



# 计算机 AI 芯片行业点评：特斯拉首发 AI 训练芯片 百度“昆仑芯”再迭代



事件：8月20日特斯拉举办 AI 日，发布 AI 训练芯片 DojoD1；8月18日，百度举办 AI 大会，宣布“昆仑 2”人工智能芯片正式量产。

特斯拉 DojoD1 对标英伟达 A100、华为昇腾 910、寒武纪思元 290，但是其表现性能弱于英伟达 A100 和华为昇腾 910，但是强于思元 290。特斯拉发布的第一款 DojoD1 训练芯片除了公布制程、算力、功耗，未向外披露更多的技术细节。从其公布的主要参数来看，DojoD1 训练芯片并不是业界最领先的 AI 训练芯片，同制程的英伟达 A100 和华为昇腾 910 的性能功耗比均要优于特斯拉 Dojo。

我们推测特斯拉自研 AI 训练芯片其主要目的其一是省去云上训练和车端模型部署之间跨平台精度损失的问题；其二是牺牲一定的通用性来获得智能驾驶场景算法训练的高性能。

(1) 特斯拉此前使用英伟达的 GPU 来训练模型，AI 日上特斯拉提到它曾用 5760 个英伟达 A100 组装了一台 1.8EFLOPS 算力的超级计算机，世界排名第五；而特斯拉车端用的是自研芯片 FSD，训练和部署用不同的芯片平台中间过程会徒增大量模型迁移的工作量以及模型精度损失的问题；如果训练和部署用同一套工具链、基于相同的指令集，则效率会大幅度提升，而且可以最大程度保证模型精度。(2) 我们推测其 D1 大概率专为智能驾驶场景打造，其架构设计大部分主要针对智能驾驶中常出现的计算机视觉等算法，不太考虑语音识别、语义理解等常用算法，因此在设计上可能会以牺牲一定通用性来换取配合上层算法特性来定制以获得高性能。

特斯拉是封闭体系，AI 芯片自产自用，所以对软件工具链打造上的要求没有像英伟达、华为、寒武纪等第三方 AI 芯片公司的要求那么高。从最上层 AI 模型所用的高级语言，到智能编程语言，再到接近 AI 芯片所用的汇编语言，中间需要经过编译器和汇编器的转化，而智能编程语言、编译器等软件需要由 AI 芯片公司开发好，而且要开发大量 AI 算子，所以一般 AI 芯片公司的软件开发人员比硬件研发人员还多。另外，即使有 AI 芯片公司将智能编程语言和编译器做好了，还存在着深度学习框架 Tensorflow、Pytorch 愿不愿意基于特定厂商的芯片和编程语言进行特别优化的问题，此即生态壁垒（类似微软针对英特尔 CPU 进行特定优化），这也是华为不仅选择做 AI 芯片还进一步拓展至自研深度学习框架 MindSpore 的原因。

第二代昆仑芯片 7nm 制程，256TOPS (INT8)，150W 功耗，自研 XPU 架构和谷歌 TPU 有共通之处。主要用在 AI 推理场景：百度两代昆仑芯 2 都是基于自研 XPU 架构。百度第一代昆仑芯片是三星 14nm 制程，2018 年 7 月正式发布，于 2020 年初量产，在百度搜索引擎部署，目前已经规模化部署超过两万片，有数十个客户。本次会议发布的第二代昆仑芯片“昆仑芯 2”采用 7nm 制程。百度做 AI 芯片的积淀可以追溯至 2010 年开始用 FPGA 做 AI 架构的研发，2017 年在百度内部数据中心、自动驾驶系统等使用。昆仑芯片自研 XPU 架构是 2017 年提出，和谷歌 TPU2 架构有比较多相似之处：有 M 层很大的编列（变量计算的小处理器）。

相关标的：特斯拉、百度、英伟达、寒武纪。

关键词: GPU 人工智能 特斯拉 芯片 计算机视觉 语音识别

预览已结束，完整报告链接和二维码如下：

[https://www.yunbaogao.cn/report/index/report?reportId=1\\_34458](https://www.yunbaogao.cn/report/index/report?reportId=1_34458)

