

阿里云CTO周靖人：

开源开放是模型生态发展的核心驱动力



图为首届魔搭开发者大会现场

本报记者 宋婧

6月30日，阿里云CTO、魔搭社区发起人周靖人在首届魔搭开发者大会上表示，今天，以模型为中心的技术栈已经形成，并成为新时代的一个核心技术体系。要促进模型技术创新发展，加速产品化进程，必须搭建一个开源开放的生态。

回顾模型发展史，从过去的CPU到今天的GPU，一代一代的新算力架构的推出为模型发展奠定了坚实的基础。与此同时，随着Transformer架构的发布，让海量数据的提取、分析、处理成为现实，模型的能力得到了突飞猛进的发展。可以看到，数据、计算与模型的协同发展，让AI(人工智能)真正成为推动技术变革的核心驱动力。

迈入AI时代，模型成为一个重要的生产元素，各行各业都在因模型的诞生发生翻天覆地的变化。“可以预料到，一方面，所有的技术体系都将围绕模型重新搭建，今后大量的算力、数据，包括云计算的创新，都将围绕模型去提供服务，这将驱动模型不断迭代升级；另一方面，模型将成为一个应用中心，各行各业的开发者将基于模型进行一系列应用创新，最大程度地把模型的力量发挥出来。”周靖人表示。

据悉，早在ChatGPT发布前，阿里云魔搭社区就提出了“模型即

服务”(MaaS, Model as a Service)的理念，将AI模型变为直接可用的服务。如今，这个理念已然成为行业共识。“从过去可能每年有一次技术迭代，到现在几乎每个月甚至每个星期都有各种各样的模型层出不穷，新产品不断发布、新技术不断迭代，整个产业生态呈现繁荣发展态势。如何让更多企业、更多人能够参与到今天的技术变革中，能够去使用最新的模型、尝试最新的技术，这是业界共同面临的重要课题。”周靖人说道。

他指出，开源开放是生态发展的核心驱动力。一方面，开源开放让模型不再是少数几家公司或者少数技术人员所能触达到的，而是让所有人都可以参与进来，共同促进技术创新发展；另一方面，开源开放有效地降低了模型使用门槛，消除了企业对于使用新技术时被迫绑定厂商的担忧，让模型技术更加安全可靠，促进产业应用落地。另外，越来越多的人使用开源模型，也会

给开发者带来更多反馈、建议，从而形成正向循环，推动模型进一步迭代升级。可以说，开源开放推动了大模型蓬勃生态的形成。

以魔搭社区为例，经过短短两年多的发展，魔搭社区已经成为了中国最活跃的AI模型的开源社区之一，从最开始的10家组织发展至2000多家组织、1600万名开发者，从最开始的300多个模型扩展到70000多个模型。记者从此次开发者大会上了解到，魔搭社区已形成全链路服务，支持开发者体验、下载、调优、训练、推理、部署模型，覆盖LLM、对话、语音、文生图、图生视频、AI作曲等多个领域，并提供4000+MCP服务和调试工具。百川智能、上海人工智能实验室书生系列大模型、零一万物Yi模型等业界头部模型均在魔搭社区率先开源。DeepSeek系列模型、阶跃星辰开源的Step-Video-T2V视频生成模型和Step-Audio语音交互模型在魔搭社区首发。

“今天，魔搭已经成为了开源模型社区领域一个风向标。”周靖人表示。据他介绍，通过将先进的开源模型汇聚在社区，魔搭让广大的模型开发者可以迅速获取到最新最好的模型；与此同时，魔搭也为这些模型提供了迅速触及潜在的用户以及下游生态的桥梁，模型贡献者与模型使用者在魔搭上的双向奔赴，使得模型应用的无限可能，得以被激发。此外，通过平台服务以及开源工具链的建设，促进优秀模型的创新，助力模型应用探索，推动中国大模型生态的迭代和发展。

“我们应该很自豪地讲，中国的企业、中国的模型在全球开源社区技术的发展过程中扮演着至关重要的角色。”周靖人表示。他指出，开源是技术创新的重要源泉，未来希望有更多开发者能够从找模型、学模型、玩模型、用模型等各个方面，真正把模型创新做实，为开源社区做出贡献。

量子国际科技产业园启动区在苏州开工

本报讯 6月27日，量子国际科技产业园启动区开工暨量子-原子制造综合创新平台签约活动在苏州市相城区举行。

据了解，量子国际科技产业园是苏州发展量子科技产业的重要抓手，产业园总规划面积约2平方公里，布局生产制造、创新研发、公共服务三大功能板块，构建量子科

技“基础研究+技术攻关+成果转化+金融赋能”全链条产业体系，目标建设量子科技创新策源地、量子产业创新集群高地、量子科技企业集聚高地、量子创新创业生态高地。

此次开工的启动区，预计总投资16亿元，占地面积约9万平方米，整体打造以生产制造为主、科

研为辅、高品质配套为支撑的现代化科技产业园区，建设国内首条量子芯片自动化智能产线。

当天，量子生态项目正式启动，苏州将与中国电子科技集团携手营造优质产业生态，共同推动量子技术创新与产业发展迈向更高能级。量子-原子制造综合创新平台签约，将以“开放协同、自主可

控、产研共融”为核心定位，覆盖创新要素集聚、公共平台建设、核心技术攻关、标杆产品制造等主要功能，着力打造量子-原子制造技术创新策源地与产业化高地。电科投资公司与苏州国投集团签署战略合作协议，将构建“资本赋能+产业成长+生态构建”的央地协同合作新模式。(苏文)

中国移动与中兴通讯“联创+”自智网络开放实验室揭牌

本报讯 7月2日，中国移动通信集团有限公司(以下简称“中国移动”)和中兴通讯股份有限公司(以下简称“中兴通讯”)在北京举行“联创+”自智网络开放实验室揭牌仪式。这是双方贯彻落实国家科技创新驱动发展战略，推进高水平科技自立自强，推动科技创新与产业创新相互促进，加速自智网络高阶发展的重要举措。

中国移动副总经理李慧镭表示，通信产业是推动现代社会发展的基础性、战略性、先导性产业，既是新质生产力的重要组成部分，也是赋能千行百业发展新质生产力的关键基础设施。面对网络结构复杂度提升、业务多形态融合发展、用户体验需求升级等多重挑战，中国移动率先提出“三零三自”的自智网络

理念，迅速凝聚了全球运营商数字化转型的共识，成为通信产业创新驱动发展的最佳实践。

李慧镭指出，自智网络正处于全面迈向L4高阶自智的关键时期，需要将人工智能技术和网络运营深度融合，加快网络运营数字化转型。自智网络开放实验室因智而生、向智而动，要立足关键技术自主可控，攻关突破引领性、原创性技术，努力拓展技术出海机会，不断强化我国在自智网络产业的领先地位。要进一步打造完善自智网络产业链生态，在技术攻关、载体装置共建、产业转化等方面，积极创造更多开放合作机遇，汇集优势力量合力攻坚，促进产业链核心能力提升。实验室工作要突出成效导向，尽快取得成果，强化科技成果从研发到现

网试点、再到全网规模推广的转化机制，组织一线单位对接开展生产环境验证和推广，加速应用落地产生价值。

中兴通讯总裁徐子阳表示，人工智能正在加速科技革命和产业变革，将深刻改变人类生产生活方式。中兴通讯将积极深化同中国移动的全面合作，自智网络是推动双方合作升级的重要战略方向。此次与中国移动携手共建“联创+”自智网络开放实验室意义重大，是进一步深化产业科创合作、实现创新驱动发展的“重要里程碑”。中兴通讯将聚焦“三个体系能力”支撑“联创+”实验室建设，包括依托全专业全栈技术积累推进协同创新，迎难而上攻克高阶自智演进中的核心难题；以价值为导

向深耕场景，实现从网络设备到网管系统的深层协同，拓展研发成果价值转化的新范式；以可组式开放架构共享产业机遇，推进产业生态共建。

中国移动“联创+”自智网络开放实验室，由中国移动集团公司网络事业部携手中兴通讯联合打造，未来将基于“联创+”协议引入更多产业合作伙伴参与共建。实验室聚焦网络与人工智能前沿技术的融合创新，协同产业伙伴共创开放创新的“平台+生态”，研发知识图谱、图神经网络、强推理大模型、多智能体协同等四大技术体系，加速实现自智网络高价值场景端到端自主闭环、自主演进，打造全球领先的高阶自智网络最佳实践。(联和)

华为发布星河AI 高品质万兆园区网络解决方案

本报讯 6月27日，以“无智联，不AI”为主题的品智联接数据通信商业市场创新峰会在深圳举办。会上，华为依托核心技术突破，发布星河AI高品质万兆园区网络解决方案，以及“无智联，不AI”星云计划，并与众多行业客户和伙伴代表一道，共同探讨了品智联接的领先技术和建设理念，深入分享了各行业成功实践与数智化转型经验，全面助力行业数智化转型迈向新高度。

华为数据通信产品线总裁王雷在致辞中表示，2024年，在伙伴和客户的鼎力支持下，数据通信商业市场高速增长。今年，随着年初DeepSeek爆发，AI大模型能力提升、企业训推成本百倍降低，AI已从“精英专属”加速演变为“大众普惠”的新质生产力核心引擎。AI产业跃升和价值释放，离不开数据通信网络的强力支撑。华为始终站在技术突破的最前沿，致力于打造最佳体验的联接。今年，“星河AI网络”实现重磅升级，AI技术深度赋能于网络设备与网络大脑，实现网络自身的智能化重塑，以“智能”提升“品质”。

以企业园区网络为例，华为为数据通信围绕“无线、安全、应用、运维”实现四大体验升级，持续引领体验建网。其中，在无线体验方面，通过创新的“Wi-Fi通感一体”技术，让酒店、楼宇秒变智慧空间，节能30%以上；在安全体验升级方面，华为独家Wi-Fi零入侵，实现“空口数据零泄露”，为政府、企业撑起主动防

御盾牌。这些创新标志着网络正在从“联接万物”向“感知万物”的全新时代演进，未来的每一次数据交互，都将成为价值创造的起点。

作为2025年商业市场Campaign的首站，华为带来了一系列创新产品和解决方案，比如能够感知人员存在的全新Wi-Fi 7-Advanced AP；一周能开通一学校的极简网络架构——PEN无源以太全光网络解决方案；一台设备搞定一门店的新一代超融合网关等。华为将持续携手伙伴，深化面向商业市场的行业和场景创新，共同为客户打造最优体验的网络。

为了更好地助力伙伴，抓住百亿园区新机遇，华为中国政企商业MKT与解决方案销售部部长浦强在《“无智联，不AI”星云计划2025发布》的分享中表示，华为将继续践行“伙伴+华为”体系，并沿着伙伴Lead的作战路径，从研营销供服全方位支撑伙伴高效作战。华为聚焦打造行业领先的星河AI园区网络解决方案和代际领先的明星产品，开放上百场营销活动、提供300万元的联合营销基金支持以及40多个样板点研学游等营销资源。

在营销服务方面，沿着项目全生命周期支撑伙伴作战，今年会向伙伴共享36亿数通产业商机，同时提供从网络规划、项目投标到运维交付的多种平台工具。除此之外，还会全阶段使能伙伴，培养7000+名数据通信产业工程师。(华文)

腾讯混元推出 开源混合推理模型

本报讯 6月27日，腾讯混元宣布开源首个混合推理MoE模型Hunyuan-A13B，总参数80B，激活参数仅13B。模型已经在GitHub和Huggingface等开源社区上线，同时模型API也在腾讯云官网正式上线，支持快速接入部署。

据介绍，这是业界首个13B级别的MoE开源混合推理模型，基于先进的模型架构，Hunyuan-A13B表现出强大的通用能力，并且在Agent工具调用和长文能力上有突出表现。

对于时下热门的大模型Agent能力，腾讯混元建设了一套多Agent数据合成框架，接入了MCP、沙箱、大语言模型模拟等多样环境，并且通过强化学习让Agent在多种环境里进行自主探索与学习，进一步提升了Hunyuan-A13B的效果。

在长文方面，Hunyuan-A13B支持256K原生上下文窗口，在多个长文数据集取得了优异的成绩。

在实际使用场景中，Hunyuan-A13B模型可以根据需要选择思考模式，快思考模式提供简洁、高效的输出，适合追求速度和最小计算开销的简单任务；慢思考涉及更深、更全面的推理步骤，如反思和回溯。这种融合推理模式优化了计算资源分配，使用户能够通过加think/no_think切换思考模式，在效率和特定任务准确性之间取得平衡。

Hunyuan-A13B模型对个人开发者较为友好，在严格条件下，只需要1张中低端GPU卡即可部署。目前，Hunyuan-A13B已经融入开

源主流推理框架生态，无损支持多种量化格式，在相同输入输出规模上，整体吞吐是前沿开源模型的2倍以上。

Hunyuan-A13B集合了腾讯混元在模型预训练、后训练等多个环节的创新能力，这些技术共同增强了其推理性能、灵活性和推理效率。

预训练环节，Hunyuan-A13B训练了20T tokens的语料，覆盖了多个领域。高质量的语料显著提升了模型通用能力。此外，在模型架构上，腾讯混元团队通过系统性构建，建模与验证，构建了适用于MoE架构的Scaling Law联合公式。这一发现完善了MoE架构的Scaling Law理论体系，并为MoE架构设计提供了可量化的工程化指导，也极大地提升了模型预训练的效果。

后训练环节，Hunyuan-A13B采用了多阶段的训练方式，提升了模型的推理能力，同时兼顾了模型创作、理解、Agent等通用能力。

为更好地提升大语言模型能力，腾讯混元也开源了两个新的数据集。其中，ArtifactsBench用于弥合大语言模型代码生成评估中的视觉与交互鸿沟，构建了一个包含1825个任务的新基准，涵盖了从网页开发、数据可视化到交互式游戏等九大领域，并按难度分级以全面评估模型的能力；C3-Bench针对Agent场景模型面临的三个关键挑战：规划复杂的工具关系、处理关键的隐藏信息以及动态路径决策，设计了1024条测试数据，以发现模型能力的不足。(讯文)

百度文心大模型4.5系列 正式开源

本报讯 近日，百度正式开源文心大模型4.5系列模型，涵盖47B、3B激活参数的混合专家(MoE)模型与0.3B参数的稠密模型等10款模型，并实现预训练权重和推理代码的完全开源。目前，文心大模型4.5开源系列已在飞桨星河社区、HuggingFace等平台下载部署使用，同时开源模型API服务也可在百度智能云千帆大模型平台使用。据悉，此次文心大模型4.5系列开源后，百度实现了框架层与模型层的“双层开源”。

文心大模型4.5开源系列针对MoE架构提出了一种创新性的多模态异构模型结构，适用于从大语言模型向多模态模型的持续预训练范式，在保持甚至提升文本任务性能的基础上，显著增强了多模态

理解能力，其优越的性能主要得益于多模态混合专家模型预训练、高效训练推理框架和针对模态的后训练等关键技术点。文心大模型4.5开源系列均使用飞桨深度学习框架进行高效训练、推理和部署。在大语言模型的预训练中，模型FLOPs利用率达到47%。

据了解，文心大模型4.5开源系列权重按照Apache 2.0协议开源，支持开展学术研究和产业应用。此外，基于飞桨提供开源的产业级开发套件，广泛兼容多种芯片。此外，百度还同步升级发布文心大模型开发套件ERNIEKit和大模型高效部署套件FastDeploy，为文心大模型4.5系列及开发者提供开箱即用的工具和全流程支持。(文编)