

云天励飞董事长陈宁:聚焦AI芯片战略

推动三大场景规模化落地

日前,在2025世界人工智能大会期间,云天励飞宣布,公司将在未来战略上全面聚焦AI芯片。云天励飞董事长陈宁在接受中国证券报记者时表示,未来,公司将围绕边缘计算、云端大模型推理、具身智能三大核心布局,打造国产算力“加速器”。

●本报记者 齐金钊

十一年铸就研发根基

云天励飞成立于2014年,2023年在上交所科创板上市。过去11年,云天励飞一直致力于推动AI的大规模落地应用,已在消费级、企业级和行业级三大领域取得多项创新成果。

据陈宁回忆,云天励飞刚成立的时候,初心就是通过NPU(嵌入式神经网络处理)降低AI算法计算成本,当时还用NPU课题申报了人才引进项目,并且获得了第一名,得到了研发资金的支持。可以说,芯片技术,正是云天励飞初期获得投资的关键。

2018年,云天励飞第二代具有自主知识产权的神经网络处理器芯片正式投片。该芯片采用嵌入式前端和边缘计算应用设计,广泛应用于智能相机、警务终端、机器人及工业智能检测等领域,并通过产品销售和IP授权等方式实现市场化应用。

凭借前两代芯片积累的技术实力,云天励飞在2019年获得了国家发改委、科技部和工信部的芯片项目,

其中科技部项目已圆满完成验收。2020年,云天励飞成为首家采用国产Chiplet(芯粒)工艺的AI芯片公司。

2023年,云天励飞发布第三代芯片产品DeepEdge10。该芯片在研发初期便针对大模型时代的推理计算需求进行布局,采用创新的“算力积木”AI芯片架构。通过支持多芯粒扩展的Chiplet技术以及C2C Mesh Torus互连技术,可以将多个DeepEdge10标准计算单元像搭积木一样,封装成不同算力的芯片和多芯片互连的可扩展计算系统,覆盖8T至256T算力需求,可实现各个不同参数量的大模型高效推理,赋能各类智算推理硬件产品。

如今,DeepEdge10芯片平台已成功适配DeepSeek R1系列模型、国产鸿蒙操作系统以及QwQ-32B模型。“经过11年发展,云天励飞已经推出五代NPU,并针对边缘推理场景、大模型推理场景、具身智能场景推出系列产品矩阵。”陈宁表示。



云天励飞研发的系列产品

公司供图

陈宁认为,2025年是人工智能发展的重要转折点。其主要表现为大模型技术达到新的成熟高度,模型调用成本显著降低,使得AI从“专家工具”逐步演变为“全民基础设施”。这不仅标志着AI从训练时代迈入推理时代,也预示着推理算力需求将迎来爆发式增长。

“如果说训练芯片是AI时代的‘发

电厂’,那么推理芯片就是‘电网’。一座城市只需几座发电厂,但用电设备却数以亿计。”陈宁称,未来基础大模型可能仅存少数,但推理算力需求将呈指数级增长。

在陈宁看来,虽然当前AI推理芯片市场仍处于蓝海阶段,其市场规模相较于数千亿美元的训练芯片市场尚显微小,但预计未来三至五年内,AI

推理芯片的增长速度将大幅超越训练芯片。

云天励飞预测,未来3至5年,AI可能全面重塑各类电子产品——从智能耳机、智能眼镜等可穿戴设备,到扫地机器人、冰箱等家用电器,再到电动汽车,几乎所有电子设备的形态与功能都将被重新定义。这些设备将具备“主动工作”的能力,能像人类一样开口交流,理解复杂

指令、主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

陈宁认为,推理芯片正在成为推动AI应用规模化部署的核心动力。在此背景下,云天励飞正围绕未来AI推理需求,聚焦三个关键场景:边缘计算、云端大模型推理加速以及具身智能,全面构建高性价比的AI推理芯片技术与产品体系。

具体来看,在边缘计算领域,云天励飞历经十余年产品实践,已积累大量Know-how(技术诀窍)。目前公司在高密度视频智能分析、小规模边缘大模

型推理等典型场景,已实现多个示范项目落地,并获得行业头部客户的高度认可。凭借在低功耗、高性能、模块化部署方面的优势,云天励飞的边缘芯片产品正持续拓展其在城市治理、智慧交通等场景的市场应用。

在云端推理加速领域,云天励飞已经推出基于国产工艺打造的推理加速卡,采用“算力积木”架构;通过Chiplet封装多颗中小算力芯粒,实现模块化扩

展与任务并行,配合自研软件栈和工具链,在算力调度、成本控制和封装灵活性方面实现突破。该架构不仅支持一次设计、多种封装,也显著提升了芯片的适配效率与产品迭代速度。

在具身智能领域,云天励飞已与十余家机器人厂商建立合作关系,边缘芯片产品应用于服务型机器人已进入实际部署阶段。随着VLA(视觉-语言-动作)大模型的不断演进,在未来3到5年内具身智能

将成为机器人产业新的增长引擎。

面向未来,陈宁表示,围绕边缘计算、云端推理和具身智能三大应用场景,云天励飞正在构建一个涵盖“高性能、低成本、强适配”三大优势的国产AI推理芯片体系。公司致力于成为中国AI推理芯片的领军企业,打造面向AI大模型时代的关键“加速器”,通过高性价比国产品算力,推动人工智能在各类场景中的规模化落地与快速发展。

陈宁认为,推理芯片正在成为推动AI应用规模化部署的核心动力。在此背景下,云天励飞正围绕未来AI推理需求,聚焦三个关键场景:边缘计算、云端大模型推理加速以及具身智能,全面构建高性价比的AI推理芯片技术与产品体系。

具体来看,在边缘计算领域,云天励

飞历经十余年产品实践,已积累大量

Know-how(技术诀窍)。目前公司在

高密度视频智能分析、小规模边缘大模

型推理等典型场景,已实现多个示范项

目落地,并获得行业头部客户的高度认

可。凭借在低功耗、高性能、模块化部署

方面的优势,云天励飞的边缘芯片产品

正持续拓展其在城市治理、智慧交通等

场景的市场应用。

在云端推理加速领域,云天励飞已

经推出基于国产工艺打造的推理加速

卡,采用“算力积木”架构;通过Chiplet

封装多颗中小算力芯粒,实现模块化扩

展与任务并行,配合自研软件栈和工具

链,在算力调度、成本控制和封装灵活

性方面实现突破。该架构不仅支持一次

设计、多种封装,也显著提升了芯片的适

配效率与产品迭代速度。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余

家机器人厂商建立合作关系,边缘芯片

产品应用于服务型机器人已进入实际部

署阶段。随着VLA(视觉-语言-动作)

大模型的不断演进,在未来3到5年内具

身智能

能力,能像人类一样开口交流,理解复杂

指令,主动回应需求甚至解决实际问题。

陈宁表示,上述场景的落地,都离不开AI推理芯片的底层支撑。它将贯穿端、边、云,形成无处不在的算力网络。正是这种“全域覆盖”的算力基础,让对话式AI能在各类设备中高效运行,最终推动电子产品从“工具”向“智能伙伴”跨越,让各类Agent(智能体)能够服务于人们的工作和生活。

在具身智能领域,云天励飞已与十余