



# “人工智能+”行动正当时

## DeepSeek 的一小步 国产 AI 芯片生态突围的一大步

中经记者 李玉洋 上海报道

在人工智能训练和推理加速的竞赛中,浮点数(Floating Point)的表示方式正成为关键突破口。

随着 DeepSeek-V3.1 的发布,算力“破壁人”——深度求索(DeepSeek)在这方面更进一步:继成功训练出世界首个使用 FP8(8 位浮点数)精度的开源大模型 DeepSeek-V3 后,在 DeepSeek-V3.1 中又使用了 UE8M0 FP8 Scale 的参数精度,并置顶留言“UE8M0 FP8 是针

对即将发布的下一代国产芯片设计的”。

UE8M0 FP8 是什么?下一代国产芯片又是什么?这个充满谜团的声明带火了 UE8M0 FP8,也迅速传导至 A 股资本市场,国产芯片及半导体上市公司股价应声大涨,其中寒武纪(688256.SH)一路高歌猛进,总市值已跃居科创板头名。

《中国经营报》记者注意到,DeepSeek V3/R1、Kimi-K2 等主流开源大模型均原生支持 FP8 的低精度格式,这代表了低精度

### 实现两倍的浮点算力提升

DeepSeek 通过 DeepGEMM 开源库实现了工程化突破,为整个国产生态在大模型中的落地提供了可能。

计算机里的小数是用科学计数法表示,只不过不是 10 的次方,而是 2 的次方。浮点数则是计算机用于表示小数的核心手段,由符号位(Sign)、指数(Exponent)和尾数(Mantissa)三部分构成。其中,符号位决定正负,指数决定小数点的位置,尾数影响精度。

在 AI 计算领域,数据精度格式长期被国际巨头所主导。比如眼下被业内推崇的 FP8(FP 代表浮点数,8 代表数据用 8bit 表示 0、1),是一种较新的数据格式,能显著降低显存占用和计算资源需求,同时保持较高的计算精度。

DeepSeek 在开源社区 Hug-ginface 放出信息:DeepSeek-V3.1 使用 UE8M0 FP8 缩放数据格式进行训练,以确保与微缩放数据格式兼容。这里所谓“微缩放数据格式”,即业界的 Microscaling FP8 (MXFP8) 标准。这是 Open

Compute Project 在 2023 年发布的 Microscaling (MX) Formats Specification v1.0 里定义的 8bit 微缩放格式。

UE8M0 是 FP8 的一种特殊的数字表示格式。U 表示没有符号,E8 表示 8 位指数,M0 表示没有尾数。DeepSeek 在 V3.1 模型中提出的 UE8M0 FP8,相当于一条别人没走过的路。有观点认为,DeepSeek-V3.1 把原本只在硬件/内核实现层面存在的 scale 表达方式,提升到模型部署和训练标准,并适配国产下一代芯片,需要说明的是,UE8M0 FP8 并非 DeepSeek 独创,但 DeepSeek 通过 DeepGEMM 开源库实现了工程化突破,为整个国产生态在大模型中的落地提供了可能。

摩尔线程方面表示,UE8M0 是指无符号 8 位指数的 FP8 格式,用于对 FP8 张量数据进行分块缩

## AI+产业发展提速:智算规模提升 应用多点开花

中经记者 谭伦 北京报道

随着新一轮科技革命与产业变革的加速演进,我国人工智能产业正走向产业应用阶段,成为

### 筑牢 AI+底座 算力基础设施加速完善

“人工智能+”行动的顺利推进,离不开强大的算力支撑。

工业和信息化部最新数据显示,截至 2025 年 6 月底,我国在用算力中心标准机架达 1085 万架,智能算力规模达 788EFLOPS(每秒百亿亿次浮点运算),干线 400G 端口数量大幅增加至 14060 个,存力总规模超过 1680EB(艾字节,存储容量的度量单位),全国算力中心平均电能利用效率(PUE)降至

推动经济社会高质量发展的新质生产力。

8 月 26 日,国务院印发《关于深入实施“人工智能+”行动的意见》(以下简称《意见》),其中,不仅

### 智能新业态剧增 AI 应用场景多点开花

如果说智算基础设施是“人工智能+”的坚实底座,那么在产业领域的深度应用,则是其价值的最终体现。当前,我国“人工智能+”正从平台能力向行业深度渗透,医疗、教育、制造、金融和消费服务等领域涌现出一批可复制的典型案例。“人工智能+”产业应用正呈现出从试点走向规模、从通用走向行业的发展走势。

在制造与矿业场景中,行业大模型正与多模态感知结合,把“经验密集型”作业转为“数据驱

动型”流程。以“盘古矿山”为代表的行业模型为例,记者从华为了解到,其围绕带式输送、井下巡检、选煤智能控制等环节进行“感知—决策—执行”的闭环重构,目标直指安全生产、良率与能耗等核心指标的提升;同时,在电力系统,国家电网与南方电网相继发布了行业大模型,聚焦设备状态评估、隐患识别、负荷预测、调度辅助与应急处理等,推进了数字化资产全生命周期管理。

在通信网络与城市基础设施

### 算力与应用并进 产业链聚焦攻克痛点

伴随算力加速部署与产业应用落地的双轮并进,我国 AI 产业也在迈入新的发展阶段。

算力方面,记者注意到,此次《意见》在强化八项基础支撑能力中明确要求,将优化国家智算资源布局,完善全国一体化算力网,充分发挥“东数西算”国家枢纽作用,加大力、算、电、网等资源协同。加强智能算力互联互通和供需匹配,创新智能算力基础设施运营模式。

对此,杨光认为,此次《意见》关注到了目前我国算力行业的最大痛点,即算力利用率不均衡问题依然突出。部分西部智算中心利用率不足,而东部算力需求高企。如何实现跨区域调度,也是全行业正在重点聚焦的难点问题。

为了加强协同,记者注意到,国家级综合性算力服务平台——中国算力平台目前已正式完成山西、辽宁、上海、江苏、浙江、山东、

河南、青海、宁夏、新疆 10 个省区市分平台接入工作。工业和信息化部副部长熊继军表示,下一步,工信部将推动完善算力布局政策体系,优化布局算力基础设施,引导各地合理布局智能算力设施,持续开展国家绿色数据中心建设。

在提升 AI 基础能力方面,记者也注意到,《意见》着重强调,未来要提升模型基础能力。加强人

工智能基础理论研究,支持多路径

技术探索和模型基础架构创新。

加快研究更加高效的模型训练和推理方法,积极推动理论创新、技术创新、工程创新协同发展。

对此,国内大模型创新企业传神语联创始人何恩培向记者解读称,目前 AI 已经到了可以开展场景化应用落地阶段,并深入赋能核心生产环节,提升效率,优化乃至重塑流程,但我们还应该清醒地看到,今天大模型依然处在人工智能

应用早期阶段,还远远没有到成熟定型的阶段,因此,在兼顾应用的同时,的确应继续加强发展基础大模型,并加强专业大模型研发。

虽然挑战重重,但总体来看,对于《意见》的出台,产业链普遍认为,随着此次政策路线的发布,我国人工智能产业自身的技术演进、应用落地和生态完善,有望在未来几年迎来加速发展,并成为我国经济高质量增长的重要引擎。

应用早期阶段,还远远没有到成熟定型的阶段,因此,在兼顾应用的同时,的确应继续加强发展基础大模型,并加强专业大模型研发。

虽然挑战重重,但总体来看,对于《意见》的出台,产业链普遍认为,随着此次政策路线的发布,我国人工智能产业自身的技术演进、应用落地和生态完善,有望在未来几年迎来加速发展,并成为我国经济高质量增长的重要引擎