

计算机

报告日期：2023年09月23日

# 华为昇腾构筑世界 AI 算力第二选择

## ——算力行业双周报第一期（2023.9.11-2023.9.24）

### 投资要点

#### □ 个股表现：近两周四川长虹上涨 18.92%

前两周涨幅前五的算力标的是：四川长虹（+18.92%）、同方股份（+12.63%）、浪潮信息（+7.79%）、曙光数创（+6.92%）、中兴通讯（+5.81%）；算力前期调整较多，近两周估值具备优势，华为产业链受近期事件催化表现较好，英伟达产业链受上游供给及下游应用影响，表现亦为可观。

#### □ 行业专题：华为推出全面智能化战略，昇腾构筑世界第二选择

1) **大模型的应用对算力基础设施的规模提出了更高的要求，算力基础设施急需迭代升级。**根据华为发布的《加速行业智能化》白皮书，万亿参数一般需要万卡规模，算力不足意味着无法处理庞大的模型和数据量，也无法有效支撑高质量的大模型技术创新，未来 AI 算力将呈现总部-区域-边缘-终端四层架构，企业对基础设施的需求呈算力集群化、算力系列化等特征，模型参数越大，越需要单一物理空间的大规模训练集群；

2) **华为提出行业智能化参考架构，昇腾 AI 算力赋能产业生态。**华为在全连接大会上提出了 All Intelligence 战略，推出了全行业通用的行业智能化参考架构，发布了可支持超万亿参数大模型训练的全新一代计算集群 Atlas 900 SuperCluster 和兼容业界主流的 AI 框架、加速库和大模型的 CANN 7.0 架构，未来华为有望成为世界算力第二选择；

3) **智算中心需求持续释放，昇腾服务器迎来千亿市场空间。**需求端，12 个省市在最新公布的政策文件中明确提及对 AI 芯片及算力的规划支持，智算中心相关规划陆续出炉，建设节奏有望加快；供应端，全国处于建成/在建状态的智算中心共有 40 个，昇腾算力集群已在华为云、东数西算的枢纽节点贵州和内蒙、中国 28 个城市的 AI 智算中心大规模商用部署，华为生态市占率领先，我们预计未来昇腾市场空间超 950 亿；

#### □ 要点新闻

1) **微软将在消费端、企业端系统及应用中深度嵌入 Copilot 功能。**Windows 11 将在 9 月 26 日的更新增加 Copilot 功能，Copilot 将集成在整个 Windows 操作系统中，可以控制电脑设置、启动应用程序或者提问问题等，能使系统自带软件得到 AI 加持；此外 Microsoft 365 Copilot 企业版将于 11 月 1 日全面上市，AI 将进入 Word、Excel、PowerPoint、Outlook 和 Teams 等协同工具中；

2) **ChatGPT 升级多模态能力，可通过文本生成图片。**9 月 21 日，OpenAI 在官网宣布，在今年 10 月份将通过 API 向 ChatGPT Plus 和企业版用户提供全新文本生成图片产品——DALL·E 3，ChatGPT 将开启多模态输出模式，用户通过文本就能直接在 ChatGPT 中生成各种类型图片。

#### □ 重点关注

- 1) 华为链：神州数码、烽火通信、拓维信息、四川长虹、中国长城、卓易信息；
- 2) 英伟达链：浪潮信息、紫光股份、中兴通讯、工业富联、华勤技术；
- 3) 中科链：海光信息、中科曙光；

#### □ 风险提示

上游关键部件供给能力及节奏不及预期的风险、下游需求释放节奏不及预期的风险、AI 应用推广速度不及预期的风险、国际局势风险、行业测算偏差风险、行业竞争加剧风险等；

### 行业评级：看好(维持)

分析师：刘雯蜀

执业证书号：s1230523020002

liuwenshu03@stocke.com.cn

分析师：李佩京

执业证书号：S1230522060001

lipeijing@stocke.com.cn

### 相关报告

- 1 《数据资产评估指导意见颁布，催化数据要素细节逐步完善》 2023.09.17
- 2 《数据资产评估指导意见颁布，推动数据资产定价落地》 2023.09.11
- 3 《信创框架报告》 2023.09.09

## 正文目录

<b>1 市场表现</b> .....	<b>4</b>
1.1 个股表现: 9月11日至9月22日四川长虹上涨18.92%.....	4
<b>2 行业专题: 华为推出全面智能化战略, 昇腾构筑世界第二选择</b> .....	<b>4</b>
2.1 大模型应用亟需 AI 算力基础设施及架构升级 .....	4
2.2 华为提出行业智能化参考架构, 昇腾 AI 算力赋能产业生态.....	6
2.3 智算中心需求持续释放, 昇腾服务器迎来千亿市场空间.....	8
<b>3 政策要闻</b> .....	<b>11</b>
<b>4 行业新闻</b> .....	<b>12</b>
4.1 国际新闻 .....	12
4.2 国内新闻 .....	12
<b>5 风险提示</b> .....	<b>13</b>

## 图表目录

图 1: AI 算力基础设施两级训练三级推理的部署范式.....	5
图 2: 华为行业智能化参考架构.....	6
图 3: 华为昇腾 AI 产业生态 .....	6
图 4: Atlas 900 SuperCluster .....	8
图 5: 华为多种算力供给模式.....	8
图 6: 华为盘古大模型 3.0 系列.....	10
图 7: 基础大模型架构.....	10
表 1: 算力行业重点个股市场表现回顾 (2023.9.11-9.22) .....	4
表 2: 华为昇腾 AI 基础硬件产品 .....	7
表 3: 今年各地 AI 芯片相关政策 .....	8
表 4: 各地智算中心统计.....	9
表 5: 各地智算中心算力需求量.....	10
表 6: 大模型厂商训练算力需求量.....	11

## 1 市场表现

### 1.1 个股表现：9月11日至9月22日四川长虹上涨18.92%

前两周涨幅前五的算力标的是：四川长虹(+18.92%)、同方股份(+12.63%)、浪潮信息(+7.79%)、曙光数创(+6.92%)、中兴通讯(+5.81%)；年初至今涨幅前五的算力标的是：鸿博股份(+473.98%)、曙光数创(+242.84%)、拓维信息(+186.43%)、中贝通信(+162.55%)、寒武纪-U(+139.17%)，算力前期调整较多，近两周估值具备优势，华为产业链受近期事件催化表现较好，英伟达产业链受上游供给及下游应用影响，表现亦为可观。

表1：算力行业重点个股市场表现回顾（2023.9.11-9.22）

涨幅前十个股				跌幅前十个股			
双周度涨幅		年度涨幅		双周度跌幅		年度跌幅	
四川长虹	18.92%	鸿博股份	473.98%	安路科技	-12.14%	安路科技	-34.03%
同方股份	12.63%	曙光数创	242.84%	复旦微电	-11.30%	复旦微电	-33.16%
浪潮信息	7.79%	拓维信息	186.43%	卓易信息	-10.67%	紫光国微	-32.76%
曙光数创	6.92%	中贝通信	162.55%	景嘉微	-7.87%	澜起科技	-17.66%
中兴通讯	5.81%	寒武纪-U	139.17%	寒武纪-U	-7.11%	华勤技术	-3.81%
工业富联	4.51%	工业富联	122.61%	神州数码	-5.97%	龙芯中科	3.19%
烽火通信	4.31%	青云科技-U	94.16%	紫光国微	-5.93%	中国长城	4.79%
拓维信息	4.28%	利通电子	92.05%	数据港	-5.87%	东华软件	17.10%
鸿博股份	3.00%	四川长虹	85.15%	海光信息	-5.21%	光环新网	20.34%
华勤技术	1.66%	同方股份	80.05%	首都在线	-4.30%	数据港	23.16%

资料来源：Wind，浙商证券研究所

## 2 行业专题：华为推出全面智能化战略，昇腾构筑世界第二选择

### 2.1 大模型应用亟需 AI 算力基础设施及架构升级

深度融合知识和数据的第三代 AI 有望成为数字经济未来发展的新灯塔和新航道。AI 的发展历经知识+算法+算力、数据+算法+算力两代的发展，已逐步迈向知识+数据+算法+算力融合的新阶段，以大模型为代表的第三代 AI 通过文本的语义向量表示和转换器的多头注意机制，能够对文本的语义，即文本中所包含的知识进行自监督学习，从而在自然语言层面向通用 AI 迈出了坚实一步，第三代 AI 将聚焦政府和企业，通过智能化提升客户的经营效率，也将催生新的业态。

大模型的应用对算力基础设施的规模提出了更高的要求，算力基础设施急需迭代升级。根据华为发布的《加速行业智能化》白皮书，端到端大模型的理论训练时间为  $8 * T * P / (n * X)$ 。其中 T 为训练数据的 token 数量，P 为模型参数量，n 为 AI 硬件卡数，X 为每块卡的有效算力。以 ChatGPT 为例，参数量为 175B（1750 亿）规模下，在预训练阶段，数据量 35000 亿，使用 8192 张卡，其训练时长为 49 天。同等条件下参数变多，计算量变大，按照业界的经验，能达到可接受的训练时长，需要百亿参数百卡规模，千亿参数千卡规模，万亿参数万卡规模，这对算力资源的规模提出了极高的要求，算力不足意味着无法处理庞大的模型和数据量，也无法有效支撑高质量的大模型技术创新。

**AI 将成为算力的主要应用领域。**《加速行业智能化》白皮书预计，到 2025 年，中国的 AI 算力总量将超过 1800EFlops，占总算力的比重将超过 85%；到 2030 年，全球通用计算算力将达到 3.3ZFLOPS (FP32)，AI 计算算力将超过 105 ZFLOPS(FP16)，增长 500 倍。

**未来 AI 算力将呈现总部-区域-边缘-终端四层架构。**目前企业获取大模型能力通常通过自研、AI 相关合作伙伴、云服务平台订阅三种模式实现，并通过高阶模型构建专属的 L1、L2 等行业低阶模型。算力方面需要两级训练+三级推理的架构，训练端由总部建设大规模 AI 算力、区域构建小规模 AI 算力，推理端由总部大算力提供服务于全公司共享业务的推理能力、区域构建小算力来服务于自身区域的推理需求、边缘侧就近部署推理算法以降低网络时延、边缘侧布局超小规模 AI 算力。

图1: AI 算力基础设施两级训练三级推理的部署范式



资料来源:《加速行业智能化》白皮书, 浙商证券研究所

在上述四层架构下，企业对基础设施的需求呈算力集群化、算力系列化等特征：

**算力集群化：**总部及区域层需要做模型训练和推理资源池，有建设大规模 AI 算力的需求，要求算力基础设施 1) 具备大规模集群能力，并支持弹性扩展，灵活支持不同阶段大模型的算力需求；2) 实现低时延、大带宽、高可靠的网络，训练集群网络的丢包率会极大影响算力效率，千分之一丢包，算力降低 30%，优先卡内通讯、其次单机通讯、尽量减少集群通讯是关键，模型参数越大，越需要单一物理空间的大规模训练集群；3) 大容量、高带宽、高 IOPS、高可靠的数据存储能力。

**算力系列化：**针对自上而下的算力需求，需要建设对应的 AI 算力集群、训练服务器、推理服务器、训推一体机、边缘侧算力，进行不同的搭配以更符合成本效益。

## 2.2 华为提出行业智能化参考架构，昇腾 AI 算力赋能产业生态

华为提出了具备分层开放、体系协同、敏捷高效、安全可信等特征的，全行业通用的行业智能化参考架构。其中智能底座提供大规模 AI 算力、海量存储及并行计算框架，支撑大模型训练，提升训练效率，提供高性能的存算网协同。根据场景需求不同，提供系列化的算力能力。适应不同场景，提供系列化、分层、友好的开放能力。另外，智能底座层还包含品类多样的边缘计算设备，支撑边缘推理和数据分析等业务场景。

图2：华为行业智能化参考架构



资料来源：《加速行业智能化》白皮书，浙商证券研究所









华为昇腾 AI 产业生态包括昇腾 AI 基础软硬件平台，即 Atlas 系列硬件、异构计算架构 CANN、全场景 AI 框架昇思 MindSpore、昇腾应用使能 MindX 以及一站式开发平台 ModelArts 等。基于昇腾 910 系列板卡，华为推出了 AI 训练集群 Atlas900、AI 训练服务器 Atlas800、智能小站 Atlas500、AI 推理与训练卡 Atlas300 和 AI 加速模块 Atlas200，完成了 Atlas 全系列产品布局，支持万亿参数大模型训练，同时覆盖云、边、端全场景。

图3：华为昇腾 AI 产业生态



资料来源：《加速行业智能化》白皮书，浙商证券研究所

表2: 华为昇腾 AI 基础硬件产品

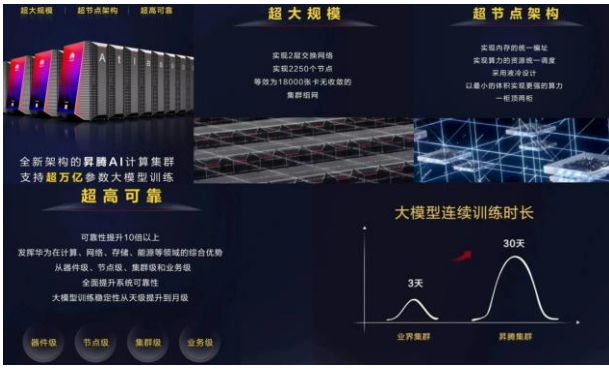
	Atlas 900 AI 集群	Atlas 900 PoD	Atlas 800 训练服务器	Atlas 800 训练服务器	Atlas 800 推理服务器	Atlas 800 推理服务器	Atlas 300T 训练卡	Atlas 300I 推理卡
型号	9000	9000	9000	9010	3000	3010	9000	3000/3010
图示								
形态	-	47U 机柜	4U 服务器	4U 服务器	2U 服务器	2U 服务器	全高, 3/4 长, 双槽位	半高半长 PCIe 卡
CPU	-	32 * 鲲鹏 920	4 * 鲲鹏 920	2 * Intel V5 Cascaded Lake 处理器	2 * 鲲鹏 920	1/2 个 Intel® Xeon® SP Skylake 或 Cascade Lake 处理器, 最高 205W	-	-
AI 处理器	数千颗昇腾 910 AI 处理器互联	64 * 昇腾 910	8 * 昇腾 910	8 * 昇腾 910	最大支持 8 个 Atlas 300I 推理卡	最大支持 7 个 Atlas 300I 推理卡	昇腾 910	昇腾 310
HBM	-	2048 GB	32 GB, 1228GB/s	32GB, 1228GB/s	-	-	-	-
AI 算力	256 ~ 1024 PFLOPS@FP16	14.08 ~ 20.48 PFLOPS @FP16, 最大可扩展至 1 EFLOPS FP16	2.24 PFLOPS FP16 1.76 PFLOPS FP16	2.24 PFLOPS FP16 1.76 PFLOPS FP16	最大 704 TOPS INT8	最大 616 TOPS INT8	内置 30 个达芬奇 AI Core 280 TFLOPS FP16 (Pro) 220 TFLOPS FP16	88 TOPS INT8
网络及接口	集成 HCCS, PCIe 和 100G RoCE 三种高速接口	-	8 * 100GE+ 4 * 25GE/2 * 100GE	8 * 100GE 1 * OCP NIC 3.0 标卡, 支持 2 * 25GE	最多支持 9 个 PCIe4.0 PCIe 接口	10 个 PCIe Gen3.0 接口	1*100GE QSFP-DD 接口, 出口总带宽 56.5 Gb/s	PCIe x16 Gen3.0
功耗	单柜 50KW	单柜 46 kW, 根据采购的设备配置不同, 功耗会有所差异	最大功耗 5.6 kW	最大功耗 5.6 kW	-	-	最大 300W	最大 67W
散热方式	混合液冷	液冷	风冷/液冷	风冷	风冷	风冷	被动风冷	-

资料来源: 华为, 浙商证券研究所

华为发布了全新一代计算集群 Atlas 900 SuperCluster 和 CANN7.0 异构计算架构。SuperCluster 可支持超万亿参数大模型训练, 采用了最新一代的华为星河 AI 智算交换机 CloudEngine XH16800, 该交换机具备高密度的 800GE 端口能力, 使得只需要两层交换网络即可构建一个拥有 2250 个节点的超大规模集群, 相当于 18000 张计算卡的规模; CANN 7.0 是一个更加开放、易于使用的异构计算架构, 不仅与业界主流的 AI 框架、加速库和大模型兼容, 还深度开放底层能力, 使得 AI 框架和加速库可以更加灵活地调用和管理计算资源, 为开发者提供了更多的自定义高性能算子的可能性。

华为可为客户提供多种算力供给模式, 满足行业客户的差异化需求。裸算力模式包括智能感知、智能联接和智能底座, 多租户模式增加 HCS/HCSO 基础云平台, 云算力模式叠加了 ModelArts 一站式 AI 开发平台, MaaS (ModelasaService) 模式面向千行万业的中小企业, 提供开箱即用模型即服务。

图4: Atlas 900 SuperCluster



资料来源: 华为全连接大会, 浙商证券研究所

图5: 华为多种算力供给模式



资料来源:《加速行业智能化》白皮书, 浙商证券研究所

### 2.3 智算中心需求持续释放, 昇腾服务器迎来千亿市场空间

需求端, 各地智算中心相关规划陆续出炉, 建设节奏有望加快。在大模型引发的智能化浪潮趋势下, 江苏、南京、北京、深圳、成都、无锡、四川、上海、福建、重庆、杭州、宁夏等至少 12 个省市在最新公布的政策文件中明确提及对 AI 芯片及算力的规划支持:

北京在 5 月印发了《北京市加快建设具有全球影响力的人工智能创新策源地实施方案(2023-2025 年)》, 明确分别在海淀区、朝阳区建设北京人工智能公共算力中心、北京数字经济算力中心, 在人工智能产业聚集区新建或改建升级一批人工智能商业化算力中心, 加强国产芯片部署应用, 推动自主可控软硬件算力生态建设; 上海在 7 月发布了《上海市推动人工智能大模型创新发展的若干措施》, 提出智能算力加速计划, 强化大模型智能算力建设力度, 建立绿色通道。

表3: 今年各地 AI 芯片相关政策

省市	时间	政策名称
江苏	1 月 19 日	《关于进一步促进集成电路产业高质量发展若干政策的通知》
南京	2 月 23 日	《南京市加快发展新一代人工智能产业行动计划(2023-2025)》
北京	5 月 25 日	《关于征集 2023 年度“中央引导地方”专项人工智能领域储备课题的通知》
	5 月 30 日	《北京市加快建设具有全球影响力的人工智能创新策源地实施方案(2023-2025 年)》、《北京市促进通用人工智能创新发展的若干措施》
	6 月 13 日	《用人工智能创新发展的若干措施》《北京市机器人产业创新发展行动方案(2023-2025 年)》
深圳	5 月 31 日	《深圳市加快推动人工智能高质量发展高水平应用行动方案(2023-2024 年)》
成都	6 月 5 日	《成都市关于进一步促进人工智能产业高质量发展的若干政策措施(征求意见稿)》
无锡	6 月 14 日	《无锡市人工智能产业创新发展三年行动计划(2023-2025)》
四川	7 月 3 日	《关于深入推进新型工业化加快建设现代化产业体系的决定》
	8 月 21 日	《四川省元宇宙产业发展行动计划(2023-2025 年)(征求意见稿)》
上海	7 月 8 日	《上海市推动人工智能大模型创新发展的若干措施》
福建	7 月 12 日	《福建省新型基础设施建设三年行动计划(2023-2025 年)》
重庆	7 月 24 日	《重庆市以场景驱动人工智能产业高质量发展行动计划(2023-2025 年)》
杭州	7 月 27 日	《杭州市人民政府办公厅关于加快推进人工智能产业创新发展的实施意见》
	8 月 16 日	《关于高标准建设“中国视谷”高质量发展视觉智能产业的实施意见》
宁夏	8 月 13 日	《促进人工智能创新发展政策措施》

资料来源: 芯东西, 各地政府官网, 浙商证券研究所

**供应端，华为占据主要份额。**据我们统计，全国处于建成/在建状态的智算中心共有40个，主要由地方政府出资，单次规划算力均在100P以上，根据华为，昇腾算力集群已在华为云、东数西算的枢纽节点贵州和内蒙、中国28个城市的AI智算中心大规模商用部署，华为生态市占率领先。

表4: 各地智算中心统计

序号	省份	智算中心名称	出资主体	规划总算力	规划总投资额(亿)	芯片供应商
1	广东	深圳鹏城云脑	政府			华为昇腾
2	北京	北京昇腾人工智能计算中心	政府	1000P		华为昇腾
3	天津	天津人工智能计算中心	政府	300P	12.7	华为昇腾
4	河北	河北人工智能计算中心	政府		5.9	华为昇腾
5	山东	济南人工智能计算中心	政府	400P	5.47	华为昇腾
6	山东	青岛人工智能计算中心	政府	100P	3.9	华为昇腾
7	江苏	南京鲲鹏·昇腾人工智能计算中心	政府	140P	5.7	华为昇腾
8	江苏	南京智能计算中心	政府	1250P	13	寒武纪
9	江苏	太湖量子智算中心	高校/企业			图灵智算、上海交大
10	上海	腾讯长三角人工智能超算中心	企业		>450	腾讯
11	上海	商汤人工智能计算中心	企业		96	商汤科技
12	浙江	杭州人工智能计算中心	政府	140P	3.3	华为昇腾
13	安徽	淮海智算中心	政府	300P	10	浪潮信息
14	广西	中国-东盟人工智能计算中心	政府			华为昇腾
15	福建	福建人工智能计算中心	政府	400P		福州电信集团
16	广东	深圳人工智能融合赋能中心	政府			华为昇腾
17	广东	广州人工智能公共算力中心	政府	1000P		华为昇腾
18	江苏	昆山智能计算中心	政府			寒武纪
19	河北	阿里云张北超级智算中心	企业	12000P		阿里云
20	浙江	浙江省青田县元宇宙智算中心	政府	100P		浪潮信息、谷梵科技
21	上海	上海有孚临港云计算数据中心	政府		10	
22	上海	阿里云华东智算中心	企业		180	阿里
23	山西	山西先进计算中心	政府			寒武纪
24	山西	百度阳泉智算中心	企业	4000	61	百度
25	河南	中原人工智能计算中心	政府	300P	15	华为昇腾
26	湖南	湖南人工智能算力数据中心	政府	>2000P	50.5	寒武纪
27	湖南	长沙昇腾人工智能创新中心	政府	1000P	180	华为昇腾、寒武纪
28	湖北	武汉人工智能计算中心	政府	400P	10	华为昇腾
29	广东	横琴先进智能计算中心	企业		18	寒武纪
30	安徽	合肥先进计算中心	政府			
31	安徽	合肥人工智能计算中心	政府			华海智汇
32	安徽	合肥智算中心	企业		8.5	中贝通信
33	四川	成都人工智能计算中心	政府	1000P	109	华为昇腾
34	四川	新川人工智能创新中心	政府			
35	陕西	未来人工智能计算中心	政府		19	华为昇腾
36	重庆	重庆人工智能创新中心			11.8	华为昇腾
37	甘肃	甘肃庆阳智算中心	政府		1.97	
38	辽宁	大连人工智能计算中心	政府			华为昇腾
39	黑龙江	哈尔滨人工智能先进计算中心	政府	5.5P	4.3	华为昇腾
40	辽宁	沈阳人工智能计算中心	政府	300P	10.6	华为昇腾

资料来源: 置顶智库, 各中心官网等, 浙商证券研究所

昇腾的主要下游客户为各地智算中心以及大模型及应用厂商，我们预计市场空间超 950 亿元：

1、地方智算中心算力需求规模为 3.5 万 P

城市数量：根据第一财经发布的 2023 年《城市商业魅力排行榜》，一线/新一线/二线/三线/四线/五线城市数量分别为 4/15/30/70/90/128 个；

单城市算力规模：根据我们统计的各地智算中心情况，较发达地区或城市算力规模相对较大，因此假设一线/新一线/二线/三线/四线/五线城市算力需求分别为 1500/500/300/100/50/10P；

表5：各地智算中心算力需求量

城市级别	城市数量	单城市算力需求 (P)	本级城市合计算力需求 (P)	合计所需算力 (P)
一线城市	4	1500	6000	35280
新一线城市	15	500	7500	
二线城市	30	300	9000	
三线城市	70	100	7000	
四线城市	90	50	4500	
五线城市	128	10	1280	

资料来源：华为，云头条，浙商证券研究所

2、行业大模型训练需求规模为 2 万 P

根据华为，7 月发布的盘古大模型 3.0 是一个面向行业的大模型系列，包括 100 亿参数、380 亿参数、710 亿参数和 1000 亿参数四个版本，预训练使用了超 3 万亿 tokens，架构分为三层：

L0：基础大模型，包括自然语言、视觉、多模态、预测、科学计算；

L1：N 个行业大模型，如政务、金融、制造、矿山、气象等；

L2：更细化场景的模型，提供“开箱即用”的模型服务，可专注于政务热线、网点助手、先导药物筛选、传送带异物检测、台风路径预测等具体行业应用或特定业务场景；

图6：华为盘古大模型 3.0 系列



资料来源：华为，量子位，浙商证券研究所

图7：基础大模型架构



资料来源：华为，量子位，浙商证券研究所

考虑到盘古大模型 3.0 系列的特点，我们假设各行业模型累计参数量将达到 2 万亿，tokens 为 3 万亿，单次训练需要 20 天，则对应训练算力需求

$$= \text{大模型数量} * \text{平均参数数量} * 6 * \text{训练 Tokens 数量} / \text{单次训练所需时间}$$

$$= 1 * 20000 * 6 / 10000 * 3 * 10^{12} / (20 * 24 * 3600) / 1000 = 20833.33P$$

表6: 大模型厂商训练算力需求量

参数	数据
假设大模型数量	1
平均参数数量 (亿个, N)	20000
训练 Tokens 数量 (亿个)	30000
单个模型单 Token 训练所需运算次数 (TFLOPS, 6N)	6
单模型所需算力 (PFLOPS)	$1.8 * 10^{10}$
假设单次训练所需时间 (天)	20
训练算力需求 (PFLOPS)	20,833.33

资料来源: 华为, 浙商证券研究所

### 3、按算力需求换算，昇腾的市场空间超 950 亿元

根据长沙麓谷城市发展建设有限公司发布的《长沙人工智能创新中心项目模块一项目》招标公告，100P 的昇腾算力总预算 4 亿元，其中 AI 算力硬件系统采用 7 套湘江鲲鹏“兆瀚”CA9900AI 集群基础单元，单套机柜金额为 2400 万元，合计金额 1.69 亿，因此我们假设 100P 的算力对应机柜投资金额为 1.70 亿，单套机柜可提供 14.29P 的算力，单价为 0.24 亿元，由此我们可得地方智算中心市场规模 599.76 亿元，行业大模型训练市场规模 354.17 亿元，合计市场规模 953.93 亿元。

## 3 政策要闻

1、9月5日，工信部、财政部联合印发《电子信息制造业 2023-2024 年稳增长行动方案》并指出，推动先进计算产业发展和行业应用，开展先进计算在工业、城市管理等领域应用案例征集和应用对接，举办先进计算技术创新大赛等活动，加快先进技术和产品落地应用。鼓励加大数据基础设施和人工智能基础设施建设，满足人工智能、大模型应用需求。(来自证券日报网)

2、9月6日，山东省工业和信息化厅牵头起草了《山东省加快元宇宙产业创新发展的指导意见》并指出，到 2025 年，元宇宙相关产业规模年均增长 15% 左右，达到 1500 亿元左右，重点企业累计申请国际、国内专利 3000 项以上，引进培育 100 家以上元宇宙生态链特色优势企业，支持建设 30 个以上元宇宙公共应用体验中心，打造 100 个元宇宙特色应用场景。(来自 PANews)

3、9月8日，工信部等五部门对外发布《元宇宙产业创新发展三年行动计划（2023~2025 年）》，提出到 2025 年，元宇宙技术、产业、应用、治理等取得突破，成为数字经济重要增长极，产业规模壮大、布局合理、技术体系完善，产业技术基础支撑能力进一步夯实，综合实力达到世界先进水平；培育 3~5 家有全球影响力的生态型企业和一批专精特新中小企业，打造 3~5 个产业发展集聚区；工业元宇宙发展初见成效，打造一批典型应用，形成一批标杆产线、工厂、园区；元宇宙典型软硬件产品实现规模应用，在生活消费和公共服务等领域形成一批新业务、新模式、新业态；(来自北京日报)

4、9月15日,《中关村科学城通用人工智能创新引领发展实施方案》发布,提出要在2025年建成具有全球影响力的人工智能创新区、产业集群区,打造2300亿元核心产业规模,集聚100家大模型企业机构,以及60家国家级专精特新小巨人企业,以及新培育5-10家独角兽企业;(来自智东西)

## 4 行业新闻

### 4.1 国际新闻

#### 1、微软将在消费端、企业端系统及应用中深度嵌入 Copilot 功能

微软 Windows 11 将在 9 月 26 日的更新增加 Copilot 功能, Copilot 将集成在整个 Windows 操作系统中,可以控制电脑设置、启动应用程序或者提问问题等,能使系统自带软件得到 AI 加持;此外 Microsoft 365 Copilot 企业版将于 11 月 1 日全面上市, AI 将进入 Word、Excel、PowerPoint、Outlook 和 Teams 等协同工具中。(来自快科技)

#### 2、ChatGPT 升级多模态能力,可通过文本生成图片

9 月 21 日凌晨, OpenAI 在官网宣布,在今年 10 月份将通过 API 向 ChatGPT Plus 和企业版用户提供全新文本生成图片产品——DALL·E 3, ChatGPT 将开启多模态输出模式,用户通过文本就能直接在 ChatGPT 中生成各种类型图片;

2021 年, OpenAI 基于 GPT-3 大语言模型和变分自编码器 (VAE), 开发了文本生成图片产品 DALL·E, 使用方法和功能与 Midjourney、百度文心一格等产品类似, 2022 年 4 月, OpenAI 又发布了 DALL·E 2, 在功能、资源消耗方面进行了大幅度优化, DALL·E 3, 将比前两代功能更加强大并且可集成在 ChatGPT 中使用, 使其实现多模态输出。(来自 AIGC 开放社区)

#### 3、亚马逊发布新款 Alexa 语音助手 接入大模型

9 月 21 日, 亚马逊推出新款 Alexa 智能语音助手, 接入为语音交互定制的大模型, 可提供更自然的对话功能和智能家居控制, 预计在未来三个月陆续定向测试。(来自 Techweb、站长之家)

### 4.2 国内新闻

#### 1、总规划 4000P 的北京 AI 公共算力平台开建, 明年 Q1 建成

9 月 15 日, 北京人工智能公共算力平台正式在中关村科学城北区投入建设, 由北京电信和京能集团共同承建, 将于今年内分别上线 500P、合计 1000P 公共算力, 力争明年一季度建成总计 4000P 公共算力。(来自智东西)

#### 2、中贝通信与济南超算签署 AI 算力服务合同

9 月 18 日中贝通信(603220)公告称, 公司与济南超级计算中心有限公司(下称超算公司)签订《AI 算力服务合同》及其附件《算力服务清单》, 鉴于中贝通信拥有 AI 算力服务能力, 超算公司向中贝通信采购智能计算中心的 AI 算力服务。双方约定, 中贝通信提供包括算力服务、带宽、存储在内的相关服务(含 A800 算力服务器 60 台), 服务周期 60 个月, 合同总金额为 1.8 亿元。算力支撑方面, 支持最高等级 GPU 集群, 大幅度缩短计算时间, 提高效率, 适用于人工智能、生命科学、高校科研、工程仿真、药物研发等场景; 数据安全方面, 用户环境物理隔离、相互独立, 充分保护用户隐私, 同时加以防火墙等安全措施。(来自 wind、证券时报)

### 3、华为全连接大会开幕，华为提出 All Intelligence 战略，目标加速千行万业的智能化转型，为世界构建算力第二选择

9月20日，华为副董事长、轮值董事长、CFO孟晚舟在大会上发表了“打造中国坚实的算力底座，为世界构建第二选择”的主题演讲。她表示，“华为将持续打造坚实的算力底座，使能百模千态，赋能千行万业...在此进程中，通过算力底座、AI平台、开发工具的开放，支持主流大模型在智能化时代的‘百花齐放’，华为努力做好‘百花园’的黑土地。”

她进一步指出，华为将持续提升通用大模型的能力，做好工程化工具，让客户和伙伴可以更便捷、更高效地调用模型，“我们支持每个组织使用自己的数据训练出自己的大模型，让每个行业用自己的专业知识发展出自己的行业大模型。我们的客户和伙伴已经能够基于盘古大模型，打造出了自己的政务、金融、矿山、电力、铁路等行业大模型。”

9月20日，华为云正式面向全球企业和开发者上线昇腾 AI 云服务“百模千态”专区，华为云昇腾 AI 云服务可提供多元、高效、长稳的算力服务选择，千卡训练 30 天长稳率 90%，训练作业故障自动恢复，非自动场景下恢复时长小于 30 分钟。华为云昇腾 AI 云服务“百模千态”专区收录了业界主流开源大模型，并全面基于昇腾 AI 云服务进行适配和优化，精度和性能显著提升；提供丰富的应用开发的工具链，开发工具全部云化，免去繁琐的配置流程，实现一键接入。此外，专区提供丰富的使用指导，企业和个人开发者可自助申请体验，审核通过后可通过点击相应的开源大模型进入控制台进行部署，并发布成在线服务进行推理调用。（来自华为官网、智东西）

## 5 风险提示

**上游关键部件供给能力及节奏不及预期的风险：**产业链仍面临供给不足导致算力不足的风险，可能会影响 AI 应用推广的进程；

**下游需求释放节奏不及预期的风险：**算力需求受下游应用带来的训练和推理需求直接影响，接受亦受影响；

**AI 应用推广速度不及预期的风险：**目前 AI 应用处于初期阶段，需求尚不明确，推广速度和节奏影响整个产业链的需求情况；

**国际局势风险：**国际局势影响算力的供需；

**行业测算偏差风险：**行业市场空间测算基于一定前提假设条件，存在实际达不到、不及预期的风险；

**行业竞争加剧风险等；**

## 股票投资评级说明

以报告日后的6个月内，证券相对于沪深300指数的涨跌幅为标准，定义如下：

1. 买入：相对于沪深300指数表现+20%以上；
2. 增持：相对于沪深300指数表现+10%~+20%；
3. 中性：相对于沪深300指数表现-10%~+10%之间波动；
4. 减持：相对于沪深300指数表现-10%以下。

## 行业的投资评级：

以报告日后的6个月内，行业指数相对于沪深300指数的涨跌幅为标准，定义如下：

1. 看好：行业指数相对于沪深300指数表现+10%以上；
2. 中性：行业指数相对于沪深300指数表现-10%~+10%以上；
3. 看淡：行业指数相对于沪深300指数表现-10%以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

## 法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，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，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

## 浙商证券研究所

上海总部地址：杨高南路729号陆家嘴世纪金融广场1号楼25层

北京地址：北京市东城区朝阳门北大街8号富华大厦E座4层

深圳地址：广东省深圳市福田区广电金融中心33层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：<https://www.stocke.com.cn>