

华为昇腾：国产 AI 算力“扛旗者”

2023 年 09 月 04 日

► **AI 算力需求快速增长，未来市场大有可观。**根据 IDC 预计，国内 2026 年智能算力规模有望进入每秒十万亿亿次浮点计算级别，2021-2026 年国内智能算力规模年复合增长率有望达 52.3%。在 AI 芯片方面，国内芯片市场 22-24 年复合增速有望达到 46%。在 AI 服务器方面，预计在 2026 年中国 AI 服务器市场规模达到 123.4 亿美元，市场空间广阔。

► **华为昇腾芯片为 AI 体系提供强大算力。**华为昇腾芯片是华为发布的两款人工智能处理器，包含昇腾 310 用于推理和 910 用于训练业务，均采用自家的达芬奇架构。昇腾 310 整数精度 (INT8) 算力可达 16TOPS，主要应用于边缘计算产品和移动终端设备等低功耗的领域。昇腾 910 整数精度 (INT8) 算力可达 640TOPS，在业界其算力处于领先水平，性能水平接近于英伟达 A100。

► **软件释放硬件性能，推动 AI 生态繁荣。**1) **CANN**：一种异构计算架构，位于芯片使能层，功能类似于英伟达 CUDA，可实现硬件加速。通过高性能调优引擎和 1400 个高性能算子等，支撑开发者快速部署神经网络业务。在 CANN 6.0 版本下，模型迁移成功率可达 90%。2) **MindSpore**：新一代全场景 AI 框架，帮助开发者孵化出各种 AI 创新算法和应用。具有能够实现动态图和静态图之间的切换、满足全部 AI 计算需求、降低用户使用模型并行时所面临的难度以及全场景快速部署等功能。根据 2023 年 Omdia 的调研数据，MindSpore 以 11% 的市场占比排名全球第三，逆势进入了 AI 框架的第一梯队。3) **MindStudio**：是一套开发工具链平台，工具链是指一系列用于制作软件的工具。其提供了应用开发、调试、模型转换、网络移植、优化和分析功能。开发者能够高效地完成端到端开发，支持训练和推理业务，可使训练速度提升 25%，推理提速 47%。

► **五大类生态伙伴，覆盖众多优质主流合作伙伴，繁荣生态助力产业长期发展。**昇腾生态伙伴包含整机硬件伙伴、IHV 硬件伙伴、应用软件伙伴、一体机解决方案伙伴以及生态运营伙伴五大类。其中，整机硬件伙伴有 13 家、IHV 硬件伙伴有 6 家、软件伙伴大约 1200 家。

► **投资建议：**AI 发展下算力产业发展迎来重大机遇，是 AI 大潮下确定性最高的细分领域之一；华为昇腾打造性能领先的算力处理器，同时打造软硬产品开发和合作伙伴良好生态，未来有望成为国产 AI 算力的“扛旗者”。建议重点关注科大讯飞、恒为科技、拓维信息、神州数码、软通动力、润和软件、四川长虹、海量数据等。

► **风险提示：**行业竞争加剧风险，底层算力市场格局不确定性风险，技术路线变革风险。

重点公司盈利预测、估值与评级

代码	简称	股价 (元)	EPS (元)			PE (倍)			评级
			2022A	2023E	2024E	2022A	2023E	2024E	
002230	科大讯飞	54.37	0.24	0.63	0.83	227	86	66	推荐
603496	恒为科技	19.10	0.24	0.42	0.60	80	45	32	推荐
002261	拓维信息	17.00	-0.81	0.09	0.18	/	195	95	推荐
000034	神州数码	30.95	1.50	1.83	2.21	21	17	14	推荐
301236	软通动力	20.36	1.02	1.29	1.57	20	16	13	推荐
603138	海量数据	19.58	-0.23	0.05	0.15	/	392	131	推荐

资料来源：民生证券研究院预测；(注：股价为 2023 年 9 月 1 日收盘价)

推荐

维持评级



分析师 吕伟

执业证书：S0100521110003

电话：021-80508288

邮箱：lvwei_yj@mszq.com

分析师 郭新宇

执业证书：S0100518120001

电话：010-85127654

邮箱：guoxinyu@mszq.com

相关研究

1. 计算机周报 20230901: 2023 年中报总结：低点已现，春天不远-2023/09/02
2. 计算机周报 20230827: 华为“王者归来”：从手机、自动驾驶到 AI 芯片-2023/08/27
3. 计算机行业事件点评：政策指引数据要素入表，行业加速发展可期-2023/08/22
4. 计算机周报 20230820: 再次强调证券 IT 有望成为“活跃资本市场”先锋-2023/08/20
5. 计算机周报 20230813: 再次聚焦信创与大安全-2023/08/13

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

目录

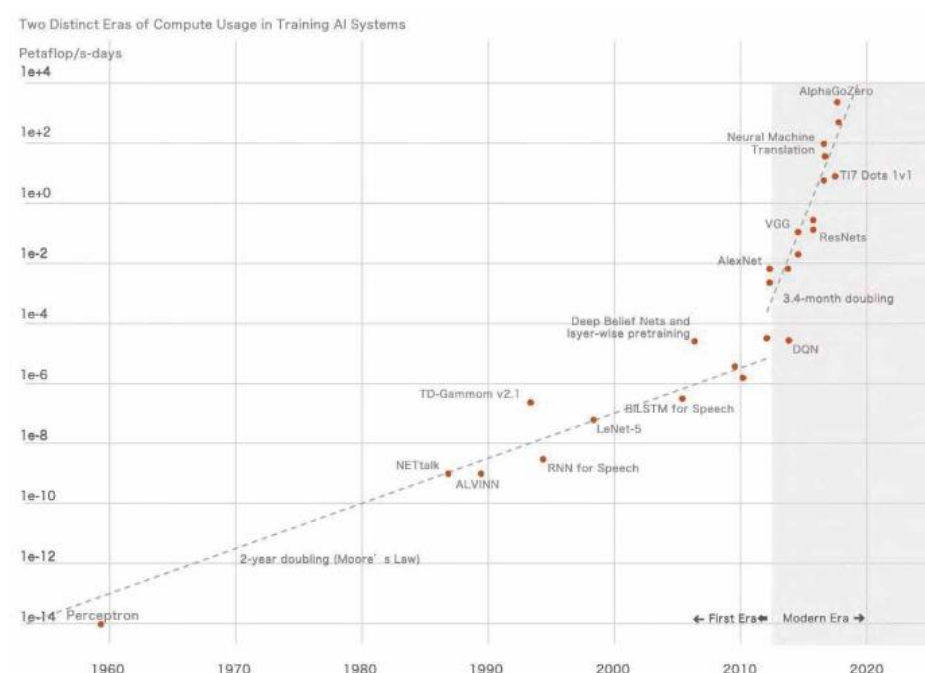
1 AI 算力需求快速提升，未来市场大有可观	3
1.1 数据量增长带来算力需求提升，智能算力规模有望快速增长	3
1.2 AI 芯片是提供算力的基础，国内市场 22-24 年增速有望超 40%	4
2 昇腾生态羽翼渐丰，芯片扛旗者初露锋芒	7
2.1 昇腾 AI 软硬件平台，构建智能世界基石	7
2.2 打造开发者生态，提供一站式服务	16
3 携手生态伙伴，共同赋能千行百业	18
3.1 昇腾五类生态伙伴，覆盖众多主流厂商	18
3.2 打造昇腾计算产业，为千行百业增砖添瓦	20
4 华为昇腾产业链重点公司梳理	26
4.1 科大讯飞：与华为昇腾共建国产算力底座，为赋能行业大模型奠定主要基础	26
4.2 恒为科技：AI 算力可视化的“光模块”，有望受益于华为算力底座建设	27
4.3 拓维信息：昇腾重要的硬件生产合作厂商	28
4.4 神州数码：华为第一批人工智能服务器伙伴	30
4.5 软通动力：昇腾 AI 生态圈践行者	31
4.6 润和软件：人工智能领域与华为强强联合	32
4.7 四川长虹：高性能服务器领军企业	33
4.8 海量数据：Vastbase G100 有望打开新的成长空间	34
5 投资建议	35
6 风险提示	36
插图目录	37
表格目录	37

1 AI 算力需求快速提升，未来市场大有可观

1.1 数据量增长带来算力需求提升，智能算力规模有望快速增长

数据海量增加，算法模型愈加复杂，应用场景的深入和发展，带来了算力需求的快速提升。根据白皮书数据显示，从 2012 年开始的 6 年中，AI 计算的需求增加了 30 万倍。

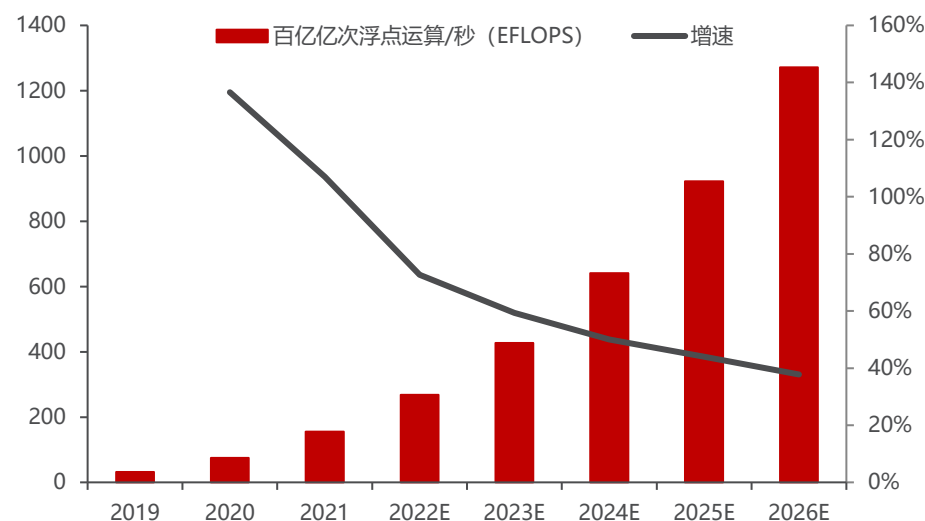
图1：AI 计算的需求上升



资料来源：《昇腾计算产业发展白皮书》，民生证券研究院

1) 根据 IDC 测算，国内智能算力规模正在高速增长，2021 年中国智能算力规模达 155.2 每秒百亿亿次浮点运算 (EFLOPS)，2022 年智能算力规模将达到 268.0EFLOPS，预计到 2026 年智能算力规模将进入每秒十万亿亿次浮点计算 (ZFLOPS) 级别，达到 1,271.4EFLOPS。2021-2026 年期间，预计中国智能算力规模年复合增长率达 52.3%。

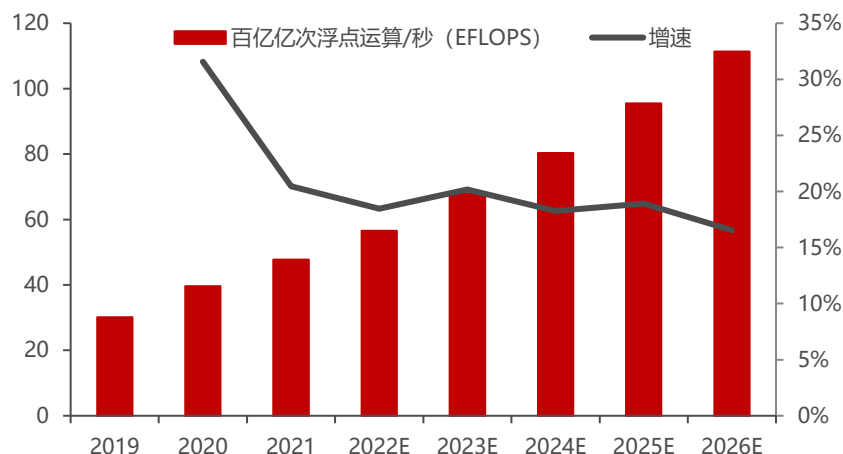
图2：中国智能算力规模（EFLOPS）及预测



资料来源：IDC，民生证券研究院

2) 通用算力规模也正在高速增长，根据 IDC 测算，2021 年中国通用算力规模达 47.7EFLOPS，预计到 2026 年通用算力规模将达到 111.3EFLOPS。2021-2026 年期间，预计中国通用算力规模年复合增长率为 18.5%。

图3：中国通用算力规模（EFLOPS）及预测



资料来源：IDC，民生证券研究院

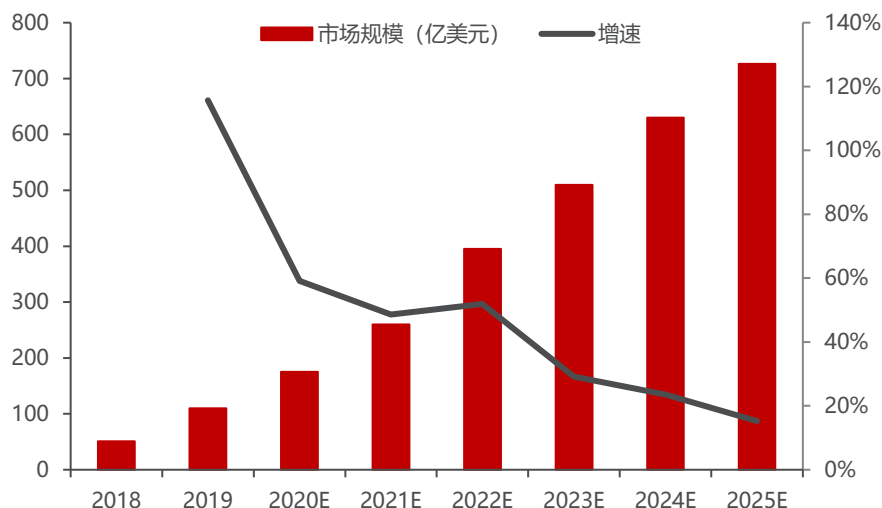
1.2 AI 芯片是提供算力的基础，国内市场 22-24 年增速有望超 40%

AI 包括人工智能芯片、服务器、计算架构、算法及应用等方面，本文主要关注算力层面，即 AI 芯片和服务器。

1) 全球市场：全球人工智能技术发展逐渐成熟，数字化基础设施不断建设完善，推动全球人工智能芯片市场高速增长。IDC 全球范围调研显示，人工智能芯片搭载率 (attach rate) 将持续增高。根据 Tractica、寒武纪招股书相关数据，全球

人工智能芯片 2022 年市场规模预计约 395 亿美金，预计到 2025 年将达到 726 亿美金，年复合增长率达到 22%。

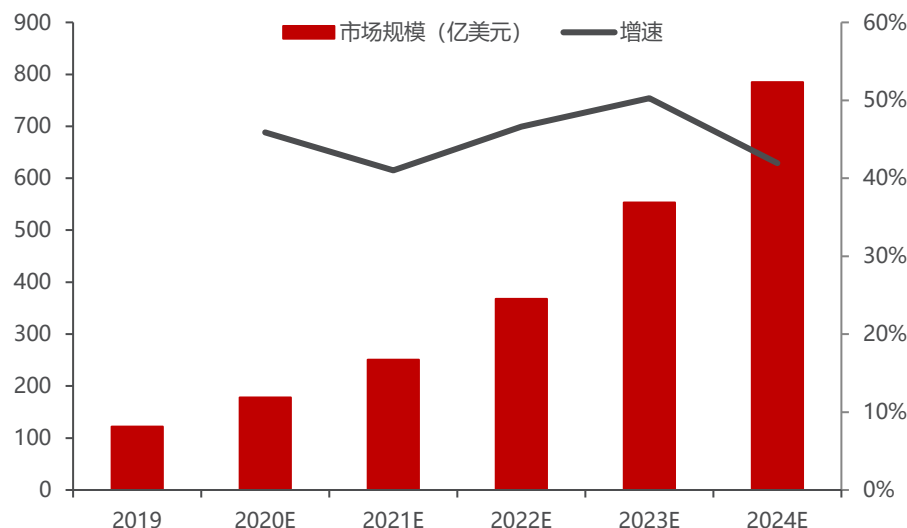
图4：全球 AI 芯片市场规模（亿美元）



资料来源：Tractica、寒武纪招股书，民生证券研究院

2) 国内市场：22-24 年复合增速有望达到 46%。随着大数据的发展和计算能力的提升，根据寒武纪招股书，2022 年中国人工智能芯片市场规模预计达 368 亿元，预计 2024 年市场规模将达到 785 亿元，复合增速有望达到 46%。

图5：国内 AI 芯片市场规模（亿美元）



资料来源：寒武纪招股书，民生证券研究院

AI 服务器市场保持快速发展：国内 2021 年市场规模超 300 亿元，增速近 70%。1) 全球市场：根据 IDC 发布的《全球人工智能市场半年度追踪报告》显示，2021 年全球人工智能服务器市场规模达 156.3 亿美元，约合人民币 1045 亿元，全球年度人工智能服务器市场首次突破千亿元人民币，同比 2020 年增速达 39%。其中，浪潮信息、戴尔、HPE 分别以 20.9%、13.0%、9.2% 的市占率位列前三，

三家厂商总市场份额占比达 43.1%。人工智能服务器市场预计将继续高速增长，预计 2026 年全球人工智能服务器市场规模将达到 347.1 亿美元，五年复合增长率为 17.3%。

2) 国内市场: 根据 IDC 数据, 2021 年中国人工智能服务器市场规模达到 59.2 亿美元, 与 2020 年相比增长 68.2%, 其中, 浪潮信息、新华三、宁畅、安擎、华为等诸多中国厂商正加速推动人工智能基础设施产品的优化更新。预计到 2026 年, 中国人工智能服务器市场规模将达到 123.4 亿美元。

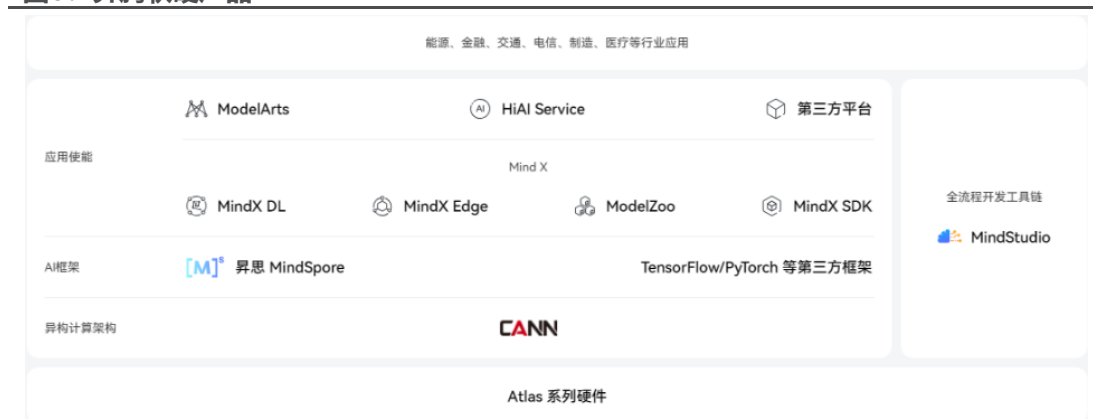
从工作负载角度而言: 企业将把精力更多从人工智能训练转移到人工智能推理工作负载上, 根据 IDC 数据, 2021 年中国数据中心用于推理的服务器的市场份额占比已经过半, 达到 57.6%, 预计到 2026 年, 用于推理的工作负载将达到 62.2%。

2 昇腾生态羽翼渐丰，芯片扛旗者初露锋芒

2.1 昇腾 AI 软硬件平台，构建智能世界基石

全栈 AI 软硬产品，覆盖“端边云”。昇腾计算产业凭借着以昇腾 AI 处理器为核心，通过系列硬件和基础软件构建全栈 AI 计算基础设施，为各行各业赋能。在硬件方面，包括模块、标卡、小站、服务器、集群等产品形态；在软件层面，包括异构计算架构、AI 框架、应用使能、全流程开发工具链等产品。

图6：昇腾软硬产品



资料来源：华为昇腾官网，民生证券研究院

2.1.1 硬件是 AI 体系的基础，提供强大计算能力

基于达芬奇架构的两颗“芯”。2018 年 10 月，代号为 310 和 910 的昇腾处理器系列产品推出。1) **昇腾 310 处理器：**本质上是人工智能片上系统，主要应用于边缘计算产品和移动端设备等低功耗的领域。该芯片采用 12nm 制造工艺，最大功耗仅为 8W，半精度 (FP16) 算力可达 8TFLOPS，整数精度 (INT8) 算力可达 16TOPS，同时还集成了 16 通道全高清视频解码器。2) **昇腾 910 处理器：**该芯片计算密度大，领先全球，相比于同时代的英伟达 Tesla V100 GPU 还要高出一倍，其主要应用于云端，可以为深度学习的训练算法提供强大算力。在算力方面，昇腾 910 表现非常出色，半精度 (FP16) 算力可达 320TFLOPS，整数精度 (INT8) 算力可达 640TOPS，功耗只有 310W，同时采用了 7nm 先进工艺进程，支持 128 通道全高清视频解码。从算力上看，昇腾 910 和英伟达 A100 性能基本上相当。

图7：昇腾 310 关键特性

自研华为达芬奇架构NPU		在8W数据精度下算力可达16TOPS	高性能3D Cube
Architecture	• HUAWEI Da Vinci		
Computing Engine	• 3D Cube		
Performance	• 16 TOPS@INT8 and 8 TOPS@FP16		
Max Power	• 8W		
Process	• 12nm FFC		

资料来源：华为海思官网，民生证券研究院

图8：昇腾 910 关键特性

自研华为达芬奇架构NPU		640 TOPS@INT8, 320TFLOPS@FP16	最大功耗310W
Architecture	• HUAWEI Da Vinci		
Computing Engine	• 3D Cube		
Performance	• 320 TFLOPS @FP16 and 640 TOPS @INT8		
Max Power	• 310W		
Process	• N7+		

资料来源：华为海思官网，民生证券研究院

表1：主流芯片特性对比

公司名称	产品型号	运算能力（理论峰值性能）	性能功耗比	峰值半精度（FP16）性能	峰值双精度（FP64）性能	架构	显存带宽	面向训练/推理任务
英伟达	A100	624TOPS@INT8（非稀疏模式）	约 2TOPS/W（非稀疏模式）	312TFLOPs	19.5TFLOPs	Ampere	1935GB/s	推理和训练
AMD	MI 100	184.6TOPS@INT8	约 0.6TOPS/W	184.6TFLOPs	11.5TFLOPs	CDNA	1.2TB/s	推理和训练
寒武纪	思元 590					MLUarch05		推理和训练
海光	深算一号						1024GB/s	推理和训练
百度昆仑	昆仑 2	256 TOPS@INT8	约 2.1TOPS/W	128TFLOPS		XPU-R		推理和训练
华为海思	Ascend 910	640TOPS@INT8	约 2TOPS/W	320TFLOPs		HUAWEI Da Vinci		推理和训练

资料来源：nvidia 官网、AMD 官网、寒武纪官网、海光招股书、昆仑芯官网，华为海思官网，民生证券研究院

行业性能领先的服务器。昇腾 AI 服务器具有超强算力，适用于中心侧 AI 推理以及深度学习模型开发和训练场景。1) **Atlas 800 推理服务器（型号：3000）**：具有高算力和高能效特点。最大可支持 8 个 Atlas 300I/V Pro，提供强大的实时推理能力，广泛应用于中心侧 AI 推理场景。2) **Atlas 800 推理服务器（型号：3010）**：具有灵活配置，适配多项负载特点。最大可支持 7 个 Atlas 300I/V Pro，广泛应用于中心侧 AI 推理场景。3) **Atlas 800 训练服务器（型号：9000）**：具有高算力密度等特点。拥有 4 颗鲲鹏 920 芯片，广泛应用于深度学习模型开发和训练，适用于智慧城市、智慧医疗、天文探索、石油勘探等需要大算力的行业领域。4) **Atlas 800 训练服务器（型号：9010）**：具有高算力密度等特点。可提供 2.24PFLOPS FP16 算力，广泛应用于深度学习模型开发和训练。5) **Atlas 800T A2 训练服务器**：具有高速带宽等特点。双向互联带宽达 392GB/s，广泛应用于深度学习模型开发和训练。6) **Atlas 500 Pro 智能边缘服务器**：具有易于部署维护和支持云边协同等特点。最大支持 3 张 Atlas 300I/V Pro 推理卡，单卡功耗仅 72W，发挥鲲鹏架构多核、低功耗优势，在边缘场景中广泛部署。

表2：各类服务器特性

产品名称	单卡算力	能效	内存频率	CPU	AI 加速卡	网络
Atlas 800 推理服务器 (型号: 3000)	1120TOPS@INT8	约 15.5TOPS/W	3200MT/s	2*鲲鹏 920	最大支持 8 个 Atlas 300I/V Pro	
Atlas 800 推理服务器 (型号: 3010)	980TOPS@INT8	约 4.78TOPS/W	2933MT/s	1/2 个 Intel® Xeon® SP Skylake 或 Cascade Lake 处理器	最大支持 7 个 Atlas 300I/V Pro	
Atlas 800 训练服务器 (型号: 9000)	1.76/2.24PFLOPS@FP16	2.24PFLOPS/5.6kW	3200MT/s	4*鲲鹏 920		8 * 100GE+ 4 * 25GE/2 * 100GE
Atlas 800 训练服务器 (型号: 9010)	1.76/2.24PFLOPS@FP16	2.24PFLOPS/5.6kW	2933MT/s	Intel V5 Cascade Lake 处理器		100GE * 8 1 * OCP NIC 3.0 标卡, 支持 2 * 25GE
Atlas 800T A2 训练服务器	2.5/2.2 PFLOPS@FP16 0.65/0.60 PFLOPS@FP32	0.6 PFLOPS / kW	3200MT/s	4 * 鲲鹏 920		8 * 200GE QSFP 接口直出, RoCE 协议
Atlas 500 Pro 智能边缘服务器	420TOPS@INT8	约 5.8TOP/W	3200MT/s	1*鲲鹏 920	最大支持 3 个 Atlas 300I/V Pro 推理卡	

资料来源：华为昇腾官网，民生证券研究院

华为昇腾超强算力推理+训练标卡。训练及推理标卡融合“通用处理器、AI Core、编解码”于一体，需要安装到服务器上，通过服务器提供运算和模型，为数据中心提供强劲算力。**1) 推理卡：**包括 Atlas 300V 视频解析卡、Atlas 300V Pro 视频解析卡、Atlas 300I Pro 推理卡、Atlas 300I Duo 推理卡。其中，Atlas 300I Duo 推理卡在配置下可以输出整数精度 (INT8) 达到 280TOPS，半精度 (FP16) 算力达到 140TOPS。**2) 训练卡：**主要以 Atlas 300T Pro 训练卡 (型号: 9000)、Atlas 300T A2 训练卡为主，可广泛应用于智慧城市、智慧交通、智慧园区、智慧金融等诸多 AI 行业场景。

表3：各类标卡特性

产品名称	单卡算力	能效	带宽	应用场景	优势
Atlas 300V 视频解析卡	100TOPS@INT8 50TFLOPS@FP16	约 1.4TOP/W	204.8Gb/s	推理	1.超大视频解析路数 2.高性能特征检索 3.安全启动
Atlas 300V Pro 视频解析卡	140TOPS@INT8 70TFLOPS@FP16	约 1.9TOPS/W	204.8Gb/s	推理	1.超大视频解析路数 2.高性能特征检索 3.安全启动
Atlas 300I Pro 推理卡	140TOPS@INT8 70TFLOPS@FP16	约 1.9TOPS/W	204.8Gb/s	推理	1.超强算力 2.超高能效 3.高性能特征检索 4.安全启动
Atlas 300I Duo 推理卡	280TOPS@INT8 140TFLOPS@FP16	约 1.86TOPS/W	408Gb/s	推理	1.超强算力 2.超高能效 3.高性能特征检索
Atlas 300T Pro 训练卡 (型号: 9000)	280TFLOPS@FP16	约 0.9TFLOPS/W	1228Gb/s	训练	1.超强算力 2.高度集成 3.高速带宽
Atlas 300T A2 训练卡	280TFLOPS@FP16 75TFLOPS@FP32	0.8TFLOPS/W	800Gb/s	训练	1.超强算力 2.高度集成 3.高速带宽

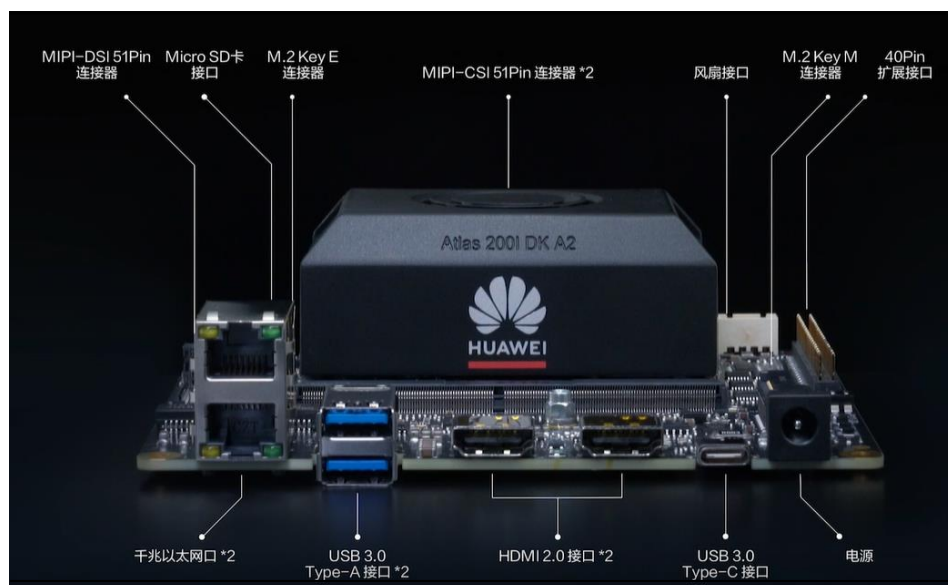
资料来源：华为昇腾官网，民生证券研究院

性能领先的训练集群。华为昇腾面向人工智能计算中心等重算力场景，构筑性能领先的训练集群，目前主要包含 Atlas 900 PoD (型号: 9000)、Atlas 900 A2 PoD、Atlas 900 AI 集群三种产品，广泛应用于深度学习模型开发和训练。**1) Atlas 900 PoD (型号: 9000):** 可提供 20.4PFLOPS FP16 算力，能效比最高可达 20.4 PFLOPS/46 kW，搭载 32 颗超强算力的鲲鹏 920 处理器，全节点 100G 网络互联，可实现高效加速应用。**2) Atlas 900 A2 PoD:** 提供 25.6/24PFLOPS FP16 算力，最大可拓展至 3.2EFLOPS，全节点 200G 网络互联，能效比可达 25.6 PFLOPS/47.5kW，具有极致算力密度、极高 A 级能效、极简交付部署等特点。**3) Atlas 900 AI:** 其总算力达到 256P~1024P FLOPS FP16，相当于 50 万台高性能 PC 的计算能力，让人类更高效地探索宇宙奥秘、预测天气、勘探石油，加速自动驾驶的商用进程。

其他华为昇腾超高算力集成产品。**1) 开发者套件:** Atlas 200I DK A2 开发者套件开发板集成了昇腾 310B 处理器，半精度为 4 TFLOPS@FP16，整数精度可达 8 TOPS@INT8，功耗为 24W。可以直接为开发者提供运算和模型等，能满足视频图像分析、自然语言处理、机器人等多个领域的需求。**2) 智能小站:** Atlas 500 AI 边缘推理小站算力可达 10 TFLOPS@FP16 或 20 TOPS@INT8，在有盘配置情况下，功耗为 44.5W，无盘配置时，功耗 32.3W。可独立部署，易于维护和支持云边协同。**3) 模块:** Atlas 200 AI 加速模块仅有半张信用卡大小，却能提供 22TOPS

的 INT8 算力。具有在边端侧实现人脸识别、图像分类等功能，因此广泛应用于智能摄像机、机器人、无人机等边端侧 AI 场景。

图9: Atlas 200I DK A2



资料来源：华为昇腾官网，民生证券研究院

2.1.2 软件释放硬件性能，推动 AI 生态繁荣

华为针对 AI 场景，设计了异构计算架构 CANN、AI 计算框架 MindSpore 和第三方适配以及全流程开发工具链 MindStudio。本文主要关注 CANN、MindSpore 和 MindStudio。

CANN 是一种异构计算架构，功能类似英伟达 CUDA。CANN 位于计算资源层和应用层之间，即芯片使能层，实现了在高性能计算硬件和 AI 应用之间架起一座桥梁。部署在昇腾服务器，包含统一的编程语言、统一网络构图接口、高性能计算引擎以及算子库。通过 AscendCL 对外提供 Device 管理、Context 管理、Stream 管理、内存管理、模型加载与执行、算子加载与执行、媒体数据处理等 API，帮助开发者实现在 CANN 平台上进行深度学习推理计算、图像预处理、单算子加速计算。在离线推理场景和训练场景都有应用。

CANN 释放 AI 硬件的极致性能，其高性能算子库和调优引擎，支撑快速部署神经网络。深度学习框架构建的模型是由一个个计算单元组成，这些计算单元被称为算子，对应着特定的计算逻辑，构成了加速神经网络的基础和核心。CANN 支持超过 1400 个高性能算子，并提供 900 多种优选模型，并且覆盖了 80% 的 DSL 算子，其 DSL 算子相比业界其他的产品，提升了 70% 开发效率。正是这些丰富的算子，筑起了澎湃的算力源泉。此外，CANN 还提供了新一代智能调优工具 AOE，取代了繁琐的手动优化操作，降低了调优的门槛，并提高了调优的效率。以 ResNet50 网络为例，使用 AOE 进行调优的效率比上一代工具提升了一倍以上，

并且性能提升超过 100%。在 CANN 6.0 版本下，模型迁移成功率可达 90%。

图10: CANN AI 异构计算架构

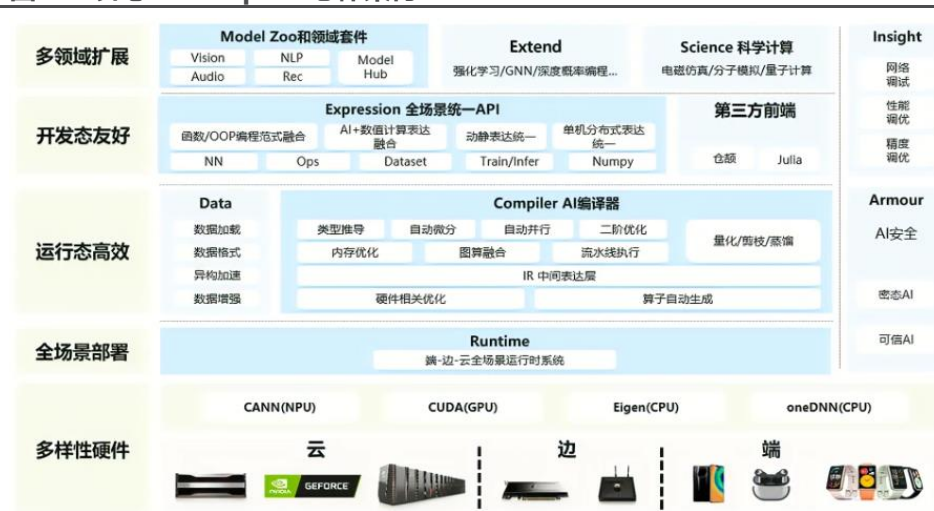


资料来源：华为昇腾官网，民生证券研究院

CANN 在各领域深受认可。1) 在学术领域方面：2022 年 3 月，基于 CANN 的 AI 论文 PLGAN 强势上榜 2022 CVPR, 5 月荣获“软件行业示范案例”，12 月，在第八届中国国际“互联网+”大学生创新创业大赛中，“基于昇腾算力及 CANN 的创新应用”突破“互联网+”国赛金奖。2) 在科研领域方面：分子动力学模拟服务在 CANN 的助力下成功商用，AI 预测性能达到现有产品的 1.5+倍，可预测规模较传统方法提升 10000+倍，为光伏材料、新能源电池、半导体材料研究带来巨大的商业应用价值。此外，昇腾 AI 平台端到端使能蛋白质结构预测基于 CANN 进行多维度性能优化，使得 AI 预测效率较 Baseline 提升 2+倍，同时具备长序列推理能力，2022 年底已支持 3800+长度的蛋白质氨基酸序列，达到业界领先水平。

MindSpore 是华为推出的支持云边端灵活部署的深度学习框架。人工智能领域的发展，离不开深度学习框架。从 2012 年以前的 Torch、OpenNN 等原始深度学习框架的发布，到后面形成了 TensorFlow 和 PyTorch 双头垄断，再到现在中国国产框架有了一席之地，深度学习框架市场新品更新不断。2020 年，华为研制的 MindSpore 正式开源，和 TensorFlow、PyTorch、飞桨等框架共同为人工智能领域发力。

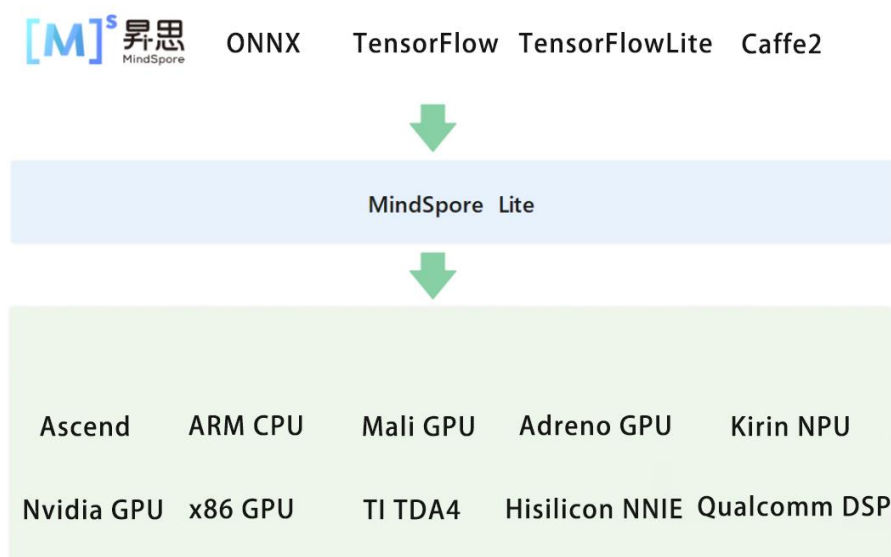
图11：昇思 MindSpore 总体架构



资料来源：华为昇腾官网，民生证券研究院

MindSpore 帮助开发者孵化出各种 AI 创新算法和应用，大幅提升效率。在 2018 年“华为全联接大会”上提出了人工智能面临的十大挑战，其中提到长部署周期，高开发成本、高技术门槛等问题，为了应对这些问题，昇腾 MindSpore 专注于实现开发友好、运行高效、全场景按需协同三大目标，以有效降低开发门槛。一般而言，开发深度学习软件框架能够把程序员从烦琐细致的具体编程工作中解放出来，从而将主要精力集中在人工智能算法的调优和改进上。昇思 MindSpore 作为新一代全场景 AI 框架，亦是如此，同时具有图算融合、分布式并行、企业安全可靠等功能，能够实现模型的训练-推理-全场景部署。

图12：支持全场景统一推理，大幅提高部署效率和推理性能



资料来源：昇思 MindSpore 官方公众号，民生证券研究院

MindSpore 易用性和性能全国领先。根据 Omdia 发布的 2023 年《中国人

《人工智能框架市场调研报告》可知，有 40% 的开发将“易用性”作为选择框架的首选因素，35% 的开发将“性能”作为选择框架的首选因素。其易用性是指产品被用户使用时，能够被用户理解、学习、使用和吸引用户的能力，易用性是产品的基本自然属性，标志着最终产品的可用性和成熟度。MindSpore 在易学习低门槛、易开发高效率、问题快速闭环三方面，提升易用性，并且成立了易用性 SIG。《中国人工智能框架市场调研报告》中提到，在国外的端到端框架中，开发者认为 TensorFlow，PyTorch 的易用性更好，而本土框架昇思 MindSpore 和飞桨 PaddlePaddle 在易用性方面紧随其后。在端到端人工智能框架性能方面，MindSpore 以 13% 的占比排名第三。

图13：融合编程新范式优势突出，实现易用性和性能的完美结合

```
def train_step(inputs, targets):
    loss, grads = grad_fn(inputs, targets)
    optimizer(grads)
    return loss
```



```
import mindspore as ms

@ms.jit
def train_step(inputs, targets):
    loss, grads = grad_fn(inputs, targets)
    optimizer(grads)
    return loss
```

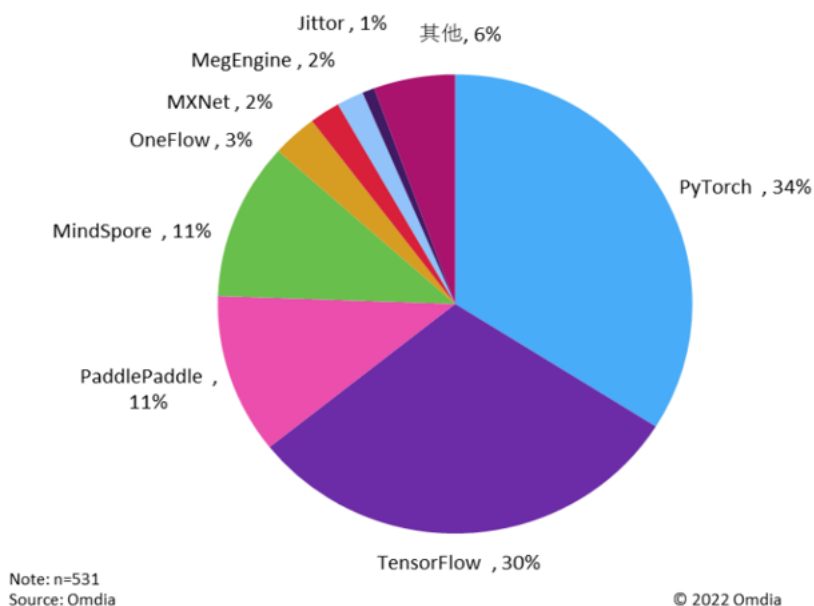
ms.jit修饰器：

- 一行代码切换动静态图
- 即时编译，被修饰函数转为整图

资料来源：昇思 MindSpore 官方公众号，民生证券研究院

MindSpore 在全球 AI 框架使用率方面处于第一梯队。根据 Omdia 的调研数据，在社区活跃度方面，MindSpore 以 11% 的占比，排名第四；在中国开发者心中，MindSpore 在国产框架中认知度排第一，全球框架中认知度排第三。在人工智能框架使用率方面，TensorFlow、PyTorch、昇思 MindSpore 和飞桨合计占了 86% 的市场份额，其中，MindSpore 以 11% 的占比在全球框架里排名第三，逆势进入了 AI 框架的第一梯队。

图14：中国开发者主流人工智能框架使用率排名



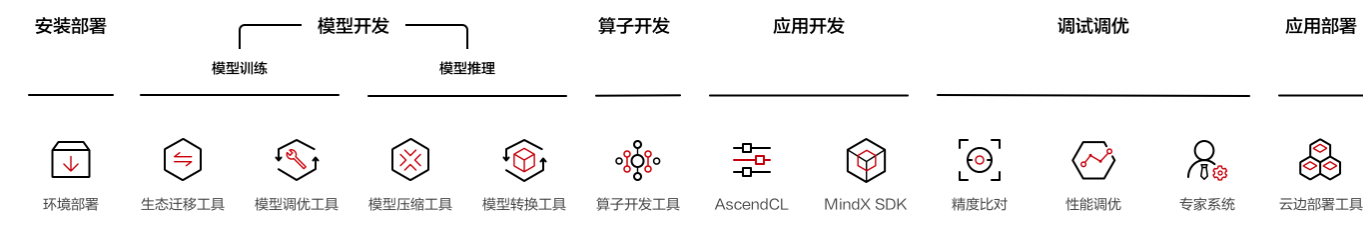
资料来源：《中国人工智能框架市场调研报告》，民生证券研究院

MindSpore 在国内市场广受认可。昇思 MindSpore 与超过 200 所高校和科研机构展开了创新合作，截至 2022 年，获得了超过 800 家企业的昇思技术认证，涵盖了制造、金融、电信运营商等国家重要领域。此外，昇思 MindSpore 在 Gitee AI 开源项目中综合排名第一，软件下载速度排名第一，已经超过 370 万次下载总量。

图15：昇思 MindSpore 合作伙伴

资料来源：华为昇思官网，民生证券研究院

提供全流程开发的工具链 MindStudio。开发者可以利用 MindStudio 提供的简单易用的开发工具，高效地完成端到端开发，让开发者能够在安装部署、模型训练、模型推理、算子开发、应用开发、调试调优和应用部署全流程中一站式完成，无需切换不同的工具，从而显著降低了开发门槛。MindStudio 支持训练、推理场景。**1) 训练场景：**MindStudio（昇腾训练加速工具）利用独有的算法，使训练速度提升 25%。**2) 推理场景：**MindStudio（昇腾模型压缩工具）利用独有的智能算法，加速推理进程，可大致提速 47%。

图16: MindStudio 开发流程


资料来源: 华为昇腾官网, 民生证券研究院

2.2 打造开发者生态, 提供一站式服务

数字时代, 开发者是时代的弄潮儿。华为将构建开发者生态视为最重要的工作之一, 从五个方面围绕开发者生态进行展开。

1) 多生态、多领域能力全面开放: 利用华为技术底座和开放能力, 华为开发者官网集结了各种生态和领域的开发平台和工具套件, 全面开放华为在人工智能、数据库、物联网、云原生、HarmonyOS、HMS、鲲鹏、昇腾等领域的技术能力。

2) 旗舰赛事汇聚, 赋能应用创新: 精选华为各生态领域的顶级赛事, 向开发者全面开放华为在各产业领域取得的技术成果, 鼓励开发者探索前沿技术。

3) 线上线下活动, 链接生态商机: 根据企业、高校、个人等不同开发者群体的需求, 华为提供丰富的线上活动、线下沙龙、技能培训和产品体验等活动。凭借华为专业的产品能力和大量的学习资源, 全方位赋能开发者在云上开展实际项目, 帮助企业开发者链接生态商机, 共同推动产业生态的发展。

4) 丰富的项目促进开发者成长: 华为拥有覆盖多个领域、多个生态圈的 30+ 开发者计划, 联结全球各地的开发者。针对企业、个人和高校, 华为提供全方位的支持, 包括在开发、推广和变现等环节给予帮助, 助力全球开发者成长, 加速开发者生态的成熟发展。

5) 打造一站式开发者服务: 围绕开放能力、学习赋能、应用构建、商业变现、服务支持等一站式的开发者服务, 全新升级的开发者官网覆盖了产品、活动、program、社区、学堂和支持菜单, 便于开发者从自身兴趣和目标出发, 快速检索获取相关产品服务。

图17: 昇腾相关的开发者计划

开发者成长计划	HAE与HAE计划	AI人才培养计划
帮助昇腾开发者学习和构建基于昇腾的技术能力	- 经华为认证的, 对赋能全球昇腾开发者有突出贡献的个人 - 为各地昇腾开发者提供学习交流、互动成长的平台	帮助高校利用华为AI资源和平台, 开展AI学课建设和人才培养

资料来源: 《昇腾计算产业发展白皮书》, 民生证券研究院

昇腾开发者社区是昇腾计算产业的门户网站，承担了资源中心，赋能中心和支持中心的角色。开发者可通过资源中心提供的资源，降低开发门槛，通过赋能中心的课程和实践平台快速掌握 AI 领域知识，通过支持中心，准确高效得到专家的回复。华为通过分层开放能力，打造昇腾开发者社区，提供最佳开发者体验，5 年内发展超过 100 万 AI 开发者。

图18：昇腾开发者社区三个中心

资源中心	赋能中心	支持中心
•提供超过100种工具与样例 •提供超过50种数据集	•超过50门直播课程 •超过70门的个人课程	•超过20名专家轮值服务 •开发者平均1小时内可得到专家的答复

资料来源：《昇腾计算产业发展白皮书》，民生证券研究院

华为坚持“硬件开放、软件开源、使能伙伴、发展人才”生态战略，据华为总裁周军介绍，昇腾 AI 开发者已经超过 180+万，合作伙伴超过 1200 家，行业 AI 解决方案认证超过 2500 个多。

3 携手生态伙伴，共同赋能千行百业

3.1 昇腾五类生态伙伴，覆盖众多主流厂商

昇腾生态伙伴包含整机硬件伙伴、IHV 硬件伙伴、应用软件伙伴、一体机解决方案伙伴以及生态运营伙伴五大类。其中，整机硬件伙伴有 13 家、IHV 硬件伙伴有 6 家、软件伙伴大约 1200 家，以萨技术是其一体机解决方案伙伴，极视角科技以及中软国际是其生态运营伙伴。

1) **整机硬件伙伴**：在昇腾计算整机硬件的产品方面，华为通过包括宝德、华鲲振宇、神州数码等众多主流的厂商合作，内容主要涵盖服务器，共同打造适应产业发展的产品。

图19：整机硬件伙伴



资料来源：华为昇腾官网，民生证券研究院

以黄河信产为例，黄河信产专注于鲲鹏系列通用计算和昇腾系列 AI 计算的服务器、微型计算机以及软硬一体化解决方案，并与生态伙伴紧密合作，共同推动计算生态的创新。其产品包括 Huanghe OceanAI 900H 训练服务器、Huanghe OceanAI 800H 推理服务器、Huanghe OceanAI 500 Pro 智能边缘服务器等。

表4：黄河信产相关产品参数

产品名称	AI 算力	AI 处理器	AI 加速卡
Huanghe OceanAI 900H 训练服务器	2.56/2.24/2/1.76 PFLOPS FP16	8 个昇腾 910	最大支持 8 个 Atlas 300I 推理卡
Huanghe OceanAI 900 训练服务器	2.56/2.24/2/1.76 PFLOPS FP16	8 个昇腾 910	
Huanghe OceanAI 800H 推理服务器	最大 980 TOPS INT8	英特尔	最多支持 7 张 Atlas 300I/300I Pro/300V Pro 推理卡
Huanghe OceanAI 800 推理服务器	最大 1120 TOPS INT8	昇腾 AI+鲲鹏 920	最大支持 8 张 Atlas 300I/300I Pro/300V Pro 推理卡
Huanghe OceanAI 500 Pro 智能边缘服务器	最大 420 TOPS INT8	鲲鹏 920	最大支持 3 张 Atlas 300I/300I Pro/300V Pro

资料来源：华为昇腾官网，民生证券研究院

2) **IHV 硬件伙伴**：凌华科技、研扬科技、研华科技、飞途、东声智能、全爱科技 6 大公司成为 IHV 硬件伙伴。该合作伙伴基于华为昇腾部件进行二次开发，形成自有品牌硬件产品并进行销售。

表5：IHV 硬件伙伴相关产品参数

伙伴名称	产品名称	AI 算力	功耗
凌华科技	AES-200 1U 边缘服务器	16 核 1.9GxHz ARM CPU 188TOPS 算力 NPU	98 W
研扬科技	研扬边缘智能盒子 ARES-500AI	22/16 TOPS INT8 11/8 TFLOPS FP16	
飞途	RTSO-7001		典型功耗 11W
	飞云智盒 Z507		典型功耗 11W
东声智能	DS-ATS200-506A 系列智能相机	最高支持 22TOPS INT8	
全爱科技	QA200EP 推理卡	22/16TOPS INT8 11/8TFLOPS FP16	20W
研华科技	嵌入式边缘计算平台 MIC-770Q	88 TOPS INT8	180W
	通用型边缘计算平台 IPC-NS01	88 TOPS INT8	850W
	短款边缘计算平台 HPC-7420	264 TOPS INT8	1200W

资料来源：华为昇腾官网，民生证券研究院

图20：IHV 硬件伙伴


资料来源：华为昇腾官网，民生证券研究院

3) 应用软件伙伴：对接昇腾服务器、智能小站、集群、加速模块和卡以及 MindSpore 等产品, 华为昇腾应用软件伙伴开发、销售自有知识产权的应用程序、软件、垂直细分应用等产品。该合作伙伴属于数字政府和制造行业的居多。

图21：应用软件伙伴


资料来源：华为昇腾官网，民生证券研究院

4) 一体机解决方案伙伴：以萨技术成为华为昇腾唯一的一体机解决方案伙伴, 基于昇腾系列的产品, 打造了以萨智慧交通和以萨智慧交通两大解决方案。

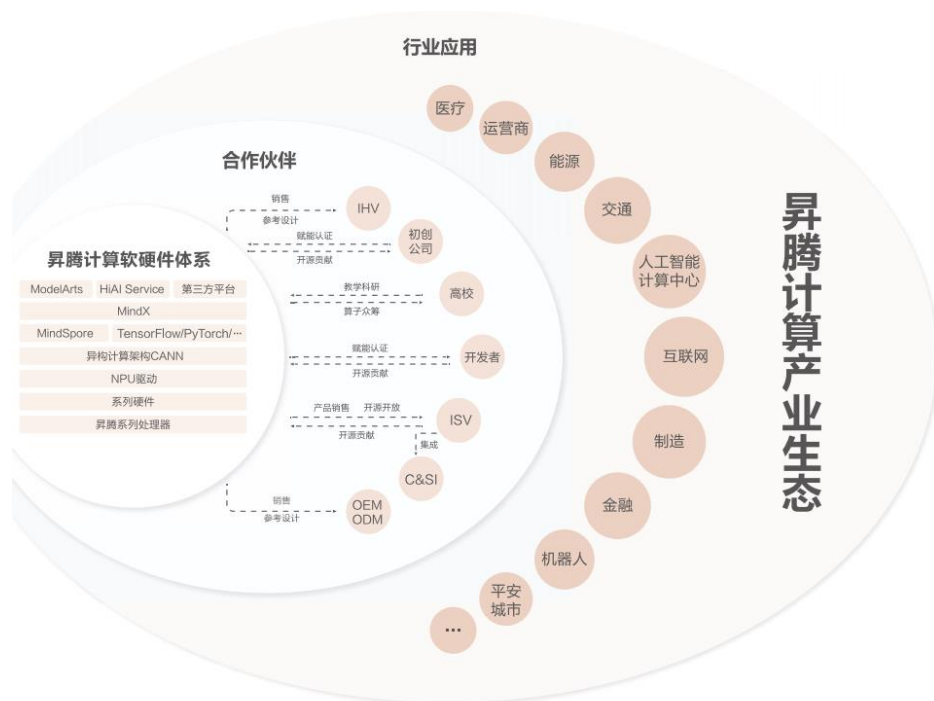
5) 生态运营伙伴：华为昇腾和极视角科技以及中软国际达成合作。其中, 极视角通过昇腾产品构建了“智慧城市”解决方案, 具有部署方便、一键式安装、一站式管理和数据不外泄功能; 而中软国际年蝉联中国软件和信息服务业综合竞争力百强企业前 15 位, 致力于昇腾算力企业合作、方案构建、算力运营等领域。

3.2 打造昇腾计算产业，为千行百业增砖添瓦

3.2.1 昇腾计算产业，助力 AI 发展

昇腾计算产业基于昇腾系列处理器和基础软件构建的全栈 AI 计算基础设施、行业应用及服务。其合作伙伴生态体系包括围绕着昇腾计算技术和产品体系所开展的学术、技术、公益及商业活动，产生的知识和产品以及各种合作伙伴，同时，还包含围绕昇腾相关产品对外提供服务交付的服务类伙伴，提供培训服务的人才联盟伙伴，提供投融资和运营服务的投融资运营伙伴等，而高校和开发者成为整个昇腾生态的重要组成部分。昇腾计算产业相互配合，共同促进技术进步和产品优化，推动以 AI 实现客户价值的理念，达到共建共赢的效果。

图22：昇腾计算产业生态

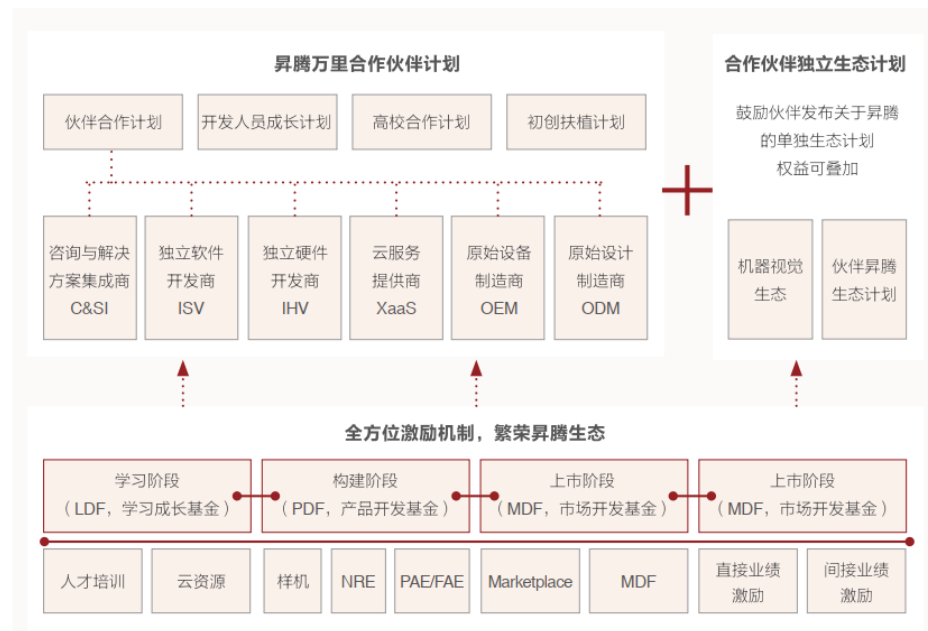


资料来源：《昇腾计算产业发展白皮书》，民生证券研究院

昇腾万里伙伴计划是基于昇腾 AI 基础软硬件平台推出的一项计划，提供培训、技术、营销和市场的全面支持。从伙伴、开发者、高校、初创企业等方面，开展相应计划，进行全面激励。1) 伙伴合作计划：根据伙伴的技术和市场能力，以及业绩贡献，将伙伴分为注册级、认证级、领先级、优先级，享有不同权益。2) 开发人员成长计划：根据开发者在昇腾开发者社区的活动，给与不同积分，可获得相应权益。3) 高校教学合作计划：与国内“双一流”建设高校合作，或校内设有 AI 学院或研究院、或设有独立的 AI 相关学科均可以加入 AI 人才培养计划。4) 初创扶持计划：分为孵化、成长和云腾三个阶段，企业参加该计划后，获得相应的扶持，随着与 Atlas AI 计算解决方案的进一步深入融合和裁员，进入成长和云腾阶段，

获得更高额度的扶持，并且会优先向伙伴和客户推荐销售。

图23：昇腾万里伙伴计划

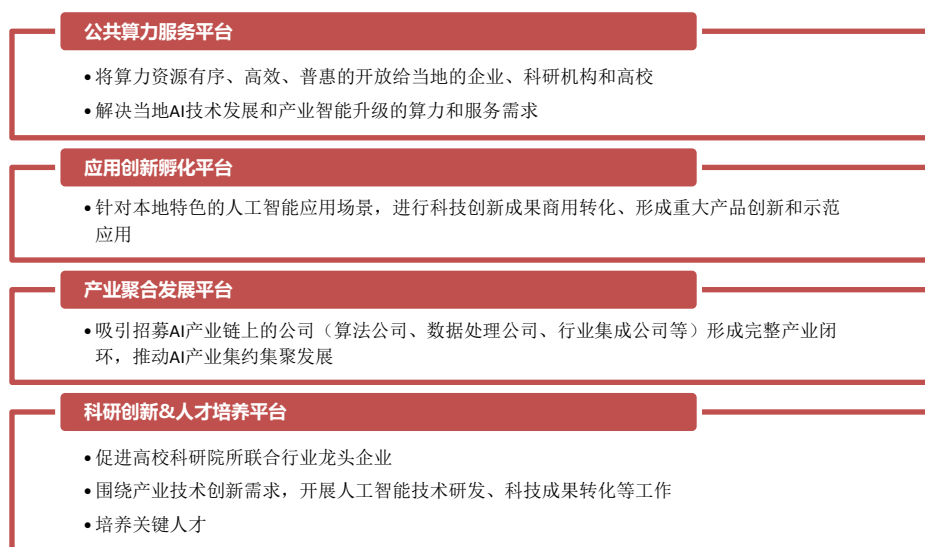


资料来源：《昇腾计算产业发展白皮书》，民生证券研究院

3.2.2 昇腾赋能行业，AI 算力无处不在

华为基于硬软件产品，推出众多 AI 解决方案。在人工智能计算中心、能源、金融、交通、电信、智慧城市、制造和医疗等行业广泛应用，并且创造了行业高价值。

AI 计算中心赋能产业集群。AI 计算中心提供从底层芯片算力释放到顶层应用使能的人工智能全栈能力，重点是构建“一中心四平台”，实现政府、产业、学术研究和应用场景的全方位连接和合作。通过这种方式，推动人工智能产业在区域乃至全国范围内的集聚和发展，促进“政产学研用”五位一体的协同发展。

图24：AI 计算中心的“四平台”发展模式


资料来源：华为昇腾官网，民生证券研究院

根据客户需求和 AI 计算中心，华为提供三种不同规模的解决方案：全栈 AI 计算中心解决方案、云 AI 计算中心解决方案和轻量化 AI 计算中心解决方案。

1) 全栈 AI 计算中心解决方案：提供多样化的算力，从 L0 到 L3 全栈交付。通过应用驱动的多样性算力融合、超高密度硬件架构创新、应用感知的存算协同以及全栈能效管理，实现倍数级的全栈优势。适用于需要构建全栈智能计算中心的客户。

2) 云 AI 计算中心解决方案：提供 Atlas 硬件和华为云 HCSO (ModelArts) 软件，高效、可知识复用的一站式 AI 模型开发平台。在数据处理阶段可节省 50%-80% 的人力，智能标注可提升 70% 的标注效率，基于图像分类的 AI 训练性能可达到业界的 4.5 倍，推理性能可达到业界的 1.8 倍。适用于需要构建硬件和软件平台的智能计算中心的客户。

3) 轻量化 AI 计算中心解决方案：提供基于 MindX DL 的具有行业属性的深度学习平台，具备 web 页面、一键安装部署、操作简单等特点。只需 5~10 步即可完成训练，非常方便快捷。适用于需要聚焦软件平台的行业。

图25：AI 计算中心解决方案

	全栈 AI 计算中心 解决方案	云 AI 计算中心 解决方案	轻量化 AI 计算中心 解决方案
L4 应用平台	AI 应用 (合作伙伴 / 开发者 / 高校)		
L3 软件平台	应用使能、AI 框架、芯片使能	华为云 HCSO (ModelArts)	合作伙伴深度学习平台
L2 硬件平台	通用计算：鲲鹏、X86 AI 计算：昇腾、GPU	Atlas 800 训练服务器 Atlas 900 AI 集群	Atlas 800 训练服务器 Atlas 900 AI 集群
L1 机房设施	机柜 / 电源 / 冷机 / 油机	可选	可选
L0 楼宇平台	预制式机房	可选	可选

资料来源：《昇腾计算产业发展白皮书》，民生证券研究院

华为昇腾是国内政府主导智算中心建设的最重要力量。我国超大规模预训练模型的发展如火如荼，算力需求持续攀升，人工智能计算中心的建设保持快速增长。政府统筹建设的人工智能计算中心，多采用国产通用处理器和 AI 加速器技术，以华为昇腾、寒武纪思元等国内 AI 芯片为主。目前，在国内政府为主导的智算中心中，华为昇腾签约合作的占一大部分。重庆、武汉、西安等城市均已建成人工智能计算中心并投入运营，淮海、大连等城市正在建设中，北京、青岛等地的人工智能计算中心建设也在陆续规划中。

表6：华为昇腾智算中心情况

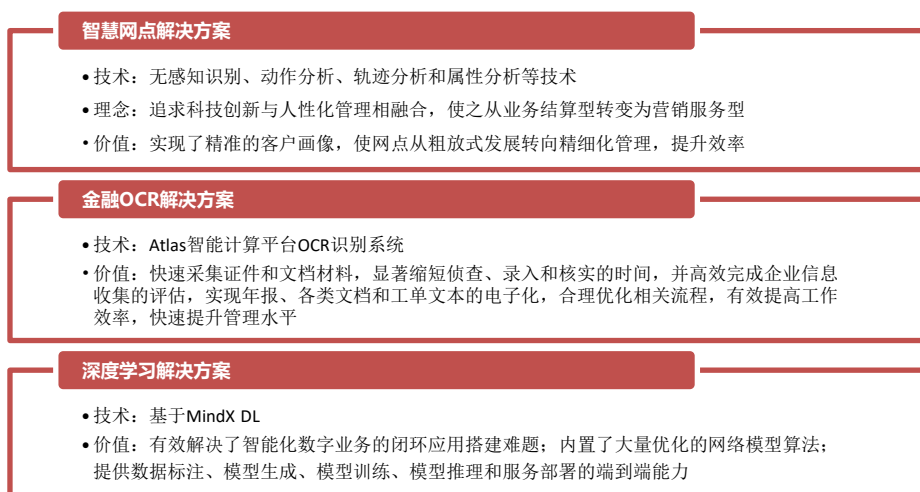
智算中心名称	运营状态	算力情况 (Pflops)
武汉人工智能计算中心	2021 年 5 月 31 日投入运营 2022 年 2 月 7 日完成扩容	总算力 200P
西安人工智能计算中心	2021 年 9 月 9 日上线运营	一期 AI 算力 300P；HPC 算力 8P
中原人工智能计算中心	2021 年 10 月 21 日上线运营	一期算力 100P；未来将达 300P
南京人工智能计算中心	2022 年 5 月完成基础建设	一期规划 40P
成都智算中心	2022 年 5 月 10 日正式上线	算力达 300P
杭州人工智能计算中心	2022 年 5 月上线运营	一期 40P；二期 140P
沈阳人工智能计算中心	2021 年 10 月启动建设	一期 100P；总规划 400P
天津人工智能计算中心	2023 年 3 月 18 日上线运营	一期 100P；可提供 300P
青岛人工智能计算中心	2023 年 2 月 23 日正式上线	首期 100P

广州人工智能计算中心	2022 年 9 月 15 日建设完成	对外提供 99PAI 算力服务
重庆人工智能计算中心	2023 年 5 月 14 日正式使用	一期 400P
北京晟腾人工智能中心	2023 年 2 月 13 日上线	一期 100P; 短期 500P; 远期 1000P
河北人工智能计算中心	2022 年 2 月 14 日揭牌	计划 100P
济南人工智能计算中心	接入算力网	
淮海智算中心	在建	300P
中国·东盟海人工智能计算中心	2022 年 9 月 23 日揭牌	一期 40P 训练, 1.4P 推理
深圳人工智能计算中心	2021 年 12 月 13 日启动建设	
长沙人工智能计算中心	2022 年 11 月 4 日正式运营	首期 200P; 2025 年 1000P
未来人工智能计算中心	2021 年 9 月 9 日正式上线	一期 300P; 2025 年 1000P
大连人工智能计算中心	在建	100P; 未来 300P

资料来源: 华为昇腾官网, 中国政府网, 大连市人民政府网, 宿迁市人民政府网, 中国网, 新华网, 腾讯网, 澎湃新闻网, 中国新闻网, 民生证券研究院

携手 AI 拥抱创新, 引领金融智能化。在金融行业, 华为基于 Atlas 智能计算平台, 形成智慧网点、金融 OCR、深度学习三大解决方案, 与高重科技、合合、眼神科技、依瞳科技、第四范式等企业达成合作。有效提高金融行业工作人员的效率, 为该行业提供了更好的服务。

图26: 金融行业三大解决方案



资料来源:《昇腾计算产业发展白皮书》, 民生证券研究院

加速医疗智能化服务。面对目前医疗资源分布不均、供需不平衡、医疗数据资源存在孤岛的现象，华为合作伙伴基于 Atlas AI 计算解决方案开发部署智能辅助诊断方案，采用 Atlas 800 推理服务器搭建 AI 辅助诊断平台，与接院方 PACS 对接。在新冠疫情期间，智能辅助诊断解决方案使得识别新冠肺炎的诊断时间由传统方式的 10-15 分钟下降至 10-15 秒，准确率由 90% 上升到 99%。

图27：新冠肺炎医疗影像辅助诊断解决方案



- 提升50倍检测效率
- 10-15秒/人 诊断时间
- 99%准确性

资料来源：华为昇腾官网，民生证券研究院

4 华为昇腾产业链重点公司梳理

4.1 科大讯飞：与华为昇腾共建国产算力底座，为赋能行业大模型奠定主要基础

与华为联合发布星火一体机，合作共建全国产化算力底座。华为与科大讯飞共同在昇腾 AI 的软硬件平台和软件的支撑工具上，把高算力 AI 芯片、高性能的算子库、多卡高速互联以及分布式存储结合起来，尤其是联合针对人工智能所需要的最重要的算子库进行甄别和打磨，然后在此基础上，架构科大讯飞的训练和数据闭环全流程设计，以及训练和推理一体化设计的自研大模型训练平台，中间是支持大规模的异构算力兼容，也支持混合云架构易拓展。在单卡算力上，平台对标英伟达 A100，同时公司还正在跟华为一道打造面向超大规模大模型训练的国产算力集群，形成集群化优势。

图28：科大讯飞联合华为昇腾打造算力底座



资料来源：科大讯飞官方公众号，民生证券研究院

星火一体机有望成为助力行业和企业快速部署私有化大模型的重要抓手。在应用端，科大讯飞给出了 5 种特别需要又方便的定制优化模式，支持场景开发，也给出了十几种直接可用的场景包，以及 7000 多个小助手，可以使每一家企业、每一个行业、每一个学校、每一个医院，都有机会构建自己的专属大模型，企业可以从头到尾训练大模型，也可以基于科大讯飞的星火大模型来定制。科大讯飞有 130 亿、650 亿、1750 亿的模型参数可选，在这基础上针对企业的场景进行快速优化，一周左右就可以在原有应用上增加 20% 技术指标。因此，星火一体机的发布，可以帮助每一个企业、每个行业快速部署私有化大模型。

图29：科大讯飞联合华为昇腾打造算力底座



资料来源：科大讯飞官方公众号，民生证券研究院

风险提示：技术路线变革具有不确定性；行业格局变化

4.2 恒为科技：AI 算力可视化的“光模块”，有望受益于华为算力底座建设

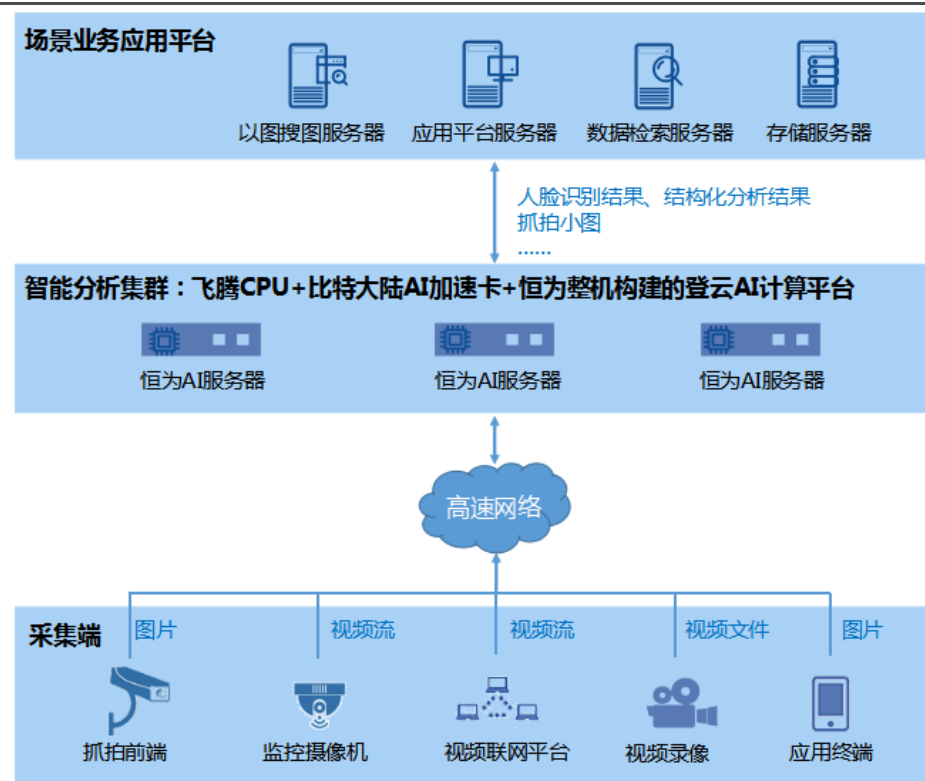
由“东数西算”到“一体化算力算网调度平台”，国家打造如同电网一样可以任意调度的算网体系。2021年5月，国家发展改革委等四部门联合印发的《全国一体化大数据中心协同创新体系算力枢纽实施方案》提出“东数西算”，实现我国系统化合理布局数据中心资源，全国数据中心的一体化发展；2023年6月5日，中国信通院联合中国电信共同发布我国首个实现多元异构算力调度的全国性平台“全国一体化算力算网调度平台（1.0版）”，旨在落地国家东数西算战略，通过“三跨四互联”，解决算力可管可控可调度的问题，通过汇聚通用算力、智能算力、高性能算力、边缘算力等多元算力资源，设计异构资源池调度引擎，实现不同厂商异构资源池的算力动态感知与作业智能分发调度。特别在AI训练作业调度流程中，可在智算资源池上进行训练推理，在通用算力资源池部署，从而实现跨资源池/跨架构/跨厂商的异构算力资源调度，目前已接入天翼云、华为云、阿里云等。

公司网络可视化对于流量采集、分发、分析等能力与智能计算业务各类服务器、AI一体机等技术相结合，能够很好满足智算可视化技术需求。据《新一代智算中心网络技术白皮书》，智算中心关键技术包括超大规模网络技术、超高性能网络技术以及网络可靠性与智能运维技术等，恒为科技技术储备能够较好满足智算中心关键技术需求。

公司2020年8月就已经推出“登云”AI计算平台解决方案，2023年一季度已有一部分与科研院所合作的AI一体机和AI服务器项目落地，营收贡献约1500万元左右。该解决方案以恒为高性能AI服务器平台为载体，具有领先业界的计算性能，特别适用于超大运算规模的深度神经网络的推理运算，可实现对视频结构化、人脸识别、物品识别、视频内容增强等典型应用的AI加速运算。同时，

公司表示，与华为在网络可视化等业务上有深入的合作关系。

图30：恒为“登云”AI 计算平台解决方案



资料来源：飞腾微信公众号，民生证券研究院

智算中心业务模式有望沿着算力出租、算力增值再到模型训练和 MaaS 的趋势逐步升级，当前恒为科技与中贝通信的合作已经进入到算力增值的环节。据中贝通信公告，公司、恒为科技及第三方拟共同出资 20,000 万元，设立合资公司为客户提供优质高性能的智算算力及专业的技术服务，恒为科技智算可视化相关业务有望落地。据《AIGC 算力发展趋势》，公司 MaaS 与模型一体机是更前沿的技术，恒为科技有望在算力增值业务稳定后进一步向二者发展。

风险提示：行业格局变化；新技术路线变化给行业发展带来不确定性。

4.3 拓维信息：昇腾重要的硬件生产合作厂商

拓维实现覆盖软+硬的 AI 全面布局。拓维连续中标了重庆、长沙人工智能计算中心项目以及全国一体化算力网络国家（贵州）主枢纽中心；在教育、交通行业率先发布开源鸿蒙行业发行版并联合相关部委制定行业标准等。根据拓维公开调研纪要，拓维已经成为了 A 股唯一华为“鲲鹏/昇腾 AI+开源鸿蒙”软硬一体战略合作伙伴。

华为昇腾重要的硬件生产合作厂商，积淀深厚。拓维 2021 年 3 月成为华为首批昇腾硬件生产合作厂商，2021 年 7 月成为华为首批昇腾人工智能创新中心运营类合作伙伴，并成为华为昇腾智行 ISV、昇腾智造 ISV，在交通、制造等领域打造

了首批基于昇腾 AI 的行业标杆产品与解决方案。面向 AI 领域，拓维基于 27 年行业数字化技术沉淀、大量行业自有数据以及丰富的案例应用经验，未来有望构建从自研硬件到行业大模型到 AI+行业应用的全方位 AI 差异化产品矩阵。

打造多款 AI 产品，不断拓展自身能力。在 AI 产品方面，拓维控股子公司湘江鲲鹏拥有华为昇腾整机伙伴授权，基于鲲鹏处理器和昇腾处理器，先后发布 AI 推理、AI 训练、AI 小站、AI 集群、智能边缘等数十款兆瀚 AI 产品，并携手华为参与全国各地的人工智能计算中心建设。拓维拟改的兆瀚 AI 推理服务器是一款基于鲲鹏平台+昇腾 310 处理器的推理服务器，能为用户提供强大的实时推理能力，广泛应用于中心侧 AI 推理场景，帮助用户应对较重的计算压力。

图31：公司兆瀚 AI 推理服务器产品优势

 超强算力 支持8张Atlas300I推理卡，满足多场景推理需求；整机可提供640路高清视频实时分析(1080P25FPS)。搭载64核架构、具有超强算力的鲲鹏920处理器，支持端边云不同场景的差异化场景算力需求，并具备从几十毫瓦IP到几百瓦芯片的平滑扩展，覆盖了端边云全场景部署的能力。	 超高效能 发挥鲲鹏架构多核、低功耗优势，为推理场景构建高效能、低功耗的AI计算平台。	 低能耗 在边缘侧，模组能效比达 2 TOPS/W，适应低功耗和边缘部署需求。Atlas 300I 单卡功耗仅 67 W，为AI服务器算力加速同时提供更优的能效比。
--	---	--

资料来源：拓维信息公司官网，民生证券研究院

表7：拓维信息兆瀚 AI 服务器梳理

产品型号	产品形态	处理器	内存	AI 加速卡	AI 算力
RA2300-A	2U 推理服务器	2*鲲鹏 920 处理器	32 个 DDR4 内存插槽	支持 Atlas 300I Pro 推理卡和 Atlas 300V Pro 视频解析卡	最大 1.12 POPS INT8 最大 560 TFLOPS PF16
RA2302-B	2U AI 服务器	2*64 核青松处理器	32 个 DDR4 内存插槽，最高 3200 MT/s，支持 ECC	最大支持 4 个 Atlas 300I/V Pro	最大 560 TPOS INT8
RA5900-A	4U 训练服务器	4*鲲鹏 920	最多 32 个 DDR4 内存插槽	8*昇腾 910	2.56 PFLOPS FP16
RA5900-B	4U AI 服务器	4*鲲鹏 920	32 个 DDR4 内存插槽，最高 3200Mb/s	8*昇腾 910	2.5/2.2PFLOPS FP16 0.65/0.60 PFLOPS FP32
CA9900	47U 机柜	32*48 核鲲鹏 920	最多 256 个 DDR4 内存插槽，支持 RDIMM 单根内存条 容量支持 32GB/64GB	64*昇腾 910	最高 20.48 PFLOPS FP16
SA200	轻量边缘设备	昇腾 310	LPDDR4X，8GB/4GB，最大 51.2 GB/s	昇腾 310	22/16 TOPS INT8 11/8 TFLOPS FP16
SA300	2U 智能边缘服务器	1*鲲鹏 920	4 个 DDR4 内存插槽	支持 Atlas 300I Pro 推理卡/Atlas 300V Pro 视频解析卡	最大 420 TOPS INT8 或 384 路 1080P 30 FPS 视频解析

资料来源：湘江鲲鹏官网，民生证券研究院整理

表8：拓维信息盈利预测表

项目/年度	2022A	2023E	2024E	2025E
营业收入（百万元）	2,237	2,801	3,452	4,185
增长率（%）	0.3	25.2	23.3	21.2
归属母公司股东净利润（百万元）	-1,013	109	223	343
增长率（%）	-1337.4	110.8	104.6	53.6
每股收益（元）	-0.81	0.09	0.18	0.27
PE	-21	195	95	62
PB	8.6	8.2	7.6	6.8

资料来源：wind，民生证券研究院预测（股价为9月1日收盘价）

风险提示：行业下游需求波动；行业格局变化。

4.4 神州数码：华为第一批人工智能服务器伙伴

神州数码 2020 年开启了与华为昇腾产品线的合作，是华为第一批人工智能服务器三家合作伙伴之一，也是昇腾优选级合作伙伴。根据神州数码公开调研纪要，神州数码人工智能服务器收入在自主品牌业务中的占比大约是 10%。

神州数码在 AI 服务器领域主要包括 KunTai A222 2U 单路边缘型 AI 服务器、KunTai A722 2U 双路推理型 AI 服务器和 KunTai A924 4U 四路训练型 AI 服务器。

表9：神州数码鲲泰 AI 服务器梳理

产品型号	产品形态	处理器	AI 加速卡	AI 算力
KunTai A222	机架式服务器	1*鲲鹏 920 处理器，24 核，主频 2.6GHz	最大支持 3 张 Atlas 300V 视频解析卡或 Atlas 300I Pro 推理卡或 Atlas 300V Pro 视频解析卡	最大 420 TOPS INT8
KunTai A722	机架式服务器	2*鲲鹏 920 处理器，支持 32、48、64 核可选，主频 2.6GHz	最大支持 8 张，Atlas 300V 视频解析卡或 Atlas 300I Pro 推理卡或 Atlas 300V Pro 视频解析卡	最大 1120 TOPS INT8
KunTai A924	机架式服务器	4*鲲鹏 920 处理器，支持 48 核，主频 2.6GHz	8*昇腾 910，支持直出 100G RoCE 网络接口	最大 512Tops Int8 或 256Tops FP16

资料来源：神州鲲泰官网，民生证券研究院整理

神州数码在昇腾领域持续取得进展。在 WAIC2022 昇腾人工智能生态大会上，神州数码发布了三款产品。在运营方面，8 月 29 日，神州数码与沈阳昇腾人工智能生态创新中心正式签署辅助运营协议，双方围绕“昇腾”底座，共同打造人工智能联合创新实验室，在生态合作、软硬件适配、解决方案创新、人才培养等多个维度展开合作，构建自主创新、端到端的人工智能计算生态，推动产业的数字化变革，助力区域数字经济高质量发展。

图32：WAIC2022 昇腾人工智能生态大会上神州数码发布的新产品

Kuntai AT200中心训练服务器	Kuntai AI200 推理服务器	Kuntai AE200 边缘服务器
整机具有强算力密度、超高能效与高速网络带宽等特点，可广泛应用于深度学习模型开发和训练，适用于智慧城市、智慧医疗、天文探索等需要大算力的行业领域，具有自我管理能力	可广泛应用于城市治理、疫情防控、语音识别、金融大脑等场景，并提供强大的实时推理能力，并具有超强的计算性能、低能耗、易于部署维护和支持云边协同等特点	可支持CPU通过总线连接协同工作，兼具超强计算性能、高环境适应性、易于部署维护和支持管理协同等特点。体积小，可以在边缘场景中广泛部署

资料来源：神州数码官方公众号，民生证券研究院

表10：神州数码盈利预测表

项目/年度	2022A	2023E	2024E	2025E
营业收入（百万元）	115,880	125,136	133,538	146,274
增长率（%）	-5.3	8.0	6.7	9.5
归属母公司股东净利润（百万元）	1,004	1,227	1,481	1,716
增长率（%）	303.1	22.1	20.7	15.9
每股收益（元）	1.50	1.83	2.21	2.56
PE	21	17	14	12
PB	2.7	2.4	2.1	1.9

资料来源：wind，民生证券研究院预测（股价为9月1日收盘价）

风险提示：细分领域行业地位波动；技术路线变革对公司影响具有不确定性。

4.5 软通动力：昇腾 AI 生态圈践行者

软通动力作为中国领先的软件与信息技术服务商，长期深耕数字创新服务+AI赛道，拥有完整的技术创新体系，并成立了专注于AI领域的人工智能研究与创新中心(AIC)，为客户的AI服务体系提供支撑。同时软通动力积极融入昇腾AI生态圈。

AI领域的重点产品：软通动力打造出了具有端云协同、软硬一体、安全可信三大优势的AI端云一体化产品（昇腾版），该产品已与华为Atlas 300I/V Pro推理卡完成兼容性测试。此外，软通动力在各地的昇腾AI生态创新中心，均扮演多样化算力能力服务商的角色，协助生态伙伴和用户，共同完成芯片适配、系统迁移、应用创新、人才培养等工作。同时，软通动力作为首批合作企业签约入驻北京昇腾人工智能计算中心，并将参与到“人工智能安全可信护航计划”的理论研究、标准制定、评估评测等相关工作中，与昇腾一道共同推进AI规模应用与产品创新。

图33：软通动力与昇腾结合的相关产品及优势

AI端云一体化产品

- 在中铁十一局集团有限公司财务共享中心RPA+AI智能化项目中，软通动力从集团财务共享中心业务顶层设计开始，对现有管理模式、业务系统进行改造升级，实现了单位日处理数据10倍以上的提升

知识图谱技术

- 通过知识图谱技术软通动力为某物联网公司的5G定制网能力魔方系统进行智能化革新，通过对数据采集、知识库管理、推荐模型管理、评估管理等四大模块升级，提升工作效率

自研echo智能客服

- 软通动力形成以自研echo智能客服为代表的生态产品服务，通过iRSP智能机器人软件平台，TTS+ASR语音识别与合成及RPA机器人流程自动化三种产品，解决了企业客户广泛面临的人员成本高等问题

资料来源：软通动力官方公众号，民生证券研究院

风险提示：AI 技术路线具有不确定性；公司新产品研发进度具有不确定性。

4.6 润和软件：人工智能领域与华为强强联合

润和软件聚焦“金融科技”、“智能物联”和“智慧能源”三大业务领域，拥有全球化的技术整合、客户响应与精准服务能力，帮助客户实现价值提升。2021年底润和软件加入昇腾万里伙伴计划，致力于在人工智能领域与华为强强联合，共建昇腾 AI 生态。

以石油冶炼行业为例，润和软件基于昇腾 AI 基础软硬件平台，打造危化品存储室智能安全管控解决方案，帮助危化品存储操作实现智慧安全分析。

1) 润和软件打造的危化品存储室智能安全管控解决方案借助于昇思 MindSpore AI 框架快速训练视觉算法模型，集成华为昇腾强大 AI 算力与润和智慧视觉分析引擎，快速落地作业安全规范的智能化。对防护要求、作业流程等纸质化规范进行智能化改造，预防此期间可能发生的违规操作，为作业人员的安全提供坚实保障。

2) 智能安全管控解决方案，在规定的监管区域内，进行人员检测，并检查作业人员装备规范性，包括作业服、呼吸器、氧气面罩等，同时分析穿戴序列，对作业时间进行监管。目前，该方案已经实现规模化应用，在某大型石化企业上线后，迅速将其月均违规作业次数降为零。

图34：润和软件的智能视觉分析引擎



资料来源：润和软件官方公众号，民生证券研究院

4.7 四川长虹：高性能服务器领军企业

四川长虹的子公司四川虹信是中国领先的创新型软件及 IT 解决方案提供商和服务商，经过十多年发展和成长，已经发展成为拥有四大核心能力的高新技术企业，专注于智慧企业、智慧城市等业务领域前沿 IT 技术研发、咨询与实施。目前活跃在国内的能源化工、电子高科技、芯片（IC）、快速消费品、汽车配件、装备制造、教育、农业等各个行业，具备端到端“软件+硬件+咨询+服务”综合业务能力和强大的纵深服务优势，在数十个重要行业服务过世界 500 强等数百家知名企业，为各领域客户创造价值。

根据四川长虹的子公司四川虹信官网，四川虹信打造基于昇腾的天宫系列 AI 服务器，包括 AT800（Model 3000）等。

表11：四川虹信软件 AI 服务器产品梳理

产品型号	产品形态	处理器	AI 加速卡	AI 算力
AT800 (Modle3000)	2UAI 服务器	支持 2 颗 32/48/64 核 CPU 主频支持 2.2GHz/2.6GHz/3.0GHz	最大支持 8 块推理卡或 2 块训练卡	推理卡算力：最大支持 560 TFLOPS FP16(单颗 AI 处理器算力 70 TFLOPS FP16) 训练卡算力：最大支持 560T/440 TFLOPS FP16 (单颗 AI 处理器算力 280T/220 TFLOPS FP16，可选)
AT800 (Modle3010)	2UAI 服务器	1/2 个 IntelXeon SP Skylake 或 Cascade Lake 处理器，最高 205W	最大支持 7 个 Atlas， 300I/VPro 或 2 个 Atlas300T	最大 980TOPSINT8
AT800 (Modle9010)	4UAI 服务器	2 个 IntelV5 CascadeLake 处理器	8 个昇腾 910	2.56/2.24/2/ 1.76PFLOPSFP16
AT800 (Modle9000)	4UAI 服务器	4 个鲲鹏 920	8 个昇腾 910	2.24/1.76PFLOPSFP16(单颗 AI 处理器算力 280T/220TFLOPSFP16，可选)

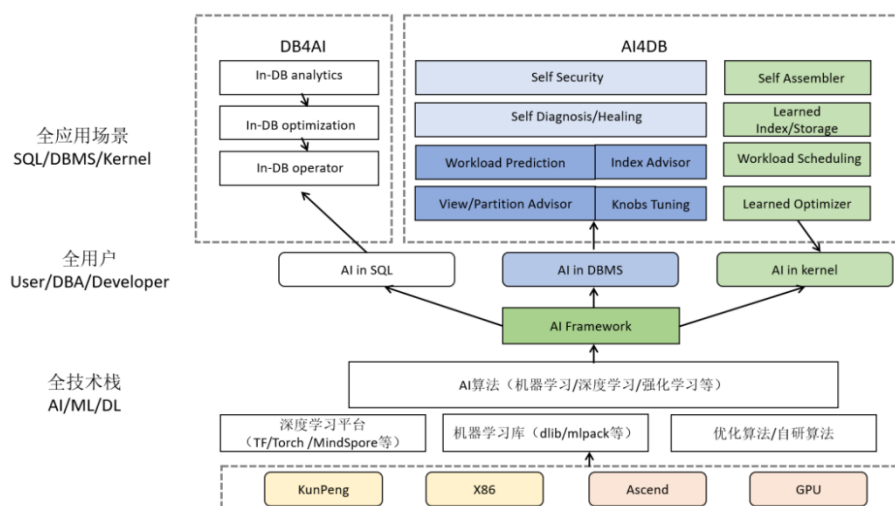
AT500PRO (Modle3000)	2UAI 服 务器	1 个鲲鹏 920	最大支持 3 个 Atlas, 300I/VPro 推理卡或 4 个 Atlas300I	最大 420TOPSINT8
-------------------------	--------------	-----------	--	----------------

资料来源：四川虹信软件官网，民生证券研究院

4.8 海量数据：Vastbase G100 有望打开新的成长空间

Vastbase G100 数据库作为海量数据的核心产品，引入了人工智能（AI）特性，通过 AI 技术使得用户操作和管理数据库变得更加简单高效，实现了 Vastbase G100 数据库自运维、自管理和自优化的能力。AI 在 Vastbase G100 中的应用，主体上分为两部分：AI4DB 和 DB4AI。1) AI4DB，即通过 AI 技术对数据库进行优化。例如，参数调优、索引推荐和分区键推荐等都可以通过 AI 技术来进行优化。2) DB4AI，即数据库提供 AI 的能力，内置一些 AI 算子，用户直接通过 SQL 语法实现 AI 模型训练和推理。例如，把内置的 AI 算子和数据库内普通的算子（JOIN 算子等）进行关联查询，提升查询速度。

图35：Vastbase G100 数据库的 AI 框架全景图



资料来源：海量数据官方公众号，民生证券研究院

其中，DB4AI 在数据库内提供 AI 计算能力，通过 SQL 降低 AI 使用门槛，可以快速推广 AI 的使用场景，基于 AI 芯片降低 AI 开发成本，实现训练推理验证加速一体化。在关键技术方面，孵化基于 CPU、GPU、昇腾芯片的异构计算框架，实现基于异构芯片的数据统一封装与最优化调度；基于 GPU、昇腾芯片实现数据库重负载算子 Agg、Join 等的加速；结合 Tensorflow/MindSpore 机器学习框架，在数据库内部实现 CNN、DNN 等神经网络算法，并探索基于昇腾芯片的对接和加速。

风险提示：数据库领域竞争加剧；技术路线变革具有不确定性。

5 投资建议

AI 发展下算力产业发展迎来重大机遇, 是 AI 大潮下确定性最高的细分领域之一; 华为昇腾打造性能领先的算力处理器, 同时打造软硬产品开发和合作伙伴良好生态, 未来有望成为国产 AI 算力的“扛旗者”。建议重点关注领军企业科大讯飞、恒为科技、拓维信息、神州数码、软通动力、润和软件、四川长虹、海量数据等。

6 风险提示

1) 底层算力市场格局不确定性风险：目前底层算力领域存在一定数量的竞争者，未来随着 AI 发展，可能会有更多参与者参与，进而给市场发展带来一定不确定性。

2) 行业竞争加剧风险：目前国产软硬件尚未呈现出清晰的格局，若后续行业竞争加剧，可能会影响相关公司的毛利率水平，进而影响相关公司的盈利能力。

3) 技术路线变革：目前 AI 行业尚处于发展初期，很多技术路线并未定型，后续可能由于技术路线改变带来行业格局端的变化。

插图目录

图 1: AI 计算的需求上升.....	3
图 2: 中国智能算力规模 (EFLOPS) 及预测.....	4
图 3: 中国通用算力规模 (EFLOPS) 及预测.....	4
图 4: 全球 AI 芯片市场规模 (亿美元)	5
图 5: 国内 AI 芯片市场规模 (亿美元)	5
图 6: 昇腾软硬产品.....	7
图 7: 昇腾 310 关键特性	8
图 8: 昇腾 910 关键特性	8
图 9: Atlas 200I DK A2.....	11
图 10: CANN AI 异构计算架构.....	12
图 11: 昇思 MindSpore 总体架构	13
图 12: 支持全场景统一推理, 大幅提高部署效率和推理性能.....	13
图 13: 融合编程新范式优势突出, 实现易用性和性能的完美结合	14
图 14: 中国开发者主流人工智能框架使用率排名	15
图 15: 昇思 MindSpore 合作伙伴	15
图 16: MindStudio 开发流程	16
图 17: 昇腾相关的开发者计划.....	16
图 18: 昇腾开发者社区三个中心.....	17
图 19: 整机硬件伙伴	18
图 20: IHV 硬件伙伴	19
图 21: 应用软件伙伴	19
图 22: 昇腾计算产业生态.....	20
图 23: 昇腾万里伙伴计划.....	21
图 24: AI 计算中心的“四平台”发展模式.....	22
图 25: AI 计算中心解决方案.....	23
图 26: 金融行业三大解决方案.....	24
图 27: 新冠肺炎医疗影像辅助诊断解决方案.....	25
图 28: 科大讯飞联合华为昇腾打造算力底座.....	26
图 29: 科大讯飞联合华为昇腾打造算力底座.....	27
图 30: 恒为“登云”AI 计算平台解决方案.....	28
图 31: 公司兆瀚 AI 推理服务器产品优势	29
图 32: WAIC2022 昇腾人工智能生态大会上神州数码发布的新产品	31
图 33: 软通动力与昇腾结合的相关产品及优势.....	32
图 34: 润和软件的智能视觉分析引擎	33
图 35: Vastbase G100 数据库的 AI 框架全景图	34

表格目录

重点公司盈利预测、估值与评级	1
表 1: 主流芯片特性对比.....	8
表 2: 各类服务器特性.....	9
表 3: 各类标卡特性	10
表 4: 黄河信产相关产品参数	18
表 5: IHV 硬件伙伴相关产品参数	19
表 6: 华为昇腾智算中心情况	23
表 7: 拓维信息兆瀚 AI 服务器梳理.....	29
表 8: 拓维信息盈利预测表	30
表 9: 神州数码鲲鹏 AI 服务器梳理.....	30
表 10: 神州数码盈利预测表	31
表 11: 四川虹信软件 AI 服务器产品梳理	33

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并登记为注册分析师，基于认真审慎的工作态度、专业严谨的研究方法与分析逻辑得出研究结论，独立、客观地出具本报告，并对本报告的内容和观点负责。本报告清晰地反映了研究人员的研究观点，结论不受任何第三方的授意、影响，研究人员不曾因、不因、也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

评级说明

投资建议评级标准		评级	说明
以报告发布日后的 12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。	公司评级	推荐	相对基准指数涨幅 15%以上
		谨慎推荐	相对基准指数涨幅 5% ~ 15%之间
		中性	相对基准指数涨幅-5% ~ 5%之间
		回避	相对基准指数跌幅 5%以上
	行业评级	推荐	相对基准指数涨幅 5%以上
		中性	相对基准指数涨幅-5% ~ 5%之间
		回避	相对基准指数跌幅 5%以上

免责声明

民生证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。

本公司境内客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅为参考之用，并不构成对客户的投资建议，不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，客户应当充分考虑自身特定状况，不应单纯依靠本报告所载的内容而取代个人的独立判断。在任何情况下，本公司不对任何人因使用本报告中的任何内容而导致的任何可能的损失负任何责任。

本报告是基于已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及预测仅反映本公司于发布本报告当日的判断，且预测方法及结果存在一定程度局限性。在不同时期，本公司可发出与本报告所刊载的意见、预测不一致的报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问、咨询服务等相关服务，本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

若本公司以外的金融机构发送本报告，则由该金融机构独自为此发送行为负责。该机构的客户应联系该机构以交易本报告提及的证券或要求获悉更详细的信息。本报告不构成本公司向发送本报告金融机构之客户提供的投资建议。本公司不会因任何机构或个人从其他机构获得本报告而将其视为本公司客户。

本报告的版权仅归本公司所有，未经书面许可，任何机构或个人不得以任何形式、任何目的进行翻版、转载、发表、篡改或引用。所有在本报告中使用的商标、服务标识及标记，除非另有说明，均为本公司的商标、服务标识及标记。本公司版权所有并保留一切权利。

民生证券研究院：

上海：上海市浦东新区浦明路 8 号财富金融广场 1 幢 5F； 200120

北京：北京市东城区建国门内大街 28 号民生金融中心 A 座 18 层； 100005

深圳：广东省深圳市福田区益田路 6001 号太平金融大厦 32 层 05 单元； 518026