2026年中国AI芯片发展趋势报告

亿欧智库 https://www.iyiou.com/research

关于亿欧智库（EO Intelligence）

亿欧智库

用第三方视角和专业服务
助力产业科技升级和价值创造

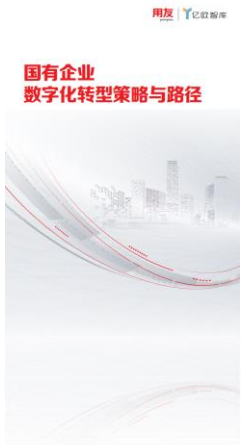
- ✓ 研究领域：覆盖人工智能、未来产业、**汽车出行、大健康、消费生活、智能制造、电商零售、数字农业、智慧城市、金融科技、物流供应链、企业服务、双碳**等多行业领域
- ✓ 服务对象：包含国家部委、地方政府、央国企、互联网科技型公司以及外资500强和民营500强
- ✓ 独创模型：亿数合创团队在10余年产业研究和咨询经验的基础上联合科研单位，研发了诊断企业数字化和创新力水平的TOIPO模型。模型从5大维度，30个细分维度对企业的战略、产品、技术、供应链、经营等方面进行全面诊断。

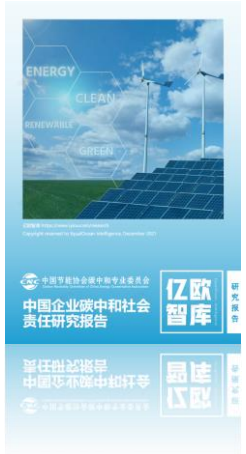
亿欧智库历史服务项目

累计发布自研型研究报告 600+

定制型研究与白皮书项目 300+

战略规划型项目 100+





亿欧智库服务项目类型

定制型研究服务	T-技术驱动	• 技术扫描与对接、技术选型、路径设计等
	O-组织与战略	• 组织架构、体系搭建、流程设计等
	I-产业与定位	• 战略规划、招股书、对标分析、尽职调查等
	P-产品与服务	• 产品定位、可行性研究等
	O-经营与品牌	• 资源扫描与对接、白皮书等
自研型研究服务	联合撰写	• 客户作为亿欧智库某专题研究报告的联合编写单位，进行品牌露出
	专题报告	• 亿欧智库将针对客户所在的细分领域，撰写相应的专题报告
	案例植入	• 在公开发布的报告中，客户将有名称、产品名、品牌logo等形式提及
	大咖说	• 在公开发布的报告中，在大咖说部分体现客户高管的行业观点
	配套传播	• 在配套报告预热文章/报告解读文章中，植入客户品牌相关的观点文字

亿欧智库

目录

CONTENTS

1.1.1 国内“十四五”收官与“十五五”展望下的芯片扶持政策

- “十四五”期间，政策核心围绕“技术突破”与“全产业链自主可控”展开，通过多维度组合拳为产业跨越式发展奠定了基础,实现“从追赶到并跑”的系统性突破。其中，2025年，AI芯片国产化率突破40%，国产供应链逐步成形，头部企业实现盈利。
- “十五五”规划将集成电路产业的战略地位提升至前所未有的高度，政策导向从“全面扶持”转向“聚焦核心、全链条突破”。

亿欧智库：“十四五”回顾：奠定基础，实现“从追赶到并跑”的系统性突破

战略与财税双轮驱动

- 纳入国家战略性新兴产业重点目录，实施国家重大科技项目。
- 对符合条件的集成电路企业实施企业所得税“两免三减半”、“五免五减半”乃至“十年免税”等阶梯式优惠，并免征部分进口关税和增值税。

多元投融资体系

- 国家集成电路产业投资基金（大基金）一二期重点投资制造、设计等
- 政策性金融配合
- 资本市场助力

研发创新与人才培育

- 通过国家重点研发计划、国家科技重大专项等，聚焦高端芯片、装备等核心研发。
- 推进集成电路一级学科设置，建设示范性微电子学院，强化人才培养。

区域协同发展

- 全国超14个省市出台集成电路专项政策，形成区域差异化布局。例如，长三角、珠三角、京津冀主攻高端设计与先进制造，中西部地区聚焦特色工艺与封测。

亿欧智库：“十五五”展望：聚焦核心，全链突破

战略定位跃升

- 明确定位为六大新兴支柱产业之首，定位为数字经济与高端制造的算力底座
- 目标：实现全链条决定性突破， 具体路径是“成熟制程做精做细、先进制程突破瓶颈”

资金投入持续加码

- 大基金三期投资超4000亿元，重点投向晶圆制造、先进封装等，以及核心环节与薄弱领域。
- 央行科技创新再贷款额度从8000亿元提升至 1.2万亿元，以降低半导体企业融资成本。

全链条攻关路径

- 政策覆盖从基础架构（RISC-V）、核心产品（GPU/NPU/ASIC）、前沿技术（存算一体/类脑计算）到制造环节（先进封装） 的全链条。
- 重点突破高端芯片设计、先进制程（7nm及以下）、封测技术等关键领域

场景驱动与生态构建

- 鼓励央企、国企带头开放应用场景，推动在汽车电子、工业控制等领域的供需对接，以应用需求牵引国产AI芯片的技术迭代和生态构建。
- 明确支持突破高端训练芯片、端侧推理芯片、人工智能服务器等关键技术

数据来源：《政府工作报告》、国家发改委等、亿欧智库

1.1.2 全球半导体管制加剧背景下的国产替代新阶段

- 当前管制呈现体系化升级和范式改变的特点，已经从针对特定企业拓展到全产业链生态隔离，如2025年12月生效的《半导体制造设备出口管制修正案》将管控范围延伸至先进封装等领域，并引入“架构导向”和“最终用途穿透”的双重审查机制。
- 国产替代已告别零散的技术突破，进入全产业链协同攻坚的新阶段，AI芯片已完成“设计-制造-设备-材料”全链条布局。根据三方机构预测，到2030年，中国AI芯片自给率将升至86%，市场规模突破670亿美元。

亿欧智库：全球半导体管制升级

从算力到全生命周期封锁

- “算力总量控制”：确保AI总算力不超过美国10%。
- “滚动式技术门槛”：基于中国当前的AI技术水平和本土芯片量产能力，动态设定出口管制门槛。依据4项关键参数，定义每代AI芯片的“风险评估周期”（RTT），并结合情报精准设定管制标准。
- 针对中国AI发展的核心企业和关键产品进行精准打击。

管制持续加剧升级

- 《MATCH法案》：扩大设备禁运范围；明确禁止设备供应商为已在中国运行的受限设备提供售后服务；计划对五家核心芯片制造企业及其所有关联公司的芯片设施实施类似“实体清单”的全面出口限制。
- 美国盟友协同：如荷兰全面禁止向中国出口DUV光刻机并叫停在途设备发货，同时限制对已售设备的售后支持。

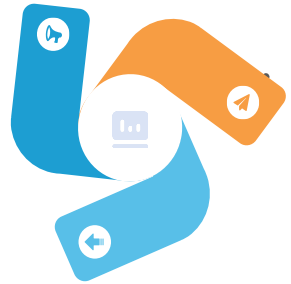
亿欧智库：国产替代新阶段：体系化创新和生态构建

国家战略与政策强力驱动

- 战略定位升级：列为“十五五”六大新兴支柱产业之首，采取“超常规措施”
- 资金保障：大基金三期规划超4000亿元重点投向AI芯片等关键环节
- 产能扩张：计划将先进芯片（7nm/5nm）产能扩大到当前的5倍

技术路径创新突破性能封锁

- 系统级创新：通过“超节点”集群计算和高速互联技术，用系统优势弥补单卡性能差距。
- 架构创新：积极拥抱 RISC-V开源指令集，规避专利壁垒
- 聚焦先进封装，弥补先进制程短板



应用生态与市场闭环加速形成

- 强制替代与需求牵引：中国首次将华为、寒武纪等国产AI芯片纳入官方采购清单。同时，要求新建智算中心国产芯片占比需超50%。
- 软硬件协同：头部AI公司开始将算力底座从CUDA迁移至国产平台，如华为昇腾的CANN软件平台已实现与DeepSeek等头部大模型的“Day0适配”

数据来源：中国科学院、芯域前沿、摩根士丹利等、亿欧智库

1.1.3 地方产业集群（如北京、上海、深圳、合肥）的差异化布局与成效

- 中国AI产业已形成以京津冀、长三角、珠三角引领，中西部城市多中心崛起，区域协同的集群化发展格局。其中，三大核心区集中了全国88%的AI产业。各集群基于自身资源，形成了差异化的定位与竞争优势。
- 以北京、上海、深圳、合肥为代表的中国AI芯片产业集群，已形成定位清晰、优势互补的差异化发展格局。分别代表了国家创新策源、全产业链生态、硬科技融合应用、制造与设计协同四种典型模式。

亿欧智库：北京、上海、深圳、合肥差异化布局与成效

- 全国AI创新策源地与算力基建引领者，聚焦算法研发、通用大模型、高端芯片设计及超大规模智算集群建设。
- 2025年AI核心产业规模约4500亿元，占全国近四成；聚集超2500家AI企业。率先建成3万卡级国产智算集群，智谱、面壁等国产大模型已经完成与国产芯片深度适配。

- “AI+制造”融合与智能终端创新主阵地。凭借强大的电子制造产业链和华为、腾讯等巨头，推动AI技术与硬件、场景深度融合。
- 2025年人工智能核心产业营收约2200亿元，AI规上企业超过2600家，深圳的训练芯片代表性企业已经成为国产AI芯片的主力军；在端侧AI芯片及推理芯片方面，多家公司相关芯片，已在细分场景实现商业化落地。

- 全产业链生态构建者与资本高地。形成从芯片设计、制造、封测到设备材料、EDA/IP的完整产业链，尤其在AI芯片设计领域优势突出。
- 2025年上海集成电路产业全市规模预计超4600亿元，同比增长24%。规上人工智能产业规模预计超5500亿元，增速超30%。已基本构建形成集成电路与AI全产业链体系。

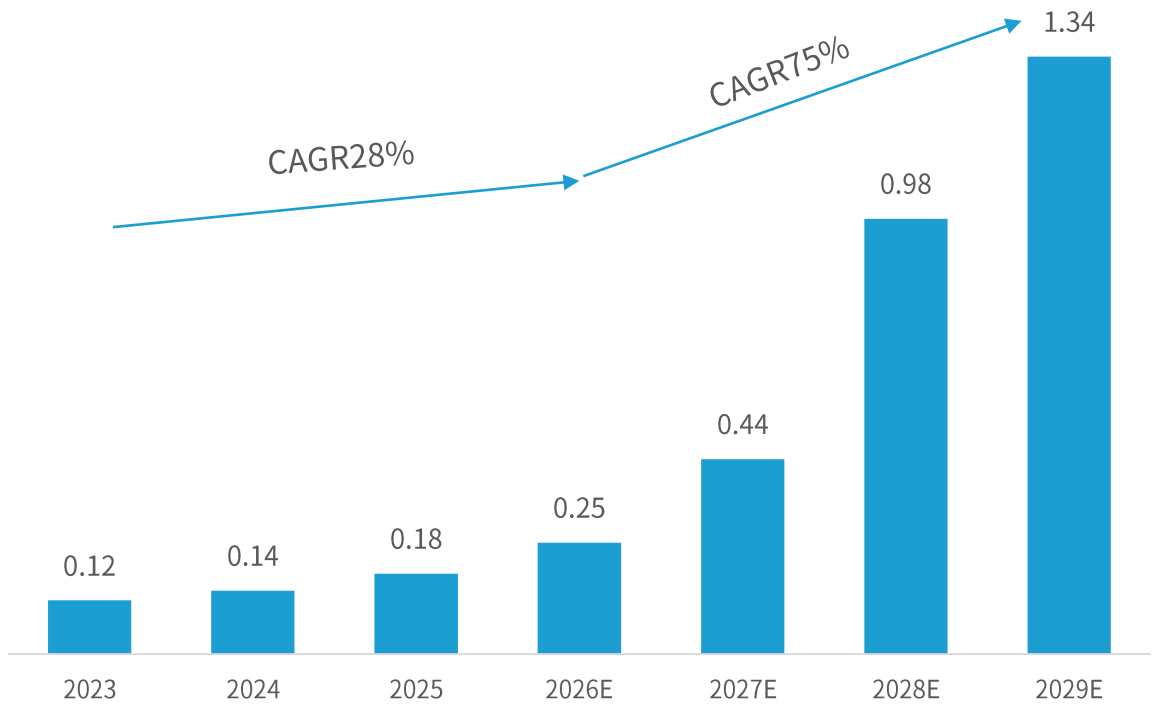
- 从“制造强”迈向“制造、设计双强”的AI赋能高地。依托中科大科研优势及“芯屏汽合”产业基础，聚焦存储芯片制造，并大力推动 “AI for Chip Design”（AI赋能芯片设计）。
- 2025年，合肥集聚集成电路企业超500家，形成龙头企业驱动格局，合肥高新区发力AI与高端通用芯片设计，

数据来源：东湖产业通、北京政府工作报告、上海经信委、深圳市发改委、安徽与合肥半导体行业协会、亿欧智库

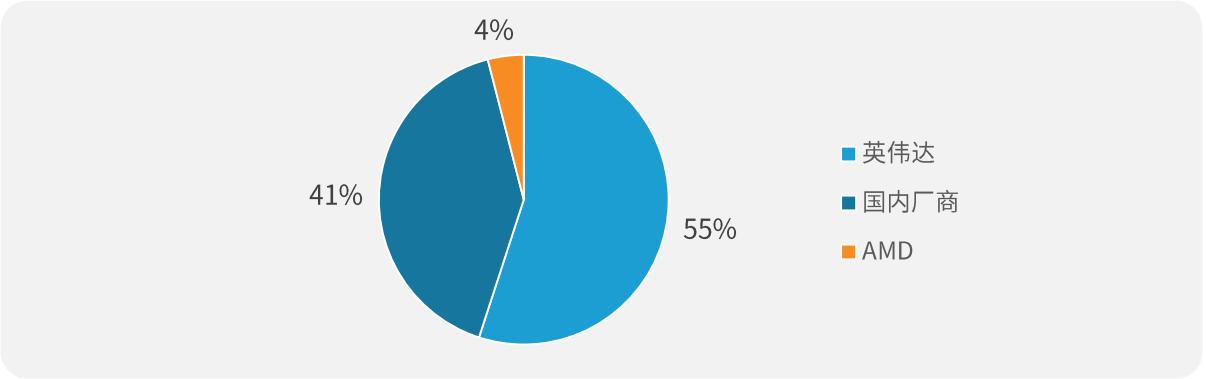
1.2.1 2023-2029年中国AI芯片市场规模及细分结构

- 根据三方机构数据，24年中国AI芯片市场规模约1425亿元，预计2026年将达到2500亿元，年复合增长率为32%。随着AI芯片供应链自主化进程加速，AI芯片市场规模将进一步加速增长，至29年激增至1.34万亿元。
- 2025年中国云端AI芯片出货量占41%的市场份额，英伟达市场份额跌至55%。AI算力需求从“训练主导”转向“训练+推理并重”，国产芯片在推理领域优势明显，预计中国推理芯片市场占比62%。

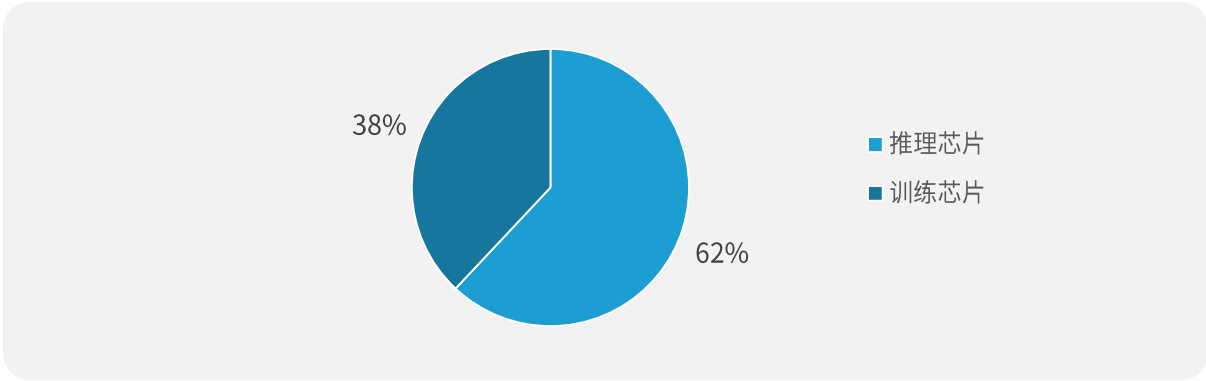
亿欧智库：2023-2029年中国AI芯片市场规模（万亿元）



亿欧智库：中国云端AI芯片市场份额占比



亿欧智库：中国AI芯片细分结构

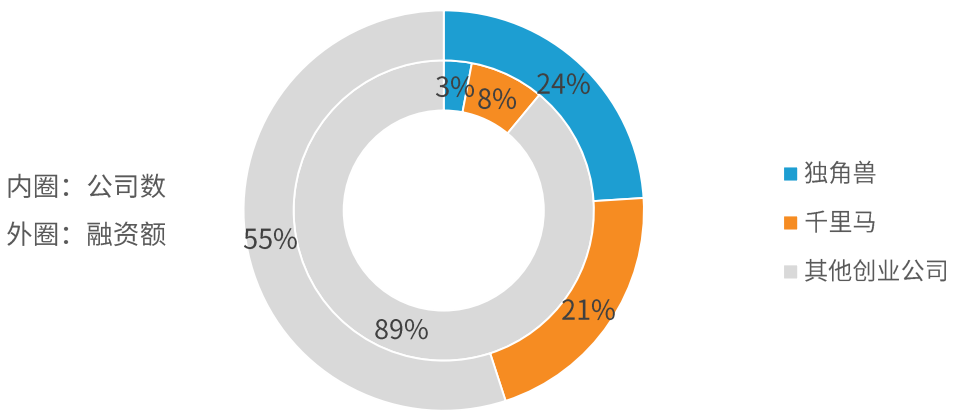


数据来源：弗若斯特沙利文、IDC、头豹研究院、半导体产业研究院等、亿欧智库

1.2.2 一级市场投融资热点：从通用GPU向DSA、存算一体、硅光等细分赛道转移

- 通用GPU作为AI算力的传统主力，其市场格局保持英伟达主导，叠加管制政策，投资窗口逐渐收窄。随着大模型推理需求的爆发和边缘AI的兴起，通用GPU难以在所有场景中保持最优。AI算力需求从训练向推理和边缘端扩散，需要更高能效比，更低延迟的专用解决方案。
- 因此融资已经不局限于通用GPU，而是朝着更加多元化、更具颠覆性的“专用架构”和“前沿技术”方向发展，多个前沿细分领域受到资本青睐，如DSA、存算一体、硅电等细分赛道。

亿欧智库：2025年国内芯片半导体创业公司融资生态



2025年中国芯片一级市场投资市场呈现“量增价减”的特征，投资事件数超1200起，但融资金额降至1244亿。占比九成的其他创业公司合计获得55%的融资额。充分说明了资本正从明星头部转向更广泛的细分环节，采取广覆盖策略，投资具备技术优势和落地优势的中小企业。

亿欧智库：DSA、存算一体、硅电获资本青睐

DSA及类DSA：产业资本深度入局

- 奕行智能核心路线"RISC-V+AI DSA"，获中国移动链长基金数亿元战略投资
- 浦软创投与隼瞻科技共同设立DSA领域计算实验室
- 全球首个基于RISC-V的DSA产业创新合作组织（RDSA产业联盟）在珠海成立

存算一体：资本加码进入千万级到亿元级规模

- 无锡产业集团发行存算一体科技创新债券，规模5亿元
- 蓝驰创投、中芯聚源、锦秋基金等联袂加持，微纳核芯完成超亿元B轮融资
- 感存算一体化技术产品及解决方案提供商铭芯启睿完成超亿元Pre-A轮融资

硅光：IPO引爆赛道，多家初创再获资本

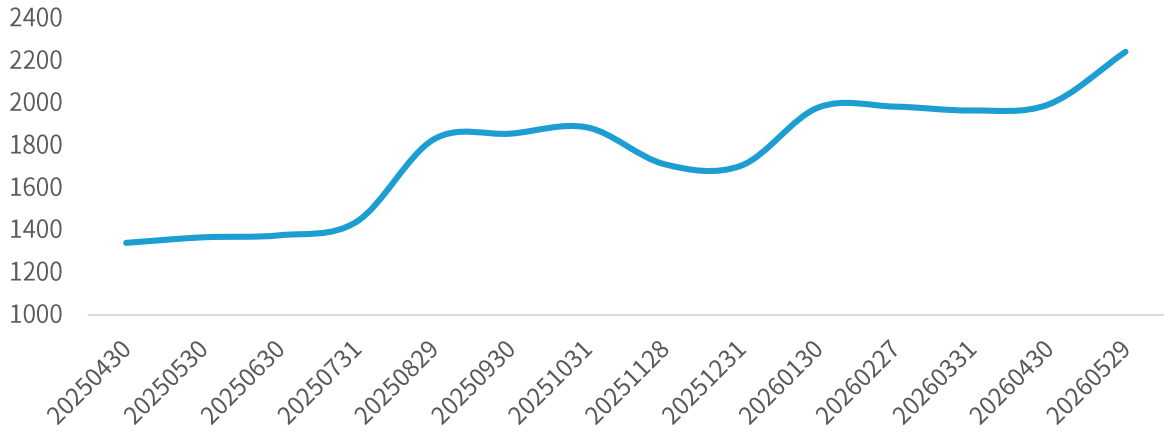
- 曦智科技2026年4月港交所上市，成为“全球AI光算力第一股”
- 灵动芯光完成数千万元天使++轮融资，聚焦芯片间光互连核心技术研发与产品落地
- 共进股份以5000万元增资硅光芯片设计企业弘光向尚，约持有其7%股权

数据来源：IT桔子、全球半导体观察、亿欧智库

1.2.3 二级市场表现与并购整合趋势

- 二级市场整体板块表现强势，内部分化显著，产业从概念普涨进入业绩验证与价值重估的深水区。2026年5月，AI芯片板块（BK1127）突破2100元，国产AI芯片龙头寒武纪因业绩爆发，股价突破1900元，创历史新高，带动产业链全线大涨。
- 国内产业资本以战略投资、参股与产业并购为主。企业产业整合主要围绕场景和产业链进行，并购以战略投资和参股为主。

亿欧智库：2025-2026年AI芯片板块市场趋势



亿欧智库：AI芯片部分龙头公司2026年Q1营收情况

寒武纪	海光信息	摩尔线程
2026年Q1	2026年Q1	2026年Q1
营收28.85亿元，同比+159.56% 归母净利润10.31亿，净利润同比+185.04%	营收40.34亿元，同比+68.06% 归母净利润6.87亿元，净利润同比+35.82%	营收7.28亿元，同比+155.35% 归母净利润0.29亿，实现盈利

数据来源：东方财富、亿欧智库

亿欧智库：国内产业资本战略投资情况

并购转向构建全场景能力，围绕主业“补链强链”

- 自动驾驶芯片公司黑芝麻智能拟收购端侧AI SoC公司亿智电子
- 芯片IP龙头芯原股份收购图像处理芯片公司逐点半导体
- 安凯微电子（物联网SoC）拟收购思澈科技（超低功耗物联网芯片）

产业链上下游强强联合

- 材料企业跨界：禾盛新材以2.5亿元增资ARM架构CPU设计公司熠知电子，持股约10%
- 芯片厂商投资核心场景：数模混合芯片龙头艾为电子战略投资Rokid（AR眼镜）

跨界并购追求长期价值

- 户外用品公司探路者并购贝特莱与上海通途两家芯片公司
- 包装企业大胜达参股芯瞳半导体，持股约23%

1.3.1 先进制程受限下的“后道工艺”突围

- 当更小纳米制程的芯片受到技术或外部限制时（如摩尔定律逼近物理极限或美国出口管制等），Chiplet（芯粒/小芯片）路线作为高性能芯片的新架构范式，带动先进封装进入架构层。
- 国内先进封测厂已在2.5D/3D等先进封装技术路线上实现对国际主流方案（如CoWoS）的技术对标，并开始提供可行的替代解决方案。然而，在高端工艺的成熟度、产能规模以及经济性良率方面，与国际头部企业相比仍存在显著差距。

亿欧智库：先进制程受限下AI芯片技术演进

传统大芯片

单die面积大

- 良率下降
- 成本上升
- 面积受限

Chiplet架构

拆分模块+不同制程+先进封装集成

- 提升良率，降低成本
- 灵活选用最佳制程
- 易于迭代，快速混合

芯片越做越大，单die的良率、成本和面积都会遇到限制。AI芯片已发展成由GPU、HBM、硅中介层、ABF载板、互连、散热和测试等组成的系统。AI芯片越复杂，封装越靠近架构。

Chiplet已成为高性能芯片绕开单芯片极限的重要方法：把计算、I/O、缓存、互连、模拟或射频等模块拆开，用不同制程分别制造，再通过先进封装组合起来。

亿欧智库：先进封装是后道工艺突围的核心引擎

传统封装：芯片和内存距离远，延迟高、带宽受限、功耗大。

先进封装：将GPU和HBM直接“贴身”堆叠在一起，或者垂直堆叠，数据实现短距离传输，速度极快。如2.5D/3D封装，是实现Chiplet互连的封装技术。

2.5D封装

将多个芯片并排放置在同一"中介层"上,通过中介层完成高密度互连, 再统一连接到底部基板。目前AI芯片最主流路线。

台积电CoWOS产能占据全球85%以上的市场份额，产能持续增长。国内长电科技、盛合晶微、通富微电等已实现规模化量产。

3D封装

将多个芯片在垂直方向上堆叠, 通过"超短距离互连(TSV或混合键合)"实现高速通信。随着算力需求提升，3D封装已成为产业演进方向。

台积电SoIC混合键合密度是国内的5-10倍，国内仍处于研发和小批量阶段。

数据来源：半导体产业纵横、电子工程、亿欧智库

1.3.2 本土晶圆厂与封测厂的协同创新进展

- 国内AI芯片产业链围绕晶圆厂前道工艺+封测厂后道技术的深度耦合展开的，在前沿技术、商业和生态层面深度融合。如2025年，27 家中国芯片企业联合组建的先进封装攻坚联盟，涵盖封装测试、设备制造、材料供应等全产业链环节。
- 产业链内的技术协同，正逐步破解 “设备 - 工艺 - 材料” 脱节的行业痛点，多家成员企业通过技术互补实现了效率提升。

亿欧智库：国内晶圆厂与封测厂形成一体化解决方案

- **垂直整合与产能协同**
  - **晶圆厂向下游封装延伸**：中芯国际成立芯三维半导体，布局先进封装领域；
  - **封测厂与晶圆厂共建**：长电科技与中芯国际合作共建了12英寸中道生产线；
  - **产业链内部深度绑定**：易卜半导体（聚焦2.5D/3D Chiplet封装）与汇成股份合作，形成“上游凸块配套+中端Chiplet封装+全链路测试”的深度产业协同。
- **技术共研与方案共创**
  - **联合开发新型封装方案**：通格微（沃格集团子公司）与AI芯片设计公司北极雄芯合作，共同推进基于全玻璃基板（TGV） 的多层互联AI芯片的产业化进程；
  - **协同攻克关键工艺与设备**：在混合键合（3D封装），国内设备商（如北方华创、华卓精科）正与封测厂协同推进设备验证与工艺开发。
- **打造自主可控的交付体系**
  - **形成“本土芯片+本土封装”交付模式**：国内头部AI芯片设计公司（如华为昇腾、寒武纪、燧原科技等）已与本土封测龙头深度绑定。例如，盛合晶微已成为华为昇腾
  - **协同提升产能与良率**：通过协同，国产先进封装的良率已提升至90%以上，长电科技的2.5D封装良率已达99.9%
- **生态构建与标准探索**
  - **建立协同创新平台**：中芯国际牵头成立了先进封装研究院，打造“政产学研用”一体化创新平台；
  - **应用端厂商深度参与**：以华为为代表的系统应用厂商，正深度介入先进封装系统集成的国产替代进程；
  - **探索国产技术标准**：国内正尝试依托创新链和产业链的“链主”企业联合上下游，共同建立涵盖接口、封装要求和场景应用的国产标准和规范。

数据来源：艾邦半导体、中国科技网、西安高新技术开发区、亿欧智库

1.3.3 供应链安全与国产EDA工具的应用现状

- 国内AI芯片供应链因外部技术封锁和国内突破进程并存，呈现非对称竞争策略，在无法突破的环节（如EUV）探索系统级替代方案；在已突破的环节（如部分刻蚀设备、封装）加速产业化导入与良率提升；在机遇窗口期（如RISC-V， AI for EDA）全力换道超车。
- 国产EDA在AI芯片供应链安全中扮演着“创新引擎”与“安全基石”的双重角色，并开始深度赋能国产AI芯片设计。

亿欧智库：AI芯片供应链安全现状

- 美国BIS要求三大EDA巨头停止向中国大陆提供核心EDA工具
- 美国通过三级算力管控、全球禁运、出口许可抽成等措施持续收紧对AI芯片的封锁
- 晶圆代工端面临限制风险

海外封锁持续加码

自主成果已现规模

- 英伟达中国市场份额约从95%跌至55%，国产替代效应开始显现
- 北京市政府提出2027年前达成半导体设计制造完全自给
- 上海计划2027年实现数据中心AI芯片“自主控制率”超七成

新的替代方案正在涌现

- 国内“技术突围”的路径逐渐清晰，Chiplet技术实现性能跃升是代表性替代路径之一
- 新技术验证了通过存算一体等新架构绕过先进制程封锁的可能
- 国产14纳米AI芯片在特定AI推理场景中,其特定精度或和单位算力成本可媲美英伟达采用4纳米产品

亿欧智库：国产EDA工具的应用现状

市场格局：梯队化发展，全流程能力构建加速

- 国内已形成近百家EDA企业的产业生态，行业正通过并购整合快速补齐工具链。全流程工具实现突破，模拟电路和平板设计工具已成熟。

技术范式：AI深度融入，从辅助工具迈向智能体自洽

- 发布全球首款模拟电路AI EDA工具，助力实际效率，效率提升10倍。中低端数字芯片工具已基本完成替代，正逐步向高端领域渗透。

关键进展：在AI设计全链条实现多点突破

- 在数字验证与仿真、模拟/混合信号与制造端EDA、系统级与先进封装EDA等均实现突破。已能支持先进工艺、Chiplet、3D集成等复杂系统设计。

数据来源：专家访谈、中国工业报、中国半导体行业协会、亿欧智库

亿欧智库

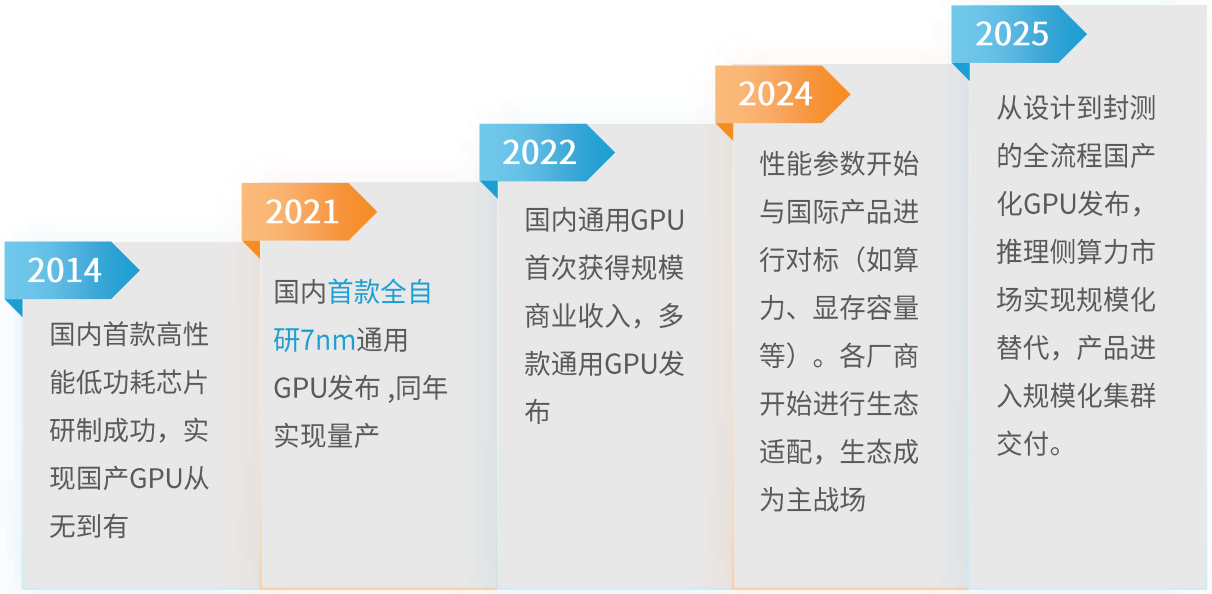
目录

CONTENTS

2.1.1 国产通用GPU的迭代进展与生态适配

- 国产通用GPU实现从“可用”到“好用”的突破，正处于从1到N规模化扩张的黄金期。产品能力从单卡扩展到超大规模集群交付，进入主力算力基础设施行列。随着"国产GPU四小龙"密集登陆资本市场，行业已从"技术验证"进入"量产迭代与商业突围"的新阶段。
- 国内企业形成CUDA兼容适配+原生重构双路径推进模式，实现开发者“零感知迁移”。国产CPU对CUDA API接口兼容性已实现显著提升，头部企业的核心算子兼容率突破95%，在此基础下，代码迁移的额外开发工作量可控制在原代码总量的20%以内。

亿欧智库：国产通用GPU迭代关键节点与里程碑



标杆化案例：智算中心落地

多地国家级、省级大型智算中心完成国产GPU批量采购与部署，形成规模化商用大单，行业认可度全面跃升，标志国产算力从技术验证迈向大规模工程化应用新阶段。

亿欧智库：国产通用GPU生态从兼容到共生

全球90%以上的AI大模型训练与推理均依赖CUDA生态。传统国产GPU集群的平均算力利用率不足40%，远低于英伟达GPU的80%以上，核心原因便是生态适配不完善，硬件性能无法充分释放。

生态壁垒主要表现在三个维度：

- 算子适配差距：稀疏算子、动态shape等冷门算子仍存在适配缺口。
- 框架兼容不足：主流AI框架深度绑定CUDA，国产GPU需通过适配插件实现兼容，初期适配效率较CUDA低30%-40%。
- 开发者惯性：国内AI从业者众多，但能熟练使用国产异构计算工具链的深度开发者尚不足CUDA生态的1/10，开发习惯切换成本较高。

目前，全栈生态逐步完善，核心环节取得多项突破。

- 国产GPU驱动稳定性持续提升，驱动程序平均无故障运行时间达到国际同类产品水平
- 国产AI框架持续优化，开发者及模型数量持续增加
- 生态适配场景持续拓展，覆盖AI训练、推理、视频渲染等领域，涉密场景适配进度加快

数据来源：信通院、中科曙光、华为、摩尔线程、亿欧智库

2.1.2 专用架构（DSA/ASIC）的爆发：针对大模型推理的极致优化

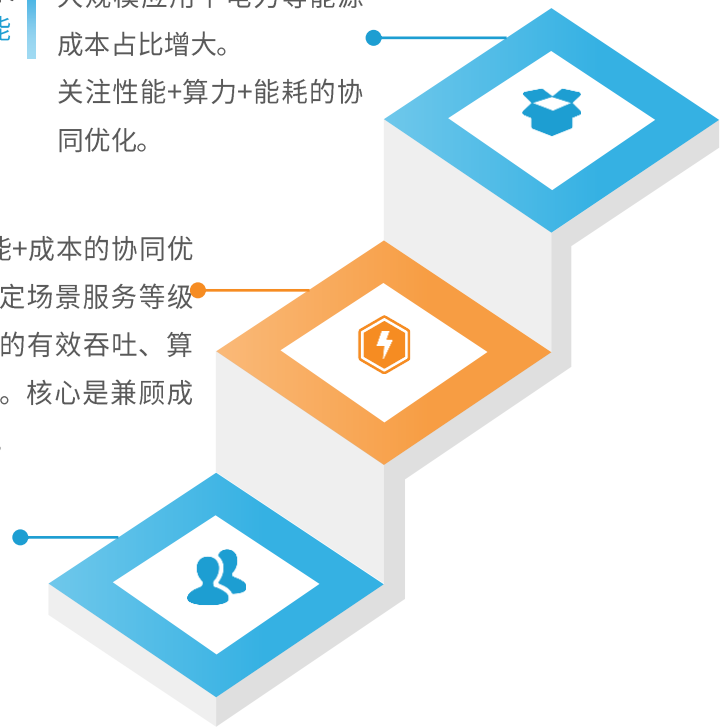
- 大模型从实验室走向商业化，算力需求从峰值训练延伸到长期推理，推理阶段的持续运营成本和实时响应速度成为关注重点。
- 专用架构的核心思想是用确定性换极致的效率。通过针对大模型推理的确定性计算图进行硬编码式优化，实现性能提升、功耗降低、成本下降，本质上是对"通用性冗余"的精准剔除与"特定负载的极致适配"，更加适合特定目标的系统性提升。

亿欧智库：大模型推理发展趋势

未来
目标：提效+降本+节能
大规模应用下电力等能源成本占比增大。
关注性能+算力+能耗的协同优化。

当前
目标：提效+降本
关注效果+性能+成本的协同优化，在满足特定场景服务等级目标下，系统的有效吞吐、算力成本压缩等。核心是兼顾成本和用户体验。

初期
目标：提效
关注时延、吞吐类指标优化。
未考虑差异化场景目标要求、算力成本、资源重组等。



亿欧智库：专用架构的核心优势

极致性能

- 计算效率暴涨，实现近100%的硬件利用率。
- 确定性数据流极大减少数据移动距离和能耗。
- 延迟极低且可预测，实现亚毫秒级、波动极小的确定性延迟。

超高能效

- 去除冗余功耗，功耗大幅降低。
- 精准匹配推理所需的低精度格式，计算能效提升数倍至数十倍。
- 支持稀疏计算，直接处理有效数据，实现“实际算力”和能效的双重提升。

显著成本优势

- 在同等工艺下，单片晶圆可产出更多芯片，制造成本下降。
- 系统级成本大幅优化：电费和配套散热设施成本减少。节省数据中心空间和机柜租赁费以及额外硬件费用。
- 规模效应放大优势，超大规模数据中心，节省数亿美元级别的支出。

垂直场景深度优化

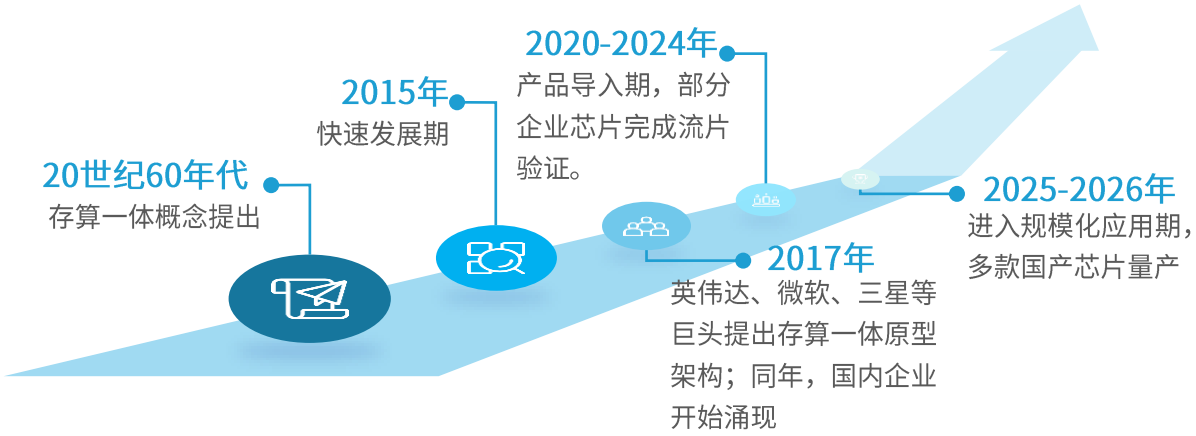
- 硬件与软件栈协同设计，实现“1+1>2”的效果。
- 集成特定功能单元，降低系统复杂度和延迟。

数据来源：信通院、专家访谈、亿欧智库

2.1.3 新型计算范式：存算一体与类脑芯片的产业化临界点

- 在AI大模型和端侧智能的快速发展下，传统冯·诺依曼架构中计算与存储分离、数据来回搬运的计算范式，使得传统芯片在内存和功耗上已无法满足现有需求。单纯的提升芯片制程、堆叠核心，只能小幅度优化，无法突破底层物理限制解决问题。
- 新型计算范式存算一体与类脑芯片，主要解决内存和功耗问题。存算一体可消除数据搬运开销，从而实现能效10-100倍提升；类脑芯片可实现超低功耗、高实时性以及拥有自适应学习能力。两者在技术、市场、资本的叠加作用下，已逐步进入产业化，其中存算一体已经来到产业化临界点，类脑芯片在超低功耗感知场景完成小规模验证，正高速逼近产业化临界点。

亿欧智库：存算一体产业化发展及典型代表



典型代表

- **知存科技**：基于NOR Flash的存算一体芯片已累计实现超千万颗出货量，广泛应用于智能可穿戴设备。同等算力下，能耗仅为行业同类方案的10%。

典型代表

- **后摩智能**：存算一体架构的AI芯片M50进入量产阶段，已通过整机集成的方式在B端落地，并于5月搭载联想AI主机P7发布。单芯片最高可提供算力160TOPS，典型功耗10W。

亿欧智库：类脑芯片产业化发展

国外发展：科研+商用双路线

IBM：工程化类脑芯片，2014年发布第一款，2023年迭代款较第一款产品能效提高25倍，面向嵌入式工业感知。仅供给高校、科研院所，无消费级量产。

Intel：2018年发布首款类脑芯片，2021年迭代款单芯片能效比同等GPU高1000倍；2025年最新迭代款支持多芯片互联搭建类脑超算，面向机器人、事件视觉科研，未面向消费电子大规模量产。

其他：BrainChip、SynSense 等企业实现商业化营收，工业、安防等场景标准化落地。

国内发展：科研领跑，商业化逐步落地

科研：清华大学发布的类脑互补视觉芯片“天眸芯”，带宽可降低90%、高动态视觉感知，适配车载、AR视觉；浙江大学达尔文系列，数字类脑芯片，单芯片百万级神经元，支持片上实时在线学习。

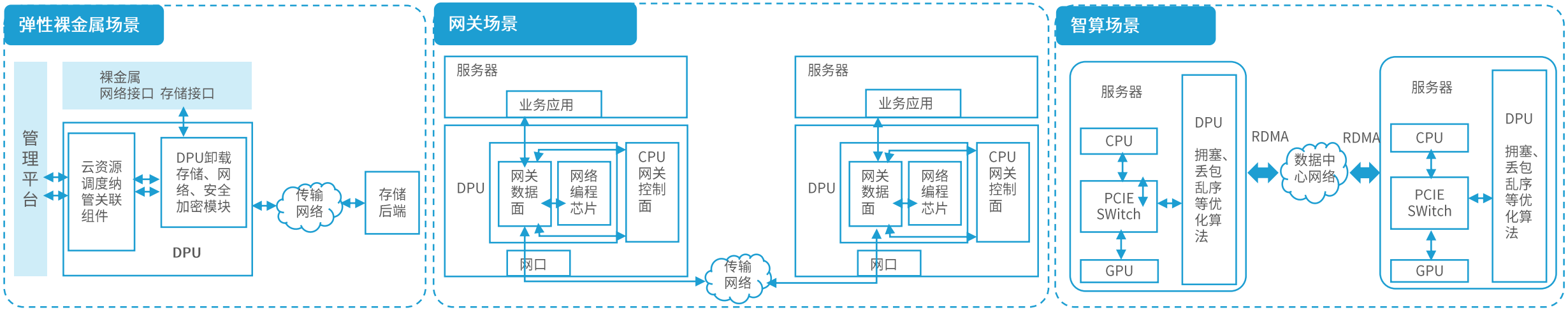
商业化：灵汐科技发布“领启® KA200”系列类脑芯片，配套完整国产化类脑软件栈；西井科技推出可商用化的类脑神经元芯片（Deepwell），主要应用于智慧医疗、智慧港口城市、无人机无人车、智能终端等；时识科技（SynSense 中国分部）批量供给工业、车载厂商。

数据来源：《存算一体技术产业发展研究》、半导体产业纵横、亿欧智库

2.1.4 DPU/智能网卡架构：数据面专用加速与AI集群算力释放新范式

- DPU已超越基础的“降本增效”，在AI推理与超大模型训练场景中，其核心作用演变为重构数据路径与内存层级，以系统级优化直接释放算力红利，负责基础设施卸载、加速和隔离，通常具有独立处理器、内存和软件运行环境。
- 同时，DPU作为智能的数据调度中枢，实现对网络、存储、安全协议及虚拟化功能的彻底卸载与硬件加速，使CPU与GPU专注于核心计算，在提升算力中心整体性能与利用效率的同时，有效降低硬件冗余与总体能耗。

亿欧智库：DPU在算网融合场景下的应用



DPU将传统由主机CPU承担的数据路径处理、加密校验等高开销任务迁移至DPU内置的存储加速引擎，降低CPU占有率。

- 在高带宽链路下娴熟降低有效数据传输量，减少CPU参与度与网络负载。
- 保障从主机内存到存储介质的数据一致性。同时，通过卸载垃圾、回收元数据管理等任务，延长存储设备使用寿命。
- 将存储服务从CPU解耦，释放主机算力用于AI训练等核心业务。

- DPU通过在硬件层面实现网络协议栈、低延时RDMA传输、存储I/O，加密解密等功能的专用加速。
- 在AI集群中承担构建推理上下文内存存储平台、卸载非计算任务释放GPU算力、承担机架内数据神经中枢等功能，实现对算力资源的细粒度调度与动态优化。

数据来源：信通院、邮电设计技术、亿欧智库

2.2.1 2.5D/3D封装在国内的落地瓶颈与突破

- 2.5D/3D先进封装正处在“瓶颈犹存，但突破显著”的战略机遇期。AI算力需求的爆发式增长，使得先进封装从可选项变为必选项，尽管全面落地仍面临一系列核心瓶颈，但是在技术、产业链和生态建设上也取得了显著突破。
- CoWoS封装是2.5D封装的典型封装技术，目前高端AI芯片使用的是国外大厂生态，国内尚无企业具备完整、高良率的CoWoS-S全栈量产能力。
- HBM是3D堆叠封装技术的典型产品，其突破是衡量国产3D封装能力的关键标尺。目前国内企业已量产HBM2e，HBM3/HBM3e正在突破。

亿欧智库：2.5D/3D封装落地瓶颈

关键设备与材料高度依赖进口

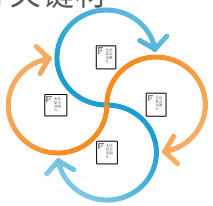
高端探针台、划片机等关键设备依赖进口，高端探针台国产化率不足10%，划片机国产化率约为10%-15%。

封装基板ABF绝缘膜、高端光刻胶等关键材料国产化率低。

技术积累与生态短板

在晶圆级TSV 技术、超高密度互联等前沿领域，与国际仍有差距。

国产Chiplet产品已实现小批量商用，部分兼容UCIe国际标准，但整体生态仍处于发展初期。



散热成为新瓶颈

随着3D堆叠密度提升，热流密度急剧增加，传统热界面材料难以满足需求，成为影响器件可靠性的关键障碍。

成本与良率挑战

2.5D封装中，硅中介层的成本占封装总成本50%以上。

3D堆叠对TSV填充、键合工艺的良率要求极高（99.99%）。

亿欧智库：2.5D/3D封装国内突破

技术自主化与量产突破

- 国内部分厂商已实现类CoWoS技术的稳定量产，良率可达75%-80%。
- 国产CoWoS-S良率可达99%，接近台积电水平，CoWoS-L实现部分量产。
- 创新技术路径涌现，直接适配HBM的先进封装工艺。

核心装备国产化突破

- 自主研发的TCB设备，已成功完成国产AI芯片的CoWoS-L测试打样，实现国产TCB设备在AI芯片CoWoS封装领域的首次突破。国产3D集成混合键合设备已完成客户端验证，对准精度达纳米级。

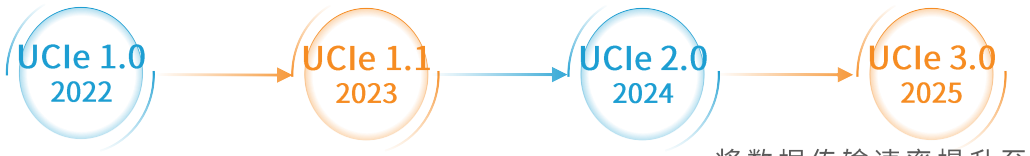
中国2.5D/3D先进封装产业已实现“从无到有、从有到用”，头部封测厂已实现稳定量产并导入头部客户。2.5D封装已形成稳定量产能力，有力支撑了国产AI芯片的自主供给。3D封装在HBM等关键产品上取得突破，整体仍处于追赶阶段。在算力自主的战略牵引下，国产先进封装将从“保障供应”向“技术引领”迈进。

数据来源：半导体产业报告、摩根大通、亿欧智库

2.2.2 异构集成技术（Chiplet）标准化进程与互联协议

- 早期各厂商全都是封闭私有互联接口，不同厂商芯粒无法互通，芯粒复用性差，产业链碎片化，生态割裂。异构集成技术需要异构混搭不同工艺芯片，UCLe通用芯粒高速互联标准打通跨厂商芯粒互联互通，是后冯·诺依曼时代多范式产业化的底层硬件基础。
- 国际UCIe完成1.0/1.1/2.0/3.0迭代，联盟覆盖全球头部芯片、封测、云厂商，已经进入规模化商用落地期，成为全球Chiplet事实通用接口标准；国内完成行业联盟标准发布、国家级国标立项，形成自主芯粒互联标准体系，兼顾国际UCIe兼容与国产化自主可控。

亿欧智库：国际互联协议标准化进程



在 UCIe 1.0 基础上优化电源管理、信号完整性、多芯粒故障校验与封装兼容性规范，补齐量产落地所需工程细则，解决初代标准稳定性短板，成为芯粒原型批量验证的基准版本。

将数据传输速率提升至64 GT/s，边带信道长度扩展至100毫米等，引入多项关键技术，旨在提升多芯片系统封装（SiP）设计的能效与灵活性。

异构集成技术是把不同工艺、不同功能的裸片分开单独流片，再通过先进封装封装到同一封装内，因此需要依靠高速互联协议实现片间高速通信。英特尔、AMD、台积电、日月光、三星、微软、谷歌等十家行业巨头联合成立 UCIe 产业联盟，发布UCIe互联协议，可以实现跨公司Chiplet生态。

亿欧智库：国内和其他互联协议标准化进程

中国标准化进程	具体内容
HiPi联盟《芯粒互联接口规范》系列（国标）	集成电路芯粒领域的首批国家标准，标志着我国芯粒标准化建设取得里程碑式突破，为产业规范化、规模化发展提供了顶层指导。
CCITA《小芯片接口总线技术要求》（团标）	该标准规定了小芯片接口的总体架构、协议层、链路层、物理层及封装要求。其一大特点是同时定义了并行和串行接口，总连接带宽可达1.6Tbps，以灵活适配不同应用场景和国内技术供应商的能力。标准在设计上与UCIe兼容，但在封装层面更侧重于采用国内可实现的技术

其他互联协议	具体内容
BoW（Bunch of Wires）	更侧重于物理层和链路层的“裸线”标准，设计更轻量、灵活，被视为UCIe的补充或替代，尤其在对成本敏感或特定优化的场景中，如AI加速器Die-to-Die
AIB（Advanced Interface Bus）	采用单向信号通道，设计简单，虽然生态不如UCIe广泛，但已向业界开放，主要在英特尔生态内使用

数据来源：奎芯科技、电子信息、 UCIe、亿欧智库

2.2.3 DPU与AI芯片的Chiplet合封趋势

- 单靠GPU算力的堆叠，已经越来越难以解决AI计算的瓶颈，真正的限制来自内存带宽、互联延迟、以及计算与存储之间的协调效率。英伟达将芯片竞争转向系统博弈，发布Vera Rubin平台，该平台将CPU、GPU、DPU等芯片深度耦合，将计算、存储、网络在系统层面协同优化。在训练混合专家模型（MoE）时，所需的GPU数量减少至原来的四分之一，每个Token的推理成本最高可降低10倍。
- DPU不仅能接管原本由CPU/GPU处理的数据搬运、网络协议、存储管理等基础设施任务，将CPU/GPU算力释放给核心AI计算。还为AI提供了一个巨大的外部“工作记忆”，使其能处理超长上下文的对话。DPU与AI芯片的Chiplet合封获得超高速片内互联，使DPU能够高效协同CPU、GPU和存储系统，实现数据流的高效调度与卸载，有效突破AI计算的“内存墙”和“通信墙”瓶颈。

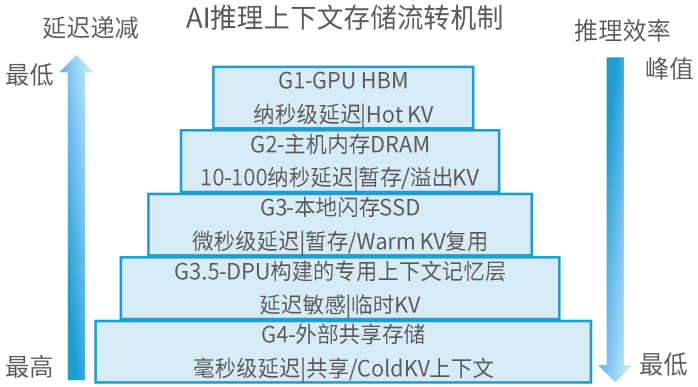
亿欧智库：DPU功能

释放CPU与加速器算力

- DPU通过硬件级卸载网络、存储及安全任务，有效释放CPU与加速器的算力。
- DPU可采用芯粒(Chiplet)设计，集成专用网络处理单元、可编程计算核心与高速I/O接口，通过片上互连网络实现内部功能模块的高效协同。

实现GPU间低时延、高吞吐数据交换

- 在互联架构层面，DPU重点支持单节点内Scale-Up与跨节点集群Scale-Out 互联，依托远程直接内存访问等先进协议，实现GPU间的低时延、高吞吐数据交换，缓解万卡级AI集群的通信瓶颈。



DPU可赋能推理上下文记忆存储平台(ICMS)通过卸载KVCache的I/O路径与元数据管理，构建高速、可共享的“记忆层”，有效加速大模型推理中的上下文复用，减少GPU间数据搬运开销。如NVIDIABlueField-4 DPU。

亿欧智库：DPU与AI芯片合封逻辑，减少片外互联瓶颈

突破“内存墙”瓶颈

AI训练需要频繁读写海量数据。DPU与GPU、HBM（高带宽内存）等近距离封装，可极大地缩短数据路径，提供远超传统板级连接的带宽和更低的延迟。

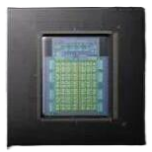
降低系统功耗

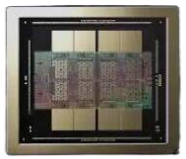
芯片间近距离通信，其能耗远低于长距离的板级传输，有助于构建更绿色节能的计算系统。

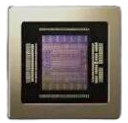
实现“异构计算”

AI计算任务多样，各芯片负责不同，如GPU擅长并行计算，DPU擅长处理网络和数据。将它们合封，能各司其职、高效协同。

英伟达Vera Rubin平台


Vera CPU


Rubin CPU


NVLink 6 Switch


ConnectX-9


BlueField-4 DPU


Spectrum-6

英伟达的Rubin平台是Chiplet设计理念的集大成者，通过将计算、网络和存储等不同功能（六款专用芯片）的芯粒“合封”在一起，整合成一个完整的“AI工厂”基础设施，构建面向AI时代的超级计算平台。

数据来源：英伟达、信通院、专家访谈、亿欧智库

2.3.1 从“兼容”到“原生”：国产AI芯片软件栈的成熟度评估

- 国产AI芯片软件栈已分化为三条清晰的技术路径：兼容软件栈、垂直软件栈、原生软件栈；其成熟度评估需从“生态开放性”、“性能深度”与“商业效率”三个维度综合考量。而路径的选择，取决于应用场景和市场窗口期的权衡。
- 短期看，“兼容路径”是抢占市场的现实选择；长期看，“原生路径”是构建自主可控数字根基的必然。当前，国产AI芯片软件栈整体正处于从“可用”迈向“好用”的关键阶段，全行业仍处于攻克“工具链完备性”和“算子库深度”两大核心短板的攻坚期。

亿欧智库：国产AI芯片软件栈成熟度评估

	兼容软件栈	垂直软件栈	原生软件栈
核心逻辑	优先兼容主流生态（如CUDA），以最低迁移成本快速切入市场	不追求全场通用性，聚焦特定场景（如大模型推理、自动驾驶等），打造极致效率。	构建自主技术体系，追求软硬件深度协同与长期生态壁垒
优势	• 用户迁移成本相对较低 • 主流模型可快速部署 • 降低开发者切换门槛	• 目标场景性能与能效比突出 • 软硬件耦合度高 • 市场空间相对聚焦，快速实现商业闭环	• 自主可控上限高 • 软硬件协同优化空间大 • 长期生态粘性强
核心瓶颈	• 受制于上游生态迭代节奏 • 兼容层可能带来一定性能损耗 • 差异化优势相对优势	• 生态扩展性差，跨场景拓展需要重新投入 • 技术路线风险高，对特定场景依赖度较高 • 市场天花板明显	• 生态建设周期长 • 初期应用匮乏 • 开发者需要重新学习
成熟度关键指标	• API兼容覆盖率 • 性能损耗比 • 新CUDA特性跟进速度	• 目标场景市占率 • 场景专用工具链深度 • 客户定制化支持能力	• 原生框架采用率 • 开源社区活跃度 • 第三方应用数量
适用场景	希望快速验证、降低迁移风险的互联网公司、高校及中小型AI企业	对功耗、成本、实时性有极端要求的边缘推理、车载、安防等终端场景	对自主可控有强需求、有长期投入决心的国家重大工程、大型云服务商

数据来源：《国产AI芯片软件生态白皮书》、亿欧智库

2.3.2 开源社区贡献度与开发者生态建设

- 中国在开源生态建设方面已完成从学习使用向深度参与的跨越，从贡献量、采用量到社区规模，中国开源生态已实现结构向跃升，成为全球开源体系的重要力量。同时行业协同也开始出现，如制定生态兼容性标准，地方政府主导的智算中心在实际部署中采用多厂商芯片混合架构等。
- 不同开发者社区规模差异化明显，中国开源社区贡献度和社区活跃度均呈头部集中态势，贡献度结构不均衡，重应用轻基础。

亿欧智库：开源社区贡献度现状

01 贡献度分布与社区活跃度头部集中

- 国内以昇腾为代表的生态贡献度最高，体现在官方文档完备、社区问答活跃、开源代码仓库更新频繁等，且对全球开发者的吸引力强。部分国产AI芯片在基础文档和代码开源上已有覆盖，但在高阶调优文档、社区讨论活跃度及开源代码的持续维护度相对较弱。
- 国内社区昇腾/MindSpore依托华为云开发者社区和异思论坛，技术干货征集与问答活动频繁，形成相对稳定的用户群体。多数厂商仍处于“社区建设期”，比如部分厂商的官方论坛或讨论区的规模多为几十到数百主题，部分讨论被分散在微信群、公众号和合作社区中。而NVIDIA长期维持十万级主题与回复，几乎任何问题都能在社区中找到类似案例。

02 贡献度结构不均衡

- **重应用轻基础**：国内AI开源项目多集中在应用层，底层基础软件（如底层算子库、编译器、核心计算库）的原创贡献相对较少，开发者在核心开源项目中的决策参与度偏低。
- **主体贡献差异**：企业贡献多集中在代码和工具链的开放，个人开发者贡献多集中在文档编写、问题解答和社区维护，缺乏将开源贡献与商业价值深度结合的高端人才。

亿欧智库：AI芯片开发者生态建设模式

- **建设路径**：依托“芯片+框架+云服务”全栈协同，提供完整的工具链（如CANN、MindSpore）和丰富的行业案例，通过官方社区（如昇思社区）提供全栈培训与认证，降低开发者迁移成本，形成高粘性的开发者闭环。
- **贡献度**：通过开源核心模块、提供丰富的示例代码，吸引开发者进行应用开发和模型迁移，实现生态“反哺”。

垂直聚合与产业联盟模式

- **建设路径**：通过聚合产学研用多方资源，提供算力、数据集、模型和工具链的共享平台，降低开发者的入门门槛，实现资源的普惠。
- **贡献度**：通过项目共建共享，吸引开发者参与跨架构迁移、高质量数据集标注和模型优化，实现生态的协同共创。

兼容生态模式

- **建设路径**：在接口和工具上对标国际主流（如CUDA/ROCm），降低开发者的学习成本和迁移成本，通过提供兼容的开发体验吸引原有CUDA生态的开发者。
- **贡献度**：通过开源兼容层（如torch_musa）和提供丰富的入门示例，吸引开发者进行跨平台适配和兼容性开发，扩大生态的覆盖面。

数据来源：《国产AI芯片软件栈白皮书》、亿欧智库

2.3.3 编译器等基础软件的优化对实际性能的影响

- 在AI芯片系统中，编译器负责将上层训练好的模型转换为适配芯片的高效执行方案，包括计算图优化、算子融合和指令调度等。编译器、算子库、部署工具等共同决定了AI模型是否能稳定、高效地跑在芯片上（将理论性能转成真实性能）。
- 相关数据显示编译器优化后算子融合可减少10-30%的执行时间，内存减少20%-50%的显存使用，可提升2-4倍的推理速度、对于图地优化，可以减少5%-15%的计算开销。

亿欧智库：编译器等基础软件对性能的主要影响

性能提升，释放硬件潜力

- 算子融合优化：编译器通过全局分析，将多个连续的算子融合为单个内核，大幅减少内存访问和内核启动开销，通常可减少10%-30%的执行时间，部分场景下可实现2-4倍的推理加速。
- 内存与访存优化：减少内存碎片和系统调用开销，提升内存带宽利用率，部分场景下可减少20%-50%的显存使用。
- 并行与调度优化：通过自动并行化、循环重排、同步优化，合理分配GPU/NPU的线程和计算资源，提升硬件利用率，提升整体吞吐率。
- 精度与计算路径优化：通过量化和稀疏化等编译优化，在保持模型精度的前提下，大幅减少计算量，实现推理速度的成倍提升。

开发与部署效率，降低门槛，提升迭代速度

- 跨平台与自动优化：编译器通过硬件抽象和机器模型构建，可实现“一次编译，跨平台优化”，降低不同芯片架构的适配成本，使模型能快速迁移并达到接近原生硬件的优化性能。
- 自动调优与智能化：通过自动调优（AutoTVM）和机器学习驱动的编译优化，自动搜索最优的编译配置，减少人工调优成本，缩短模型部署周期，适应快速变化的AI算法。

资源优化，降低资源消耗

- 显存与内存节省：通过编译时静态内存分配与运行时动态内存分配的混合优化，减少中间激活值的存储，降低大模型推理时的显存占用，内存可节省25%，性能提升18%，缓解“内存墙”瓶颈。
- 功耗与能效优化：通过减少不必要的数据搬运（存算协同优化）和动态编译自适应，降低芯片的功耗，在端侧设备或高并发场景下，显著提升能效比。

实际性能与理论算力“弥合”

- AI芯片的理论算力受限于内存带宽、数据调度、计算单元利用率等系统瓶颈，无法完全转化为实际性能。编译器通过系统级的全局优化，将芯片的理论算力高效的转化为实际的有效算力，弥补硬件代差。

数据来源：《国产AI芯片软件生态白皮书》、《硅基大脑进化论：AI芯片全景解析》、亿欧智库

亿欧智库

目录

CONTENTS

3.1.1 互联网大厂自研芯片的规模化应用与对外输出

- 国内主要互联网大厂自研芯片规模化应用与对外输出阶段各不相同，但对外出售基本不卖裸片，通常以加速卡/板卡、云实例/云服务提供算力、整机/行业一体机、IP授权等方式进行商业化变现。
- 阿里、百度、快手（凌川科技）已实现商业化规模化应用，阿里全栈自研芯片真武系列2025年出货26.5万张，真武系列AI芯片累计出货56万片，除自用外服务于多家客户；百度昆仑芯片25年出货11.6万张；字节SeedChip推理芯片纯自用，不对外。腾讯主要以采购芯片为主，快手集团异构计算与芯片事业部拆分独立运营，为凌川科技，其团队在快手内部自研的SL200智能视频SoC芯片，持续在快手部署数万颗、稳定服务7亿用户。

亿欧智库：国内主要互联网大厂自研芯片规模化应用与对外输出情况

公司	芯片名称/型号	规模化应用情况	对外出售情况	售卖方式
阿里	含光、真武系列	真武系列AI芯片截至2026年4月累计出货56万片	已经服务电信、中国一汽、浦发银行等20多个行业的400多家客户	主要通过阿里云对外租算力或IP授权；真武高端款仅以整机形式销售，不单卖裸片
百度	昆仑芯	出货稳定在数十万颗级别	支撑文心大模型内部训练推理，对外开放算力租赁服务，同步向第三方智算中心输出芯片产品；目前外部业务的比重已经超过向百度内部供货	以加速卡 / 整机销售；可被集成至服务器厂商方案中
字节	SeedChip推理芯片	2026年3月完成工程样片，计划年内产出10万颗，逐步提升至35万颗	纯自用，不对外	/
腾讯	无成熟量产产品	方案设计阶段	/	/
快手（凌川科技）	SL200智能视频SoC芯片	已销售近十万颗	部署至快手、阿里云、百度云、B站等互联网公司，覆盖快手99.7%的直播转码业务	加速卡＋边缘盒子＋一体机；联合硬件厂出方案

数据来源：阿里、百度、字节、腾讯、亿欧智库

3.1.2 智算中心建设热潮下的国产芯片“入场券”与运营现状

- 智算中心建设热潮下，国产AI芯片正迎来国产替代的黄金期，在“东数西算”战略、大模型爆发及国产化政策多重驱动下，国产芯片从“单点替代”走向“体系化突围”。国产AI芯片要进入智算中心需要满足政策合规、技术达标、集群/软件适配、稳定交付等多重条件。
- 智算中心的运营现状呈现出“建设火热、利用不足，结构错配”的复杂局面。目前国家级建成/在建25个国家级智算中心，250+地方级智算中心，整体利用率不足40%，行业机构测算显示，国内全行业智能算力需求规模约123.6 EFLOPS，但市场有效供给仅57.9 EFLOPS，供需结构性矛盾突出，算力闲置与短缺并存。

亿欧智库：智算中心国产芯片“入场券”

政策合规门槛

国测Ⅰ级认证：政府及关键领域采购的前置条件

国产化率要求：国家智算中心要求优先采用国产芯片，部分项目要求100%国产化

供应链安全：芯片设计、制造、封测全链路自主可控（长期目标）

技术性能门槛

算力密度：单芯片算力需达数百TFLOPS级别

集群拓展：支持千卡/万卡级集群互联

显存容量：支持大模型参数加载

推理性能：INT8精度下高吞吐、低延迟

算法适配：支持主流大模型架构的完整训练与推理

软件生态门槛

模型适配：保证软硬件模型适配，软件兼容性适配国产硬件需求

算子覆盖率、编译器支持图优化、算子融合、自动并行

商业供应门槛

产能供应：稳定产能与交付能力

稳定可靠性：支持宽温域（某个温度长时间）工作环境，保证7×24小时稳定运行

价格竞争力：维护成本能耗、单位token成本（推理型）、运维成本、故障维修成本

不同地区省份准入要求不同，不同级别入场要求亦不同，国家级要求最高

亿欧智库：智算中心运营现状

全国智算中心GPU利用率分化极大，2026年一季度，已投运的商用智算中心GPU平均利用率只有30%，其中头部云厂商的专属集群利用率在75%-98%；省会城市的中型公共智算中心，25%-40%；三四线城市千卡级小型集群利用率仅10%-22%，严重闲置。

需求侧

京津冀、长三角、粤港澳占全国算力需求的65%以上，同时AI企业、高端制造业智算需求占比较高，时延要求严苛。同时八大枢纽节点未覆盖中部地区，湖北等中部省份需自行填补算力空白。

供给侧

大量算力建在西部，但西部及三四线城市缺乏AI产业生态支撑，自建智算中心闲置严重。西部算力枢纽因跨区调度未贯通，只有30%的智算中心能做跨省的算力调度，无法以富余通用算力弥补东部智算缺口。

运营良好

- 武汉人工智能计算中心，共建成200PAI算力，服务300+家机构，算力申请量同比增长900%，处于饱和运行状态
- 盐城超级计算中心实现机柜上架率100%、整体出租率73.8%

资源闲置

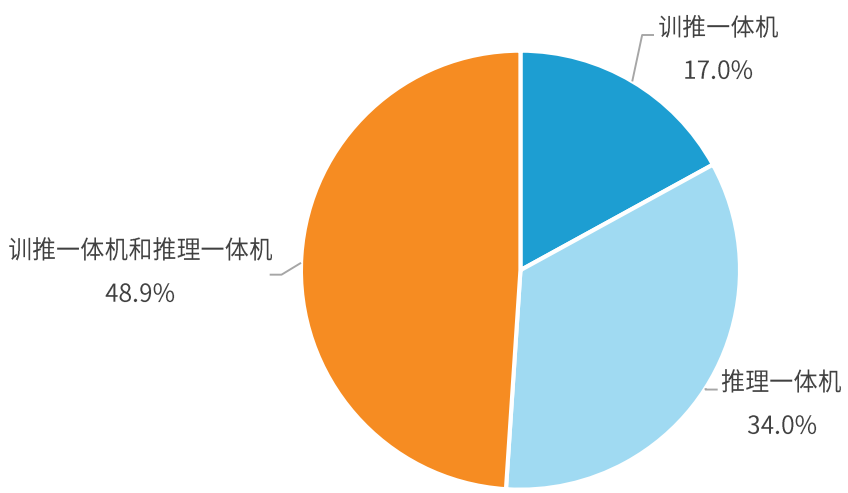
- 中部某省规划建设3个智算中心，总规划1200PFLOPS（每秒千万亿次浮点运算），但当地数字经济企业不足百家，实际算力需求不足200PFLOPS。
- 东部某市的两个区县建设了聚焦生物医药研发和主打工业互联网的智算中心，但两地产业基础薄弱，实际入驻企业不足10家，设备利用率低。

数据来源：《算力基础设施发展情况报告》、专家访谈、浪潮人工智能研究院、亿欧智库

3.1.3 大模型训推一体机的部署落地进展

- 2025年3月，工信部发布“算力强基揭榜行动”，将“训推算力一体机”列为重点揭榜任务，明确到2026年，研发支持至少3种指令集芯片的训推一体机，覆盖至少5个行业，至少10种场景。
- 2026年大模型一体机出货量预计39万台，同比增速超130%。从市场规模看，2026年训推一体高端机型市场规模约1760亿元，约占整体市场规模的60%，主要原因是由于其单价较高，单台多为30-200万。从供给端看，目前产业落地大模型仍以推理一体机为主。政策推动叠加市场需求，如政务、金融、安全等对数据安全及模型定制有强需求的领域，训推一体机仍将以较好势头增长。

亿欧智库：不同类型大模型一体机厂商分布



数据显示，产业中仅推出推理一体机的企业约占34.0%，仅推出训推一体机的企业约占17.0%，同时推出推理一体机和训推一体机的企业约48.9%。从类型看，目前产业落地大模型一体机仍以推理一体机为主。主要原因：

- 模型推理是AI落地的主战场，许多企业不再自己训练模型。
- 与训练相比，推理对算力的绝对要求低，一些中小型或创业公司能够凭借其在特定硬件或垂直行业优化上的优势切入市场。

亿欧智库：大模型训推一体机典型应用场景

政务

- 某市级政务平台通过部署训推一体机，在交通管理场景中，整合全市20万路摄像头数据，训练出多模态交通预测模型，使重点路段拥堵预警准确率达到92%
- **某政务大模型+70名“数智员工”，公文审核时间降低90%**

金融

- 某金融机构利用一体机构建反欺诈系统，将模型训练周期从72小时缩短至8小时，单笔交易推理延迟控制在2毫秒以内，误报率降低37%
- **某银行通过预训练语言模型与实时推理结合，客服系统响应延迟低于200毫秒，人力成本降低40%**

医疗

- 某医院一体机，千亿参数，140+科室，医学知识覆盖率98%，精准率超90%
- 美年健康109家体检中心"健康小美"，审核39万+份体检报告，重点指标精准率90%+

工业制造

- 为某企业搭建智能问答系统，在特定知识领域的回答准确度高达99%，端到端4K输入输出场景下实现首字时延秒级控制

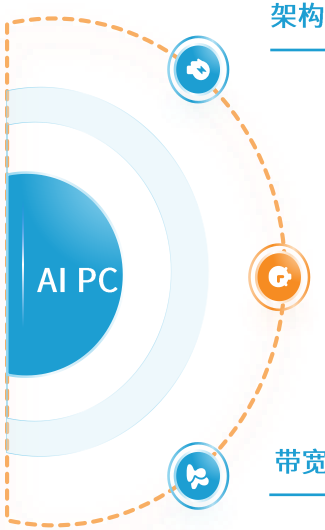
除此外，还在科研教育、物流、能源、文旅、军工等多领域应用

数据来源：信通院、亿欧智库

3.2.1 端侧大模型驱动下的AI PC与AI手机芯片升级

- 端侧大模型正推动芯片从通用计算向AI原生架构跃迁，AI PC和AI手机作为端侧AI的两大核心入口，其芯片升级不仅体现在NPU算力的提升，更关键的是系统架构的重构，以解决内存和带宽瓶颈。
- AI PC和AI手机芯片升级大方向相同，但约束条件和设计取舍不同。AI PC注重能做更多更复杂的事，AI手机追求极致的能效比，“够用且省电”。

亿欧智库：AI PC芯片升级



架构融合

异构总算力>单NPU，CPU负责调度与轻量推理，GPU承担矩阵运算，NPU专攻持续推理。

算力突破

NPU算力从个位数TOPS迈向数十TOPS级别，NPU≥40 TOPS，可推动OEM全面升级散热、内存、存储。

带宽升级

256-bit内存带宽 + 128GB统一内存，支撑千亿参数模型。

- 预计2026年中国Gen AI PC出货587万台，市场占比11.9%，销量同比提升166.4%。
- 落地能力：支持本地流畅运行1200亿参数大模型。

亿欧智库：AI手机芯片升级

双NPU架构

部分厂商推出超性能NPU（峰值）+超能效NPU（日常），动态调度降低功耗42%，峰值性能翻倍提升，NPU算力3年增长超20倍。

存算一体（CIM）

将计算单元嵌入存储阵列，可显著提升等效带宽并降低功耗。理想情况下，等效带宽提升10倍，同时算力功耗降至1/3。

低位宽固化

为提升效率，NPU指令集开始原生支持INT4/FP8等低位宽计算，减少数据搬运总量。通过INT4等量化技术可将模型体积压缩至原来的1/4。

内存革新

通过3D堆叠将逻辑晶粒与DRAM垂直集成，缩短数据路径，打破“内存墙”瓶颈，为端侧大模型释放极致算力。

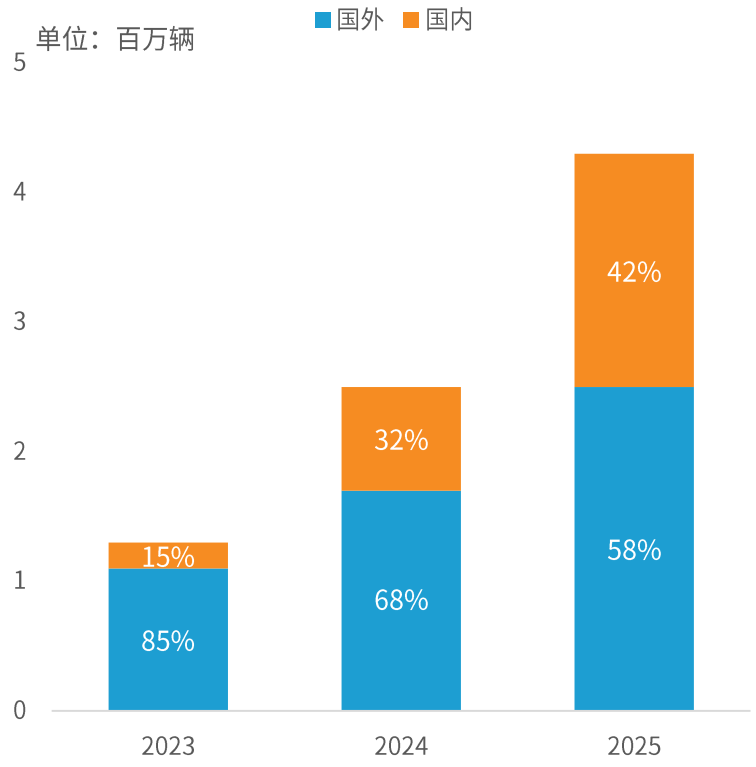
- 2026年国内生成式AI手机渗透率突破50%，正式进入普及周期。预计2026年中国AI手机出货量达1.47亿台，同比增长31.6%。
- 落地能力：端侧3B模型实现200+token/秒输出

数据来源：IDC、头豹、信通院、亿欧智库

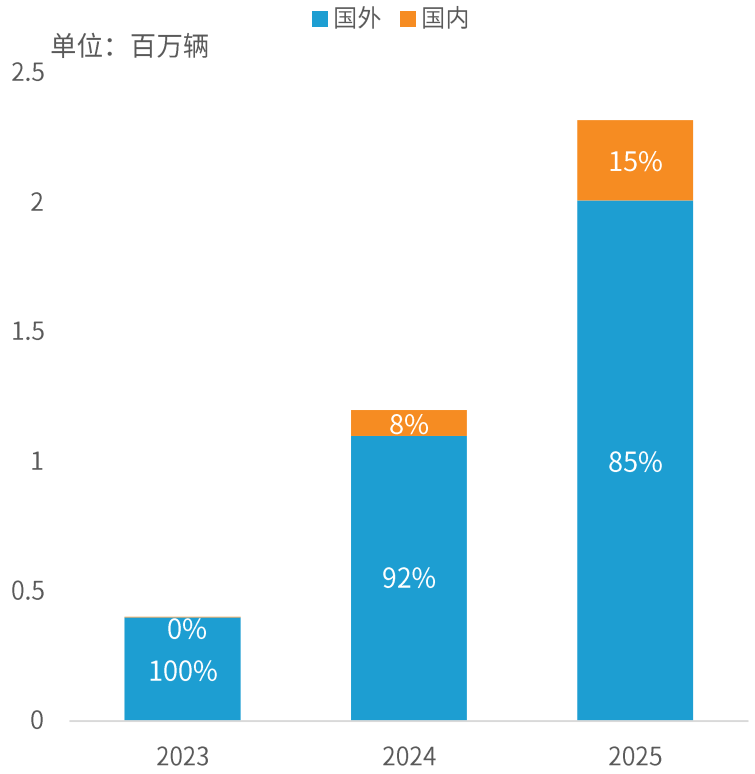
3.2.2 智能汽车：舱驾融合趋势下的高算力SoC竞争格局

- 2025年，中国市场（不含进出口）乘用车前装标配舱驾一体（含多芯、单芯以及单板、多板）计算单元交付156.36万辆，同比增长49.93%。预计到2030年，舱驾一体架构渗透率将突破30%。
- 2025年，国内>100TOPS算力区间，国产芯片市占率不断提升；>300TOPS算力区间市场，国产芯片渗透率低，仍以国外巨头主导，国产替代更具挑战和潜力。行业数据显示，从整个市场格局看，车企自研芯片的结构性优势在>300TOPS算力区间尤为突出，特斯拉、蔚来、小鹏三家自研芯片合计占比超35%。

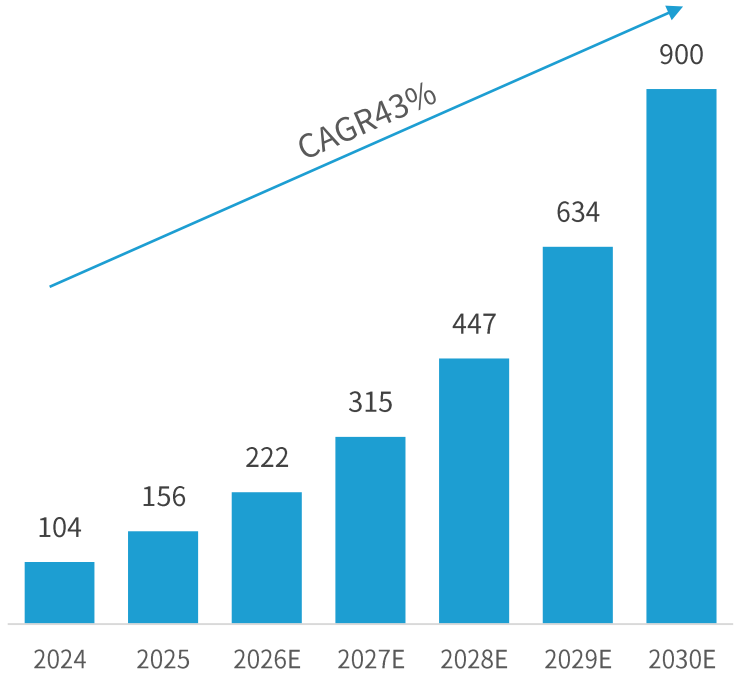
亿欧智库：2023-2025年中国乘用车高算力SOC（>100 TOPS）装车量变化



亿欧智库：2023-2025年中国乘用车高算力SOC（>300 TOPS）装车量变化



亿欧智库：中国市场（不含进出口）乘用车舱驾一体趋势（万辆）

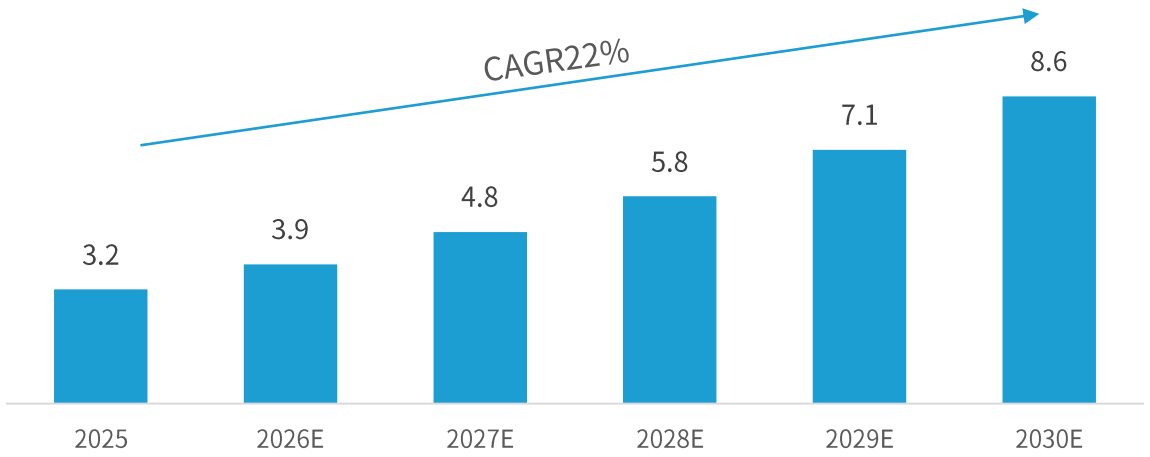


数据来源：GGII、亿欧智库

3.2.3 AIoT与具身智能：专用低功耗边缘芯片的市场机遇

- AIoT正在从“万物互联”迈向“万物智连”，应用场景不断拓展，从消费电子拓展到工业、汽车、农业等。不同场景的差异化需求、数据本地化处理、实时处理等需求，专用低功耗边缘芯片使得AIoT在特定场景下使用规模化落地的关键。
- 具身智能正加速产业化落地，其必须在动态、非结构化的真实环境中实时感知、决策和行动，对芯片的要求也十分明确：低功耗、低延迟、可靠性、多模态、高集成。在规模化制造中，需要保证芯片性能、功耗、成本的三角平衡。

亿欧智库：2025-2030年中国AIoT市场规模（万亿）



智能家居

要求芯片在极低待机功耗下，快速唤醒并完成本地语音识别。

工业制造/检测

要求芯片在高温、震动环境下，能实时处理振动传感器数据，进行边缘AI分析。

智慧农业

要求芯片在太阳能供电下，能长时间工作并处理图像识别病虫害的任务

亿欧智库：具身智能产业化发展市场机遇

01 市场需求增长加速，整机量产

2026年全球具身机器人芯片市场规模预计突破120亿美元，国内人形、四足机器人等出货规模居全球首位，具身智能正从实验室走向工业、商用、家庭等全场景落地，端侧芯片需求呈指数级增长。

02 大模型端侧部署

VLA（视觉语言动作模型）及轻量化世界模型在端侧落地，要求芯片具备低功耗、高算力及端侧轻量化推理能力

03 政策与国产替代红利

国内半导体自主可控政策加码，国产专用芯片正加速实现对海外通用芯片的规模化替代

具身智能实体需要同时满足多模态感知、模型推理、高频物理闭环等要求，对专用低功耗边缘芯片的需求是由物理本质、任务特性和数据环境共同决定的。
同时，具身智能的终极目标是大规模普及，这要求整体BOM成本降低，单颗AI芯片成本需要降低。

多模态感知

多路摄像头、足底压力、关节力矩、IMU等十几路传感器并发吞吐。

模型推理

“看-懂-动”端到端打通，动作生成有强时序依赖和频繁数据访问。

高频物理闭环

灵巧操作、动态平衡、避障需要毫秒甚至亚毫秒级闭环。

数据来源：IDC、中国电子报、工信部、亿欧智库

3.3.1 金融、医疗、工业制造等关键行业的国产芯片替代案例

◆ 中国信息安全测评中心与国家保密科技测评中心联合发布安全可靠测评结果公告，公布了国产CPU的最新安全评级。如今已经有包括海光处理器C86-4G等17款CPU获得最高安全等级II级认证，自主程度和国产化技术实力得到了充分验证，足以应对现阶段金融信创市场的复杂需求。目前已在金融、医疗、工业制造等领域进行国产芯片替代。

亿欧智库：金融、医疗、工业制造国产芯片替代案例

金融行业

- 工商银行：公示30亿元海光芯片服务器采购项目
- 平安银行：雷神科技独家全额中标平安银行2025年海光路线框架采购项目
- 建设银行：联想中标海光芯片服务器项目

工业制造行业

- 光伏： 隆基绿能自建芯片实验室：光伏逆变器用国产SiC芯片替代英飞凌，良率提升至92%，2025年目标降本30%。
- 新能源：比亚迪自研IGBT4.0芯片，电流输出能力较市场主流产品高15%，综合损耗降低约20%，特斯拉Model 3国产版已采购。宁德时代BMS电池管理芯片国产化进行中

医疗行业

- 联影医疗PET-CT用自研“北斗星”芯片，使PET-CT系统时间分辨率达到190皮秒级，成像速度提升50%（上海瑞金医院实测），维修响应从90天缩至24小时
- 政策强推：公立医院采购国内产品，核心部件国产化率≥90%；2025年前县级医院医学影像设备国产化率≥50%

数据来源：东方财富、中信国际、亿欧智库

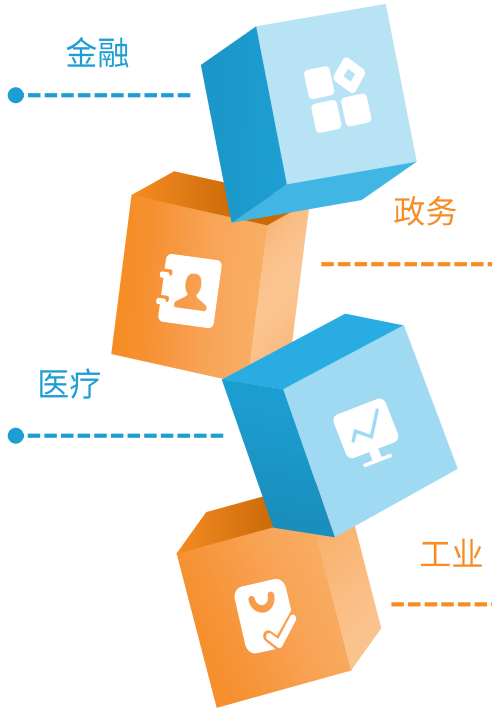
3.3.2 特定行业大模型对芯片架构的定制化需求

- 随着大模型进入产业深水区，芯片架构正在从通用到行业专用的范式转移，通过行业的业务逻辑，将行业特点转化为芯片架构的原生功能。
- 金融行业必须保证极致安全、极致性能、极致可靠、全栈自主可控。政务行业需要全栈自主可控、硬件级安全可信、适配多样化。医疗行业安全、精准与实时必须融为一体。工业行业需要保证在严苛环境下，提供确定、可靠、实时的计算能力。

亿欧智库：金融、政务、医疗、工业行业大模型需求

金融
- 极致性能：低延迟、高吞吐、算力效率极致化
- 极致安全：数据隐私与合规、硬件级安全隔离、模型资产保护、国产化与信创要求
- 极致可靠：系统达到99.999%的可用性（RTO<30秒，RPO=0），高可用设计、确定性延迟

政务
- 全栈自主可控：100%国产化
- 内生安全与国密加密：需集成TEE（可信执行环境），并结合硬件级国密引擎
- 数据主权与隐私合规：100%本地化部署，数据物理隔离不可出境。数据“可用不可见”，满足相应密评密改要求
- 分层适配：面对部级、区县、基层单位、个人等进行多模态部署



医疗
- 极致安全：数据本地化、芯片级TEE加密，满足医疗级合规要求
- 毫秒级延迟：要求芯片低延迟推理（典型要求<50ms）
- 多模态融合与专用库：专用AI加速核；专用算子库；异构SoC
- 私有化与边缘部署：高算力密度、低功耗、软硬一体
- 国产化合规：医疗设备芯片需通过NMPA Class II认证

工业
- 实时性：推理任务本地完成，需要支持边缘毫秒级响应
- 可靠性：适配工业环境，支持宽温工作采用无风扇、宽温设计
- 低功耗：解决多任务并发下的功耗与散热问题，支持边缘部署

数据来源：信创纵横、河北省卫健委、亿欧智库

亿欧智库

目录

CONTENTS

4.1.1 摩尔线程——全功能GPU芯片研发及AI计算解决方案提供商

- 摩尔线程是一家全功能GPU芯片研发及AI计算解决方案提供商，成立于2020年，以全功能GPU为核心，致力于向全球提供加速计算的基础设施和一站式解决方案，为各行各业的数智化转型提供强大的AI计算支持。
- 基于摩尔线程夸娥（KUAE）智算集群打造的三大“AI工厂”——“模型训练工厂”“词元生产工厂”“智能体生产工厂”，能够凭借显著的全场景通用优势，支撑从训练到推理、从数字世界到物理世界的多元算力需求。MTT C256超节点为万卡乃至十万卡级超大规模训推集群打造算力底座。

亿欧智库：摩尔线程MTT C256超节点

MTT C256超节点
核心技术优势

超高部署密度

两柜集成256张GPU，压缩机房占地面积，削减数据中心建设投入

极致互联性能

单层架构释放并行训练优化空间，长文本推理、MoE混合专家模型运行流畅

全生态兼容

2U标准节点适配市面主流软硬件，现有万卡集群可平滑扩容至十万卡，无需大规模改造基础设施

摩尔线程MTT C256超节点采用一层Scale-up网络，将256张GPU聚合为数据中心级计算系统，并采用计算与交换一体化的高密设计。单层Scale-up网络架构解决传统智算集群多层转发带来的带宽损耗、通信延迟痛点，为十万卡级大模型训练提供自主可控算力底座。公司称该架构可把卡间通信延迟压缩至亚微秒级，用于大模型分布式训练和推理。在工程化设计上，通过高效液冷散热、三盲插设计、完善的RAS（可靠性、可用性、可服务性）机制以及配套的集群管理系统，保障超节点在大规模部署下的长期稳定运行与便捷维护。

训练

作为十万卡集群核心组件，支撑大模型预训练、后训练与强化学习，保障长稳训练。

推理

原生支持打磨规模MoE架构与百万级超长上下文推理，兼具低延迟、高吞吐能力。

推理解码

支持高并发场景，单用户输出速度可超100 token/s，适用于AI Coding等对实时性要求极高的应用。

数据来源：摩尔线程、亿欧智库

亿欧智库：摩尔线程三大“AI工厂”

01 模型训练工厂
全功能、高性能、高拓展、高精度、高稳定

“模型训练工厂”为以摩尔线程全功能GPU为根基的夸娥智算集群。在高扩展方面，随着集群训练规模持续增大，夸娥智算集群可保持高达95%的线性扩展效率；在高精度方面，夸娥智算集群训练精度与国际主流计算卡持平，训练损失曲线高度一致；在高稳定方面，夸娥智算集群支持断点续训，有效训练时长占比超90%。

02 词元生产工厂
全适配、高性能、低延迟、高性价比

“词元生产工厂”为摩尔线程高性能推理套件。摩尔线程高性能推理套件深度适配SGLang、vLLM等主流推理框架，覆盖大语言模型、视觉、语音、DiT及视频生成等多模态推理场景。

03 智能体生产工厂
高效连接数字世界与物理世界

“智能体生产工厂”为摩尔线程自研全域数字智能体“小麦”。其掌握60余项技能（Skills），支持38款APP的跨应用控制，并能通过前端多模型、丰富知识库的灵活配置等，做到更快、更准、更懂你。

三大“AI工厂”能够凭借显著的全场景通用优势，支撑从训练到推理、从数字世界到物理世界的多元算力需求

4.1.1 摩尔线程——高能效AI基础设施建设者

- MUSA作为贯穿摩尔线程全功能GPU硬件与全栈软件体系的底层架构，已全面实现对业界主流CUDA生态的深度兼容。让国产GPU真正具备“即插即用”的开放能力。
- 夸娥（MTT KUAE）是摩尔线程智算中心全栈解决方案，是以全功能 GPU 为底座，软硬一体化、完整的系统级算力解决方案。夸娥集群管理平台除了包含 Kubernetes 集群的标准能力，还针对智算场景创新地提供了大量功能。

亿欧智库：摩尔线程MUSA生态建设

全面实现对业界主流CUDA生态的深度兼容

兼容接口数达761，核心数学库的100%对齐，100%支持PyTorch全部3194个算子，MUSA软件栈全链路覆盖了底层驱动、编译器、算子加速库、训练与推理框架

开源生态与关键场景取得突破

- 推理生态上，MUSA正式成为vLLM官方后端，成功合入SGLang官方主线并获得“原生支持”；
- 底层编译上，TileLang-MUSA成功合入开源主线，升级支持Trition 3.6最新版本，FlashAttention3等热点算子在MUSA上达到95%的极致效率；

正引入AI技术加速生态自我演进

- 依托Automusify智能迁移工具实现“零干预”自动化转化，实现加速仓库的100%自动迁移。
- MUSACODE AI编程助手通过大模型智能体协同，成功开发并交付超10000个Kernel算子，基于TileLang自动调优Group GEMM算子实现60%性能提升

云端训练	端侧智能	具身智能
官方数据显示其在训练Dense模型时MFU（模型浮点运算利用率）可达60%，MoE模型达到40%，有效训练时长超过90%	发布了集成智能体、AI PC和AI NAS功能的AICUBE计算设备，以及可同时运行多个智能体的AIBOOK笔记本电脑	“小飞”机器狗，其运动策略在仿真平台训练后，可零调参直接部署到搭载“长江”SoC的实体上。实现“物理+渲染+AI”的零拷贝数据通路。

数据来源：摩尔线程、亿欧智库

亿欧智库：夸娥核心能力

模型覆盖	推理加速	断点续训	集群就监控
200+ 主流大模型	MT Transform 千亿大模型推理加速	+15% 训练速度	万卡 最大集群规模

CUDA兼容	分布式	高性能通信	高性能存储
700+ 算子优化 Torch MUSA	DP|TP|PP|EP|CP Megatran-LM DeepSpeed/CoiossalAI	MCCL MTLink/PCLe RDMA over IB/RoCE	<2分钟 130B模型 Checkpoint写入

亿欧智库：夸娥集群管理平台功能

- 深度集成全功能 GPU 计算、网络和存储，可批量管理 GPU 驱动，降低适配和运维成本
- 通过企业空间、项目为不同组织及人员提供多维度的隔离方式
- 支持 GPU 共享，内置多 GPU 感知调度最佳实践，提升资源利用率并最大化业务性能
- 提供物理机、存储、网络、集群组件、工作负载的统一可观测平台，加快问题定位，降低解决成本
- 深度整合业务与设备数据，通过诊断管理及细粒度的监控告警，提前发现潜在问题

4.1.2 曦望Sunrise ——“All-in 推理”的GPU芯片及全栈解决方案提供商

- 曦望Sunrise是一家人工智能算力芯片企业，专注于高性能GPU及多模态场景推理芯片的研发与商业化。凭借八年技术沉淀及两代量产芯片的工程化验证，曦望已成为中国AI推理芯片领域的重要参与者。是中国首家确立“All-in 推理”战略、并跻身国内首批实现推理GPU规模化量产交付的芯片公司之一。公司提供AI推理GPU芯片及全栈推理解决方案，自芯片架构定义阶段即专为推理而设计，致力于持续降低大模型单位Token推理成本及TCO，推动AI推理算力的普惠化落地。
- 曦望Sunrise以启望系列推理GPU为核心，构建了智望系列计算卡、辰望系列服务器、寰望系列超节点、熙望系列工作站的多层次产品组合，满足云边端全场景高算力需求，不断实现极致推理性价比。其自研软件栈SIRE，兼容主流生态并支持代码无缝迁移，适配国内外主流大模型，构建面向Agent时代的国产推理算力基础设施，打造“硬件载体+软件优化+定制服务+长期运维”的完整服务体系。

亿欧智库：曦望GPU产品发展及核心优势

启望S1

场景驱动推理起点　多模态推理GPU　规模化量产交付

- 2021年量产，首款推理GPU
- IP授权索尼和小米
- 广泛应用于智慧城市、智慧交通等场景，适配主流CV模型

启望S2

主流生态无缝迁移　稳定量产持续加速　性能对标国际主流GPU

- 2024年量产，大模型推理GPU
- 主流生态兼容，适配DeepSeek、Qwen等主流大模型
- 文生图、文生视频、文生3D流畅运行
- 实测性能对标国际主流GPU，位居国内第一梯队

启望S3

极致推理性价比　精确度无损切换性能跃迁　LPDDR6显存突破

- 为Agent时代量身打造的原生推理GPU
- 国内首用LPDDR6的GPGPU芯片（向下兼容LPDDR5X）
- 国产GPU中首用PCIe Gen6的芯片
- 云智算、边缘服务器、端侧设备全场景适配

时代机遇
更懂AI的GPU芯片公司

- 10年以上真实AI业务场景打磨
- 押注推理，避开训练红海

顶尖团队
兼具“技术远见+商业模式”

- 高层具备多年行业经验
- 约300人精锐团队，汇聚业界龙头企业人才

产品内核
为“推理性价比”重写的芯片架构

- 为大模型推理重新设计芯片架构
- 统一的软件栈和集群调度系统

技术自研
全栈突破，专利护航

- 全栈自研指令集，100+核心专利
- 自研验证平台和性能模拟器支撑，实现“一次流片成功”

量产成功
多代芯片成功量产，性能领先

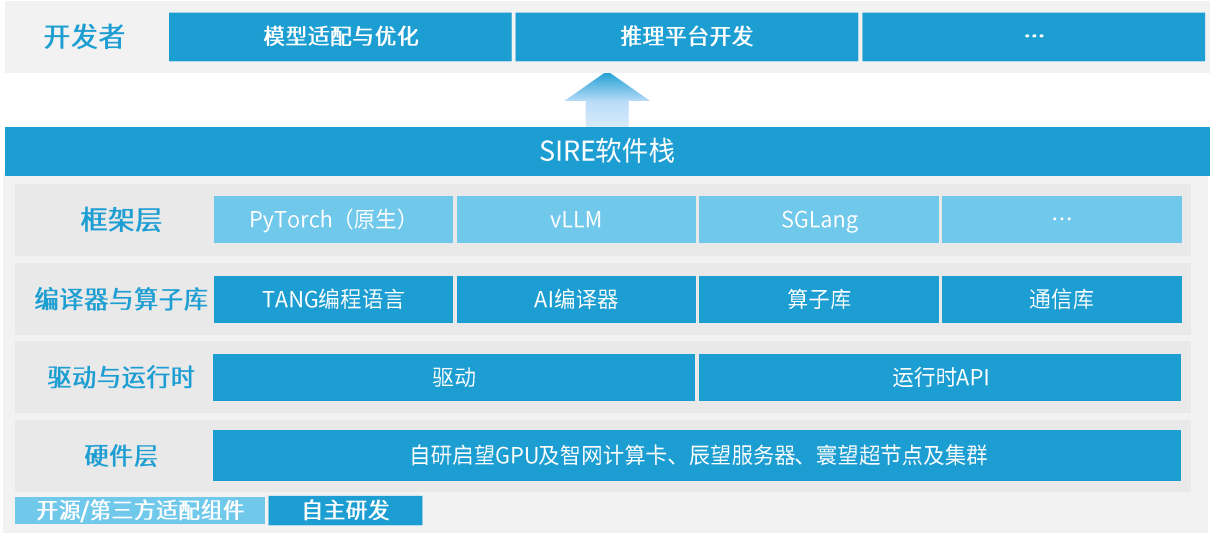
- S1(2021)：多模态推理GPU
- S2(2024)：大模型推理GPU，跻身国内推理GPU第一梯队
- S3(2026)：大模型与智能体推理GPU

商业加速
业务爆发性增长

- 2025年实现芯片量产交付
- 2026年S3发布，降低大模型推理成本一个数量级，释放AGI商业化能力

亿欧智库：曦望多层次产品组合

亿欧智库：曦望SIRE软件栈



数据来源：曦望实验室、亿欧智库

4.1.2 曦望Sunrise ——“芯片+解决方案+生态”三位一体

- 曦望Sunrise作为推理GPU独角兽，现已规模化量产并具备从芯片研发、产品量产到解决方案交付的全链路能力，并得到客户的广泛信任和认可。
- 曦望Sunrise 围绕自研核心技术资产，构建“芯片+解决方案+生态”三位一体的完整业务布局，协同赋能产业链上下游，构建了覆盖AI平台企业、算力厂商、行业龙头及科研机构的完整生态体系。将推理算力硬件产品与软硬一体解决方案，嵌入制造、能源、C 端、机器人等千行百业的真实场景，形成“生态合作-场景落地-服务升级”的正向循环，进一步提升软硬一体解决方案的规模化交付能力。

亿欧智库：曦望Sunrise 解决方案

全栈AI国产化算力平台 | 显著降低AI应用总体拥有成本（TCO）

从芯片到模型全栈AI优化

- 零代码改造：统一API屏蔽底层差异，实现主流模型零代码改造
- 开箱即用：大语言模型、文生图/视频等热门模型开箱即用，高效部署

灵活交付，支撑规模化落地

- 安全合规：满足敏感行业数据安全合规，私有化部署
- 按需提供弹性服务：具备提供MaaS服务能力，为中小企业提供弹性、按需的AI推理算力

本解决方案从底层算力应用落地进行全栈优化，能够显著降低AI应用总体拥有成本（TCO），助力企业AI解决方案规模化应用。解决企业在AI应用落地中，算力成本高、模型适配难、规模化部署门槛高的痛点。

- 高性能推理算力：专为AI推理定制，提供业界领先的性能与能效比，从源头降低算力成本
- 全栈AI落地方案：统一API屏蔽底层差异，实现零代码改造、主流模型开箱即用，大幅缩短落地周期
- 显著降低TCO：从芯片到方案全栈优化，大幅压缩AI应用总体拥有成本，助力规模化应用

空天地一体化全栈算力生态 | 加速攻克太空算力不足、天地协同低效等瓶颈[1]

全链路 算力覆盖

创新“太空算力节点-云侧智算中心-边侧响应节点-端侧感知设备”协同架构，破解星算计划太空算力不足、天地协同低效、边缘响应滞后等核心痛点

毫秒级 实时响应

空-云-边毫秒级端到端延迟，支撑暴雨预警、交通调度等应急场景，目标延长灾害预警窗口期24小时[2]

多场景 深度赋能

目标推动AI气象预报误差降低25%[2]、促进交通通行效率提升，防汛灾害损失大幅减少，激活精准农业、立体交通、空天经济、低空经济等万亿级赛道

本解决方案打造“云-边-端-太空”四侧协同的全栈算力解决方案，赋能全球空天地。解决企业旨在构建空天地一体化AI算力网络过程中，面临太空算力不足、天地协同低效、高性能与高性价比推理算力需求提升等痛点。

- 国产替代+成本优化：端到端软硬协同，优化落地成本，优化产品迭代与部署效率，保障项目快速落地
- 保障算力性能与低延迟：自研GPGPU架构针对Transformer类模型优化，支持KV Cache硬件加速单元、支持FP16和INT8精度，能高效支撑星载大模型协同计算、AI推理等高频场景的算力和低延迟需求
- 生态和落地能力强：软件栈兼容主流推理引擎；具备MaaS平台解决方案能力

亿欧智库：曦望Sunrise 部分场景应用举例

科学计算场景

- **现状**：复杂PointNet++ 架构风洞大模型，长期依赖国外旗舰算力。
- **能力**：全链条迁移与内核融合优化，实现单图推理≤1.5秒，比肩国际主流旗舰GPU。
- **价值**：首次国产芯片实现垂直领域大模型生产级商用。

大规模视频监测场景

- **现状**：视频监测存在数据体量大、无效帧冗余多、实时解析吞吐不足等问题，传统设备解码能力有限，难以支撑高帧率视频实时结构化分析。
- **能力**：百路级高清视频并发分析，跨区域视频汇聚，实时结构化输出，保证7×24不间断处理。
- **价值**：以一台辰望系列一体机替代传统多设备方案，大幅降低单路视频处理成本，满足大规模部署需求。

AI for Science场景

- **现状**：国产半导体先进工艺良率不足，缺乏高效提升工具，面临“卡脖子”技术瓶颈。
- **能力**：“硬件基座 + 软件栈 + 场景应用”三位一体的智能计算联合创新方案，构建光互连GPU超节点架构，支撑千万级像素并行计算与万亿参数模型推演。
- **价值**：开发计算光刻专用软件栈与AI气象计算库，解决“国产 GPU 硬件强、软件弱”痛点，形成“技术研发-场景验证-产业转化”良性循环，巩固营收基座，打造国产先进计算标杆生态。

OPC 场景

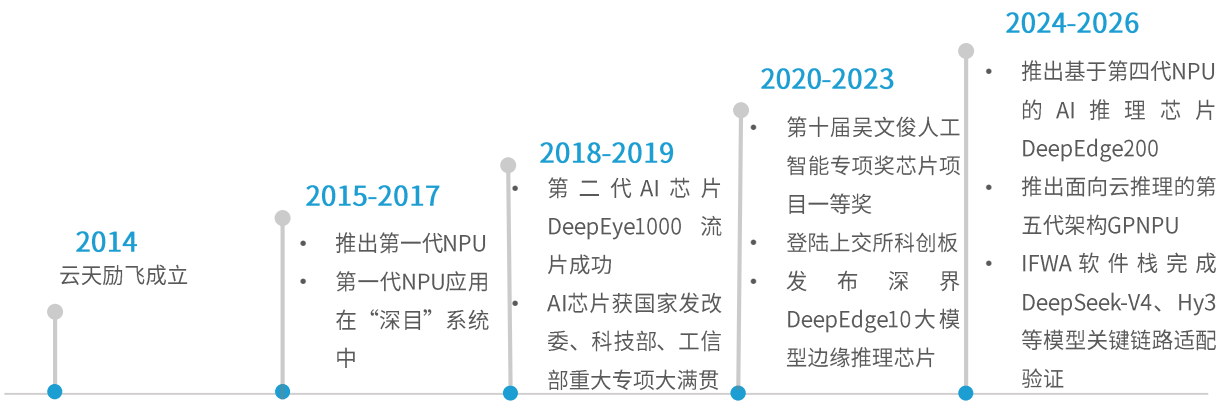
- **现状**：生成式AI与智能体工具推动个体生产力指数级提升，筑牢“一人成军”的技术底座。
- **能力**：打造一站式OPC服务平台，以"算力+消纳"双轮驱动，构建AI时代的数字基础设施。
- **价值**：激活OPC社区新动能，以“四大支柱”为基座、端到端链路为引擎，构建OPC从入门到退出的完整闭环，协同本地化生态，实现价值共享。

数据来源：曦望实验室、亿欧智库，注释1：空天地一体化解决方案仍处于研发验证阶段，涉及数据仅对测试时点有效；注释2：数据来源于曦望实验室

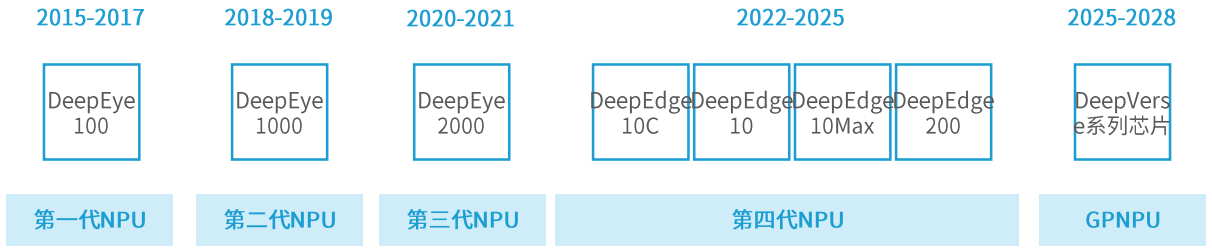
4.1.3 云天励飞——AI推理芯片产品服务提供商

- 云天励飞（688343.SH）创立于2014年，是AI推理芯片领域的领军企业。公司依托自主创新的GPNPU架构，以“百亿 Token 1 分钱”为目标，打造高性能、高性价比的AI推理算力底座，推动人工智能实现规模化普惠应用。
- 围绕“云、边、端”AI推理场景，云天励飞打造了覆盖多元应用的芯片产品体系，包括面向云端大算力推理的AI芯片“深穹”（DeepVerse）、面向边缘高能效推理的AI芯片“深界”（DeepEdge），以及面向具身智能机器人的AI芯片“深擎”（DeepXBot）。基于自研AI推理芯片，公司已全面适配国产主流大模型、国产主流操作系统和国产推理框架，广泛赋能智算中心、智慧城市、具身智能等重点领域。

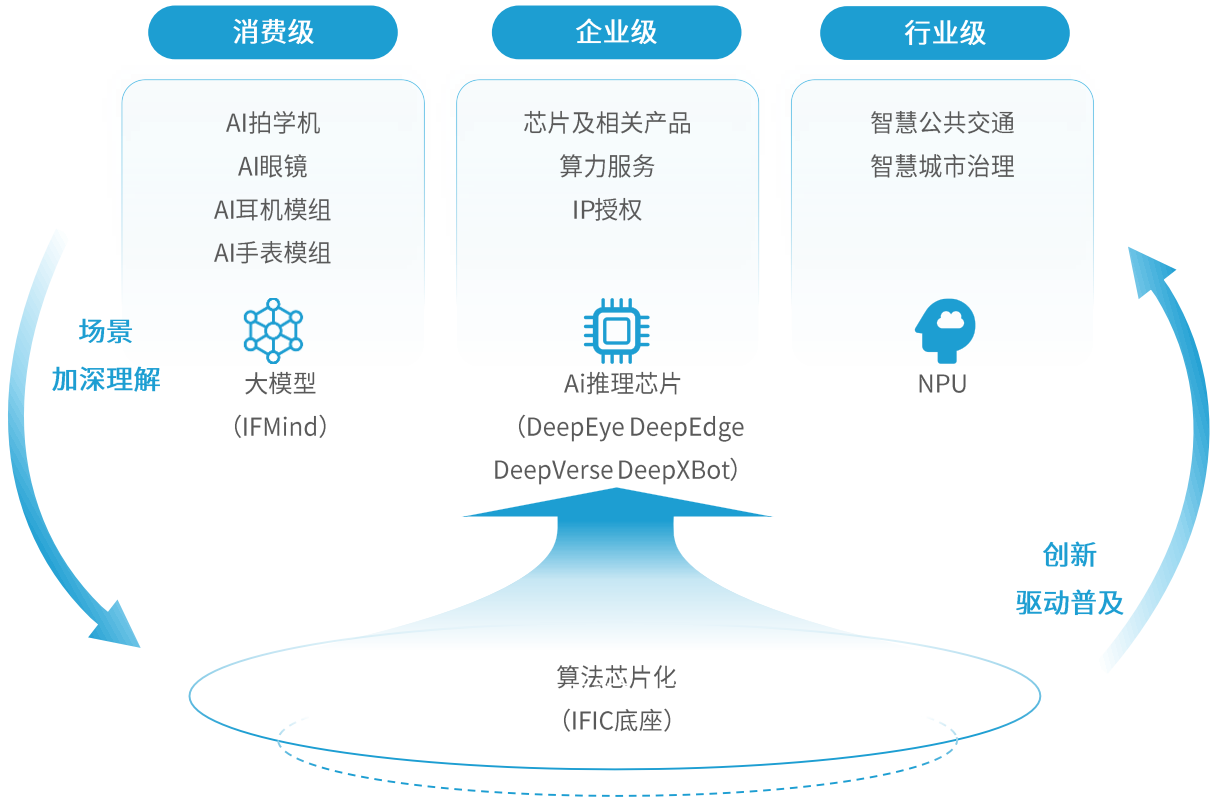
亿欧智库：云天励飞及其芯片发展历程



云天励飞结合算法架构演进趋势，打造相关指令集、专用算子、存算一体架构、低比特混合量化等芯片硬件技术，10年5代NPU。



亿欧智库：云天励飞“算法芯片化”多场景产品体系

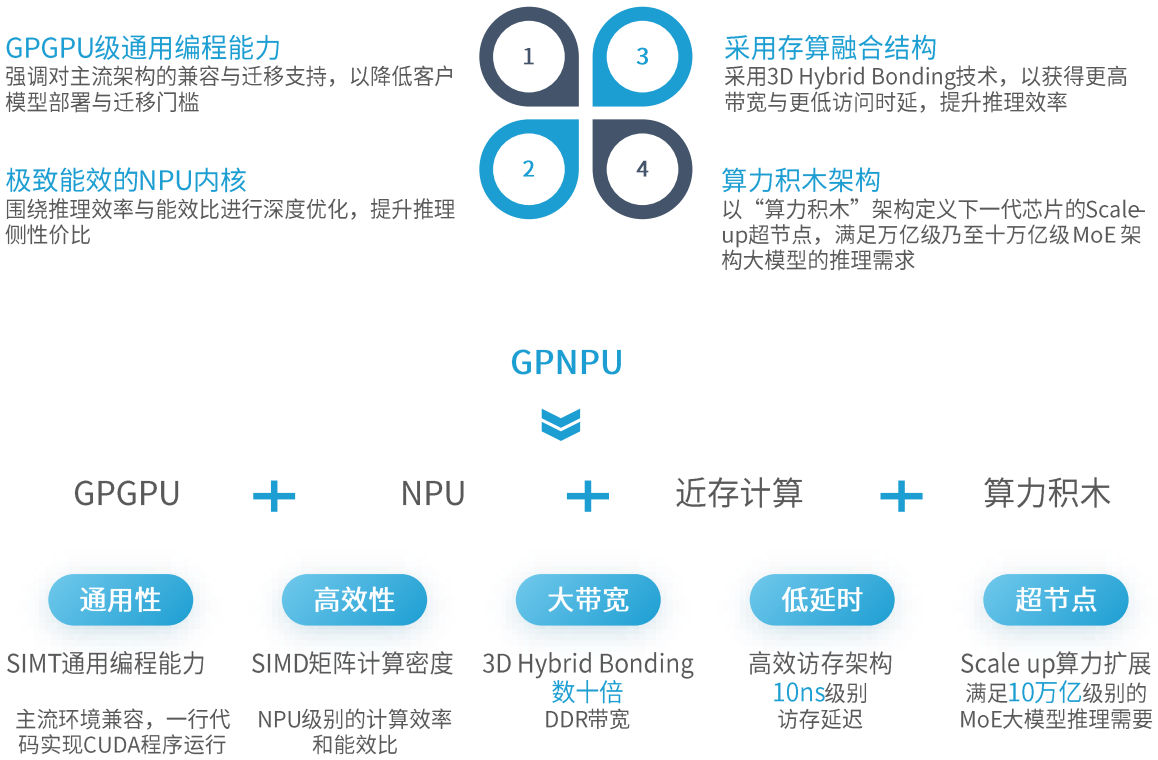


数据来源：云天励飞、亿欧智库

4.1.3 云天励飞——战略方向“训练追赶、推理超车”，致力于打造极致性价比的Token

- 云天励飞将以GPNPU架构为核心，大力推进云端大算力芯片，强化软硬协同与存储体系攻坚，致力于以极致性价比释放Token价值，推动大模型从示范应用走向规模化落地。
- 云天励飞以“百亿 Tokens 一分钱”为目标，持续推动 Token 成本下降。围绕这一目标，云天励飞已规划未来三年的芯片产品迭代路线，包括第一代超节点P芯片、第一代超节点D芯片、第二代超节点L芯片。

亿欧智库：云天励飞GPNPU技术亮点与架构



亿欧智库：云天励飞未来三年大算力芯片规划

针对大模型时代云推理场景的需求，云天励飞按照“PD 分离”的系统架构规划三类大算力芯片：

P芯片（Prefill）：面向计算密集型需求设计，满足Prefill 阶段的高算力要求；

D芯片（Decode）：面向访存密集型需求设计，满足Decode 阶段的高带宽需求。

L芯片（Decode FFN）：面向Decode阶段最重的FFN计算需求设计，大幅提升FFN计算单元利用率

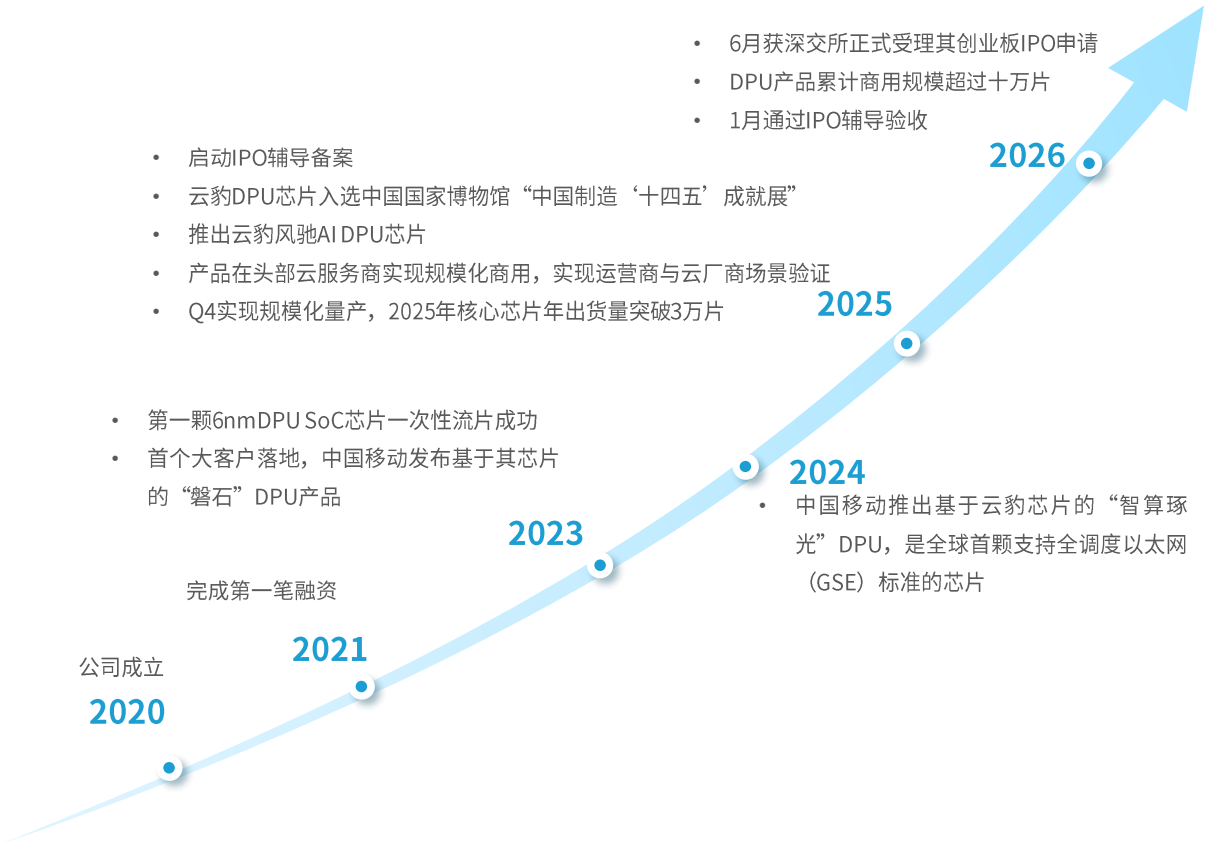
- 打造第一代超节点P芯片，面向百万级长上下文场景进行Prefill推理优化；
- 研发第一代超节点D芯片，聚焦 Decode 推理的低时延目标；
- 推出第二代超节点D芯片，面向毫秒级推理时延目标进一步优化，带动Prefill与Decode性能提升。

数据来源：云天励飞、亿欧智库

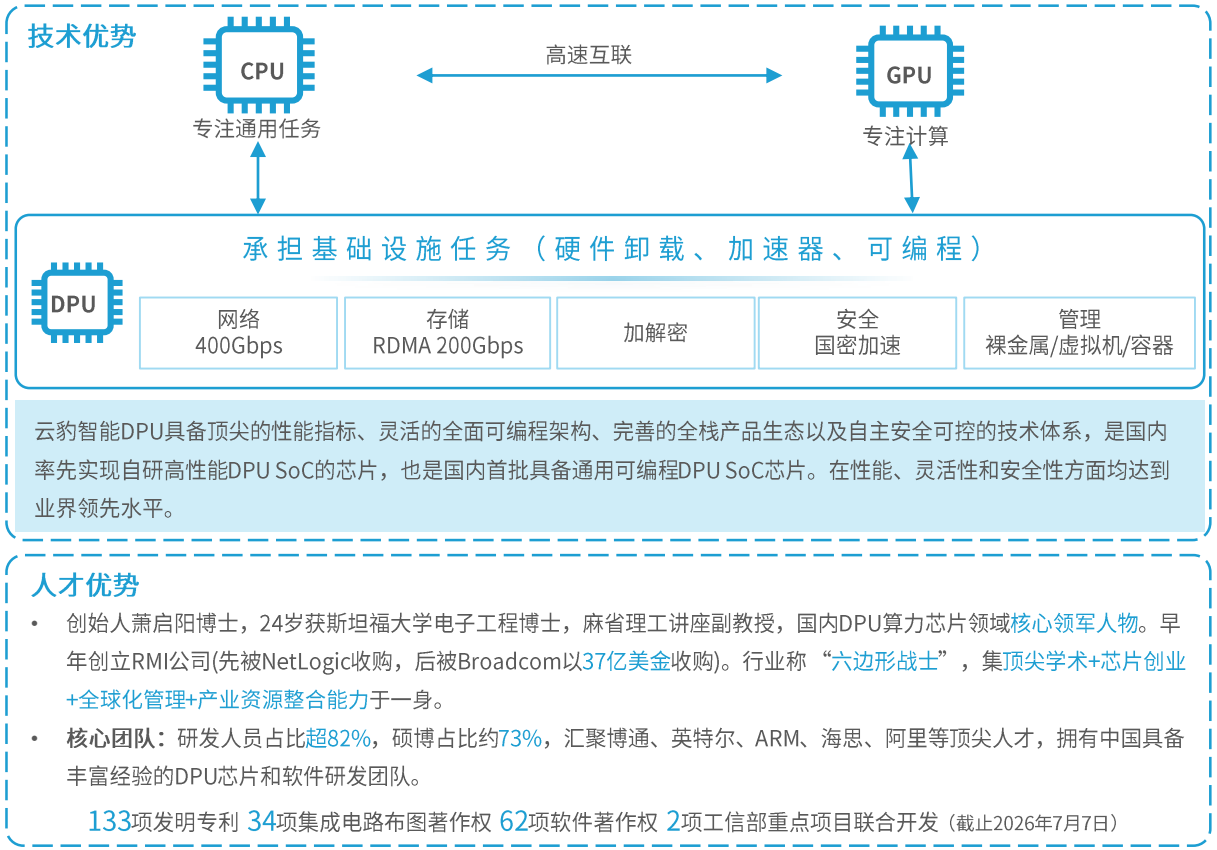
4.1.4 云豹智能——国产DPU领跑者，算力网络的核心基石

- 云豹智能成立于2020年，专注于云计算和数据中心数据处理器芯片（DPU）和解决方案，是国内先进工艺基础大芯片 DPU（数据处理器芯片）的行业龙头及独角兽企业。2026年有望成为“国产DPU第一股”。
- 云豹DPU是国内历史上首颗达到400Gbps吞吐量且已量产的国产高性能、全功能DPU芯片，主要应用在AI网络及通算云计算市场。其产品也是国内首批具备通用可编程的DPU SoC芯片。在性能、灵活性和安全性方面均达到业界领先水平。历经国家工信部严苛评审，成功入选中国制造 “十四五” 成就展。

亿欧智库：云豹智能发展历程



亿欧智库：云豹智能核心能力优势



数据来源：云豹智能、亿欧智库

4.1.4 云豹智能——国际顶尖的国产高性能DPU芯片企业

- 云豹智能已完成两款DPU芯片产品，达到国际顶尖水平，为客户提供芯片、网卡及系统解决方案。系列产品均已成功入选中国制造“十四五”成就展，亮相国家博物馆并跻身“国之重器”阵列，是入选该成就展的三大芯片企业之一。云豹DPU已在国内头部云服务商、头部电信运营商、知名AI大模型厂商、金融及电力等行业头部客户开展合作，并实现大规模量产。
- 云豹云霄DPU芯片是一款高性能通用可编程处理器（DPU），采用分层可编程架构集成P4与RISC-V核心，通过硬件加速实现400Gbps网络带宽与200Gbps RDMA性能，实现数据中心性能、能效与安全性的三重跃升。是国内首款400Gbps全功能DPU芯片，目前已在头部客户场景中实现超过十万片规模化商用，应用于高性能计算、存储与网络卸载等业务。
- 云豹风驰AI DPU芯片是专为万卡级AI智算场景打造的AI DPU芯片，可满足万亿参数大模型高速互联需求，是国内智算中心建设的关键底层硬件。

亿欧智库：云豹云霄DPU

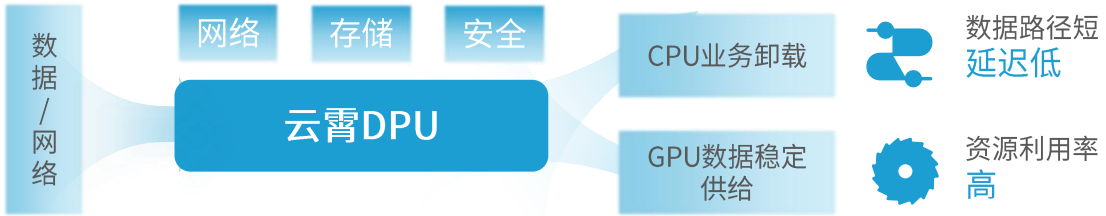
- 国内首颗达到400Gbps的高性能全功能DPU芯片
- 性能效率较传统方案提升2-4倍
- 功耗降低50%以上
- 成本大幅下降，极具性价比

该产品已实现规模化商用，并入选中国国家博物馆“中国制造‘十四五’成就展”，跻身“国之重器”阵列。

实际应用：

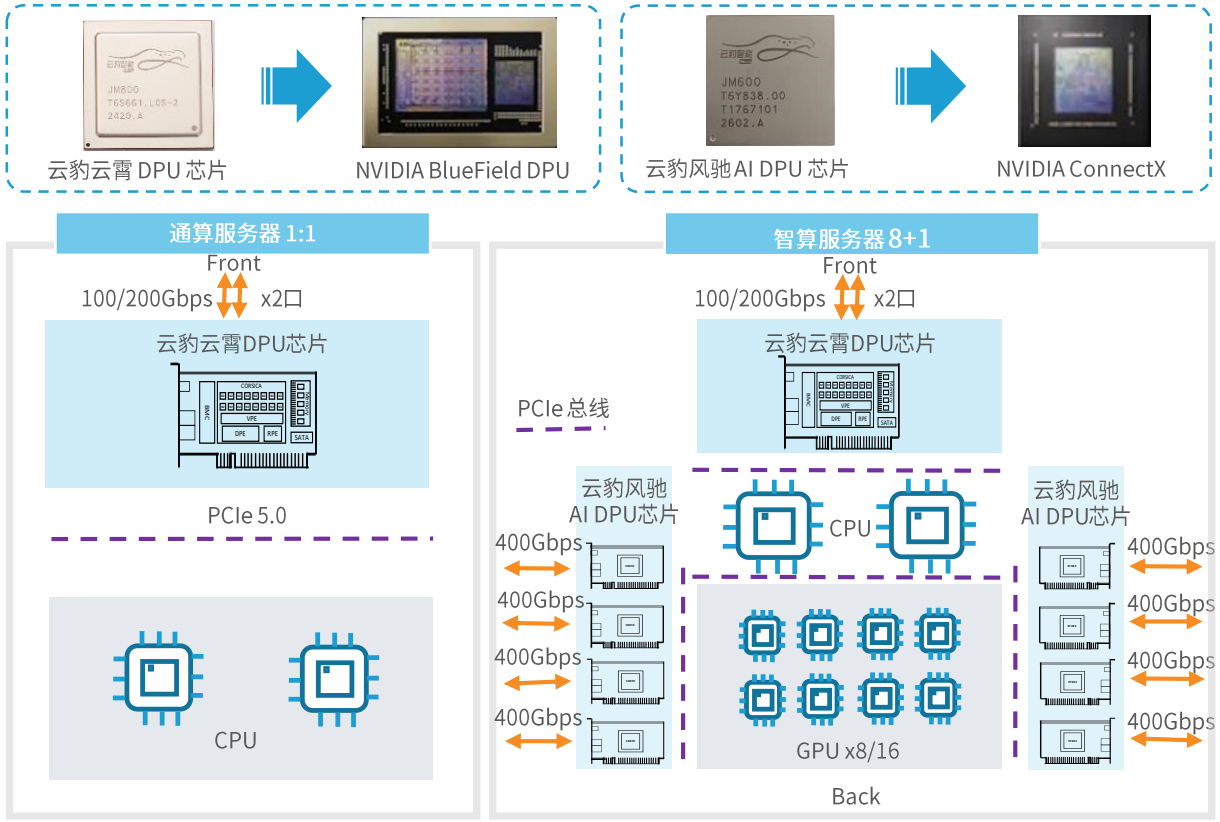
✓ 中国移动基于云豹DPU发布大云“磐石”DPU、AI网络“智算琢光”DPU和5G网络“灵耀”DPU三款产品，并于2025年实现量产。

✓ 头部云服务商于2025年春节期间大批量上线云豹DPU，实现了春保期间零故障的稳定运行。



数据来源：云豹智能、亿欧智库

亿欧智库：云豹智能对标英伟达BF/CX网卡解决方案



4.2 产业图谱

上游

EDA/IP

Empyrean 华大九天 | 概伦电子 PRIMARIUS | Semitronix | SYNOPSYS 新思 | VeriSilicon | 澜起 MONTAGE TECHNOLOGY | cadence | SIEMENS

材料

ShinEtsu | JSR 株式会社 | ANJI | 华特气体 HUATE GAS | 江丰电子 KFMI | NSIG | Lion | Jingrui 晶瑞电材 | 金宏气体 JINHONG GAS | TCL 中环

设备

ASML | ACM RESEARCH | AMEC 中微公司 | HWATSING 华海清科 | 新凯来 SICARRIER | APPLIED MATERIALS | NAURA 北方华创 | 拓荆科技 | KINGSEMI 芯源 | SMEE

中游

芯片设计

NVIDIA | Cambricon 寒武纪 | sunrise 曦望 | intellifusion 云天励飞 | 地平线 Horizon Robotics | LOONGSON 龙芯 | 云豹智能 | Witmem 知存科技 | LYNXI | 昆仑芯 KUNLUNXIN | EVAS INTELLIGENCE 奕行智能

intel | 壁仞科技 BIREN TECHNOLOGY | 景嘉微 JINGJIA MICRO | Rockchip 瑞芯微电子 | Ascend | 兆芯 | Phytium 飞腾 芯科技 共飞腾 | 中科驭数 YUSUR | REXEN technology | SynSense | 平头哥 | 黑芝麻智能 BLACK SESAME TECHNOLOGIES

AMD | METAX 沐曦 | 摩尔线程 MOORE THREADS | 紫光展锐 UNISOC | Ingenic 君正 | 中威 | HiSilicon | NEBULAMATRIX 星云智联 | 后摩智能 HOUMO.AI | 清微智能 | 国科微 | Axera 爱芯元智

HYGON | 天数智芯 Iluvatar CoreX | INNOSILICON 芯动科技 | Enflame 燧原科技 | ALLWINNER 全志科技 | C*Core C*Core Technology Co.,Ltd. | Kunpeng | 云脉芯联 YUNSILICON | 亿铸科技 | PolarBeatTech 北极雄芯 | Google | 元川微 ARCHEFLOWX

制造

Micron | SK hynix | tsmc | 华虹集团 HUAHONG GROUP | SMIC 中芯国际集成电路制造有限公司 | Nexchip

封装测试

日月光 | ANKURA 安靠 | TF 通富微电 | JCET | 长电科技 | SJSEMI | 华天科技

下游

智算 / IDC | 消费电子 | 智能驾驶 | 工业 AIoT

备注：图谱举例仅展示部分企业，未穷尽

亿欧智库

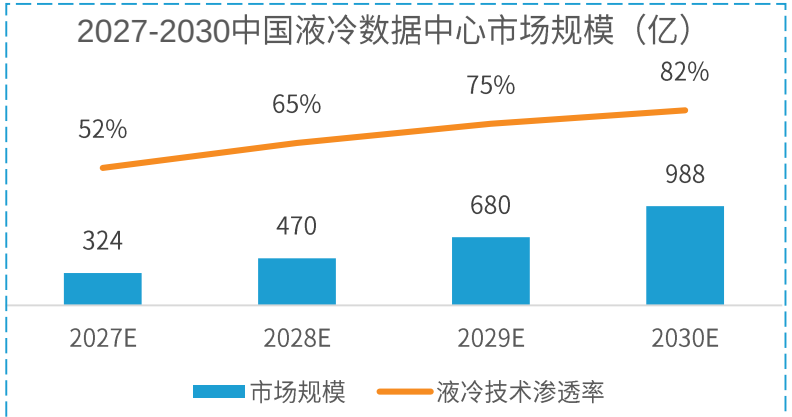
目录

CONTENTS

5.1 2027-2030年技术趋势研判

- 能效比取代峰值算力成为芯片选型的首要指标，AI芯片正式进入“绿色算力”时代，在这一战略目标牵引下，液冷散热将成为基础架构，渗透率不断提升。
- 行业将CPO（共封装光学）视为终极方向，2026年英伟达Spectrum-X以太网硅光技术已全面量产，标志着第一阶段Scale Out（机架间互联）落地，27-28年CPO有望向Scale Up（机架内互联，光入柜内）场景延伸，2028-2030年实现大规模部署。
- 量子计算与AI的融合（简称“量智融合”）正成为算力产业演进的核心方向之一，当前整体正从理论走向实践，属于早期探索阶段。

亿欧智库：能效比成核心，液冷散热普及



- 物理限制：目前主流的单机柜功耗已达130-140kw，风冷系统的物理散热极限是15-20kW。
- 可靠性：数据显示，约55%的电子元器件故障源于温度过高，温度每升高2°C，电子元器件可靠性下降10%。
- 政策推进：要求新建大型数据中心PUE不高于1.25，国家枢纽节点不高于1.2，风冷PUE通常在1.4以上
- 能效：液冷导热效率是空气25倍以上，PUE可降至1.2以下，制冷能耗从43%降至24%.

亿欧智库：硅光与光电融合技术的渗透

Scale Out（机架间互联）场景 — 第一阶段 2026-2027

2026年英伟达Spectrum-X以太网硅光技术已全面量产，第一阶段Scale Out（机架间互联）正式落地，2027年Scale Out系统出货量预计将较26年翻5倍。

Scale Up（机架内互联，光入柜内）场景 — 第二阶段 2027-2028

向Scale Up（机架内互联，光入柜内）场景延伸，是CPO发挥最大价值的场景。其市场空间比Scale Out大出数个数量级。

大规模部署，渗透率跃升 — 第三阶段 2028-2030

2028-2030年实现大规模部署，CPO在AI数据中心光通信模块中的渗透率将从2026年的0.5%攀升至2030年的约35%。

亿欧智库：量子计算与AI融合的早期探索

AI加速量子计算

- 谷歌Willow等新型量子芯片将逻辑门错误率降至1%的临界阈值以下，为稳定、可靠的量子计算奠定了关键基石
- 国产中性原子量子计算机“清河一号”内置自研量子经典混合操作系统，依靠AI算法自动修正量子运算过程中产生的误差，解决了长期制约中性原子设备商用的纠错难题

量子计算增强AI

- 未来的机器学习不再运行在传统图灵机架构上，而是以量子机器作为基础算力架构，直接去处理、学习量子数据时，将获得超越现有图灵机极限的全新认知工具。

量智融合典型案例

- 在石油勘探领域，中国移动联合中石油勘探开发研究院提出量子-经典混合神经网络架构，相比传统AI模型预测误差降低20%。
- 光伏功率预测量子算法产品，利用量子-经典混合算法优化预测模型。在安庆虚拟电厂项目中部署应用，覆盖23个光伏场站。

数据来源：IDC、赛迪、信通院、英伟达、《人工智能的威力及其局限》、专家访谈、亿欧智库

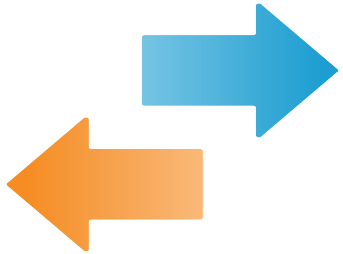
5.2 市场格局演变与企业竞争策略

- 英伟达芯片供给受管制，政策、市场需求、技术发展等对国内的芯片市场带来了极大的机遇，头部芯片厂商的竞争已从展示单个芯片，转向比拼算力底座、整机集群、软件全栈与行业生态的一体化交付能力，初创公司选择差异化布局，首先是架构差异化，其次是场景聚焦细分。
- 国内互联网厂商与芯片厂商的关系已进入深度绑定阶段，既是市场竞争的必然，也是产业走向成熟和追求算力自主的标志。
- 国内AI芯片企业潜在的并购机会主要体现在产业资本并购补齐自身产业链短板、行业企业扩大市场份额、消除竞争并购优质企业、国资与产业协同并购。

亿欧智库：头部企业与初创公司的差异化生存策略

头部企业

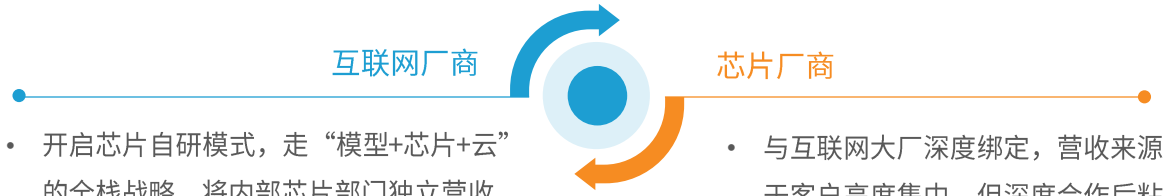
- 2026年英伟达在国内的市场份额预计下降至个位数，为国产AI芯片的市场份额带来极大机遇。
- 头部企业的竞争焦点全面转向软硬件协同、集群建设、运维服务一体化交付能力。客户采购的算力目标也倾向于整套智算解决方案。



初创公司

- 初创公司普遍选择差异化路线突围或者选择从实际需求出发。
- 各类国产自研架构差异化发展，可适配超大模型训练、高密度推理、边缘推理规划、海量数据吞吐等差异化场景。
- 随着AI产业从训练全面转向推理，众多初创公司专注于能效更高、成本更低的专用推理芯片或者端侧芯片上发力
- 与互联网、通信等产业链下游或生态伙伴的深度合作

亿欧智库：互联网厂商与芯片厂商的深度绑定与生态联盟



- 开启芯片自研模式，走“模型+芯片+云”的全栈战略。将内部芯片部门独立营收
- 通过战略投资与芯片公司结成深度利益共同体
- 采取生态互补模式，开放采购与合作生态链企业，如腾讯采取生态结合的战略，与更多芯片厂商展开合作

- 与互联网大厂深度绑定，营收来源于客户高度集中，但深度合作后粘性较强
- 通过深度合作，获得商业验证和市场准入

亿欧智库：国内AI芯片潜在的并购机会

- **纵向整合：**产业资本（如互联网大厂、汽车/消费电子/通信等产业链巨头）实现技术整合与产能控制，补强自身产业链短板，对具备特定核心技术的AI芯片初创企业存在并购需求；
- **横向整合：**行业整合并购，头部芯片企业对细分赛道的优质企业或初创企业存在并购需求；
- 国资与产业链系统：地方国资为扶持本地重点产业，对具备核心技术且具备落地场景的AI芯片企业，通过并购或产业基金协同进行整合。

数据来源：腾讯、摩根士丹利、专家访谈、亿欧智库

5.3 风险提示与战略建议

- AI芯片发展目前及未来仍将面临地缘政治、合规与贸易壁垒、市场与周期、技术与供应链、商业化风险等。
- AI芯片产业正从“技术驱动”向“价值驱动”转型，呈现多层次需求格局。科技企业、投资机构、政府/产业园区需要协同，其中科技企业以弹性配置对冲供应链风险，芯片企业以差异化切口避开巨头垄断，投资机构以产业链再平衡思维捕捉价值洼地，政府错位布局防内卷。

亿欧智库：AI芯片行业发展风险提示

地缘政治

- 美国等西方国家对华AI芯片出口管制持续升级，限制高端芯片、先进制程设备及EDA工具出口，导致国内企业获取先进算力受限，供应链面临“卡脖子”风险，推高研发与运营成本。

合规与贸易壁垒

- 国际合规审查趋严，第三方转口渠道受限，增加了企业跨境业务的不确定性及合规成本

商业化风险

- AI芯片的终端应用（如AI手机、自动驾驶、AI PC）若未能形成大规模放量，或终端消费电子需求疲软，将导致芯片企业业绩兑现不及预期，高投入难以转化为实际回报

市场与周期风险

- 需求波动与资本开支回调风险，若云厂商因AI投资回报不及预期而缩减算力采购，将直接导致上游芯片订单锐减
- 产能过剩与价格战风险，行业高利润吸引大量资本扩产，若AI大模型落地进度不及预期或推理需求增速放缓，可能导致芯片产能集中释放，价格下降

技术与供应链风险

- 端侧AI芯片面临功耗、内存墙和散热等物理限制，若技术突破不及预期，将制约终端产品的普及和算力释放。
- AI芯片研发周期长、流片成本极高，中小芯片设计企业面临研发不及预期、商业化落地难及资金链断裂风险。
- 供应链“卡脖子”风险依然存在，且产能瓶颈可能制约行业扩产节奏

亿欧智库：AI芯片行业发展战略建议

科技企业

- 非芯片企业：采取“多云+边缘+租赁”的弹性算力策略，分散地缘政治风险，提前锁定关键零组件产能，降低成本与供应压力；提前布局国产芯片，并培养算法与芯片优化的跨界人才
- 芯片初创企业：以“推理/边缘/特定算法加速”为差异化切口，绑定1-2家头部客户做深度场景优化但严控单一客户收入占比；将东南亚、中东等新兴市场作为“第二战场”提前布局

投资机构

- 区分全局趋势与局部泡沫，规避纯题材和高拥挤方向
- 建议关注国产AI算力、存储、半导体设备等，可适度分散布局
- 提前布局有望在产业链再平衡中受益的中下游企业

政府/产业园区

- 科学规划各地算力布局，逐步降低全社会算力使用成本，推动地方从“补算力”向“拓场景”转变，促进AI产业可持续发展
- 优化算力与数据资源配置，夯实产业底座
- 根据自身产业基础和资源禀赋进行差异化定位，避免同质化竞争，如封装测试和模组制造
- 对于本地的芯片企业加以扶持和投资，建立入园企业信用评级

数据来源：中关村产业研究院、专家访谈、亿欧智库

致谢

- 亿欧智库经过桌面研究，结合相关公开报道及对相关企业、专家访谈后作出此份报告。在此，亿欧智库感谢相关企业及业内专家的鼎力支持。
- 未来，亿欧智库将持续密切关注人工智能领域，通过对于行业的深度观察，持续输出更多有价值的研究成果，助力产业可持续创新发展。欢迎报道读者与我们交流联系，提出报告建议。
- 特别鸣谢

关于亿欧

◆ 团队介绍：

亿欧智库（EOIntelligence）是亿欧旗下的研究与咨询机构。为全球企业和政府决策者提供行业研究、投资分析和创新咨询服务。亿欧智库对前沿领域保持着敏锐的洞察，具有独创的方法论和模型，服务能力和质量获得客户的广泛认可。

亿欧智库长期深耕新科技、消费、大健康、汽车出行、产业/工业、金融、碳中和等领域，旗下近100名分析师均毕业于名校，绝大多数具有丰富的从业经验；亿欧智库是中国极少数能同时生产中英文深度分析和专业报告的机构，分析师的研究成果和洞察经常被全球顶级媒体采访和引用。

以专业为本，借助亿欧网和亿欧国际网站的传播优势，亿欧智库的研究成果在影响力上往往数倍于同行。同时，亿欧内部拥有一个由数万名科技和产业高端专家构成的资源库，使亿欧智库的研究和咨询有强大支撑，更具洞察性和落地性。

◆ 报告作者：

周慧慧

亿欧智库 分析师

Email：zhouhuihui@iyiou.com

◆ 报告审核：

孙毅颂

亿欧智库研究总监

Email：sunyisong@iyiou.com

关于亿欧

◆ 版权声明：

本报告所采用的数据均来自合规渠道，分析逻辑基于智库的专业理解，清晰准确地反映了作者的研究观点。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。本报告的信息来源于已公开的资料，亿欧智库对该等信息的准确性、完整性或可靠性作尽可能的追求但不作任何保证。本报告所载的资料、意见及推测仅反映亿欧智库于发布本报告当日之前的判断，在不同时期，亿欧智库可发出与本报告所载资料、意见及推测不一致的报告。亿欧智库不保证本报告所含信息保持在最新状态。同时，亿欧智库对本报告所含信息可在不发出通知的情形下做出修改，读者可自行关注相应的更新或修改。

◆ 关于我们：

亿欧是一家专注科技+产业+投资的信息平台和智库；成立于2014年2月，总部位于北京，在上海、深圳、南京、纽约设有分公司。亿欧立足中国、影响全球，用户/客户覆盖超过50个国家或地区。

亿欧旗下的产品和服务包括：信息平台亿欧网（iyiou.com）、亿欧国际站（EqualOcean.com）、研究和咨询服务亿欧智库（EOIntelligence），产业和投融资数据产品亿欧数据（EOData）；行业垂直子公司亿欧大健康（EOHealthcare）和亿欧汽车（EOAuto）等。

亿欧服务

- 基于自身的研究和咨询能力，同时借助亿欧网和亿欧国际网站的传播优势；亿欧为创业公司、大型企业、政府机构、机构投资者等客户类型提供有针对性的服务。

- 创业公司

  亿欧旗下的亿欧网和亿欧国际站是创业创新领域的知名信息平台，是各类VC机构、产业基金、创业者和政府产业部门重点关注的平台。创业公司被亿欧网和亿欧国际站报道后，能获得巨大的品牌曝光，有利于降低融资过程中的解释成本；同时，对于吸引上下游合作伙伴及招募人才有积极作用。对于优质的创业公司，还可以作为案例纳入亿欧智库的相关报告，树立权威的行业地位。

- 大型企业

  凭借对科技+产业+投资的深刻理解，亿欧除了为一些大型企业提供品牌服务外，更多地基于自身的研究能力和第三方视角，为大型企业提供行业研究、用户研究、投资分析和创新咨询等服务。同时，亿欧有实时更新的产业数据库和广泛的链接能力，能为大型企业进行产品落地和布局生态提供支持。

- 政府机构

  针对政府类客户，亿欧提供四类服务：一是针对政府重点关注的领域提供产业情报，梳理特定产业在国内外的动态和前沿趋势，为相关政府领导提供智库外脑。二是根据政府的要求，组织相关产业的代表性企业和政府机构沟通交流，探讨合作机会；三是针对政府机构和旗下的产业园区，提供有针对性的产业培训，提升行业认知、提高招商和服务域内企业的水平；四是辅助政府机构做产业规划。

- 机构投资者

  亿欧除了有强大的分析师团队外，另外有一个超过15000名专家的资源库；能为机构投资者提供专家咨询、和标的调研服务，减少投资过程中的信息不对称，做出正确的投资决策。

扫码关注亿欧智库
查看更多研究报告

扫码添加小助手
加入行业交流群

亿欧智库

网址：https://www.iyiou.com/research

邮箱：hezuo@iyiou.com

电话：010-53321289

北京：北京市朝阳区保利国际广场T1-13层 | 上海：上海市徐汇区桂平路391号新漕河泾国际商务中心B座17层1703 | 深圳：深圳市宝安区华丰国际机器人产业园1期F栋110