择“往后退”, 集体转向“水电煤”式的标准化算力租赁, 按需交付弹性资源, 但把系统协同的难题留给了客户自己。

新华三的选择更为激进, 同时攥住计算、网络、存储、云、安全、运维六个维度, 试图交出一套全栈“交钥匙”般的系统级方案。

这一思路最直观的体现是: 既然瓶颈在各环节之间的缝隙, 那就把六个维度拉通来做整体优化。在算力侧, 高密度冷方案将6颗CPU集成于1U标准机箱内, 单机柜算力超过3万核。散热方案上, 整机不依赖风

扇, 全部由液冷完成热交换, 电能利用效率指标(PUE)可控制在1.04, 这意味着电费开支中, 真正用于计算的比例远高于行业平均水平; 在存储层, 用“以存换算”的思路把GPU在等待数据的空闲中解放出来, 将大模型的历史上下文缓存到扩展内存中, 将首Token延迟降低90%; 在网络层, 通过算网协同调度让带宽利用率从70%向90%以上爬升。

在今年5月, 新华三最新发布的S80000超节点也印证了他们的技术思路。该产品覆盖32卡至1024卡全系列, 可扩展至万卡以上集群, 在一个物理单元内完成算、网、存、能的协同。新华三首席技术官张强表示: “单一能力的简单叠加无法实现系统最优。”

回顾全栈路线在IT行业的演进, 其成功

往往取决于一个关键前提: 每个技术层都不能有明显的短板。对新华三而言, 如何在六线并进的局面下保持各层的技术深度, 是一个仍需交给时间验证的命题。

参数之外的答案

验证一个战略是否成立, 不能只看发布会上的参数, 还要看机房里的真实表现和客户的续约率。在这方面, 新华三给出了多个维度的落地案例, 但也不乏待解之问。

在上海, 新华三与上海智能算力科技有限公司联合建设的万卡集群已投入实际运营。据辛帅透露, 双方在技术路线上反复磨合, “2023年我们就和新华三的专家反复讨论, 最终坚定选择了以太网, 现在证明完全正确。”以太网的开放协议带来

了供应链安全和系统灵活性的双重优势, 使得运营方可以将“选择权始终留在自己手中”。

在杭州、郑州等十余个城市, 新华三推动的“图灵小镇”模式已陆续落地。这一模式试图将算力和模型包装成水电一样的公共设施, 降低中小企业使用AI的门槛。据其公布的数据, 目前已有数百家企业入驻。

但产业园区从“签约入驻”到“真正跑出商业模式”, 还有漫长的距离。贵阳贵安图灵小镇的AI算力服务中心已建成, 预计下半年才能投入使用。其运营方负责人赵文榛在峰会发言中提到, 国企投资决策中“过度强调短期经济账”, 缺乏对产业带动等长期价值的评估。这反映出“建算力”与“用算力”之间仍横亘着不小的落地阻力。而新华三的全栈路线能否在财务上真正跑通, 关键指标也尚未完全浮出水面——其AI相关业务的毛利率、客户留存率, 目前还不是公开数字。

在领航者峰会演讲的最后, 于英涛将开场时的主题词“亢奋与焦虑”替换成了“冷静与激情”。他说, 缺了冷静的激情叫“冲动”, 缺了激情的冷静是“平庸”。

这番话既可以看作对全行业的告诫, 也可以视为对新华三自身的提醒。他列出的两个“冰冷真相”值得思考: 一是7万亿美元的全球AI基础设施账单不会自动转化为价值, 押注不等于回报; 二是Gartner预测的超40%智能体项目取消率不是诅咒而是提醒——“变革前夜, 道阻且长”。

对于新华三而言, 2026年是一个关键点。公司刚迈过成立十周年的门槛, 年收入已达760亿元规模, 但传统通信与IT硬件市场的天花板清晰可见。AI是其必须抓住的战略级机遇, 而选择“全栈筑基”这条路径, 意味着要同时承受多重考验: 技术整合的复杂性、研发投入的持续性, 以及市场对投资回报周期的耐心。

在这场关乎未来的AI马拉松中, 新华三已选定了自己的配速与跑道。起跑阶段的声势已经足够响亮, 但真正决定成败的, 是三年到五年后客户机房里持续运转的机器轰鸣声, 以及那些今天尚在“试用算力”的企业最终选择留下的背影。

智能经济催生“新饭碗”

(上接第1版) 由于生成结果具有随机性, 过程如同拆盲盒, 便被业内形象地称为“抽卡”, 而负责这一环节的从业者, 也有了专属称号——“抽卡师”。

在AI短剧制作流程中, 前期分镜设计决定“拍什么”, 后期剪辑决定“怎么讲”, 而“抽卡师”则决定“有没有可用素材”。“我们通常会先上传角色与场景图片, 再详细输入对白、情绪及镜头要求。”公司“抽卡师”徐明明介绍, 若生成结果不理想, 要么重新输入指令, 要么将指令拆解得更细致、更具体, 直到获得满意素材。

“AI不是来‘抢饭碗’的, 而是来发新Offer的。”专为全球电商卖家打造的AI设计生产平台Skild Art CEO黄露凤, 对此有着深刻体会。

在传统模式下, 从产品拍摄到完成详情页、场景图、广告素材设计, 再到最终上线平台, 往往需要三天以上时间, 不少企业因此错失销售窗口期, 最终陷入存货过剩的困境。在位于广西柳州的上汽通用五菱汽车股份有限公司, 黄露凤花了三个月时间, 成功完成从调研到内测的全流程, 目前内测商家已突破1000家, 能够实现这样的结果, 正是对AI创造力与效率的充分挖掘。

市场在发展中出现新问题, 自然也会催生新机会。以Skild Art所在的跨境电商领域为例, 全新职业诸如AI电商视觉设计师、AI训图师、AI产品体验师、AI工作流设计师等岗位需求正持续攀升。

记者采访发现, AI相关新岗位的细分趋势日益明显。过去, 行业分工多以流程为边界, 各岗位各司其职; 如今, 部分岗位可串联、替代多个传统环节, 职业体系正从粗放向精细演化。

腾讯研究院的一份调查报告显示, AI原生企业的全部在招聘岗位可分为五类——赋能者(做模型)、协作者(做产品和交付)、推广者(做销售和营销)、治理者(管安全和合规)、支持者(管后勤和运营)。

“新职业不断细分出新工种, 意味着职业不再是模糊的大类, 而是被拆解为具体、专业、细腻的岗位结构。”中国新就业形态研究中心主任张成刚指出, 工作模式也在发生深刻重构, 从“岗位导向”转向“任务导向”, 人们的关注点正从“在哪家公司上班”, 转变为“完成了什么样的工作成果”。

传统产业“人机共舞”

眼下, AI不仅在新兴领域催生新职业, 更在传统行业中落地生根, 一批新工种走进工厂车间、实验室, 推动传统岗位转型升级, 形成“人机共舞”的新场景。

在位于广西柳州的上汽通用五菱汽车股份有限公司, 黄露凤介绍了智能网联汽车测试、运维等相关就业, 要求员工适应人机协同的新工作模式。”广西柳州的上汽通用五菱汽车股份有限公司党委书记姚佐平表示。

《稳岗扩容提质行动方案》明确, 发挥AI助力传统产业改造升级作用, 研究出台专门意见, 推动工业全要素智能化发展, 鼓励企业在应用AI时同步开展转型转岗培训, 发展新质生产力创造就业。扩大数据标注员、AI训图师等新职业新工种需求。

走进四川中鸿数智信息技术有限公司, 200余名数据标注员在电脑前, 对图片、视频、语音等数据进行处理, 为互联网企业、智能驾驶汽车公司、具身智能机器人企业提供数据服务。

数据标注是通过人工按照场景规则, 对不同内容进行分类标注, 给AI灌输规则, 让AI认识现实世界, 是模型训练的核心, 也是企业构建差异化优势的关键。高素质的数据标注员, 对从业者专业素养提出极高要求。以人形机器人数据训练为例, 需依托人类第一视角完成实景采集, 全方位收录肢体动作、关节角度、力学反馈、视觉感知等多维度信息, 精细化标注流程极大抬高了从业门槛。

在讯飞医疗科技股份有限公司, AI医学研究员胡萍萍正在评估医疗大模型在复杂病例诊治中的表现, 尝试发现模型在回应患者时缺乏“同理心”的细微问题。她和团队的目标是让AI不仅“懂治病”, 更要“懂人”。

胡萍萍曾是一名消化内科医生。5年前, 她转型成为AI训图师, 用医学专业知识参与训练医疗大模型, 使其更好地服务基层医生和居民健康管理。“我们既是医学经验的传承者, 也是技术应用的推动者。”胡萍萍说道。

“我一半时间在医院, 从临床实际中捕捉医患需求; 另一半时间在办公室, 设计符合临床需求的大模型, 用AI为诊疗赋能。”刚刚通过成都市首批人工智能训图师社会化职业技能等级认定考试的邓瑶告诉记者, 如今, 越来越多人看到了AI发展带来的职业新机遇, 像她这样深耕“AI+”跨界岗位的从业者正不断增多。

在科大讯飞研究院院长刘聪看来, AI技术正在深刻重构就业生态与职业图景, 推动传统劳动者技能结构升级, 向更高附加值方

向演进, 形成以人机协同、智能赋能为特征的就业新范式。

当AI大模型从实验室走向亿级用户, 其反馈是否得体、有无偏见、是否符合监管要求, 这些曾经的“软问题”, 如今已成为行业发展的“硬指标”。在此背景下, AI伦理官应运而生。这一新职务, 主要负责对AI系统从构思到部署的全生命周期进行伦理审查, 将伦理要求融入产品团队、工程师和数据科学家的日常工作。

复合型、应用型人才更加吃香

2026年一季度, 全球科技行业裁员约7.8万人, 其中近半数被归因于AI替代。然而, 全球AI原生公司的岗位总量非但没有减少, 反而在8个月内从718个涨到了1570个, 增长了超过一倍。

这一反差说明, 智能经济时代的“新饭碗”, 并非简单的存量岗位替换。一边, AI替代了大量的重复性、机械性岗位; 另一边, 它创造出全新的高附加值岗位——这场变革的本质, 不是“岗位消亡”, 而是“岗位分化”。

“新岗位的增长并不是简单‘堆人’, 而是利用AI杠杆提升人数, 才是AI公司核心追求的目标。”腾讯研究院资深专家吴朋阳举例说, 美国AI初创公司Anthropic年化收入增长了约4倍, 在招聘岗位只增长了约1.6倍, 收入增速远超岗位扩张速度, 说明人均产出大幅提升。

有研究表明, AI对就业的影响呈现明显的层级差异: 入门级岗位受冲击较大, 而资深从业者几乎不受影响。斯坦福数字经济实验室发表的《关于人工智能近期就业影响的六个事实》报告显示, AI更容易替代那些

“书本知识”, 但还不太能取代那些随着经验累积而形成的“技巧和窍门”。

“目前, AI公司缺的不是做模型的人, 而是把模型变成外部价值的人。”吴朋阳说道, 同时, 这些新岗位的能力要求也在打破传统边界, 复合能力的需求更广泛, “技术+业务”的双栖能力更受企业欢迎。

记者了解到, 不少企业的AI部署工程师要求“8年售后工程经验+2年软件开发”; 商业分析师要求“精通SQL+Python+推理成本和毛利模型”; 甚至一些AI原生公司的实习生, 都要“熟练使用5种以上AI创作工具”。这一点, 从国内数据标注行业的薪资差异中也能窥见一斑: 一些基础的重复性标注岗位月薪可能在3000~5000元, 而具备医疗、金融、交通法规等专业背景的高级标注员, 年薪可高达50万元。

行业需求的变化, 也推动着岗位结构的调整。去年, 模型开发的核心是训练模型人员; 而现在, 让模型高效落地的推理工程师, 成为业务新增的主力。

“模型训练完成只是第一步, 如何让它在实际部署中跑得更快、更经济, 才是规模化应用阶段的核心瓶颈。”业内人士表示, 无论身处哪个行业, 技术最终要服务于用户, 既懂市场、懂用户, 又能熟练运用AI工具的复合型人才, 正成为市场的“香饽饽”。

为适配复合型人才的需求, 国内高校与企业正加速联动。上海交通大学获批增设具身智能本科专业, 与小米机器人、穹御智能等20余家行业企业开展合作, 推动课程教学、科研训练、工程实践与产业需求有机衔接; 北京理工大学具身智能专业与华为共建人工智能实践实验室, 打造集先进算力、智能开发环境、真实应用场景和工程化训练资源于一体的实践教学方案。