

科技前沿信息

第21期 (总第93期)

上海科技成果转化促进会
上海科学技术情报研究所 主办

2017年11月 上旬刊

【专题报道】

人工智能（AI）芯片发展趋势与布局

[导读] 随着云计算的广泛推广，人工智能对于计算能力的要求不断快速提升，人工智能核心计算芯片应运而生。当前人工智能大热，给芯片带了机遇，也充满了挑战。

人工智能（AI）芯片发展趋势与布局

2017年7月，国务院颁布的《新一代人工智能发展规划》中明确提出，要重点突破智能芯片与系统等一批关键技术，建立新一代人工智能关键共性技术体系；8月中旬，由中科院计算所孵化的人工智能芯片初创公司寒武纪科技获得1亿美元的A轮融资，成为该领域的第一个“独角兽”；9月，华为发布全球首款人工智能手机芯片麒麟970（Kirin 970），通过集成寒武纪1A处理器，使华为旗舰手机Mate10具备了强大的本地智能处理能力；11月6日，寒武纪科技发布了最新研发的三款可以深度学习的新一代人工智能芯片……在这短短半年时间里，人工智能芯片领域亮点频现，正成为人们争相热议的焦点。可以说，核心芯片正是未来人工智能时代的战略制高点。

一、人工智能芯片定义

人工智能（AI）的三大关键基础要素是数据、算法和硬件计算能力。2007年以前，受算法、数据等限制，人工智能对芯片并没有特别强烈的需求，通用中央处理器（CPU）即可提供足够的计算能力；但随着云计算的广泛推广，特别是深度学习成为当前人工智能研究和运用的主流方式，人工智能对于计算能力的要求不断快速提升，人工智能核心计算芯片应运而生。

人工智能核心计算芯片，主要用来处理人工智能应用中大量的训练（Training）或执行/推理（Inference）的计算任务。训练指利用大量样本数据训练出一个复杂的深度神经网络模型（算法）；推理指执行训练好的模型解读现实世界的的数据，从而推理出各种结论。

二、人工智能芯片发展现状与趋势

回顾计算机行业发展史，新的计算模式往往催生新的专用计算芯片。人工智能时代新计算的强大需求，正在催生出新的专用计算芯片。目前，根据计算模式，人工智能芯片的发展分为两个方向：

一是从功能层面模仿大脑的能力，仍采用传统的冯诺依曼结构。主要人工智能芯片产品就是通常人们熟知的中央处理器（CPU）、现场可编程门阵列（FPGA，是一种新型的可编程逻辑器件。其设计初衷是为了实现半定制芯片的功能，即硬件结构可根据需要实时配置灵活改变）、专用定制芯片（ASIC）等。

二是从结构层面去模拟大脑，参考人脑神经元结构和人脑感知认知方式来设计芯片，俗称“类脑芯片”。目标是开发出打破冯诺依曼体系的芯片，如IBM的TrueNorth，它的内存、中央处理器和通信部件完全集成在一起，因此信息的处理在本地进行，克服了传统计算机内存

与中央处理器之间的瓶颈。

从现在这个发展阶段来看,类脑芯片距离现实应用仍有较大差距,图形处理器/现场可编程门阵列(GPU/FPGA)等通用芯片是目前主要的人工智能芯片,此外低功耗、低延迟、低成本的专用定制芯片也有很大市场。

三、中美主要厂商人工智能芯片布局

可以预见,核心芯片将决定一个新的计算时代的基础架构和未来生态,因此,谷歌、微软、IBM、Facebook等全球巨头都投巨资加速人工智能核心芯片的研发,旨在抢占新计算时代的战略制高点,掌控人工智能时代主导权。

美国巨头企业凭借芯片领域多年的领先地位,迅速切入人工智能领域,积极布局,四处开花,目前处于引领产业发展的地位,并且在中央处理器和现场可编程门阵列方面是完全垄断地位。中国人工智能芯片厂商以中小公司为主,没有巨头,多集中于设备端人工智能专用定制芯片的开发,并已有所建树。例如,寒武纪成为全球人工智能芯片领域第一个独角兽初创公司,其专用处理器芯片(NPU)已被应用于首款手机人工智能——麒麟970;上海酷芯微电子开发的无人机芯片得到了业界广泛认可,是大疆无人机的芯片供应商,全球一半以上的中高端无人机,采用了酷芯微电子的芯片。

在类脑芯片领域,我国也有异军突起之势,如上海西井科技,已开发出一款具有目前全球已知最大神经元规模——100亿数量的神经元仿真器,实现实时仿真,可用于脑科学研究,包括帕金森氏综合症、阿兹海默症等神经元引起的疾病、神经性受损创伤后的术后康复治疗。

但是,我国在现场可编程门阵列、图形处理器领域缺乏有竞争力的原创产品,而是基于图形处理器/现场可编程门阵列(GPU/FPGA)

做进一步开发，这主要与我国在芯片领域一直缺乏关键核心自主技术有关。图形处理器/现场可编程门阵列的技术壁垒已很高，很难有所突破。

据腾讯研究院发布的《中美两国人工智能产业发展全面解读报告》，美国投资者更关注基础层，芯片/处理器领域融资308.18亿元，占比排名第一，占比31%。中国投资者则更关注应用层，在芯片/处理器领域的投资仅为13.28亿元，占比排名在最后。这也在一定程度上揭示了中美芯片实力和资金吸引力的差距。

四、总结与建议

1、在面向智能手机、智能摄像头、机器人/无人机、自动驾驶、虚拟现实（VR）等设备端人工智能芯片方面，目前多采用专用定制芯片，还未形成一家独大的态势。我国在这一领域掌握了一定的核心技术，也基本与美国先进企业处于同一起跑线，所以应着重发挥市场作用，以应用带动芯片，并研究推动应用/终端企业与芯片企业协同创新发展的政策和机制。

2、在云端/数据中心的人工智能芯片方面，目前主要由现场可编程门阵列占领市场，已被美国企业垄断，技术壁垒高耸，且市场有限。我国可与实际应用需求结合，借鉴谷歌（Google）的思路，充分发挥高校、研究所的科研力量，发挥政策引导作用，产学研合作，集中力量推动。

3、信息技术时代，技术热点相继不断，新概念层出不穷，人工智能芯片研发最重要的还是要关注自身的基础技术，深耕基础核心技术，坚持原创自主研发。只有增强了原创技术实力，并与市场应用实际结合，才有可能在不断切换的信息技术热点中立于不败之地。