从基础设施到具身智能 中国科技力量闪耀GTC 2026

◎记者 孙小程

3月16日，英伟达公司首席执行官黄仁勋在英伟达GTC大会上发表主题演讲

素有“AI界春晚”之称的英伟达GTC 2026已拉开帷幕。

与往年相比，中国科技领域代表企业参与的广度和深度明显提升，在主题演讲、技术展示和生态合作等多个维度中展现创新实力。

联想集团、吉利汽车、宇树科技……此次与会的中国企业，业务范围涵盖AI基础设施硬件、垂直行业应用及前沿具身智能等多个领域。在与英伟达的合作中，它们实现了从应用到生态的系统性参与。可以预见，中国科技企业正将产业动能与全球创新技术紧密联接，共同加速智能技术的普惠与落地。

助力AI基础设施建设

“今年将会是属于你们的一年，我感觉到了。”在GTC 2026展区上，英伟达首席执行官黄仁勋对联想集团董事长兼CEO杨元庆说。

杨元庆回应称：“联想集团的业务板块非常强劲，甚至比以前更强。”

联想集团与英伟达合作多年，是每年GTC的核心参与者之一。在GTC 2026上，联想集团与英伟达联合发布新一代联想Hybrid AI Advantage（混合式AI优势集）解决方案，旨在加速AI落地、缩短首token时间（TTFT），并在个人、企业和云环境中带来可量化的商业成果。

该解决方案是联想集团与英伟达三十年合作的全新成果，也是混合式AI技术走向产业化的重要标志。杨元庆表示，联想集团和英伟达共同具备独特优势，能够帮助各类机构将AI投入运营——从实验阶段到企业生产，再到AI云级巨型工厂。

除了联想集团，还有多家同样专注于AI基础设施的中国企业展示了最新布局。例如，在液冷领域，领益智造旗下子公司立敏达（Readore）进入新一代Rubin架构Manifold（分水器）生态，展示了包括UQD/MQD快接头、Inner Manifold在内的核心液冷产品。

据领益智造介绍，立敏达是一家以热管理产品为核心的服务器综合硬件方案供应商。立敏达核心客户覆盖海外算力行业头部客户及其供应链相关合作伙伴、相关服务器代工厂、电源解决方案公司，并且已建立深厚且长久的客户合作关系。

赋能汽车产业转型

以人工智能为核心的智能化体验，是近年来汽车产业的角力点。在GTC 2026上，中国汽车产业链的参与凸显了这一趋势。

吉利汽车集团发布了超级Eva——首个打通智能座舱、智能辅助驾驶、数字生态的超级智能体，并宣布极氪8X将首发搭载。超级Eva由吉利、千里科技和阶跃星辰共同研发，其核心驱动大脑来自阶跃星辰自研的Step 3.5 Flash基座模型。

同时，吉利还宣布将携手英伟达在智能辅助驾驶、智能座舱、智能制造与研发、云端与AI基础设施等领域，持续深化合作与战略协同。

除了整车企业，禾赛科技与速腾聚创等激光雷达企业也参与了GTC 2026，重点呈现与英伟达的技术及平台合作。

禾赛科技宣布，公司已加入NVIDIA Halos AI系统检测实验室，这是首个获得美国国家标准学会国家认可委员会认证的全球项目。作为成员单位，禾赛将在这一统一框架下对其激光雷达平台开展功能安全、网络安全以及AI合规方面的评估与验证。

禾赛科技联合创始人及CEO李一帆表示：“只有在安全性与可靠性达到最高标准的前提下，自主系统才能实现广泛应用。随着自动驾驶汽车和智能机器人逐步迈向规模化部署阶段，安全将成为整个系统最核心的要求。”加入NVIDIA Halos生态，将进一步提升自动驾驶系统的安全能力，并推动自动驾驶技术实现规模化应用。

速腾聚创介绍，作为全面加入NVIDIA Jetson、DRIVE、Omniverse三大生态系统的合作伙伴，公司凭借数字化激光雷达产品，与英伟达生态形成深度协同，为全球汽车与机器人头部企业共同打造更高性能的感知解决方案。

目前，搭载速腾聚创“主+补盲”数字化激光雷达产品组合与NVIDIA DRIVE AGX Thor芯片的系统方案，已成为L4级产业伙伴打造下一代自动驾驶系统的优选技术组合。在GTC 2026现场，文远知行、小马智行等均展示了搭载了这一自动驾驶系统方案的Robotaxi车型，展示无人出行服务的全球规模化、商业化运营格局。

共推具身智能落地

黄仁勋从不掩饰对具身智能的看好，他曾言：“具身智能是AI的下一浪潮”。围绕具身智能，英伟达构建了从底层算力、核心模型、仿真训练的全栈技术生态。

GTC 2026上亦出现了多家头部具身智能企业的身影，包括宇树科技、智元机器人、自变量等。

大会披露的英伟达AI生态核心伙伴中，“AI for Robotics”（即具身智能机器人板块）共有Physical Intelligence、Figure、Skild AI、智元机器人、自变量机器人等9家具身智能企业。黄仁勋宣布，将与这些企业开展物理AI的大规模部署，推动智能机器人在工厂、物流、交通、医疗和基础设施等领域应用，加速机器人在真实场景的落地。

在合作方向上，智元机器人介绍，其与英伟达围绕机载超级大脑开展联合硬件适配，支撑百亿参数VLA模型与世界模型在机器人本体本地高效推理，满足电子制造、汽车总装、物流作业等场景下0.1毫米级精密操作与长程自主规划需求。作为英伟达Isaac GR00T N系列全球首批生态伙伴，智元机器人将GR00T模型体系与自研机器人操作系统深度融合，形成可复制的通用技能迁移与灵巧操作部署流程。

宇树科技创始人兼CEO王兴兴发表了主题演讲。他表示，具身智能的ChatGPT时刻，意味着机器人可通过语音或文字指令，在80%的陌生场景中顺利完成约80%的任务，而实现这一目标需要更多全球合作。

黄仁勋：数据中心将成为Token“工厂”

◎记者 窦世平 郑维汉

当地时间3月16日，英伟达GTC 2026大会正式开幕，英伟达创始人兼CEO黄仁勋发表了主题演讲。

针对市场高度关注的订单与营收“天花板”，黄仁勋在大会上给出了极为乐观的业绩预期。“到2027年，至少有1万亿美元的需求。事实上，我们甚至会供不应求，实际的计算需求会比这高得多。”黄仁勋说。

同时，黄仁勋还介绍了Rubin、Groq、Feynman等最新数据和进展，并认为Open-Claw（“龙虾”）是新时代的操作系统，自动驾驶的“ChatGPT时刻”已经到来。

数据中心将成为Token“工厂”

“未来的数据中心不再是存储文件的仓库，而是生产Token（AI生成的基本单位）的‘工厂’。”黄仁勋说。

他解释称，每一座数据中心、每一座工厂，从定义上来说都是受电力限制的。一座1GW（吉瓦）的工厂永远不会变成2GW，这是物理和原子的定律。在固定的功率下，谁的每瓦Token吞吐量最高，谁的生产成本就最低。

Token经济学正在塑造未来的职场新形态。硅谷最新的招聘筹码是：“你的offer里带多少Token？”黄仁勋预计，在未来，公司的每一位工程师都需要一个年度Token预算，他们的基础年薪可能是几十万美元，英伟达会在此基础上再拿出大约一半的金额作为Token额度给他们，让他们实现10x的效率提升。

随着模型越来越大、上下文越来越长，AI会变得更聪明，但Token的生成速率会降低。黄仁勋表示，在这个Token工厂里，吞吐量和Token生成速度将直接转化为明年的精确收入。英伟达的架构能够让客户在免费层实现极高的吞吐量，同时在最高价值的推理层级上，将性能提升惊人的35倍。

目前，英伟达60%的业务来自排名前五的超大型云服务商，另40%的业务则广泛分布于主权云、企业、工业、机器人和边缘计算等各个领域。

LPU预计三季度发货

“过去提到Hopper，我会举起一块芯片，那很可爱。但提到Vera Rubin，大家想到的是整个系统。在这个100%液冷、完全消灭了传统线缆的系统中，过去需要两天安装的机架，现在只需两小时。”提及Vera Rubin系统时，黄仁勋表示。

据黄仁勋介绍，通过极致的端到端软硬件协同设计，Vera Rubin在同一座1GW数据中心里创造了惊人的数据跨越：在短短两年时间内，我们将Token的生成速率从2200万提升到7亿，实现了350倍的增长。

为了解决极速推理（如1000Tokens/秒）条件下的带宽瓶颈，英伟达推出了单柜容纳256个LPU的Groq 3 LPX机柜，计划与今年晚些时候交付客户的Vera Rubin机架式系统并排放置。黄仁勋表示，Groq LPX机柜可使旗下Rubin GPU的每瓦能耗性能提升35倍。

黄仁勋说，英伟达通过Dynamo软件系统，将需要海量计算和显存的“预填充（Pre-fill）”阶段交给Vera Rubin，将对延迟极度敏感的“解码（Decode）”阶段交给Groq。黄仁勋还对企业算力配置给出了建议：“如果你的工作主要是高吞吐，100%使用Vera Rubin；如果你有大量高价值的编程级别的Token生成需求，拿出25%的数据中心规模给Groq。”

据透露，由三星代工的Groq LP30芯片已在量产，预计第三季度出货，而首个Vera Rubin机架已在微软Azure云上运行。

此外，针对光互联技术，黄仁勋展示了全球首款量产的共封装光学（CPO）交换机Spectrum X。“铜缆扩展、南向光学扩展(Scale-Up)、北向光学扩展(Scale-Out)三条路线并行推进，我们需要所有合作伙伴在铜缆、光纤和CPO方面持续扩产。”黄仁勋在介绍其技术路线图时说。

黄仁勋还“剧透”了下一代计算架构Feynman，它将首次实现铜线与CPO的共同部署。同时，英伟达正在研发部署在太空的数据中心计算机“Vera Rubin Space-1”，将数据中心级AI计算能力部署到卫星和轨道数据中心（ODC），并强调其面向在轨推理、实时地理空间智能和自主航天任务。

OpenClaw是新时代“操作系统”

谈及最近风靡AI圈的“龙虾”，黄仁勋将开源项目OpenClaw形容为“人类历史上最受欢迎的开源项目”，称其仅用几周时间就超越了Linux在过去30年取得的成就。他表示，OpenClaw本质上就是Agent计算机的“操作系统”。

“每一个SaaS（软件即服务）公司都将变成AaaS（智能体即服务）公司。”黄仁勋介绍，为了让这种具备访问敏感数据和执行代码能力的智能体安全落地，英伟达推出了企业级的NemoClaw参考设计，增加了策略引擎和隐私路由器。

物理AI则是具身化的智能体。据介绍，本次GTC共有110款机器人亮相，几乎囊括全球所有机器人研发企业。英伟达提供训练计算机、仿真计算机、机载计算机和完整的软件栈及AI模型。

自动驾驶方面，黄仁勋认为，自动驾驶的“ChatGPT时刻”已经到来。“今天，我们宣布四家新合作伙伴加入英伟达RoboTaxiReady平台：比亚迪、现代、日产、吉利，合计年产量达1800万辆。加上此前的奔驰、丰田、通用，阵容进一步壮大。我们同时宣布与Uber达成重大合作，将在多个城市部署并接入RoboTaxi Ready车辆。”他说。

工业机器人方面，ABB、Universal-Robotics、KUKA等众多机器人企业与英伟达合作，将物理AI模型与仿真系统相结合，推动机器人在全球制造产线的落地。

算力大变局

（上接1版）

在GTC大会上，黄仁勋抛出一个极具冲击力的预测：到2027年，市场对英伟达Blackwell和Vera Rubin系统的订单需求将带来至少1万亿美元的营收。这一数字较去年他对2026年5000亿美元市场需求的预测直接翻倍。他甚至强调，1万亿美元只是保守预测，实际全球AI算力需求会比这一数字高得多。

其次，“上天”（太空算力）和“入地”（端侧算力）成为产业演进主线。2026年，人类一边仰望星空，突破地球物理边界；一边深耕大地，让智能渗透进每一个终端。

再次，中美走向两条侧重不同的发展道路。美国侧重于前沿突破，驱动力来自资本与顶尖研发的良性循环。目前，英伟达Vera Rubin平台七颗芯片全部投产，从3nm走向1.6nm，从地面延伸至太空，从GPU扩展至LPU，黄仁勋的每一步都在探索“下一代算力长什么样”。

中国则更重视向深处扎根，驱动力来自国家战略和产业升级的现实需求。“我们的应用空间更广阔，工业场景和商业链路更完整，所以我们是在优先解决算力如何真正‘用起来’。”方海声说。

最后，全球算力加速形成“一个世界，两套系统”的竞争格局。中国主导的生态更开放，有望覆盖新兴市场；美国主导的生态仍牢牢占据高端。

两大隐忧

持续飙涨的资本开支，是全球算力产业的“总开关”。

2026年，全球AI总支出预计达2.52万亿美元，同比增长44%。最近两年，中国各地已掀起智算中心扩建的浪潮。在A股市场，一些上市公司动辄斥资数十亿甚至上百亿元采购算力，以进军算力中心业务。

如火如荼的算力扩建背后，也有人焦虑与担心，狂热的投资会不会一地鸡毛？

近期，记者调研南方某国家级智算中心时，看到“冰火两重天”的景象。机房相关负责人指着左边几乎满负荷的机架说，这是装了英伟达GPU的服务器，出租率90%以上。他又指向右边说，这是装了国产GPU的服务器，价格便宜很多，但出租率不到50%。

记者询问缘由，“机效差距很大，尤其在生态上。”对方回答。

飞腾信息技术有限公司副总经理郭御风也有类似感触。他调研发现，不少智算中心算力利用率不足30%，大量算力资源长期闲置。他认为，行业“重算力、轻应用”“重建设、轻实效”的结构性问题突出。

“当前算力结构供需错配。低端算力过剩而高端智能算力不足；西部通用算力利用率偏低，东部产业急需的智能算力供应紧张。同时，算力孤岛现象严重，跨地域、跨主体算力资源难以高效流通。”北京国际城市发展研究院创始院长连玉明说。

还有业内人士透露：不少地方数据中心配的芯片仍以CPU为主，适配传统IT与云服务场景，难以满足AI大模型训练与推理的需求；有些地方芯片搭配的计算架构不合理，导致应用场景窄，即便电力成本低，也无法使用。

这些现象背后，是中国本轮数据中心建设热潮，既缘于AI真实需求，也与地方政府和资金方追逐热点、盲目上马项目相关。

上天入海

尽管科技巨头对AI的未来相当乐观，但要把一笔笔钱“变成”一个个智算中心，正变得越来越困难。能源供给、散热能力与耗水量正逐渐成为地面算力增长的瓶颈。

于是，人类将目光投向星空。

2月，马斯克旗下SpaceX完成了对AI公司xAI的全资并购，标志着航天与AI进入深度融合阶段。据马斯克预测，未来2至3年太空将成为全球AI算力成本最低的区域。目前，SpaceX、亚马逊、谷歌、英伟达等巨头纷纷布局太空算力。

同一时间，在浙江杭州之江实验室的“三体计算星座”指挥中心，中国工程院院士、之江实验室主任王坚透露，该星座已实现星间组网突破，在轨协同完成10个人工智能模型与应用的部署验证。

“今年‘三体计算星座’还将发射50颗卫星，计划2032年完成1000颗计算卫星组网，形成能互联互通为人工智能服务的太空算力星座。届时，总算力将达每秒百亿亿次。”在王坚看来，把算力送到太空的价值堪比电的发明，将催生诸多想象不到的新价值。

此外，中国还在悄悄试水算力“下海”。2月10日，东海之滨，全球首个“海风直联”海底数据中心在上海临港投产。它将海上风电与海底数据中心直接联通，绿电占比高达95%，利用海水自然冷却。

除了算力的“上天下海”，端侧算力（可称之为“入地”）的发展更是一场触手可及的大变革，2026年将成为驱动消费电子和汽车产业升级的核心引擎。从豆包手机形态到Openclaw带火的Mac Mini，标志性案例层出不穷。汽车具备高算力芯片、人机交互界面和充足电源，成为端侧AI硬件落地的理想场景。

“AI历经多轮演进，已迈入以推理为核心的全新阶段。”3月17日，在2026华为数据存储新春发布会上，华为存储产品线副总裁、闪存领域总裁谢黎明表示。当日，华为正式发布针对AI推理场景的全新AI数据基础设施：面向中心推理场景的AI数据平台，以及面向分支边缘推理场景的FusionCube A1000 AI超融合一体机。

国产提速

算力比拼，首当其冲的就是算力芯片。

岁末年初，国产芯片的“小龙”们——摩尔线程、沐曦股份、壁仞科技、天数智芯先后登陆资本市场，这既是资本市场助力科创企业的最新实践，也是中国算力芯片行业进入资本化的高歌时刻。

“今年芯片供给将从满足‘有没有’，向提供‘好不好、准不准’的差异化、场景化解决方案演进。”天数智芯相关负责人说。

“2026年，芯片会更聚焦易用性、安全性和高能效比。”瀚博半导体创始人兼CTO张磊对记者说，“随着下游AI推理、工业质检、数字孪生等需求释放，国产算力的竞争力正在从硬件参数向全栈解决方案能力拓展。”

国产算力芯片虽然发展迅猛，但要真正做到从“跟跑”到“并跑”，仍任重道远。

在设计与制造环节，先进EDA工具匮乏、高端工艺产能不足，仍是主要瓶颈。多家国产算力芯片公司反映，国产先进工艺产能“抢手”，即便设计出了先进芯片，大规模量产时却颇具难度。“芯片设计公司都是给晶圆厂打工的。”一位半导体行业资深分析师表示，“在‘天花乱坠’的芯片参数外，我们更关心企业能否流片及量产。”

更大的差距在软件生态。一位不愿具名的芯片业内人士坦言：“真正的壁垒不是把芯片做出来，而是让全球开发者愿意用你的软件平台。英伟达之所以能统治市场，不光因为芯片强，更因为它花了20年建立起来的CUDA软件生态。”

“这非常重要。”快思慢想研究院院长田丰说，“英伟达因为有CUDA软件层才使得它的显卡在流体力学、纳米科研算法等细分领域的算法加速表现优异。”

上述芯片业内人士还透露：目前国内大部分的AI大模型开发和训练，依然基于英伟达的CUDA生态；国产算力芯片的软件适配、算子优化、开发工具链完善度，与英伟达存在代际差距。

中国工程院院士邓中翰近日表示，国产高端算力芯片规模化应用面临三大核心挑战：一是技术适配性不足，现有芯片多对标传统架构，与AI大模型、智算集群的多元计算需求不匹配，存在“算力适配难、场景落地贵”的现实痛点；二是生态体系有短板，软硬件协同、标准统一、场景验证未形成闭环，无法满足规模化应用的稳定性要求；三是算力利用效率低，行业“暴力计算”模式大幅拉高能耗成本，也让国产芯片在工艺制程受限的情况下，难以发挥架构创新的优势。

算电协同

在太平洋的另一侧，美国算力产业也有自己的“烦恼”。

日前，一则消息从华盛顿传出：美国政府官员正要求微软、Alphabet等科技公司作出承诺，确保其数据中心不会推高电价，不会给消费者带来其他负担。

这一行动旨在应对全美范围内数据中心扩张引发的政治与公共关系问题。在部分地区，科技公司已遭遇来自民众的日益高涨的抵制声浪。一些维权团体正在积极行动，反对建设高耗能数据中心，理由是这些数据中心会挤占当地基础设施、可用水资源和电力供应。包括亚特兰大和新奥尔良在内的多个地区已经对新建数据中心实施限制措施。

美国的现状揭示了一个全球性难题：算力繁荣的社会成本，正在从隐性走向显性。

如何化解上述问题？“算电协同”概念应运而生。这是2026年中国新基建战略中的核心概念，2026年首次被写入政府工作报告。

有运营商认为，在AI进入“拼电力”时段，中国的“算电协同”模式更有可能在大规模、可持续的绿电算力供给上胜出，因为它解决了能源与算力在地理和时间上的根本性错配。而美国若不解决电网碎片化和扩容缓慢的问题，其AI发展的能源瓶颈可能会日益凸显。对于全球而言，中国的实践提供了一种将数字经济与能源转型深度融合的“中国方案”。

人工智能的浪潮已经势不可挡。作为AI核心基础设施，智能算力的研发、迭代、扩张，算力与电力的协同发展，同样势不可挡。在算力之争就是国力之争的大背景下，算力大变局的戏码，或许才刚刚开始。