

百度发布首款人工智能芯片“昆仑”

2018年7月4日，百度对外发布首款人工智能芯片“昆仑”。“昆仑”可在100W以上的功耗下提供260万亿次/秒的运算能力，达到业内领先水平。据百度介绍，“昆仑”可满足云端人工智能训练和推理需求，应用于数据中心等云端场景和自动驾驶等终端场景。“昆仑”的发布，有助于百度提升人工智能垂直整合能力，完善人工智能生态体系布局，并有望推动引领国内云端人工智能芯片的创新发展。

一、背景

人工智能芯片是人工智能发展应用的重要支撑。随着人工智能的快速发展，以中央处理器（CPU）为主的传统通用计算芯片已无法满足人工智能深度学习算法的任务需求。因此，更适于人工智能海量数据并行计算的专用芯片逐步兴起，形成了图形处理器（GPU）、现场可编程阵列（FPGA）、专用集成电路（ASIC）三大主流技术路线。其中，GPU是较为成熟的通用型人工智能芯片，FPGA和ASIC则是针对人工智能需求特征的半定制和全定制芯片。在此基础上，采用“CPU+专用芯片”的异构计算模式可基本满足现阶段的计算需求。未来，人工智能芯片将朝着突破冯·诺依曼结构的方向持续发展，推动实现计算能力的巨大跃升，以满足人工智能计算需求的急剧增长。

人工智能芯片成为各方竞相布局的焦点。近年来，人工智能芯片产业的迅速崛起，吸引了全球范围内传统芯片企业、互联网企业、初创企业等的积极布局。英伟达、英特尔等传统芯片企业凭借技术和市场积累，积极研发人工智能芯片，并快速拓展在自动驾驶、安防等领域的应用，已成为人工智能芯片产业的主导力量。谷歌、百度等互联网企业则基于自身的数据、算法和应用优势，向人工智能芯片领域跨界延伸，有力推动了人工智能芯片的

软硬融合和生态建设。美国 Cerebras 和我国寒武纪等初创企业充分利用人工智能专用芯片进入壁垒相对较低、技术路线尚未锁定的时间窗口，立足于人工智能细分领域，实现了在特定场景下的应用突破，成为人工智能芯片领域重要的技术创新力量。

百度积极布局人工智能相关领域。百度是国内最早布局人工智能的企业之一，依托其搜索引擎积累的庞大用户数据，在深度学习、人机交互、计算机视觉、语音识别、自然语言处理等方面开展基础技术研究，拥有近 3000 项人工智能相关技术专利。此外，百度推出了百度识图、百度无人车、百度无人飞行器等相关产品，在自动驾驶、智能家居、智慧城市、工业制造等领域开展应用。百度是我国第二代互联网巨头之一，在移动互联网时代承受着越来越大的竞争压力，因而已将人工智能作为打造未来核心竞争力的关键，正在推动实现由互联网公司向人工智能公司的战略转型。

二、案例介绍

事件简述。2018 年 7 月 4 日，在“2018 百度人工智能开发者大会”上，百度发布人工智能芯片“昆仑”，其中包含训练芯片“昆仑 818-300”和推理芯片“昆仑 818-100”。“昆仑”采用三星 14nm 工艺，包含数十万个核心，内存带宽为 512GB/s，可在

100W 以上的功耗下提供 260 万亿次/秒的运算能力。据百度介绍，“昆仑”基于百度已有的人工智能加速器，通过 20 多次迭代而来，可高效满足云端和终端的人工智能训练、推理需求。

后续发展。百度表示，未来“昆仑”既将实现对数据中心、公有云等云端场景的全覆盖，也将用于满足自然语言处理、语音识别、自动驾驶等终端场景的计算需求。百度在人工智能领域的布局广泛，其 DuerOS 和 Apollo 生态已经初具规模。此次“昆仑”若能成功落地，有望形成人工智能芯片与平台应用相互促进的良性循环。

三、简评

“昆仑”是百度在人工智能领域持续积累的结果，意在打造完整生态体系。早在 2011 年，百度已开始将 GPU 和 FPGA 应用于搜索、语音、图像等核心业务。2017 年，百度发布基于 FPGA 的云计算加速芯片 XPU，成为国内第一家发布自有芯片的互联网企业。与此同时，百度参与领投光学人工智能芯片初创公司 Lightelligence，并与华为、ARM、英伟达等在人工智能芯片上开展技术合作，积累了较为丰富的经验。**此次发布“昆仑”是百度厚积薄发、完善布局的必然之举。**当前，国内人工智能芯片企业大多是在终端推理所需的人工智能芯片上取得了一定突破，并致

力于向云端的训练和推理环节拓展延伸。在此背景下，“昆仑”若能基于百度的数据和算法优势，真正实现云端全场景覆盖，将显著提升百度的人工智能垂直整合能力，有助于打造融合人工智能芯片、深度学习框架和应用场景的三位一体生态体系，并可推动引领国内云端人工智能芯片的发展。

软硬结合和生态构建将成为“昆仑”未来发展的重点和难点。相较传统通用计算芯片，更为密切的软硬件结合是人工智能芯片提供更高运算能力的重要手段。例如，谷歌开发了人工智能芯片TPU，该硬件产品不单独对外销售，而是专用于谷歌的TensorFlow深度学习框架，部署在Google Compute Engine云计算引擎平台上，形成了十分强大的数据中心加速能力。类似地，未来百度“昆仑”也需与其PaddlePaddle深度学习框架、百度大脑等实现深度融合才能最大程度上发挥性能，实现对语音、图像、自然语言处理等领域的底层技术优化。**在此基础上，推动形成基于“昆仑”的应用开发生态至关重要。**当前，百度在自动驾驶、智能家居等领域持续深耕，打造了Apollo、DuerOS等平台，并集聚了一批国内外生态合作伙伴。未来，将“昆仑”植入百度人工智能生态体系，打造基于“昆仑”的应用开发生态体系，形成“芯片—云端/终端—应用—服务”生态闭环是实现“昆仑”价

值最大化和提升百度人工智能核心竞争力的关键。

应理性认识我国人工智能芯片产业现状，持续推动产业创新发展。近年来，我国在人工智能芯片领域不断取得创新突破。在市场研究和咨询公司 Compass Intelligence 近期发布的全球人工智能芯片公司 24 强排行榜中，我国已有海思等 7 家企业入围。此外，包括百度、阿里巴巴在内的平台型企业也纷纷基于数据、算法和应用优势，宣布进军人工智能芯片领域。当前，人工智能芯片的蓬勃发展将有力推动我国人工智能和集成电路领域的技术与产业创新。同时也应充分意识到，我国在集成电路领域积累薄弱，人工智能芯片技术研发和商业落地仍与国外存在差距，而底层软硬件技术、应用开发生态体系建设等方面的差距更大。因此，我国应深耕人工智能芯片基础架构技术，加大对基础和前沿领域的布局研究。同时，应构建人工智能芯片产业生态体系，提升产业应用能力，提升市场影响力和国际话语权。

本文作者：工业和信息化部赛迪研究院 冯晓辉 王哲

联系方式：18600210027

电子邮件：fengxiaohui@ccidthinktank.com

赛迪智库

面向政府 服务决策

思想从这里升华

《赛迪专报》

《赛迪译丛》

《赛迪智库·软科学》

《赛迪智库·国际观察》

《赛迪智库·前瞻》

《赛迪智库·视点》

《赛迪智库·动向》

《赛迪智库·案例》

《赛迪智库·数据》

《智说新论》

《书说新语》

《两化融合研究》

《互联网研究》

《网络空间研究》

《电子信息产业研究》

《软件与信息服务研究》

《工业和信息化研究》

《工业经济研究》

《工业科技研究》

《世界工业研究》

《原材料工业研究》

《财经研究》

《装备工业研究》

《消费品工业研究》

《工业节能与环保研究》

《安全产业研究》

《产业政策研究》

《中小企业研究》

《无线电管理研究》

《集成电路研究》

《政策法规研究》

《军民结合研究》

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：董凯

联系电话：010-68207922 13910685050

传真：010-68200534

网址：www.ccidwise.com

电子邮件：dongkai@ccidthinktank.com