

AI芯片 2023 行业简析报告

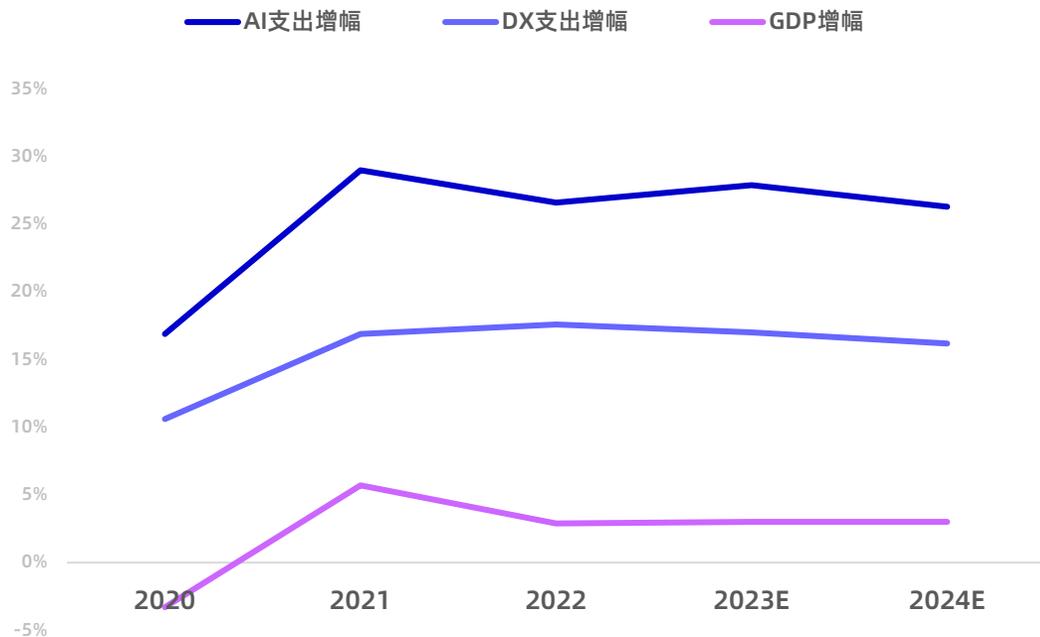
THE BRIEF MARKET ANALYSIS REPORT ON AI CHIP



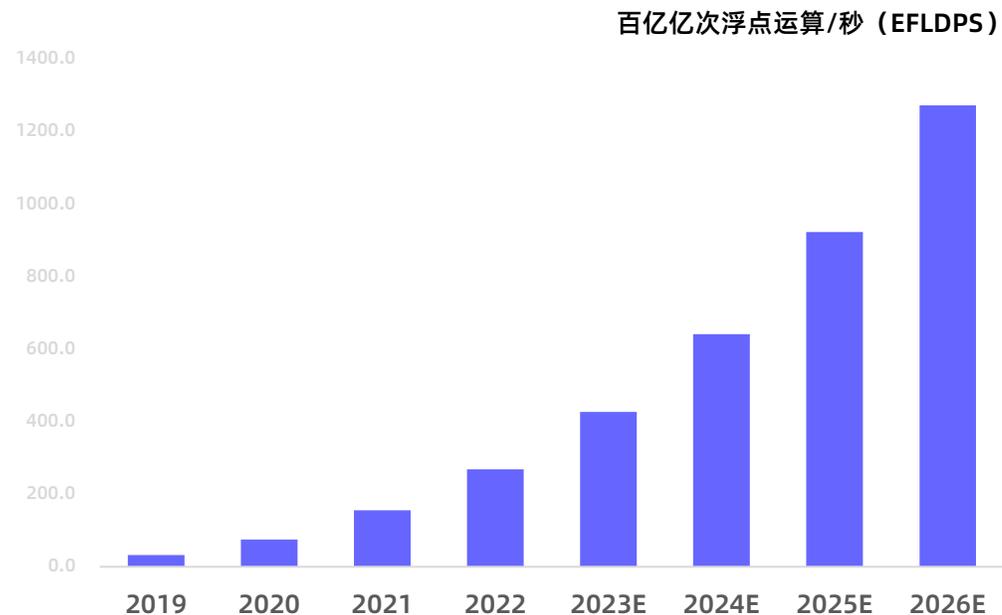
01. AI芯片：人工智能的基石

- 自2018年GPT-1.0模型首次发布以来，OpenAI不断迭代模型，今年发布的GPT-4.0模型，它拥有更大的参数量、更长的迭代时间和更高的准确性。随着数据不断增长和算法复杂度提高，人工智能对算力提出了更高的要求。算力是实现人工智能产业化的核心力量，其发展对人工智能技术的进步和行业应用起着决定性作用。
- 释放算力的价值对国家整体经济发展将发挥推动作用。算力指数每提高1点，数字经济和GDP将分别增长 3.5‰和1.8‰。可见，国家算力指数越高，对经济的拉动作用越强。2021-2026年期间，预计中国智能算力规模年复合增长率达52.3%，同期通用算力规模年增长率为18.5%。

全球AI支出、数字化转型支出及 GDP 增长趋势预测



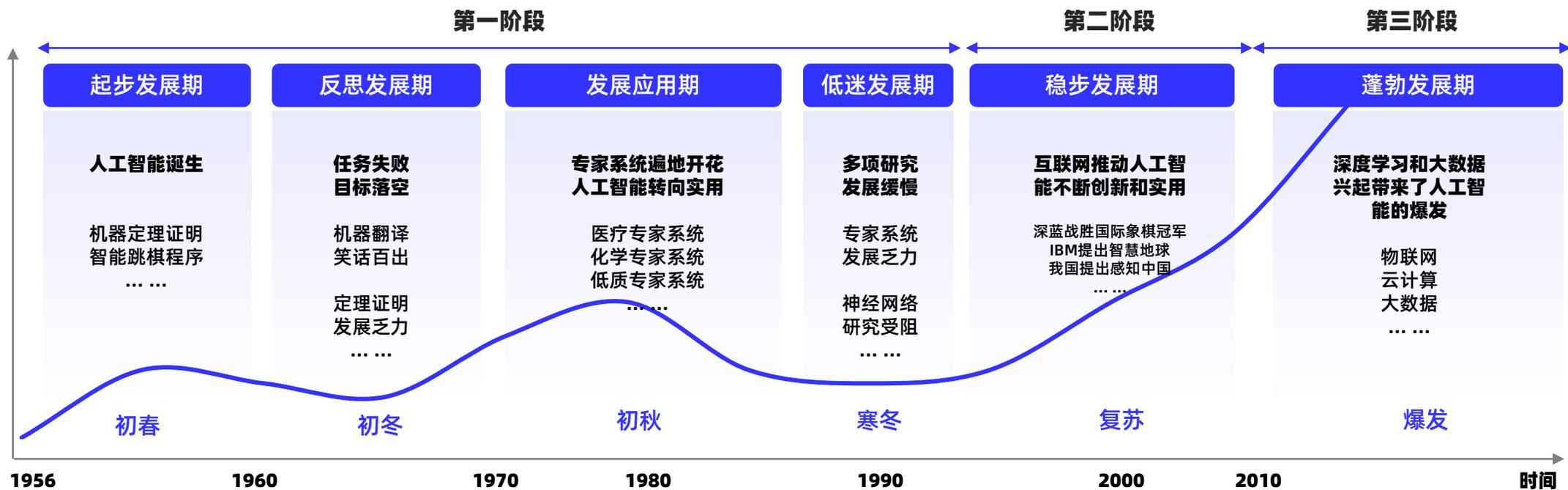
中国智能算力规模及预测



02. AI芯片算力发展的三个阶段

- 广义上讲只要能够运行人工智能算法的芯片都叫做AI芯片。但是通常意义上的AI芯片指的是针对人工智能算法做了特殊加速设计的芯片。
- AI芯片也被称为AI加速器或计算卡，即专门用于处理人工智能应用中的大量计算任务的模块（其他非计算任务仍由CPU负责）。
- 第一阶段：因为芯片算力不足，所以神经网络没有受到重视。第二阶段：通用芯片CPU的算力大幅提升，但仍然无法满足神经网络的需求。第三阶段：GPU和和新架构的AI芯片推进人工智能落地。

AI芯片算力发展阶段



03. 国内外AI芯片差距较大

- 在AI芯片领域，国外芯片巨头占据了大部分市场份额。全球范围内主要布局人工智能芯片的厂商有Intel、NVIDIA、Qualcomm、Google等。美国的巨头企业，凭借着多年在芯片领域的领先地位，迅速切入AI领域并积极布局，目前已经成为该产业的引领者。
- 我国AI芯片产业起步较晚，技术上与世界先进水平也还存在着较大的差距。国内AI芯片市场也较为分散，集中度低。随着数字经济的兴起，人工智能已经深入渗透到各个行业，特别是在互联网等科技公司中更为普及。这些公司对于计算机软件技术和存储设备的要求极高，因此对于底层技术的布局和提升更为重视，尤其是在人工智能芯片领域。

AI芯片产业图谱

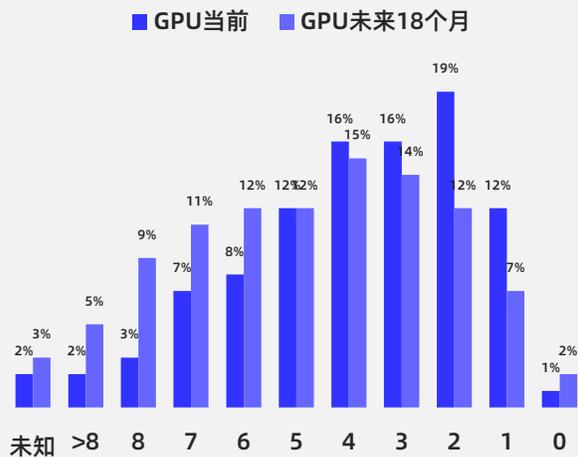


04. 全球人工智能芯片市场高速增长

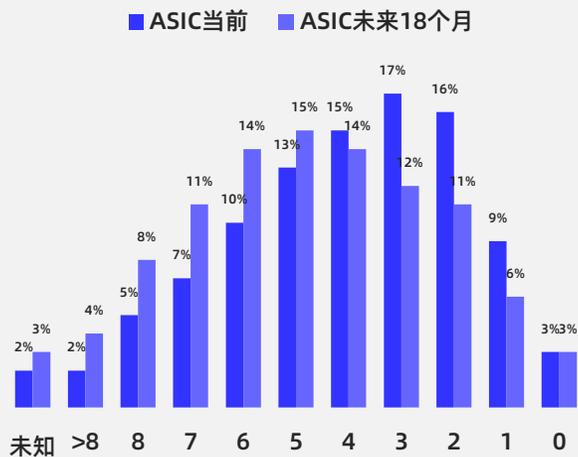
- IDC预计，到2025年人工智能芯片市场规模将达726亿美元。人工智能芯片搭载率将持续增高，目前每台人工智能服务器上普遍多配置2个GPU，未来18个月，GPU、ASIC和FPGA的搭载率均会上升。
- 2022年加速服务器市场规模达到67亿美元，同比增长24%。其中GPU服务器依然是主导地位，占据89%的市场份额，达到60亿美元。同时NPU、ASIC和FPGA等非GPU加速服务器以同比12%的增速占有了11%的市场份额，达到7亿美元。

全球人工智能服务器 GPU、ASIC 和 FPGA芯片搭载率

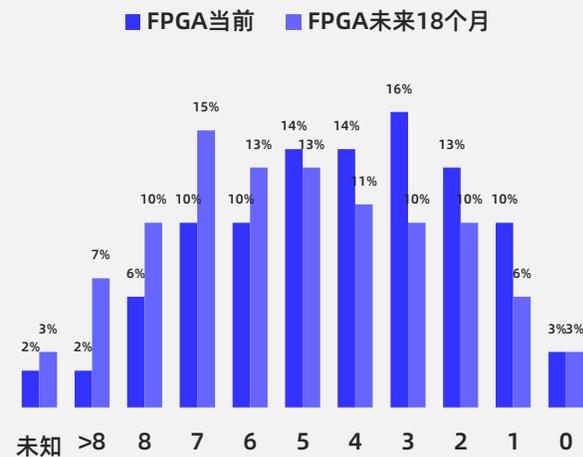
GPU



ASIC



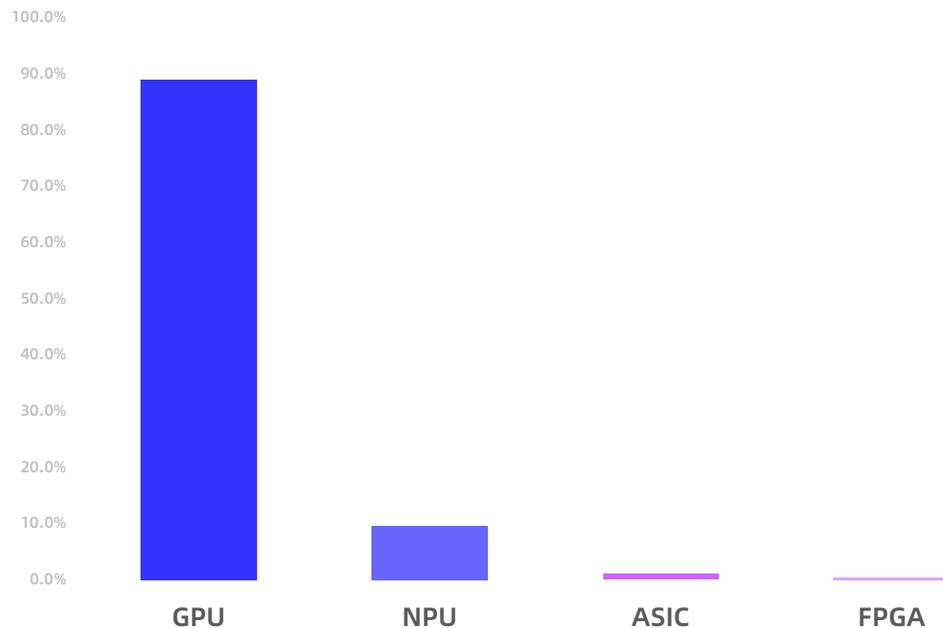
FPGA



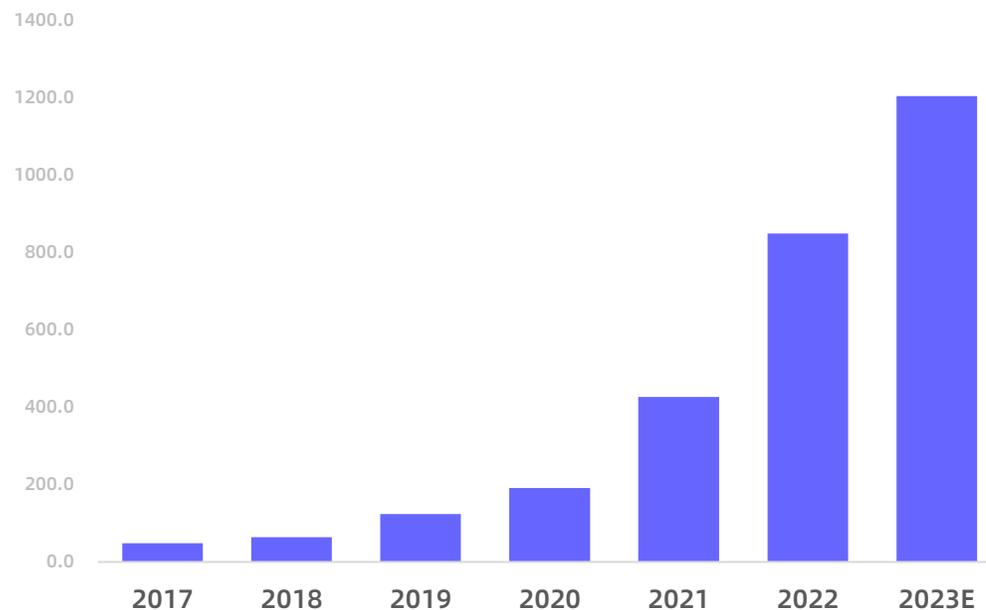
05. 中国AI芯片市场呈现出显著的增长趋势

- 根据市场规模分析，随着越来越多企业将人工智能应用于终端产品，人工智能芯片的需求快速增长。AI芯片广泛应用于云计算、数据中心、边缘计算、消费电子、智能制造、智能驾驶、智能金融及智能教育等领域。
- 近年来，我国的AI芯片行业备受关注，不断涌现出新的生产设计商，市场规模也不断扩大。数据显示，2021年我国AI芯片市场规模达到427亿元，同比增长124%。预计到2023年，市场规模将进一步扩大至1206亿元。

中国人工智能芯片市场规模占比



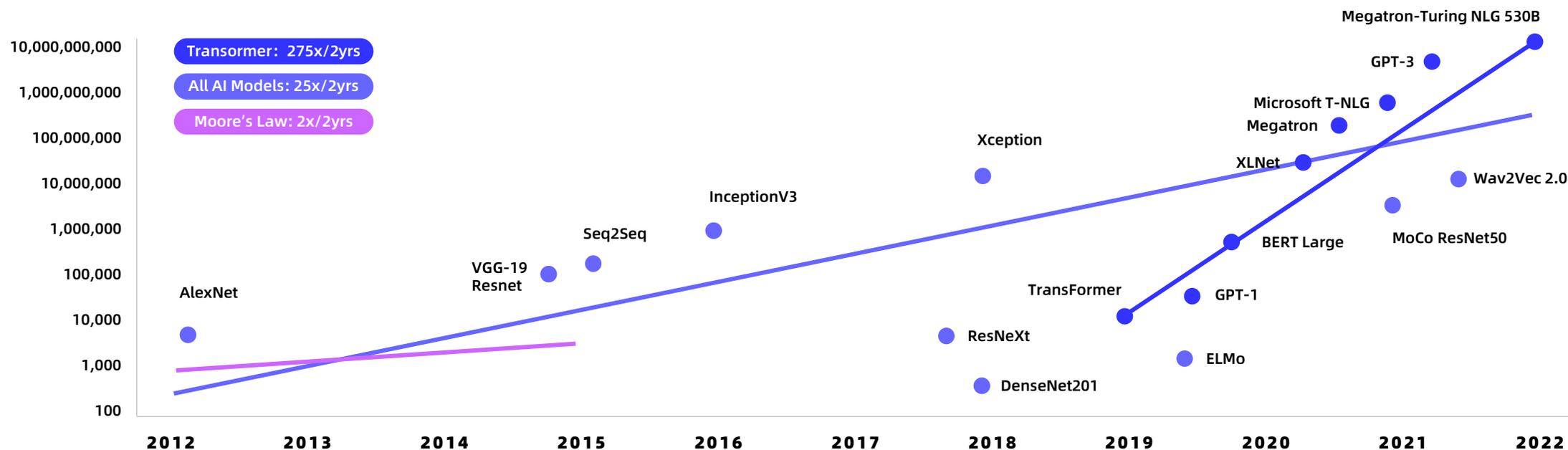
2017-2023年中国AI芯片市场规模预测趋势图（亿元）



06. 深度学习模型复杂度对芯片算力需求激增

- AI运算指以“深度学习”为代表的神经网络算法，需要系统能够高效处理大量非结构化数据（文本、视频、图像、语音等）。
- 需要硬件具有高效的线性代数运算能力，计算任务具有：单位计算任务简单，逻辑控制难度要求低，但并行运算量大、参数多的特点。对于芯片的多核并行运算、片上存储、带宽、低延时的访存等提出了较高的需求。
- 自2012年以来，人工智能训练任务所需求的算力每3.43个月就会翻倍，大大超越了芯片产业长期存在的摩尔定律（每18个月芯片的性能翻一倍）。
- 针对不同应用场景，AI芯片还应满足：对主流AI算法框架兼容、可编程、可拓展、低功耗、体积及价格等需求。

从AlexNet到GPT-3，算力增长迅速



07. AI服务器有68%价值量来自GPU

- AI算力芯片产业链可分为：最上游的EDA 软件和IP核研发；中游的芯片，包括GPU、NPU、ASIC、FPGA 等架构；下游的AI服务器及智算中心。

AI 芯片产业链



08. GPU是较为成熟的通用型人工智能芯片

- 从技术架构来看，AI芯片主要分为图形处理器（GPU）、现场可编程门阵列（FPGA）、专用集成电路（ASIC）、类脑芯片四大类。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，类脑芯片颠覆传统冯诺依曼架构，是一种模拟人脑神经元结构的芯片，类脑芯片的发展尚处于起步阶段。

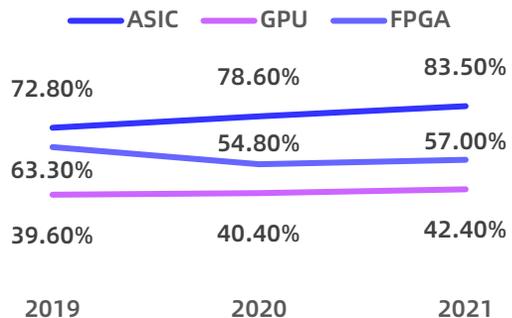
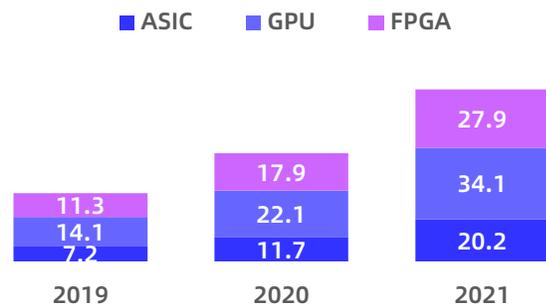
三种技术架构AI芯片类型比较

	GPU	FPGA (半定制化)	ASIC (全定制化)
定制化程度	通用性	半定制化	定制化
灵活度	好	好	不好
成本	高	较高	低
编程语言/架构	CUDA、OpenCL等	Verilog/NHDL等硬件描述语言, OpenCL、HLS	/
功耗	大	较大	小
主要优点	峰值计算能力强、产品成熟	平均性能较高、功耗较低、灵活性强	平均性能很强、功耗很低、体积小
主要缺点	效率不高、不可编辑、功耗高	量产单价高、峰值计算能力较低、编程语言难度大	前期投入成本高、不可编辑、研发成本长、技术风险大
主要应用场景	云端训练、云端推断	云端推断、终端推断	云端训练、云端推断、终端推断
代表企业芯片	英伟达Tesla、高通Adreno等	赛灵思Versal、英特尔Arria、百度XPU等	谷歌TPU、寒武纪Cambricon

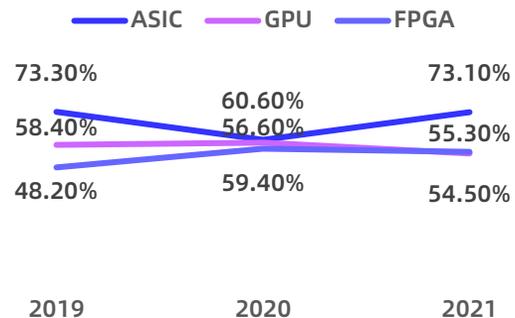
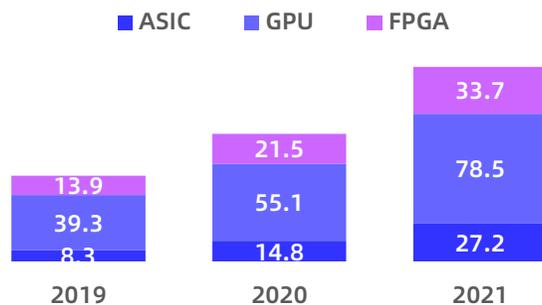
09. AI芯片三剑客将互补共享市场

- AI芯片（GPU/FPGA/ASIC）在云端同时承担人工智能“训练”和“推断”过程，在终端主要承担“推断”过程，从性能与成本来看ASIC最优；
- ASIC作为专用芯片，算力与功耗在通用芯片GPU具有绝对优势，但开发周期较长，落地较慢，需一定规模后才能体现成本优势；
- FPGA可以看做从GPU到ASIC重点过渡方案。相对于GPU可深入到硬件级优化，相比ASIC在算法不断迭代演进情况下更具灵活性，且开发时间更短。

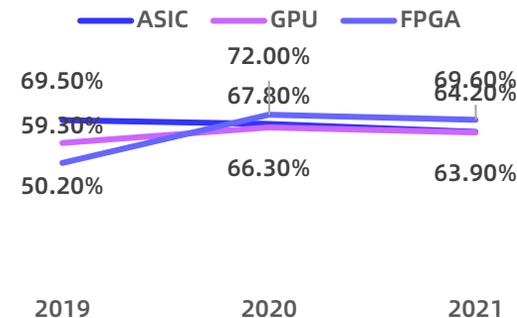
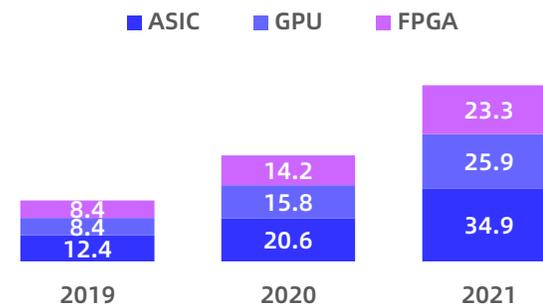
中国云端推断芯片市场结构（亿元）及增长率



中国云端训练芯片市场结构（亿元）及增长率



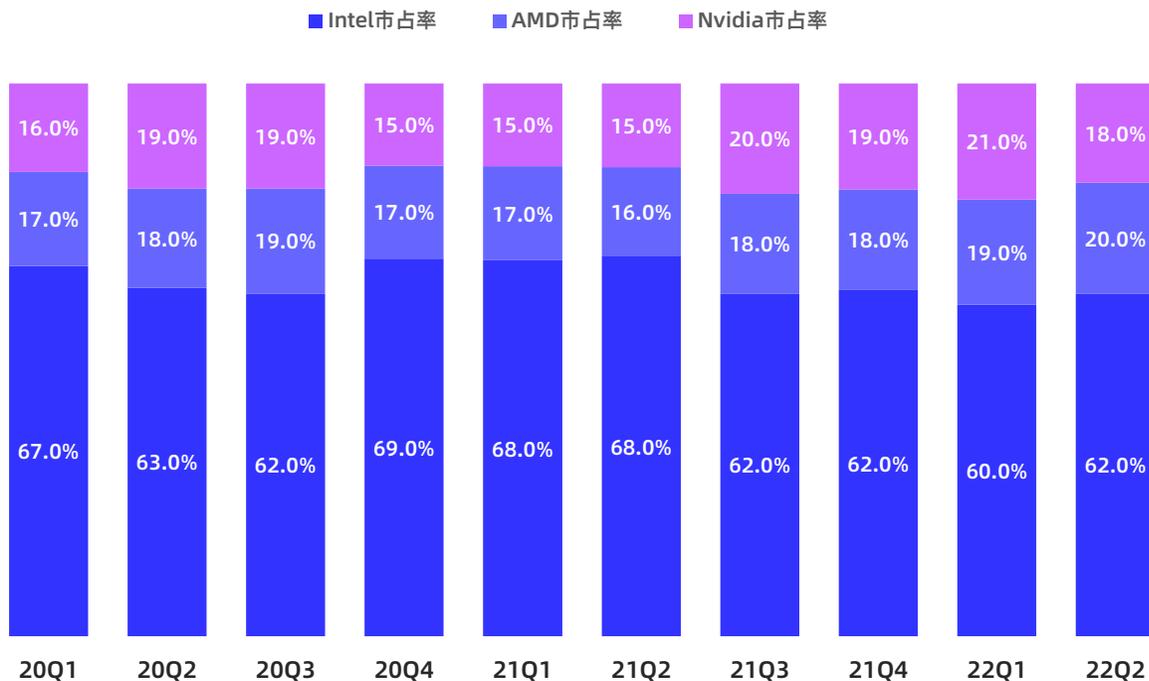
中国终端推断芯片市场结构（亿元）及增长率



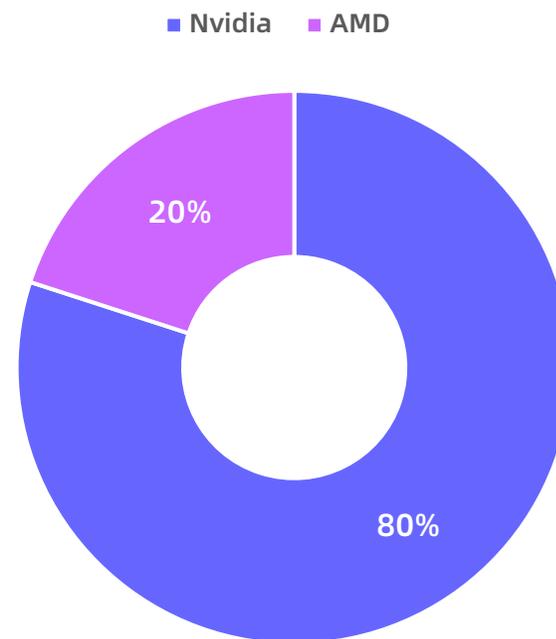
10. 全球GPU领域处于寡头垄断的局面

- 近年来GPU市场由英特尔、英伟达和AMD三分天下，其中英特尔的市场占比份额在60%以上，其次是英伟达和AMD。英特尔的身位领先主要得益于在家用机的集成GPU芯片市场的绝对主导地位，而在AI及高性能计算方面，则英伟达凭借其自身CUDA生态占据绝对主导地位。
- 从细分市场来看，在独显市场中，Nvidia与AMD双雄垄断市场，其2022年Q2市占率分别约为80%和20%，可以看到近年来Nvidia不断巩固自己的优势，其独立显卡市占率整体呈现上升趋势。

PC端 GPU 三大厂商市场份额变化



2022Q2独显GPU市场份额



11. AI芯片国产化有望加速

- 美国陆续收紧中国获得国际先进芯片的能力，不仅限制中国进口尖端芯片，还限制中国获得最新的芯片生产工具。中国在FPGA、GPU领域缺乏有竞争力的原创产品，只是基于FPGA/GPU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，FPGA/GPU的技术壁垒已很高，很难有所突破。
- 为了打破以美国为首的国家的技术封锁和制裁，中国对AI芯片领域的资源投入不断走高。当前，中国数字化的变革方向驱动了底层技术的逐渐提升，国际影响力也在逐年上涨，同时，在大数据、芯片设计及应用落地方面，逐步建立起了优势地位，产业发展也吸引更多海外人才回国创业、就业。

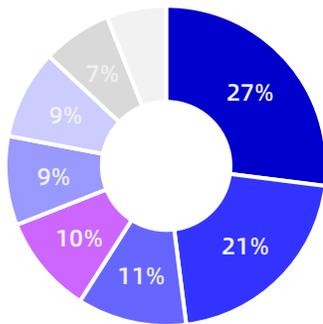
AI 相关论文发表和海外人才回国情况

2021年人工智能领域领先国家与地区论文发表

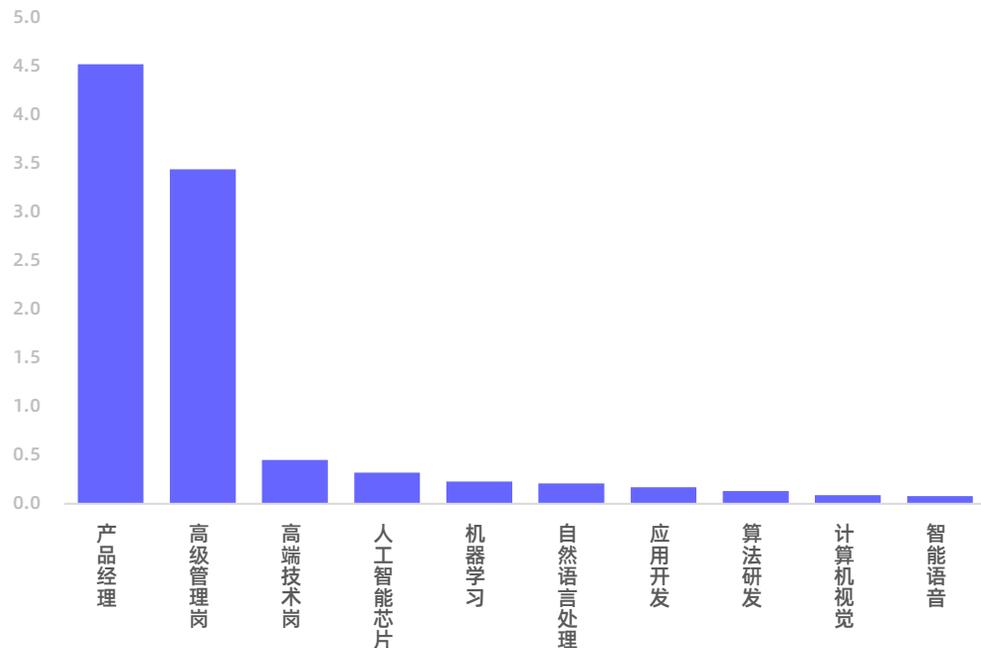
	中国	美国	欧盟
AI期刊出版文献占全球份额	18%	12.3%	8.6%
人工智能期刊引用比例	20.7%	19.8%	11.0%

2020-2021年海归投递行业分布

- 文娱/服务业等其他行业
- IT/通信/电子/互联网
- 生产/加工/制造
- 商贸/零售
- 金融
- 商业服务
- 房地产/建筑业
- 能源/环保



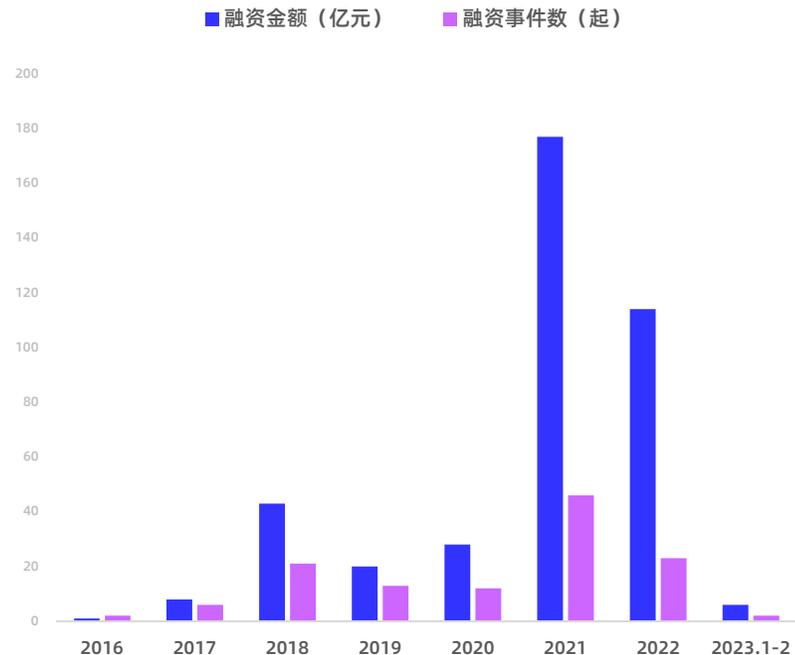
中国人工智能技术方向岗位供需情况



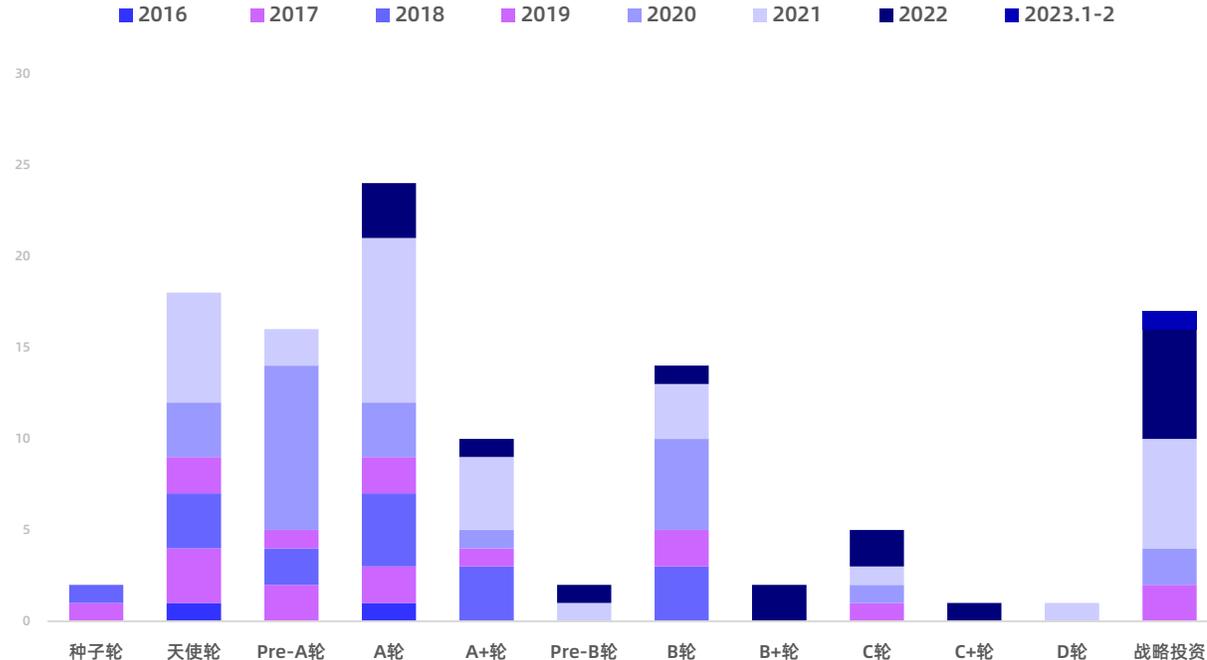
12. 资本持续进入AI芯片市场

- 2021年中国人工智能芯片相关领域融资事件共计92起，总金额约300亿人民币。企业方面，中国的科技巨头如百度、阿里巴巴、华为等都在人工智能芯片领域进行了大量的投资。
- 此外，还涌现出了许多专注于人工智能芯片研发的创业公司，如寒武纪、Horizon Robotics、Cambricon 等。这些创业公司大多得到了风险投资机构的支持，如IDG、高瓴资本、启明创投等。
- 中国的人工智能芯片市场投资情况非常活跃，政府、企业和风投机构都在加大对人工智能芯片领域的投入。这也进一步促进了中国人工智能产业的发展和 innovation。

中国人工智能芯片行业融资整体情况



中国人工智能芯片行业投融资轮次情况



13. 平台生态之争决定未来AI芯片的市场表现

- 广义上AI开源平台定义为一个深度学习的工具箱，用户可以通过开放平台，基于底层计算芯片运行算法系统。因此AI芯片厂商只有具备支持主流开源平台，才能形成稳定客户群。两大开源平台Tensorflow与Caffe，Caffe2和mxnet也逐步兴起，AI开源平台众多。
- 广义AI硬件开发环境是专门针对AI硬件推出的适应于硬件计算的开发环境，用户能利用C、C++等软件语言更方便的基于AI芯片进行顶层应用开发，同时可以起到硬件加速的效果。
- 目前比较流行的AI开发环境是Nvidia的CUDA，Xilinx推出了SDAccel开发环境，Altera推出了OpenCL SDK 开发环境。

主流AI开源平台简介

公司/高校	时间	名称	简介
Stanford	2013.10	Caffe	自2013年底以来第一款主流的工业级深度学习工具包,计算机视觉界最流行的工具包之一
DMLC	2015.09	mxnet	MxNet是一个提供多种API的机器学习框架,主要面向RPython和Julia等语言,目前已被亚马逊云服务采用。
Google	2015.11	Tensorflow	谷歌第二二代联机版人工智能深度学习系统,能同时支持多台服务器。
Facebook	2015.12	Torch	深度学习函数库Torch的框架,旨在鼓励程序代码再利用及模块化编程。
Microsoft	2016.01	CNTK	CNTK支持RNN和CNN类型的各类网络模型。
Facebook	2017.04	Caffe2	Caffe2延续了Caffe对视觉类问题的支持,且增加了对RNN和LSTM支持,功能更加完善

部分AI硬件平台开源生态支持情况

公司	平台名称	类别	支持的开源生态平台
NVIDIA	Tesla V100	GPU	Caffe, Cff2, TensorFlow, Torch, CNTK等
Xilinx	reVISION	FPGA	Caffe
Google	TPU2	ASIC	TensorFlow
Qualcomm	NPESDK	ASIC	TensorFlow、Caffe、Caffe2
华为	麒麟970	ASIC	Tensorlow、Cffe Caffe2

14. AI芯片行业发展遭遇的挑战

01 市场议价能力低

目前中国AI芯片行业面对的消费者主要包括互联网、智能手机和自动驾驶等行业，这些行业对AI芯片的市场需求量相对较大，市场依赖程度高。在当前数字化经济不断发展的背景下，这将会在一定程度上提高AI芯片行业的需求量。

03 商用前景广阔，但落地困难

AI芯片商用前景广阔，但落地困难的局面导致行业并未形成稳定的市场格局。

02 缺乏核心自主技术

与国外主流企业相比，目前我国在AI芯片开发领域仍然缺乏具有竞争力的产品，缺乏关键的核心自主技术。这将会导致关键设备落后于国际一流水平，难以突破行业发展瓶颈。

04 需要突破存储瓶颈

在以传统冯诺依曼架构方式为主的AI芯片市场中，当运算能力发展到一定程度，存储部件就决定了AI芯片的性能上限。同等条件下，“存算一体”架构能够有效降低AI芯片能耗和成本，突破“存储墙”难题，是AI芯片未来发展主流方向。

15. AI芯片行业发展关注三大趋势

01

AI场景和算法推动AI芯片走向专业化

GPU在处理大量并行计算任务中表现出色，且可通过加速设计更好地发挥AI潜能，但也存在功耗高、成本高等缺点。目前，GPU仍然是AI训练所需算力的主要硬件选择。FPGA具有较强的计算能力、较低的试错成本和足够的灵活性，但其缺点在于价格较高、编程复杂，因此在半定制化AI场景中具备优势。ASIC具有更高的处理速度和更低的能耗，并且可针对特定AI任务进行优化设计，从而在性能和能耗方面具备更好的综合素质，使其在全定制化AI场景中表现优异。

02

类脑、量子技术推动AI芯片走向多样化

随着拟态神经元、量子等前沿技术的发展，AI芯片逐渐发展出类脑、量子等多样化技术路径的新型芯片，类脑芯片更是开始走向商用化。类脑芯片拥有大规模并行计算、超低功耗和超低延迟等技术潜力，这些优势使其在未来AI应用场景中扮演重要的角色。类脑芯片和量子芯片作为新型芯片技术，拥有巨大的潜力，将在未来的AI和计算领域发挥重要作用，为我们带来更高效、更强大的计算能力。

03

AI芯片应用领域不断拓展

随着人们对生活品质要求的提高，AI芯片已应用于众多前沿及快速发展的下游领域，AI芯片应用的不断拓展。随着数字化时代的到来，我国产业也纷纷开始数字化转型，带动AI芯片需求的增长，以支持高效处理海量数据，从而推动AI芯片行业的快速发展和持续创新。可以预计，在未来我国数字化转型持续推进的背景下，AI芯片将存在广阔的应用市场，需求不断增加，推动行业快速发展阶段。

版权说明

本报告为简版报告，内容均从嘉世咨询原有完整报告中精炼提取，如需了解详细内容，请联系：
mcr@chinamcr.com。

本报告中的所有内容，包括但不限于文字报道、照片、影像、插图、图表等素材，均受《中华人民共和国著作权法》、《中华人民共和国著作权法实施细则》及国际著作权公约的保护。

本报告的著作权属于上海嘉世营销咨询有限公司所有，如需转发、转载、引用必须在显著位置标注出处，并且不得对转载内容进行任何更改。

本报告是免费报告，任何机构和个人不得将本报告用于收费为目的经营活动。

