

编者按

2026年,哪些技术将引领半导体产业创新浪潮?新年伊始,Arm发布了洞见2026年及未来发展趋势的20项技术预测,在这篇关于全球计算技术格局的文章中,有超过半数的预测与AI相关。这场颠覆计算产业的AI变革,将给产业带来何种影响?围绕AI给计算产业带来的深刻变革,《中国电子报》常务副总编辑连晓东与Arm中国区业务全球副总裁邹挺展开了深度对话。

Arm中国区业务全球副总裁邹挺:

AI竞赛进入“三重比拼”,携手中国合作伙伴拥抱AI时代机遇

本报记者 姬晓婷

“谁能在中国做好AI,谁就更能理解AI的未来。”
——邹挺

性能、能效、迭代速度

缺一不可

“真正的AI竞争,将是性能、能效和速度的三重比拼。”

“AI浪潮席卷全球,中国产业正积极拥抱这场变革,其创新活力与迭代速度令人瞩目,同时也对底层技术提出了更高的要求。”当被问及AI变革带来的影响时,Arm中国区业务全球副总裁邹挺脱口而出。

这一判断基于Arm 2025年发布的《人工智能就绪指数报告》,该报告对全球八大市场的655位企业领导者展开了深度调研,其中包括100家中国企业,覆盖智能制造、科技、能源等核心领域。在这份报告中,有一组数据颇受瞩目:在参与调研的中国企业中,100%的企业已明确制定AI策略;95%的企业计划在未来三年进一步加大对AI的投入力度;更有30%的企业直言,拥抱AI已是“刻不容缓”的战略抉择。

在这场“刻不容缓”的变革中,AI会带来什么?

邹挺指出,当前AI产业的竞争,已不再是单一维度的比拼,而是性能、能效与迭代速度的三重较量。AI加速了开发者们将想法落地的进程,也给市场竞争按下了快进键。在视时间如生命的半导体产业,技术创新的速度与质量,关乎企业能否占得先机。

同时, AI带来的高算力需求,对云、边、端全场景的芯片提出了更高挑战:既要承载更多计算负载,又要尽可能降低功耗,实现性能与效率的平衡。

以数据中心为例,为应对不断攀升的算力需求,芯片开发者不断提升CPU的核数,以提供更高的性能;与此同时,市场对AI、网络等专用加速器的需求同步增长。“这就需要将加速器与通用计算引擎高效集成,灵活搭配,以实现性能与效率的最大化。”邹挺说道。

面对这一复杂需求,邹挺结合Arm计算平台的案例,分享了经过验证的解决方案:通过预配置、预验证的Neoverse计算子系统(CSS),为芯片设计提供了高效捷径,Arm提供优化的计算核心组件,让合作伙伴能够聚焦在加速器等差异化创新领域。据介绍,Arm Neoverse CSS帮助某合作伙伴节省了80人/年的工程师时间;同时通过新的架构功能提供更高的性能和效率,为数据中心市场带来更优越的单芯片性能和单TCO性能表现。

除数据中心外,这一标准化、预集成的计算平台模式,也惠及了汽车、智能终端等领域。在汽车领域,Arm去年推出了Arm Zena CSS,这一标准化且预先集成的计算平台,能将芯片开发周期进一步缩短长达12个月,并使每个芯片项目的工程资源投入减少多达20%,可有效加速整个行业的创新进程。

“不同行业的应用场景看起来差异很大,但它们对于算力的基础诉求非常一致——高性能、高能效、高安全。Arm的计算子系统,就是把这套‘底层共性’提前打磨好,让客户不用每次都从零开始。”邹挺说道。

过去开发一颗芯片,要花大量时间在基础架构和验证上。但现在,通过标准化、预集成的平台,客户可以直接把精力放在各自的创新和差异化上,从而能够加速创新,拥抱AI时代的机遇。

云边端加速融合

端侧智能成AI前沿阵地

“端侧不是云的替代,而是用户体验的前线。”

AI时代已至,每个人都在期待其带来的实惠。而终端,是距离每一位用户最近的AI入口。如何使每位用户切实体验到端侧AI带来的便捷,是科技行业工作者的新课题。

在这样的共识下, AI运行在云端与边缘侧孰优的长期争论正逐渐平息。邹挺表示:“企业不再把云端、边缘侧和物理终端割裂看



邹挺在2025 WAIC期间分享《人工智能就绪指数报告》

待,而是将整个加速系统看作是以协同智能为核心的一体化协作体系。”

在这套体系内,云端负责大规模训练与优化;边缘侧在数据源头附近实现低延迟感知与短周期决策;机器人、汽车及工业设备等物理系统则完成最终执行。

在邹挺看来,在这套分布式的AI模式中,端侧智能的作用不容小觑。为何?

其一,云边端融合虽是行业共识,但云端模式存在持续算力成本,以及数据上云的安全隐患。对消费者而言,随着AI应用日益增多,端侧部署的安全优势会成为一大卖点;对终端厂商来说,端侧处理既能减少云算力调用开支,也能减少云端数据传输的安全风险。

其二,运行速度是端侧AI的生命线。没有用户能够接受过长的响应时长。推理速度不达预期的端侧模型,是不具备市场生命力的。

也正因如此,邹挺坦言:“AI要真正改变生活,必须从云端‘落地’,扎根在每一台设备里。端侧不是云的替代,而是用户体验的前线。”

在他看来,端侧AI落地既需要深耕底层技术,更需要立足用户需求、下沉应用场景,这也是当前产业发展的核心方向。目前已有不少企业参与其中,Arm便是其中的典型代表。

以智能终端行业为例,2025年9月,Arm推出其面向旗舰级智能手机及下一代个人电脑的Arm Lumex CSS平台,该平台支持实时端侧AI用例,覆盖智能助手、语音翻译及个性化服务等多元场景,依托搭载第二代可伸缩矩阵扩展(SME2)技术的Arm CPU,该平台可实现高达5倍的AI性能提升。

在技术应用层面,SME2技术实现了与终端、应用的同步推进。例如,Arm与支付宝及vivo密切协作,在Lumex CSS产品发布时,支付宝已在vivo新一代旗舰智能手机上完成了基于Arm SME2技术的大语言模型推理验证。结果显示,手机在预填充(prefill)与解码(decode)阶段的性能分别实现了40%和25%以上的提升。又如,Arm与腾讯GIGI-NEX针对SME2展开合作,据介绍,腾讯初步测试结果显示,启用SME2后性能提升高达2.5倍。

对于终端用户而言,应用体验变得更加迅速、本地化且安全可信。对开发者而言,鉴于实时推理不再导致电量快速耗尽或手机过热,他们能够在相同的散热与功耗限制下挖掘更多的可能性。

越是“大生态”

越有生命力

“AI时代没有单打独斗,只有共赢共生。”

AI时代,企业发展的着力点发生了何种变化?

对于这一问题,邹挺坦言:“越是具备‘大生态’的企业,越具有生命力。”在他看来,AI的复杂度,已经远远超出了任何一家企业单

资源堆叠,而是通过高效协同,成为效率的放大器。”

与广大人才一道 推动AI创新与落地

“谁能在中国做好AI,谁就更能理解AI的未来。”

人才,是一家企业对一个地区最长远的投资。对于一家有志于建立广泛生态伙伴体系的企业而言,生态伙伴的培育都是“从娃娃抓起”的。

谈及当前中国AI产业的发展支撑,“人才”成为邹挺回答中的关键词。他表示,Arm致力于人才培养与产学研合作,“让更多人才在Arm平台上成长、创新、成功。”

2026年1月12日,Arm与清华大学经济管理学院正式签署产学研合作协议。双方将在课程共建、教材开发、人才交流等多维度深化合作。Arm还将捐赠专项资金,用于采购搭载Arm架构的国产服务器,从底层基础设施支撑学院的教学科研与大模型部署工作。

据了解,Arm在中国的人才布局已持续20余年,自进入中国市场的第二年起,Arm便启动了大学计划,并在后续不断拓展和深化。截至目前,Arm大学计划已覆盖全国600多所高校,建成100余个联合实验室,开发了1700多门技术课程,每年惠及超10万名学生,助力其提升技术创新与实践能力。

“技术进步靠突破,生态进步靠共建。人才是生态中最长期、最稀缺,也是最重要的资源。”邹挺慨叹道。

在中国做好AI 共赴未来前沿

关于Arm在中国的“三年计划”,邹挺还提出了两点业务上的期许:其一,让Arm成为中国AI创新的首选平台——无论是手机、汽车、机器人还是工业设备;其二,持续推进中国的软件生态升级——让模型、框架、系统都能“即插即用”。

长期服务中国市场的经历,让邹挺深感其独特魅力。“中国市场新技术落地速度快、企业执行力强、创新密度高。AI在这里不是‘试验品’,而是推动产业升级的核心力量。这意味着中国不仅是AI技术创新的关键市场,更是AI价值实现的核心阵地。”

在采访接近尾声时,我们提了这样一个问题:“对于Arm在中国的布局、发展,有何愿景?”

“和中国伙伴一起,加速把AI的创新实践和落地成果推向全球,让AI智能体验触达每一个角落、每一个人。”邹挺满怀期待地说道。

毕竟,谁能在中国做好AI,谁就能更理解AI的未来。



图为Arm Unlocked 2025深圳站现场案例展示