

计算机行业 2024 年年度策略报告

政策、市场、技术多重驱动，看好 AIGC 和信创

强于大市（维持）

行情走势图



证券分析师

闫磊	投资咨询资格编号 S1060517070006 YANLEI511@pingan.com.cn
付强	投资咨询资格编号 S1060520070001 FUQIANG021@pingan.com.cn
黄韦涵	投资咨询资格编号 S1060523070003 HUANGWEIHAN235@pingan.com.cn



平安观点：

- **行业回顾及投资逻辑：**2023 年前三季度，计算机行业上市公司营业总收入合计实现 8113.60 亿元，同比增长 1.66%；归母净利润合计实现 227.68 亿元，同比增长 1.87%；扣非归母净利润合计实现 128.91 亿元，同比下降 28.35%，扣非归母净利润表现不佳。市场表现来看，计算机指数表现相对较好，涨幅跑赢沪深 300 指数，且在 31 个申万一级行业排名靠前。展望 2024 年，在政策、市场、技术等多重因素的驱动下，我们看好 AIGC 产业和信创产业的发展，维持对计算机行业的“强于大市”评级。
- **大模型需要大算力，国产 AI 芯片和服务器厂商迎来发展机遇。**2023 年以来，AIGC 产业引起了全球的广泛关注。我国十分重视和支持 AIGC 产业的发展，今年以来，相关利好政策相继出台，为 AIGC 的后续发展护航。大模型需要大算力，大模型算法的迭代升级将为全球和中国 AI 算力市场的增长提供强劲动力。当地时间 10 月 17 日，美国商务部工业与安全局（BIS）发布了对华半导体出口管制最终规则。受最终规则的影响，美国高端 AI 芯片的对华出口将受到非常严格的限制。在此背景下，我国国产 AI 芯片在国内市场的竞争力将进一步提高，在国内市场的市占率有望加速提升，厂商将加快发展，产业链将加快成熟，相关 AI 芯片和服务器厂商将深度受益。
- **我国国产大模型能力持续升级，AIGC 产业未来发展前景广阔。**随着大模型时代的到来，我国大模型厂商持续迭代升级算法能力，在模型能力上已可对标 ChatGPT。当前，我国国产大模型已经开始在教育、医疗、汽车、办公、工业、智能硬件等 B 端和 C 端应用场景持续落地。大模型赋能智能汽车和人形机器人，将加快高级别自动驾驶和人形机器人的大规模商用。我们认为，随着国产大模型的逐步成熟，我国大模型产品面向我国庞大的互联网 C 端用户群和丰富的行业应用场景，将与产品和应用场景深度融合，赋能我国数字经济的发展。参考我国数字经济的巨大体量，我国 AIGC 产业未来发展前景广阔。
- **信创工程市场空间巨大，华为将成为信创产业发展重要推动力。**我国信创产业的发展将按照 2+8+N 的落地体系逐步展开，党政信创先行，金融、电信等八大行业紧随其后。当前，我国信创基础软硬件产品性能提升，已实现从可用到好用。我们判断，2024 年，党政信创将有望加速，金融、电信等八大行业信创将陆续进入常态化。信创产业的发展，将为我国信创相关厂商带来良好的市场机遇。根据我们的测算，信创产业未来将有数千亿量级的市场空间，市场空间巨大。华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力。信创华为产业链值得重点关注。

- 投资建议：**展望 2024 年，在政策、市场、技术等多重因素的驱动下，我们看好 AIGC 产业和信创产业的发展，看好两条投资主线：（一）当前，全球范围内大模型领域的竞争依然白热化，这将持续拔升大模型的整体能力水平，提升大模型的多模态能力。大模型需要大算力，大模型算法的迭代升级将为全球和中国 AI 算力市场的增长提供强劲动力。在美国对华半导体出口管制升级的背景下，我国国产 AI 芯片产业链将加快成熟，相关 AI 芯片和服务器厂商将深度受益。另外，在应用端，我国国产大模型能力持续升级，AIGC 产业未来发展前景广阔。（二）信创产业的发展。我国信创产业的发展将按照 2+8+N 的落地体系逐步展开，我们判断，2024 年，党政信创将有望加速，金融、电信等八大行业信创将陆续进入常态化。华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力。信创华为产业链值得重点关注。我们看好 AIGC 和信创主题的投资机会，维持对计算机行业的“强于大市”评级。在投资标的方面：AIGC 主题，1) 在算力方面，推荐浪潮信息、中科曙光、紫光股份、海光信息，建议关注工业富联、寒武纪、景嘉微；2) 在算法方面，推荐科大讯飞；2) 在应用场景方面，强烈推荐中科创达，推荐金山办公、恒生电子，建议关注拓尔思、同花顺、彩讯股份、航天宏图、盛视科技；4) 在网络安全方面，强烈推荐启明星辰。信创主题，推荐龙芯中科、太极股份、顶点软件，建议关注神州数码、高新发展、中国软件、麒麟信安、诚迈科技、中国长城、软通动力、海量数据、普联软件、远光软件。
- 风险提示：**1) AI 算力供应链风险上升。2) 大模型产品的应用落地低于预期。3) 信创产业发展不及预期。

股票名称	股票代码	股票价格		EPS			P/E			评级	
		2023-12-14	2022	2023E	2024E	2025E	2022	2023E	2024E		2025E
中科创达	300496.SZ	83.30	1.67	1.68	2.09	2.66	49.8	49.6	39.9	31.3	强烈推荐
启明星辰	002439.SZ	27.70	0.66	1.11	1.39	1.85	41.8	25.0	19.9	15.0	强烈推荐
浪潮信息	000977.SZ	35.79	1.41	1.10	1.49	1.88	25.3	32.5	24.0	19.0	推荐
中科曙光	603019.SH	40.15	1.06	1.35	1.72	2.20	38.1	29.7	23.3	18.3	推荐
紫光股份	000938.SZ	20.56	0.75	0.77	0.95	1.16	27.2	26.7	21.6	17.7	推荐
海光信息	688041.SH	75.60	0.35	0.56	0.74	1.00	218.7	135.0	102.2	75.6	推荐
龙芯中科	688047.SH	119.01	0.13	-0.14	0.38	0.77	922.1	-850.1	313.2	154.6	推荐
科大讯飞	002230.SZ	48.90	0.24	0.36	0.48	0.64	201.8	134.0	101.7	76.9	推荐
金山办公	688111.SH	336.94	2.42	3.05	4.10	5.62	139.2	110.5	82.2	60.0	推荐
恒生电子	600570.SH	29.17	0.57	0.92	1.11	1.32	50.8	31.7	26.3	22.1	推荐
太极股份	002368.SZ	31.25	0.61	0.74	0.96	1.29	51.6	42.2	32.6	24.2	推荐
顶点软件	603383.SH	53.47	0.97	1.38	1.88	2.48	55.1	38.6	28.5	21.6	推荐

资料来源：Wind，平安证券研究所

正文目录

一、 计算机行业投资逻辑	8
二、 行业回顾与展望：行业行情震荡向上，看好 AIGC 和信创产业发展	9
2.1 行业上市公司 2023 年前三季度营收平稳增长，扣非归母净利润表现不佳	9
2.2 行业呈现明显震荡行情，整体仍处于上行态势	9
2.3 行业指数年累计涨幅跑赢沪深 300，在 31 个申万一级行业中排名靠前	10
三、 大模型需要大算力，国产 AI 芯片和服务器厂商迎来发展机遇	11
3.1 全球大模型领域的竞争依然白热化，国内政策加码助推 AIGC 产业发展	11
3.2 大模型的实现需要十分强大的算力来支持训练过程和推理过程	12
3.3 AI 芯片进入舞台中央，广泛应用于训练或推理	13
3.4 AI 服务器借助加速卡获取强大算力，市场需求快速增长	16
3.5 大模型的训练成本和推理成本高昂	18
四、 我国国产大模型能力持续升级，AIGC 产业未来应用潜力巨大	21
4.1 大模型迭代进入快速发展阶段，GPT 大模型成为关注焦点	21
4.2 大模型赋能千行百业，AIGC 未来发展前景广阔	23
4.3 汽车驶向智能化时代，大模型提升自动驾驶泛化能力	28
4.4 大模型为人形机器人注入大脑，打开通用机器人可实现性	32
五、 信创工程市场空间巨大，华为将成为信创产业发展重要推动力	37
5.1 党政信创先行，金融、电信等八大行业紧随其后	37
5.2 我国信创基础软硬件产品性能已实现从可用到好用，信创工程市场空间巨大	39
5.3 华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力	42
六、 投资建议及风险提示	49
6.1 投资建议：看好 AIGC 和信创主题的投资机会	49
6.2 风险提示	49

图表目录

图表 1	计算机行业投资逻辑	8
图表 2	2022 上半年和 2023 年上半年计算机行业上市公司业绩情况	9
图表 3	2022 前三季度和 2023 年前三季度计算机行业上市公司业绩情况	9
图表 4	年初以来计算机行业指数表现	10
图表 5	行业指数相比沪深 300 指数表现	10
图表 7	计算机行业当前估值低于历史中位数水平	11
图表 8	行业市盈率在 31 个申万一级行业排名第 1 位	11
图表 9	今年以来我国 AIGC 发展相关政策（部分）	12
图表 10	2018-2022 年大模型参数增长变化趋势	13
图表 11	各个模型所需计算量及参数量	13
图表 12	三类 AI 芯片简介	13
图表 14	全球独立 GPU 出货量市占率	14
图表 15	英伟达、AMD 主要 GPU 产品性能参数对比	15
图表 16	海光 DCU 产品-深算一号规格特点	15
图表 17	海光 DCU 产品与行业代表可比产品参数对比	15
图表 18	华为海思与谷歌芯片性能对比	16
图表 19	MLU370 系列加速卡规格	16
图表 20	寒武纪代表性 ASIC 产品	16
图表 21	2022-2026 年全球 AI 服务器市场规模（亿美元）	17
图表 22	2022-2027 年中国 AI 服务器市场规模（亿美元）	17
图表 23	2021 年上半年全球 AI 服务器市场竞争格局	17
图表 24	2021 年我国 AI 服务器市场竞争格局	17
图表 25	2022 年 AI 服务器下游采购情况	18
图表 26	2020-2023 年我国云计算支出规模及增速	18
图表 27	自建 IDC 的训练成本估算	18
图表 28	自建 IDC 的推理成本估算	19
图表 29	大模型将为国内带来的 AI 服务器市场增量测算	19
图表 30	英伟达 H20、L20、L2 性能不及部分国产芯片	20
图表 31	AI 大模型发展历程	21
图表 32	Transformer 模型系列分类	22
图表 33	Transformer 模型架构	22
图表 34	Sam Altman 和 OpenAI 推出的 ChatGPT 掀起海内外大模型布局热潮	22
图表 35	2023 年下半年国内外厂商陆续发布大模型最新版本	23

图表 36	AI 大模型按模态分类	23
图表 37	2021-2022 年我国各行业人工智能渗透率 (%)	24
图表 38	2017-2022 年我国数字经济规模	24
图表 39	Microsoft 365 Copilot 生成 PPT	24
图表 40	Microsoft 365 Copilot 全面注入 PC 端	24
图表 41	WPS AI 成为国内率先落地的 AI 办公应用	25
图表 42	万兴科技多项办公类产品推出 AI 功能	25
图表 43	昆仑万维发布 AI Agent 平台天工 SkyAgents	26
图表 44	Blue AI 帮助出海品牌创作营销文案	26
图表 45	ChatDD 给出分子亲和力计算排序结果	27
图表 46	百度发布国内首个产业级医疗大模型“灵医”	27
图表 47	同花顺 i 问财界面	27
图表 48	恒生光子·文曲为面向投顾内容创作的产品	27
图表 49	科大讯飞 AI 学习机推出 1v1 智能编程助手	28
图表 50	好未来学而思学习机进行中文作文批改	28
图表 51	自动驾驶系统模块化结构	29
图表 52	自动驾驶系统端到端结构	29
图表 53	特斯拉自动驾驶算法迭代历程	29
图表 54	BEV+Transformer 架构	30
图表 55	多摄像头下的 BEV 感知	30
图表 56	占用网络对起步中的公交车预测示意图	30
图表 57	乘用车 NOA 正从高速向城市演进	31
图表 58	部分车企城市 NOA 落地计划	31
图表 59	DriveGPT 自动驾驶生成式预训练大模型	31
图表 60	DriveGPT 的训练过程和基本效果	31
图表 61	智能座舱核心产品/功能渗透率 (标配)	32
图表 62	理想 MindGPT 上车	32
图表 63	全球机器人市场规模及增速	32
图表 64	中国机器人市场规模及增速	32
图表 65	全球工业机器人市场规模及增速	33
图表 66	全球服务机器人市场规模及增速	33
图表 67	人形机器人是机器人领域最新的发展趋势	33
图表 68	2022 年特斯拉 Optimus 外形发展历程	34
图表 69	2023 年 12 月新一代 Optimus 做深蹲	34
图表 70	全球具有代表性的人形机器人梳理	34
图表 71	大模型赋能任务级交互：将自然语言指令快速转换为代码	35

图表 72	9月特斯拉 Optimus 展示按颜色分拣	35
图表 73	特斯拉人形机器人系统移植 FSD 智驾系统	35
图表 74	谷歌机器人大模型梳理	36
图表 75	小米 CyberOne 在视觉和情绪感知方面的能力	36
图表 76	小米 CyberOne 展示打架子鼓	36
图表 77	傅利叶 GR-1 具备快速行走等运动能力	37
图表 78	傅利叶 GR-1 支持底层功能 SDK、APP 源代码和基础预训练大模型开发, 以及硬件扩展	37
图表 79	2+8+N 信创发展进程	37
图表 80	各领域信创应用落地进程	37
图表 81	2020-2022 年金融信创试点单位数量 (个)	38
图表 82	中信银行 65 亿信创订单招标细则	38
图表 83	2020-2022 年我国电信行业信创建设进展	38
图表 84	龙芯 3A6000 处理器产品性能参数	39
图表 85	龙芯生态布局	39
图表 86	公司与 Intel 服务器 CPU 的参数对比	40
图表 87	2020-2022 年我国 PC 出货量 (百万台)	40
图表 88	2020-2022 年我国服务器销售量 (百万台)	40
图表 89	信创产业 PC、服务器市场容量预测 (万台)	41
图表 90	信创 PC 市场规模预测 (亿元)	41
图表 91	信创服务器市场规模预测 (亿元)	41
图表 92	华为信创产品体系 (示意图)	42
图表 93	鲲鹏计算产业定义	42
图表 94	鲲鹏“算、存、传、管、智”芯片族	43
图表 95	鲲鹏合作伙伴整机产品	43
图表 96	华鲲振宇 2021 年-2023Q3 营收情况	44
图表 97	神州数码 2020-2022 年自主品牌业务营收情况	44
图表 98	openEuler 平台框架	44
图表 99	国内外 20 多家厂商发布 openEuler 路线的操作系统商业发行版	45
图表 100	2021-2022 年 openEuler 系服务器操作系统装机量市场份额	45
图表 101	openEuler 系服务器操作系统政府领域装机量市场份额	45
图表 102	openEuler 系服务器操作系统电信领域装机量市场份额	46
图表 103	openEuler 系服务器操作系统金融领域装机量市场份额	46
图表 104	2020-2022 年中国 openEuler 系服务器操作系统装机量	46
图表 105	2022 年中国 openEuler 系服务器操作系统装机量结构	46
图表 106	华为数据库发展历程	47
图表 107	华为数据库的公司内部配套、公有云的 GaussDB、开源 openGauss 共享代码基线	47

图表 108	基于 openGauss 开发的商业发行版	48
图表 109	2021 年下半年我国关系型数据库软件市场（公有云模式）厂商份额	49
图表 110	2021 年下半年我国关系型数据库软件市场（本地部署模式）厂商份额	49

一、 计算机行业投资逻辑

全球大模型领域的竞争依然白热化。2022年11月，由OpenAI开发的大模型聊天机器人ChatGPT火爆出圈。2023年3月，OpenAI发布GPT-4，GPT-4可以接受图片作为输入。相比ChatGPT，GPT-4在多模态处理能力、高端推理能力等方面提升明显。2023年12月，谷歌宣布推出其规模最大、功能最强的多模态大模型Gemini，谷歌作为全球人工智能龙头之一，无论在算力、算法还是商业化各个方面都积蓄较大优势，Gemini系列在传统文本领域以及多模态领域能力均对标GPT-4，全球人工智能龙头在大模型领域的竞争依然白热化。

我国高度重视和支持AIGC产业的发展。今年以来，国家陆续出台多项鼓励和支持AIGC产业发展的政策，推动我国AIGC产业的发展。同时，我国大模型技术快速发展，在算力和算法领域均取得积极进展。在算力领域，海光信息、华为等国产高端AI芯片厂商持续迭代新产品，我国国产AI芯片性能大幅提升，以浪潮信息为代表的我国AI服务器厂商竞争力全球领先。在算法领域，BAT、科大讯飞在今年9月-10月陆续推出了其大模型的最新版本，模型能力向GPT-3.5和GPT-4看齐。我国国产大模型已经开始在教育、医疗、汽车、办公、工业、智能硬件等B端和C端应用场景持续落地。

在信创产业发展方面，我国信创产业的发展将按照2+8+N的落地体系逐步展开：党政是信创产业落地的起步领域，2013年即开始在公文系统领域实行国产化替代，金融、电信等八大行业紧随其后，N个行业的信创国产化替代也有望逐步启动。当前，我国信创基础软硬件产品性能提升，已实现从可用到好用。我们判断，2024年，党政信创将有望加速，金融、电信等八大行业信创将陆续进入常态化。

展望2024年，在政策、市场、技术等多重因素的驱动下，我们看好AIGC产业和信创产业的发展，看好两条投资主线：（一）当前，全球范围内大模型领域的竞争依然白热化，这将持续拔升大模型的整体能力水平，提升大模型的多模态能力。大模型需要大算力，大模型算法的迭代升级将为全球和中国AI算力市场的增长提供强劲动力。在美国对华半导体出口管制升级的背景下，我国国产AI芯片产业链将加快成熟，相关AI芯片和服务器厂商将深度受益。另外，在应用端，我国国产大模型能力持续升级，AIGC产业未来发展前景广阔。（二）信创产业的发展。我国信创产业的发展将按照2+8+N的落地体系逐步展开，我们判断，2024年，党政信创将有望加速，金融、电信等八大行业信创将陆续进入常态化。华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力。信创华为产业链值得重点关注。我们看好AIGC和信创主题的投资机会，维持对计算机行业的“强于大市”评级。

图表1 计算机行业投资逻辑



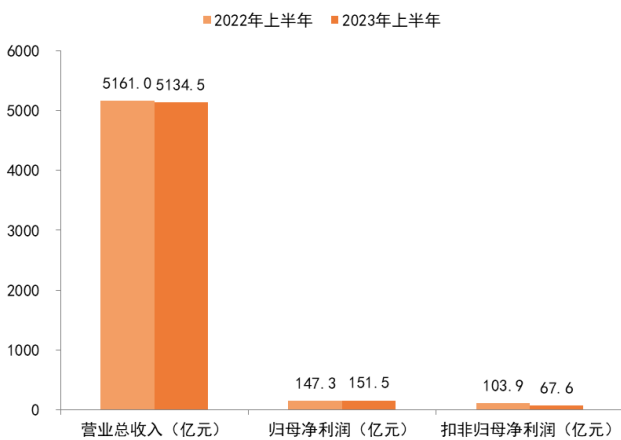
资料来源：平安证券研究所

二、行业回顾与展望：行业行情震荡向上，看好 AIGC 和信创产业发展

2.1 行业上市公司 2023 年前三季度营收平稳增长，扣非归母净利润表现不佳

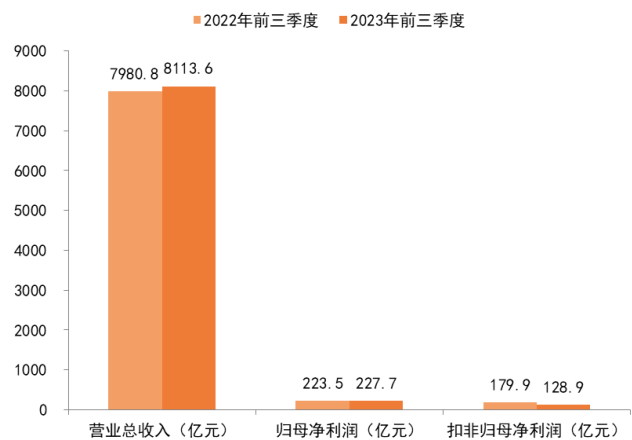
2023 年前三季度，行业上市公司营业总收入合计实现 8113.60 亿元，同比增长 1.66%；归母净利润合计实现 227.68 亿元，同比增长 1.87%；扣非归母净利润合计实现 128.91 亿元，同比下降 28.35%，扣非归母净利润表现不佳。分季度看，2023 年一、二、三季度，行业上市公司营收合计分别同比增长-1.18%、0.04%、5.65%；归母净利润合计分别同比增长 56.08%、-11.45%、-0.04%；扣非归母净利润合计分别实现由盈转亏、同比下降 13.43%、同比下降 17.64%。展望四季度及 2024 年，我们认为，在中央财政 2023 年四季度增发 2023 年国债 1 万亿元等积极因素的推动下，行业上市公司的下游需求将加快增长，行业上市公司的业绩将持续向好。

图表2 2022 上半年和 2023 年上半年计算机行业上市公司业绩情况



资料来源: Wind, 平安证券研究所

图表3 2022 前三季度和 2023 年前三季度计算机行业上市公司业绩情况



资料来源: Wind, 平安证券研究所

2.2 行业呈现明显震荡行情，整体仍处于上行态势

年初以来，受信创、数字中国、数据要素、AIGC 等主题的推动，计算机行业市场关注度较高，行业呈现明显震荡行情，整体仍处于上行态势。1 月至 4 月初，在行业市盈率处于历史低位的背景下，受信创、数字中国、数据要素、AIGC 等主题的推动，行业指数在 4 月初到达年初至今的高点。之后，随着 AIGC 合规风险的上升，叠加行业上市公司整体 2022 年业绩和 2023 年一季度业绩表现不佳的影响，行业指数高位回落。5 月 25 日，全球 AI 芯片龙头公司英伟达发布 2024 财年第一财季财报，其对 2024 财年第二财季营收的乐观预期，加强了市场对于 AI 算力需求增长的确信，提升了市场对于 AI 算力乃至整个 AIGC 主题的信心与关注度。5 月 30-31 日，北京市、上海市、深圳市相继发布政策利好通用人工智能发展，彰显我国对于 AIGC 发展的重视与支持。受益市场对 AIGC 信心的提振，行业行情开启反弹。6 月 20-21 日，行业指数高点回落。此后，叠加半年报业绩表现不佳（主要是扣非归母净利润表现不佳），行业指数持续下降。10 月 24-25 日，在中央财政将在今年四季度增发 2023 年国债 1 万亿元政策利好的推动下，叠加三季报业绩不佳的影响已逐步释放，行业行情再次开启反弹。

图表4 年初以来计算机行业指数表现

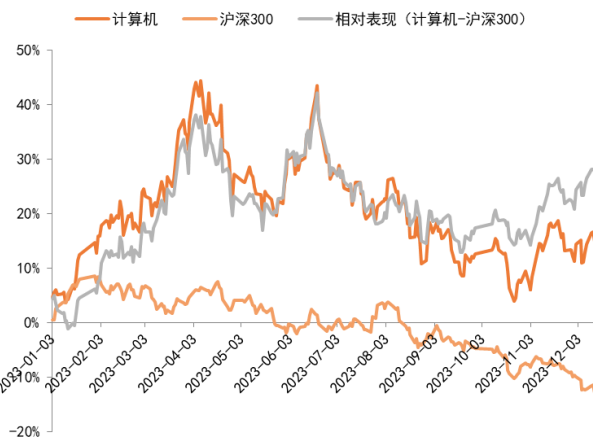


资料来源: Wind, 平安证券研究所

2.3 行业指数年累计涨幅跑赢沪深 300, 在 31 个申万一级行业中排名靠前

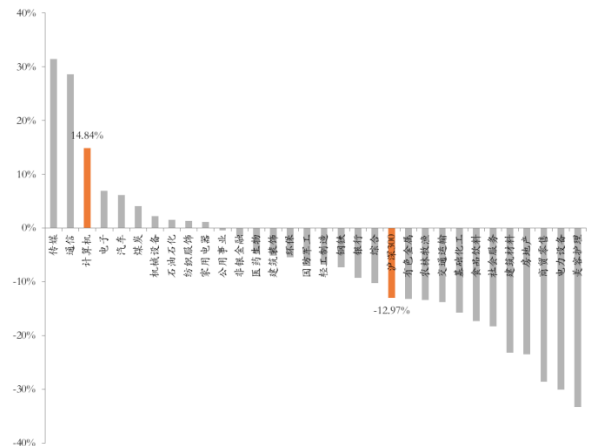
截至到 2023 年 12 月 13 日, 申万计算机指数上涨了 14.84%, 跑赢沪深 300 指数 27.81 个百分点, 在 31 个申万一级行业中排名第 3 位, 排名靠前。

图表5 行业指数相比沪深 300 指数表现



资料来源: Wind, 平安证券研究所

图表6 行业指数年累计涨幅排名靠前



资料来源: Wind, 平安证券研究所

2.4 行业估值处于历史相对较低水平, 看好 AIGC 产业和信创产业的发展

截至 12 月 13 日, 计算机行业估值处于历史相对较低水平。根据我们的统计, 2015 年以来, 申万计算机行业指数历史市盈率 (TTM, 剔除负值) 中位数为 49.7 倍。计算机行业 12 月 13 日市盈率为 48.1 倍, 在历史市盈率中位

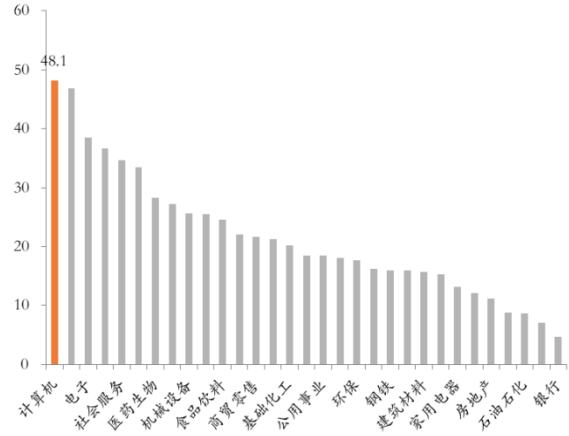
数水平之下。展望 2024 年，在政策、市场、技术等多重因素的驱动下，我们看好 AIGC 产业和信创产业的发展，维持对计算机行业的“强于大市”评级。

图表7 计算机行业当前估值低于历史中位数水平



资料来源: Wind, 平安证券研究所

图表8 行业市盈率在 31 个申万一级行业排名第 1 位



资料来源: Wind, 平安证券研究所

三、大模型需要大算力，国产 AI 芯片和服务器厂商迎来发展机遇

3.1 全球大模型领域的竞争依然白热化，国内政策加码助推 AIGC 产业发展

全球大模型领域的竞争依然白热化。2022 年 11 月，由 OpenAI 开发的大模型聊天机器人 ChatGPT 火爆出圈。2023 年 3 月，OpenAI 发布 GPT-4，GPT-4 可以接受图片作为输入。相比 ChatGPT，GPT-4 在多模态处理能力、高端推理能力等方面提升明显。2023 年 12 月，谷歌宣布推出其规模最大、功能最强的多模态大模型 Gemini，谷歌作为全球人工智能龙头之一，无论在算力、算法还是商业化各个方面都积蓄较大优势，Gemini 系列在传统文本领域以及多模态领域能力均对标 GPT-4，全球人工智能龙头在大模型领域的竞争依然白热化。

我国高度重视和支持 AIGC 产业的发展。今年以来，国家陆续出台多项鼓励和支持 AIGC 产业发展的政策，为我国 AIGC 产业的发展护航。7 月 13 日，网信办等七部门联合发布《生成式人工智能服务管理暂行办法》，为大模型的备案提供了政策依据。自 8 月份以来，百度文心一言、商汤的商量 SenseChat、百川智能的百川大模型等通过了《生成式人工智能服务管理暂行办法》备案，开始面向全社会开放服务。大模型备案工作正式启动。北京市、上海市、深圳市、成都市等陆续发布 AIGC 利好政策，推动我国 AIGC 产业发展。

图表9 今年以来我国 AIGC 发展相关政策（部分）

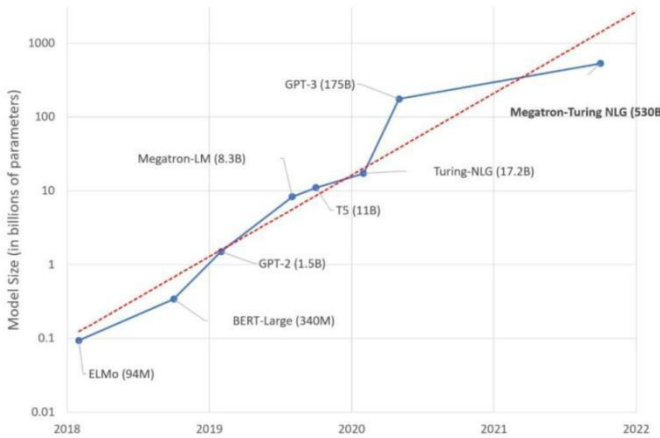
时间	部门	政策	相关内容
2023.2	中共中央、国务院	《数字中国建设整体布局规划》	系统优化算力基础设施布局，促进东西部算力高效互补和协同联动，引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局。
2023.4	网信办	《生成式人工智能服务管理办法（征求意见稿）》	对生成内容、主体责任、数据源和数据处理等方面均做出了规定。
2023.4	中共中央政治局	中央政治局会议	要重视通用人工智能发展，营造创新生态，重视防范风险。
2023.5	北京市政府	《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025年）》	2025年，人工智能算力布局初步形成，国产人工智能芯片和深度学习框架等基础软硬件产品市场占比显著提升，算力芯片等基本实现自主可控
2023.5	北京市人民政府办公厅	《北京市促进通用人工智能创新发展的若干措施》	《若干措施》针对算力资源和数据要素供给能力的提升提出了具体的措施，如加强与云厂商的合作、加快算力中心和数据训练基地的规划建设等；提出要“系统构建大模型人工智能技术体系”、“推动通用人工智能技术创新场景应用”。
2023.5	上海市发改委	《上海市加大力度支持民间投资发展若干政策措施》	分发挥人工智能创新发展专项等引导作用，支持民营企业广泛参与数据、算力等人工智能基础设施建设。
2023.5	中共深圳市委办公厅	《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）》	要打造全域全场景应用，推进“千行百业+AI”，孵化高度智能化的生产机器人。
2023.6	国务院办公厅	《国务院2023年度立法工作计划》	人工智能法草案等预备提请全国人大常委会审议。
2023.7	国家网信办、国家发改委、教育部、科技部、工业和信息化部、公安部、国家广电总局	《生成式人工智能服务管理暂行办法》	对生成式人工智能服务实行包容审慎和分类分级监管，明确了训练数据处理活动和数据标注等要求，规定了生成式人工智能服务规范。在第二章第六条提出“鼓励生成式人工智能算法、框架、芯片及配套软件平台等基础技术的自主创新，平等互利开展国际交流与合作，参与生成式人工智能相关国际规则制定”。
2023.7	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》	到2025年基本形成“高算力+强算法+大数据”的产业生态，实现大算力孵化大模型、大模型带动大产业、大产业促进大发展的良性循环将杭州打造成为全国算力成本洼地、模型输出源地、数据共享高地。
2023.7	重庆市经济信息委	《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）》	到2025年，人工智能发展整体迈上新台阶，成功创建国家人工智能创新应用先导区。培育10家亿级以上人工智能龙头企业，集聚100家人工智能规模以上企业、1000家人工智能相关企业，建成3-5个人工智能产业集聚区。
2023.8	成都市经济和信息化局、成都市新经济发展委员会	《成都市加快大模型创新应用推进人工智能产业高质量发展的若干措施》	围绕“算力”“存力”“运力”等关键领域，大力发展芯片、服务器整机、液冷设备等高端硬件，聚力推动存储芯片产业发展，打造具有全球影响力的“存储谷”。
2023.8	宁夏回族自治区人民政府办公厅	《促进人工智能创新发展政策措施》	加快实施数字宁夏“1244+N”行动计划，抢抓国家实施“东数西算”工程机遇，丰富算力基础设施，推进算力赋能应用，构建算力产业体系，加快宁夏枢纽建设，促进人工智能创新发展。
2023.9	武汉市人民政府办公厅	《武汉建设国家人工智能创新应用先导区实施方案（2023-2025年）》	到2025年，在图计算、数据治理、大模型、机器视觉、遥感图像解析等人工智能领域形成10项以上首创性技术；人工智能应用场景超过400项，产业规模达到1000亿元；初步建成全国重要的人工智能科技策源地、算力算法创新高地、产业集聚高地、场景应用高地和人7培养高地。
2023.10	工业和信息化部、中央网信办、教育部、国家卫生健康委、中国人民银行、国务院国资委	《算力基础设施高质量发展行动计划》	从算力、运载力、存储力以及应用赋能四个方面提出了到2025年发展量化指标：算力方面，算力规模超过300EFLOPS，智能算力占比达到35%。运载力方面，国家枢纽节点数据中心集群间基本实现不高于理论时延1.5倍的直连网络传输，重点应用场所光传送网（OTN）覆盖率达到80%，骨干网、城域网全面支持IPv6，SRv6等新技术使用占比达到40%。存储力方面，存储总量超过1800EB，先进存储容量占比达到30%以上。应用赋能方面，围绕工业、金融、医疗、交通、能源、教育等重点领域，各打造30个以上应用标杆。
2023.10	全国信息安全标准化技术委员会	《生成式人工智能服务安全基本要求》（征求意见稿）	为《生成式人工智能服务管理暂行办法》落地的支撑性文件，包括语料安全、模型安全、安全措施、安全评估等。
2023.10	北京市经济和信息化局	《人工智能算力券实施方案（2023—2025年）》	推动我市人工智能大模型应用落地，加快培育产业生态，以人工智能大模型高水平应用深度赋能实体经济高质量发展，加速千行百业数智化转型，特制定人工智能算力券实施方案。
2023.10	上海市经济和信息化委员会	《上海市推动人工智能大模型创新发展若干措施（2023-2025年）》	推动本市大规模预训练模型创新发展，构建开放安全创新生态，加快打造人工智能世界级产业集群；支持引进高水平创新企业，支持本市创新主体打造具有国际竞争力的大模型，鼓励形成数据飞轮，加速模型迭代，对取得重大成果的予以专项奖励。
2023.11	广东省人民政府	《关于加快建设通用人工智能产业创新引领地的实施意见》	到2025年，广东有望实现智能算力规模实现全国第一、全球领先，全省人工智能核心产业规模突破3000亿元，企业数量超2000家，力争打造成为国家通用人工智能产业创新引领地，构建全国智能算力枢纽中心、粤港澳大湾区数据特区、场景应用全国示范高地，形成“算力互联、算法开源、数据融合、应用涌现”的良好发展格局。
2023.12	中共中央	中央经济工作会议	会议强调，要大力推进新型工业化，发展数字经济，加快推动人工智能发展。

资料来源：政府网站，平安证券研究所

3.2 大模型的实现需要十分强大的算力来支持训练过程和推理过程

大模型的实现需要十分强大的算力来支持训练过程和推理过程。根据 OpenAI 数据，训练 GPT-3 175B 的模型，需要的算力高达 3640 PF-days（即以 1PetaFLOP/s 的效率要跑 3640 天）。2018 年以来，大模型的参数量级已达到数千亿参数的量级规模，对算力的需求将呈现指数级增长。

图表10 2018-2022年大模型参数增长变化趋势



资料来源:《Large Language Models: A New Moore's Law?》, 平安证券研究所

图表11 各个模型所需计算量及参数量

模型	总计算量(PT-days)	总计算量(Tops)	参数量(亿万个)	训练令牌(十亿个)	
T5模型	T5-Small	2.08	1.80E+20	60	1,000
	T5-Base	7.64	6.60E+20	220	1,000
	T5-Large	26.7	2.31E+21	770	1,000
	T5-3B	104	9.00E+21	3,000	1,000
	T5-11B	382	3.30E+22	11,000	1,000
BERT模型	BERT-Base	1.89	1.64E+20	109	250
	BERT-Large	6.16	5.33E+20	355	250
	RoBERTa-Base	17.4	1.50E+21	125	2,000
	RoBERTa-Large	49.3	4.26E+21	355	2,000
GPT模型	GPT-3 Small	2.60	2.25E+20	125	300
	GPT-3 Medium	7.42	6.41E+20	356	300
	GPT-3 Large	15.8	1.37E+21	760	300
	GPT-3 XL	27.5	2.38E+21	1,320	300
	GPT-3 2.7B	55.2	4.77E+21	2,650	300
	GPT-3 6.7B	139	1.20E+22	6,660	300
	GPT-3 13B	268	2.31E+22	12,850	300
	GPT-3 175B	3640	3.14E+23	174,600	300

资料来源: OpenAI《Language Models are Few-Shot Learners》, 平安证券研究所

3.3 AI 芯片进入舞台中央, 广泛应用于训练或推理

AI 芯片进入舞台中央, 广泛应用于训练或推理。AI 芯片是指所有能够用于人工智能的芯片, 主要包括 GPU、ASIC、FPGA 三大类, 海内外大型厂商集中于 GPU 和 ASIC 架构。AI 芯片按应用场景可以分为训练芯片和推理芯片: 训练芯片用于算法模型开发、训练, 利用标记的数据, 通过该芯片“学习”出具备特定功能的模型; 推理芯片用于应用层, 利用训练出来的模型加载数据, 通过芯片计算“推理”出各种结论。按照部署的位置可以分为云端芯片和边缘端芯片: 云端芯片部署在公有云、私有云或者混合云上, 不但可用于训练, 也可用于推理, 算力强劲; 边缘端芯片主要应用于嵌入式、移动终端等领域, 此类芯片一般体积小、耗电低, 性能要求也相对不高, 一般只需具备一两种 AI 能力, 用于推理。

图表12 三类 AI 芯片简介

AI 芯片	定义	优势	典型厂商
GPU	通用图形处理器	高并行结构, 生态体系成熟, 跨平台支持。易于编程, 成为主流的并行数据处理加速器	英伟达、AMD、海光信息、摩尔线程
ASIC	专用集成电路	专门为深度学习计算定制的芯片, 如神经网络处理器NPU、张量处理器TPU, 效率高、功耗低、体积小	谷歌TPU、寒武纪、海思昇腾
FPGA	现场可编程逻辑阵列	高度并行的结构和低延迟; 可编程和灵活性强, 能够适应模型算法迭代	赛灵思、英特尔Altera

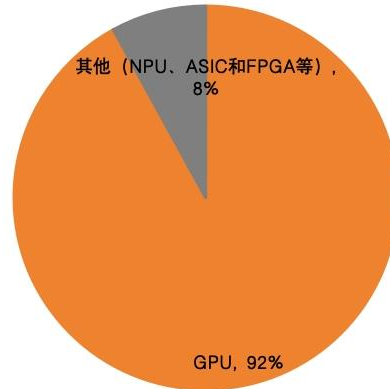
资料来源: CSDN, 平安证券研究所

GPU 大规模并行运算优势明显, 是 AI 芯片市场的首选。GPU 即图形处理器, 主要分为传统 GPU 和 GPGPU, 分别用于图形渲染和通用计算, 用于 AI 服务器的 GPU 一般指后者。GPU 中超过 80%部分为运算单元 (ALU), 擅长大规模并行运算, 主要应用于 PC、服务器、数据中心、自动驾驶等领域, 在数据中心被广泛应用于 AI 的训练、推理高性能计算等场景, 是国内数据中心加速服务器市场的首选。据 IDC 统计, 2023 年上半年, GPU 在我国 AI 服务器加速芯片市场以 92%的份额占据主导地位。

GPU 市场目前仍由英伟达、AMD 等国外厂商主导, 国内正处于发展起步阶段, 在 AI 芯片市场的竞争力较弱, 未来在大模型技术发展的催化下, 叠加美国限制向中国出口高端 GPU 芯片等因素, 国产 GPU 芯片将迎来发展机遇。据 IDC 统计, 2023

年上半年，中国加速芯片的市场规模超过 50 万张。从技术角度看，GPU 卡占有 90% 的市场份额；从品牌角度看，中国本土 AI 芯片品牌出货超过 5 万张，占比整个市场 10% 左右的份额。

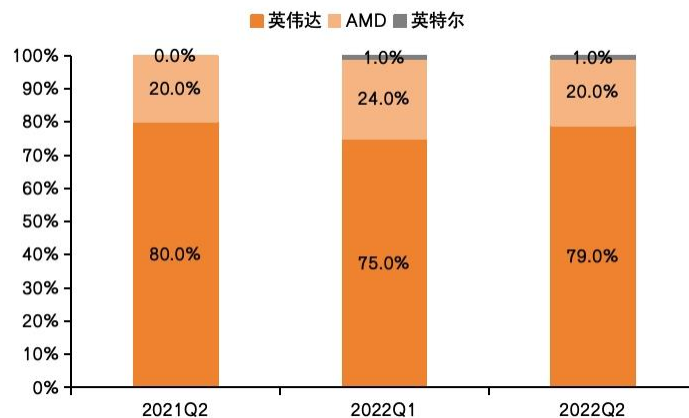
图表13 2023 年上半年 GPU 在我国 AI 服务器加速芯片市场份额为 92%



资料来源：IDC，平安证券研究所

全球 GPU 市场竞争格局：英伟达独占鳌头，AMD 跟随其后。英伟达是全球 GPU 芯片市场的绝对龙头，AMD 紧随其后。根据 JPR 数据，2022 年二季度，英伟达在全球独立 GPU 芯片市场占有率为 79%，AMD 以 20% 的市占率跟随其后。英伟达于今年 11 月份发布的新一代人工智能芯片 H200，相比 H100 内存速度和容量均实现大幅升级，在用于推理或生成问题答案时，性能较 H100 提高 60% 至 90%，预计将于 2024 年二季度上市。H200 将进一步提升大模型的训练与推理效率。近期，美国对华半导体出口管制升级，英伟达拟推出 H20、L20 和 L2 三款芯片，以替代被美国限制出口的 H100。AMD 于 2023 年 6 月推出了针对 AI 的最新款处理器芯片 MI300A 和 MI300X，12 月 AMD 举行“Advancing AI”发布会表示 MI300A 已进入量产阶段，而 Instinct MI300X 则已开始出货。海外算力能力不断提升，叠加美国对华半导体出口管制升级，将倒逼我国国产 AI 芯片产业链加快成熟。

图表14 全球独立 GPU 出货量市占率



资料来源：新浪财经，JRP，平安证券研究所

图表15 英伟达、AMD 主要 GPU 产品性能参数对比

性能参数	A100 80GB PCIe	A100 80GB SXM	H100 PCIe	H100 SXM	H200 SXM	HGX H20	L20 PCIe	L2 PCIe	MI250X	MI300X
FP32 (TFLOPS)	19.5	19.5	51	67	67	44	59.8	24.1	47.9	-
FP64 (TFLOPS)	9.7	9.7	26	34	34	1	-	-	47.9	-
TF32 (TFLOPS)	156	312	756	989	989	74	59.8	48.3	-	-
GPU 显存	80GB	80GB	80GB	80GB	141GB	96GB	48GB	24GB	128GB HBM2e	192GB HBM3
GPU 显存带宽	1935GB/s	2039GB/s	2TB/s	3.35TB/s	4.8TB/s	4.0TB/s	864GB/s	300GB/s	3.2TB/s	5.2TB/s
最大热设计功耗 (TDP)	300W	400W	300-350W	700W	700W	400W	275W	TBD	500W	-

资料来源：英伟达官网，AMD 官网，IT之家，平安证券研究所

国内 GPU 产品发展势头良好，产品性能已可对标国际主流产品。国内目前部署 GPU 赛道的厂商主要有海光信息、景嘉微、沐曦、壁仞科技、芯动科技等。当前，国内 GPU 产品发展势头良好，产品性能已可对标英伟达主流产品。以海光信息为例，海光信息 DCU (Deep Computing Unit 深度计算器，是 GPGPU 通用图形处理的一种) 产品具备强大的计算能力和高速并行数据处理能力，产品性能已可对标国际上同类型高端产品的水平，已成功实现商业化应用。目前，海光信息第二代 DCU 产品-深算二号也已经发布并实现了在大数据处理、人工智能、商业计算等领域的商业化应用，具有全精度浮点数据和各种常见整型数据计算能力。根据海光信息公司公告，公司在 2023 年三季度发布深算二号，性能相比于深算一号提升 100%以上。

图表16 海光 DCU 产品-深算一号规格特点

项目	深算一号 (海光 8100)
典型功耗	260-350W
典型运算类型	双精度、单精度、半精度浮点数据和常见整型数据
计算	(1) 60-64个计算单元 (最多4096个计算核心) (2) 支持FP64、FP32、FP16、INT8、INT4
内存	(1) 4个HBM2内存通道 (2) 最高内存带宽1TB/s (3) 最大内存容量为32GB
I/O	(1) 16 Lane PCIe Gen4 (2) DCU芯片之间高速互联

资料来源：海光信息招股说明书，平安证券研究所

图表17 海光 DCU 产品与行业代表可比产品参数对比

项目	海光	NVIDIA	AMD
品牌	深算一号	Ampere 100	MI100
生产工艺	7nm FinFET	7nm FinFET	7nm FinFET
核心数量	4096 (64 CUs)	2560 CUDA processors	120 CUs
内核频率	Up to 1.5GHz (FP64) Up to 1.7GHz (FP32)	Up to 1.53GHz	Up to 1.5GHz (FP64) Up to 1.7GHz (FP32)
显存容量	32GB HBM2	80GB HBM2e	32GB HBM2
显存位宽	4096 bit	5120 bit	4096 bit
显存频率	2.0 GHz	3.2 GHz	2.4 GHz
显存带宽	1024 GB/s	2039 GB/s	1228 GB/s
TDP	350 W	400 W	300 W
CPU to GPU互联	PCIe Gen4 x 16	PCIe Gen4 x 16	PCIe Gen4 x 16
GPU to GPU互联	xGMI x 2, Up to 184GB/s	NVLink, Up to 600GB/s	Infinity Fabric x 3, Up to 276GB/s

资料来源：海光信息招股说明书，平安证券研究所

ASIC 是一种定制芯片，可提供更高能效表现和计算效率。ASIC (专用芯片) 是一种为特定目的、面向特定用户需求设计的定制芯片，具备性能更强、体积小、功耗低、可靠性更高等优点。在大规模量产的情况下，还具备成本低的特点。ASIC 芯片主要应用于深度学习加速，在推理侧，相较于其他 AI 芯片在效率和速度方面具有明显优势。以谷歌为例，谷歌早在 2015 年就发布 TPU v1，与当时通用 CPU 和 GPU 的神经网络计算相比，TPU v1 带来了 15-30 倍的性能提升和 30-80 倍的能效提升，以较低成本支持谷歌的多项服务，仅可用于推理；2017 年发布 TPU v2，用于加速大量的机器学习和人工智能工作负载，包括训练和推理；2018 年发布 TPU v3，算力和功率大幅增长，采用了当时最新的液冷技术；2020 年和 2021 年分别发布 TPU v4i 和 v4，应用 7nm 工艺，晶体管数大幅提升，算力提升，功耗下降；2023 年谷歌发布 TPU v5e 和 TPU v5p，12 月最新发布 TPU v5p 与 TPU v4 相比浮点运算性能进一步大幅提升，高带宽内存方面是 TPU v4 的近 3 倍。

国内 ASIC 产品发展势头良好，部分国产 AISC 芯片性能已经达到国际前列水平。我国的 AISC 芯片产品发展迅速，国内主要 ASIC 芯片企业有寒武纪、澜起科技、黑芝麻、地平线、华为海思、阿里巴巴等。部分国产 AISC 技术已经达到国际前列水平，如在半精度浮点算力方面，华为海思的昇腾 910 超过谷歌的 TPU v4 和 TPU v5e。华为推出的昇腾 910B 芯片，仍然采用华为自研 Ascend 架构，其整体性能和能效均达到了国际领先水平，目前华为昇腾 910B 计算能力已经可以对标英伟达 A100。

图表18 华为海思与谷歌芯片性能对比

芯片	昇腾Ascend910	昇腾Ascend910B	谷歌TPU v4	谷歌TPU v5e	谷歌TPU v5p
FP16 (TFLOPS)	320	376	-	-	-
BF16 (TFLOPS)	-	-	275	197	459
INT8 (TOPS)	640	-	-	394	918
HBM (GB)	-	64	32	16	95
HBM BW (GB/s)	-	392	1228	820	2765

资料来源：华为官网，谷歌官网，平安证券研究所

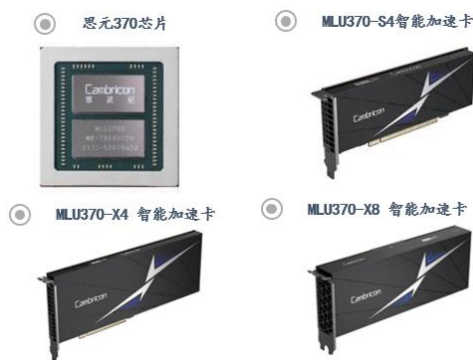
寒武纪 ASIC 产品不断迭代，最新一代产品有望承接国内 AI 算力需求。寒武纪的第三代云端推理一体芯片思元 370，最大算力高达 256TOPS (INT8)，是第二代产品思元 270 算力的 2 倍。此外，与市场主流同尺寸芯片相比，思元 370 系列加速卡在实测性能和能效方面表现出一定优势。公司的思元 370 芯片及加速卡与数家头部互联网企业完成适配工作，已经进入了批量销售环节；与金融、运营商等众多行业领域中的头部公司实现了批量销售或达成合作意向。思元 590 是寒武纪最新一代云端智能训练芯片，目前尚未正式发布。思元 590 采用 MLUarch05 全新架构，实测训练性能较在售旗舰产品有大幅提升，有望承接国内逐渐升级的 AI 算力需求。

图表19 MLU370 系列加速卡规格

性能参数	MLU370-S4	MLU370-X4	MLU370-X8
制程工艺	7nm		
INT8	192 TOPS	256TOPS	
INT16	96 TOPS	256 TOPS	128 TOPS
FP16	72 TFLOPS	96 TFLOPS	
BF16	72 TFLOPS	96 TFLOPS	
FP32	18 TFLOPS	24 TFLOPS	
内存容量	24GB		48GB
内存带宽	307.2GB/s		614.4 GB/s
系统接口	x16 PCIe Gen4		
最大热设计功耗	75W	150W	250W

资料来源：寒武纪官网，平安证券研究所

图表20 寒武纪代表性 ASIC 产品



资料来源：寒武纪官网，平安证券研究所

3.4 AI 服务器借助加速卡获取强大算力，市场需求快速增长

AI 服务器是指能够提供人工智能 (AI) 的数据服务器，具有强大的图形处理和高性能计算能力，既可以用来支持本地应用程序和网页，也可以为云和本地服务器提供复杂的 AI 模型和服务，能支持多种常用的 AI 技术，如机器学习、自然语言处理、计算机视觉、生物信息分析等。AI 服务器与普通服务器的区别主要在于计算架构的不同，AI 服务器通常根据应用场景的不同，采用 CPU+GPU/ASIC/FPGA 或其他加速卡的异构式计算架构。

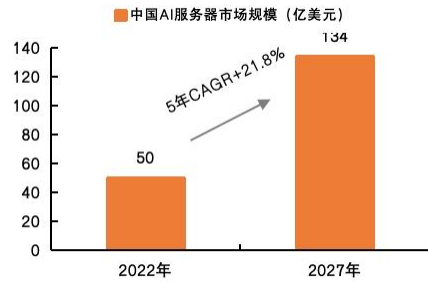
当前国内外 AI 服务器市场规模快速增长。根据 IDC 与浪潮信息发布的《2023-2024 年中国人工智能算力发展评估报告》预计，全球人工智能服务器市场规模将从 2022 年的 195 亿美元增长到 2026 年的 347 亿美元，4 年复合增长率约 15.5%。2023 年中国人工智能服务器市场规模将达到 91 亿美元，同比增长 82.5%，2027 年将达到 134 亿美元，2022-2027 年 5 年年复合增长率达 21.8%。

图表21 2022-2026 年全球 AI 服务器市场规模 (亿美元)



资料来源: IDC&浪潮信息《2023-2024 年中国人工智能算力发展评估报告》, 平安证券研究所

图表22 2022-2027 年中国 AI 服务器市场规模 (亿美元)

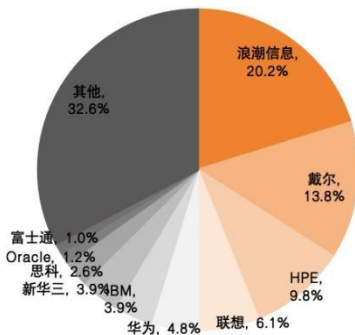


资料来源: IDC&浪潮信息《2023-2024 年中国人工智能算力发展评估报告》, 平安证券研究所

浪潮信息全球领跑，国产厂商在 AI 服务器领域大有可为。截至 2022 年，浪潮信息人工智能服务器市场份额长期居于全球前列，连续 6 年保持中国第一。根据 IDC 最新发布的《2023 年第 3 季度全球服务器市场追踪报告》，浪潮信息服务器出货量市场份额 10.3%，销售额市场份额 9.1%，均位居全球第二，销售额增速 21.8%。

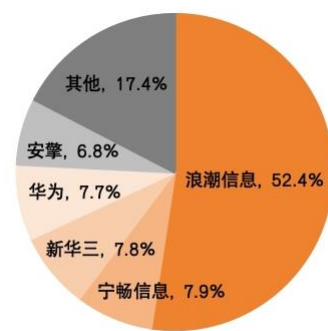
此外，新华三、宁畅、安擎等诸多国产厂商也正在加速推进人工智能基础设施产品的优化升级，并积极探索 AI 赋能传统产业的应用落地方向。据 IDC 统计，2023 年上半年，从厂商销售额角度看，浪潮、新华三、宁畅位居前三，占据了 70% 以上的市场份额；从服务器出货台数角度看，浪潮、坤前、宁畅位居前三名，占有近 60% 的市场份额，国产厂商在 AI 服务器领域大有可为。

图表23 2021 年上半年全球 AI 服务器市场竞争格局



资料来源: IDC, 中商产业研究院, 平安证券研究所

图表24 2021 年我国 AI 服务器市场竞争格局

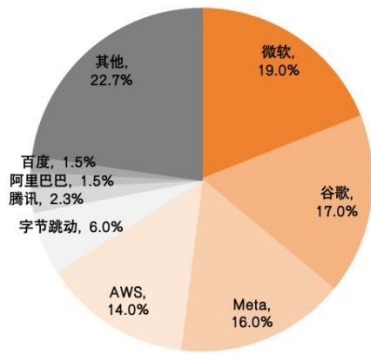


资料来源: IDC, 中商产业研究院, 平安证券研究所

互联网云巨头贡献 AI 服务器主要需求，资本投入力度有望维持增长。AI 服务器市场的下游主要是以大型云计算厂商为主。TrendForce 统计数据 displays，2022 年 AI 服务器采购量中，美国四家云厂商，微软、谷歌、Meta、AWS 的采购量位居前四，合计占比约 66%。国内企业字节跳动、腾讯、阿里巴巴、百度紧随其后，在 AI 基础设施方面的建设步伐较为领先。IDC 统

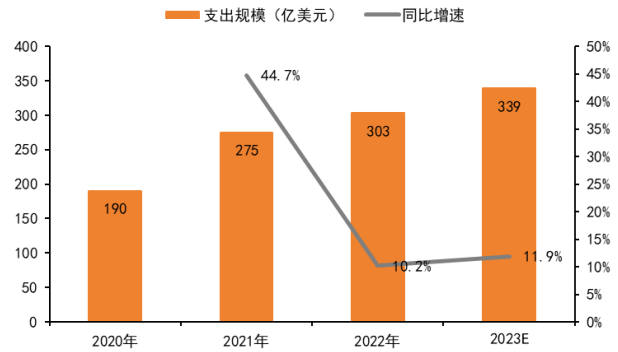
计数据显示，2023 年上半年，互联网依然是我国加速服务器最大的采购行业，占整体加速服务器市场超过一半的份额，此外金融、电信和政府行业均有超过一倍以上的增长。未来，随着 AIGC、边缘计算、自动驾驶等新兴技术和应用的不断普及，各大云厂商有望持续加大在 AI 相关基础设施方面的投入，持续为 AI 基础设施市场注入发展动力。

图表25 2022 年 AI 服务器下游采购情况



资料来源: TrendForce, 观知海内咨询, 平安证券研究所

图表26 2020-2023 年我国云计算支出规模及增速



资料来源: canalsy, 中商产业研究院, 平安证券研究所

3.5 大模型的训练成本和推理成本高昂

大模型的训练成本和推理成本高昂。以 ChatGPT 为例，在训练端：根据澎湃新闻信息，2020 年，微软宣布与 OpenAI 合作，建成了一台超级计算机，专门用来在 Azure 公有云上训练超大规模的人工智能模型。这台为 OpenAI 开发的超级计算机拥有超过 28.5 万个 CPU 核心，拥有超过 1 万个 GPU (V100GPU 芯片)。以此规格，如果自建 IDC，以 A100GPU 芯片替代 V100GPU 芯片，依照 A100 和 V100 的性能换算，需要约 3000 个 A100GPU 芯片。根据英伟达网站信息，NVIDIADGXA100 服务器搭载 8 块 A100 芯片，估算需要 375 台 NVIDIADGXA100 服务器，每台 NVIDIADGXA100 服务器的价格为 19.9 万美元，则自建 IDC 的训练服务器的算力成本为 7462.5 万美元。若在云端训练，据 Lambda Labs 首席科学官 Chuanli 介绍，拥有 1750 亿个参数的 GPT-3 单次训练成本达到 460 万美元。

图表27 自建 IDC 的训练成本估算

A100 GPU 芯片 (个)	1台NVIDIA DGX A100服务器搭载A100芯片个数	NVIDIA DGX A100服务器 (台)	NVIDIA DGX A100服务器价格 (万美元)	训练成本 (万美元)
3000	8	375	19.9	7462.5

资料来源: IT之家, 英伟达官网, 平安证券研究所

在推理 (用户访问) 端：ChatGPT 推出仅两个月月活用户数已经破亿，2023 年 1 月，全球每天约有 1300 万独立访问者使用 ChatGPT。以 ChatGPT 日活用户 2000 万估算，假设每天每用户提 10 个问题，则每天有 2 亿的访问量。若自建 IDC，假设每个问题平均 20 个字，ChatGPT 在 A100GPU 芯片上对每个字的响应时间是 350 毫秒，则 2 亿的访问量需要 A100 芯片运行 388889 个芯片小时，即每天需要 16204(388889/24) 个 A100 芯片同时工作，需要 2026(16204/8) 台 NVIDIADGXA100 服务器同时工作，则自建 IDC 的推理服务器的算力成本为 4.03 亿美元。若在云端推理，据《Fortune》杂志数据，每次用户与 ChatGPT 互动，产生的算力云服务成本约 0.01 美元，则每天 2 亿的访问量，对应的云端成本为每天 200 万美元。

图表28 自建 IDC 的推理成本估算

日活用户数 (万)	单用户每日提问次数	每个问题平均字数 (个)	A100 GPU对每个字的响应时间 (毫秒)	每日消耗GPU计算时间 (小时)	每天需要A100 GPU 芯片 (个)	NVIDIA DGX A100服务器 (台)	NVIDIA DGX A100服务器价格 (万美元)	推理成本 (亿美元)
2000	10	20	350	388889	16204	2026	19.9	4.03

资料来源: IT之家, 英伟达官网, 平安证券研究所

大模型将为全球和中国 AI 芯片和 AI 服务器市场的增长提供强劲动力。根据《2022 年北京人工智能产业发展白皮书》数据, 截至 2022 年 10 月, 北京拥有人工智能核心企业 1048 家, 占我国人工智能核心企业总量的 29%。以此计算, 我国人工智能核心企业总数约为 3614 家。初步来看, 我国参与大模型的企业大致可以分为两类: 即 C 端应用的企业和 B 端应用的企业。在 C 端应用方面, 假设其中有 20 家企业自研或与合作方共同研发通用大模型, 自建 IDC 训练和推理面向庞大 C 端月活的千亿量级参数的大模型, 算力需求可参照 ChatGPT, 即单一企业自建 IDC 推理和训练大模型的算力成本约为 4.78 (4.03+0.75) 亿美元, 按 20 家企业估算, 算力需求为 95.6 亿美元。

B 端应用方面, 假设在 3614 家人工智能核心企业中, 有 10%的企业即 361 家企业使用合作伙伴的大模型 (包括开源大模型), 使用垂直行业数据进行进一步的训练得到垂类大模型, 并将垂类大模型部署到客户的数据中心, 则训练成本为垂类大模型的训练成本, 推理成本为客户运营垂类大模型的成本。假设垂类大模型的训练成本为 ChatGPT 训练成本的 1/10 即 0.075 亿美元, 则合计训练成本为 27.08 亿美元, 假设每个垂类大模型厂商平均部署 100 个客户, 每个客户运营大模型平均需要 3 台 DGXA100 服务器, 则每个客户平均推理成本为 59.7 万美元, 合计推理成本为 215.52 亿美元, 算力需求为 242.6 (27.08+215.52) 亿美元。

以上, 根据我们的初步估算, 大模型的应用将为我国 AI 服务器市场带来 338.2 亿美元的市场需求。以 IDC 测算 2022 年我国 AI 服务器市场规模占全球 AI 服务器市场规模的份额 (26%) 估算, 则将为全球 AI 服务器市场带来约 1322.6 亿美元的市场空间。市场空间巨大, 相关芯片和服务器厂商将深度受益大模型的发展浪潮。

图表29 大模型将为国内带来的 AI 服务器市场增量测算

	企业数量 (家)	企业训练算力成本	企业 (对应客户) 推理算力成本	算力总成本 (亿美元)
C端应用的企业	20	15亿美元 (0.75亿美元*20)	80.6亿美元 (4.03亿美元*20)	95.6
B端应用的企业	361 (3614*10%)	27.08亿美元 (0.075亿美元*361)	215.52亿美元 (59.7万美元*100*361)	242.6
合计	-	-	-	338.2

资料来源: 《2022 年北京人工智能产业发展白皮书》, 平安证券研究所

美国对华半导体出口管制升级, 我国国产 AI 芯片产业链将加快成熟。当地时间 2023 年 10 月 17 日, 美国商务部工业与安全局 (BIS) 发布了对华半导体出口管制最终规则。最终规则在去年 10 月 7 日出台的临时规则基础上, 进一步加严对人工智能相关芯片半导体制造设备的对华出口限制, 同时本次将壁仞科技、摩尔线程、光线云等 13 家中国实体增列入出口管制“实体清单”。根据 10 月 24 日英伟达公告, 英伟达 A100, A800、H100、H800、L40、L40S 和 RTX4090 等高端 GPU 芯片的对华出口将受到出口管制最终规则的影响。根据 IDC 数据, 2023 年上半年, 中国加速芯片的市场规模超过 50 万张。中国本土 AI 芯片品牌出货量超过 5 万张, 占据整个市场 10%左右的份额。受最终规则的影响, 美国高端 AI 芯片的对华出口将受到非常严格的限制。

根据 IT 之家消息, 英伟达现已开发出针对中国区的最新改良版 AI 芯片, 包括 HGX H20、L20 PCIe 和 L2 PCIe, 但与国产芯片比如华为昇腾 910 等相比并不具备明显优势。在此背景下, 我国国产 AI 芯片在国内市场的竞争力将进一步提高, 在国内市场的市占率有望加速提升。根据 IDC 数据, 2023 年上半年, 中国本土 AI 芯片品牌出货量国内市场份额约 10%, 我

国产 AI 芯片市占率提升空间很大。我国国产 AI 芯片厂商将加快发展，产业链将加快成熟，相关 AI 芯片和服务器厂商将深度受益。

图表30 英伟达 H20、L20、L2 性能不及部分国产芯片

厂商	英伟达	英伟达	英伟达	英伟达	海光信息	华为	华为
芯片型号	H200 SXM	HGX H20	L20 PCIe	L2 PCIe	深算一号	昇腾910	昇腾910B
BF16 FP16 (TFLOPS)	1,979	148	119.5	96.5	-	320	376
FP32 (TFLOPS)	67	44	59.8	24.1	-	-	94
INT8 (TFLOPS或TOPS)	3,958	296	239	193	-	640	-
GPU 显存	141GB	96GB	48GB	24GB	32GB	-	64GB
GPU 显存带宽	4.8TB/s	4.0TB/s	864GB/s	300GB/s	1024GB/s	-	392GB/s
最大热设计功耗 (TDP)	700W	400W	275W	-	350W	310W	400W

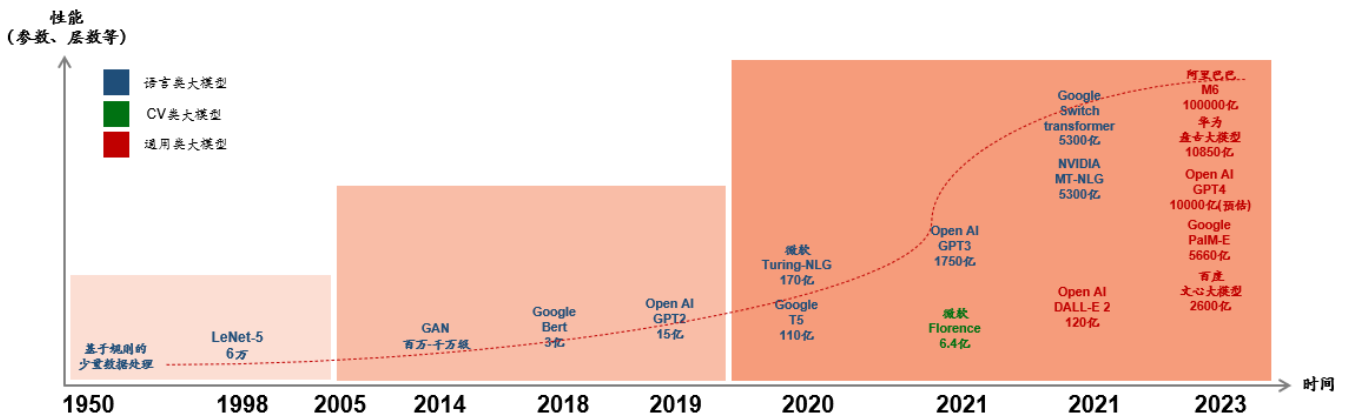
资料来源：IT之家，英伟达官网，海光信息官网，华为官网，平安证券研究所

四、我国国产大模型能力持续升级，AIGC 产业未来应用潜力巨大

4.1 大模型迭代进入快速发展阶段，GPT 大模型成为关注焦点

大模型算法的迭代可以大致分为萌芽阶段、探索阶段和快速发展阶段三个阶段。第一阶段以 CNN 为代表的传统神经网络模型为主。1980 年，卷积神经网络的雏形 CNN 诞生；1998 年，以卷积神经网络为架构的 LeNet-5 深度学习模型问世，为大模型的发展奠定基础。2006 年-2019 年，是以 Transformer 为代表的卷积神经网络模型阶段。2014 年，对抗式生成网络 GAN 诞生，开启生成模型研究新阶段。2017 年，Google 开创性地提出了 Transformer 架构，随后，OpenAI 和 Google 分别发布了 GPT-1 与 BERT 大模型，大模型的性能显著提升。2020 年以后，卷积神经网络模型的参数量激增，多种通用大模型相继问世，其中，OpenAI 在 2023 年 3 月发布的预训练大模型 GPT-4 成为目前关注焦点。

图表31 AI 大模型发展历程

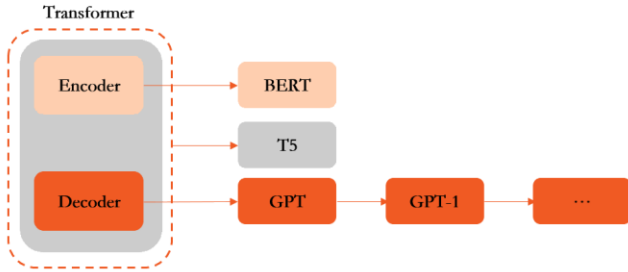


资料来源：亿欧智库，平安证券研究所

Transformer 引入注意力机制，突破 RNN、CNN 处理长序列的局限。Transformer 模型是由 Google 团队的 Ashish Vaswani 等人发表论文《Attention Is All You Need》提出的模型概念，是当前大模型领域主流的算法架构基础，其上形成了 BERT 和 GPT (Generative Pre-Trained Transformer) 两条主要的技术路线，其中 BERT 的代表性落地项目是谷歌的 AlphaGo。在 GPT3.0 发布后，GPT 逐渐成为大模型的主流路线。当前，国内包含百度文心一言、阿里通义千问等在内的几乎所有参数规模超千亿的大型语言模型均采用 GPT 模式。

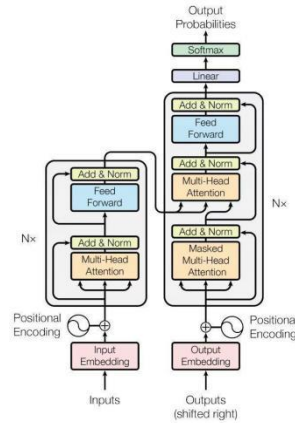
Transformer 模型是一个深度学习模型，其标志性特征是采用了 self-attention 机制，可为输入数据的各部分分配不同权重，核心是从关注全部到关注重点，从而节省资源，快速获得最有效的信息。self-attention 机制的引入突破了 RNN 处理长序列的输入，以及 CNN 解决远距离特征依赖问题的局限，使得通过提升参数量来提升预训练大模型的性能成为了可能。

图表32 Transformer 模型系列分类



资料来源：CSDN，平安证券研究所

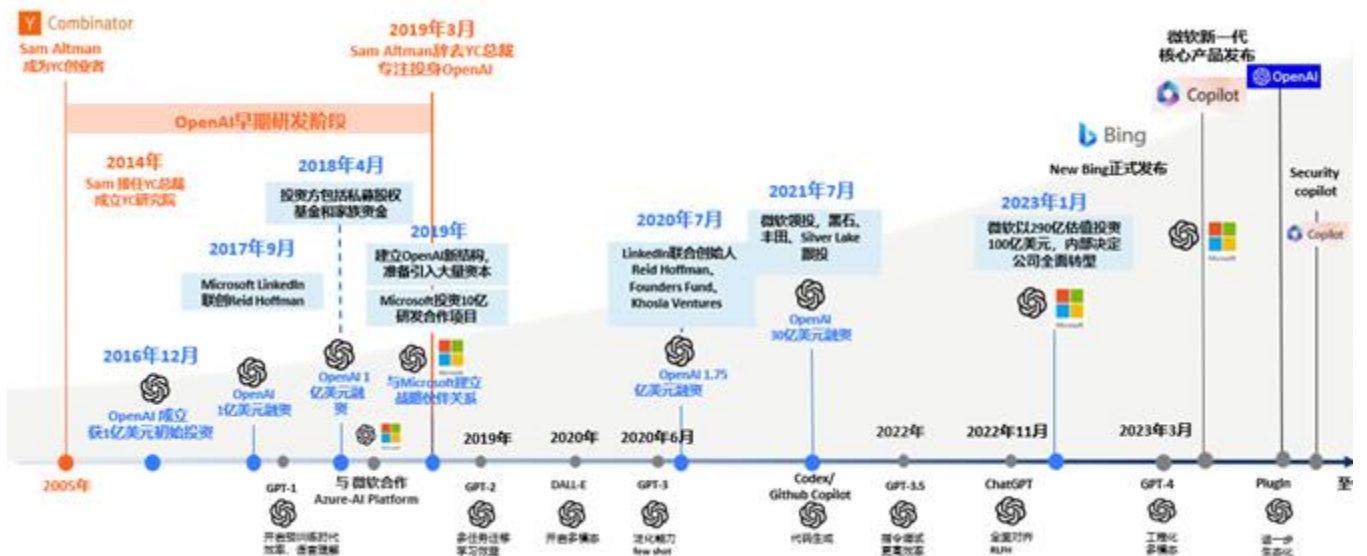
图表33 Transformer 模型架构



资料来源：谷歌《Attention is All you Need》，平安证券研究所

国外大模型起步较早，全球人工智能企业领先布局，国内 2023 年迎来爆发式增长。2018 年开始，谷歌、OpenAI、英伟达、Meta 和微软等大型科技企业纷纷推出自研大模型技术，预训练大模型逐渐成为自然语言处理领域的主流研究。2022 年 11 月 ChatGPT 发布仅两个月的时间，线上活跃用户规模超过 1 亿人，生成式大模型市场广泛关注，尤其是微软和谷歌两家人工智能领军企业迅速意识到 ChatGPT 对于未来商业的战略意义，开始围绕公司现有业务，在生成式大模型领域加速布局。2023 年受 ChatGPT 驱动，大模型发展迈向新阶段，国产大模型一时间也呈现出爆发式增长态势，据赛迪顾问统计，2023 年 1-7 月国内共计 64 个大模型发布，截至 2023 年 7 月，中国累计已经有 130 个大模型问世。

图表34 Sam Altman 和 OpenAI 推出的 ChatGPT 掀起海内外大模型布局热潮



资料来源：陆奇博士在奇绩创坛的演讲《新范式 新时代 新机会》，平安证券研究所

目前，海外大模型进展来看，ChatGPT 已迭代至 GPT-4 Turbo 版本，Google 也于 12 月推出了其最强的 Gemini 多模态大模型。当地时间 11 月 6 日，OpenAI 举办首届开发者大会“OpenAI DevDay”，Sam Altman 在会上推出 GPT-4 Turbo，

将上下文长度提升至 128K，相当于 300 多页文本内容，为此前 GPT-4 扩容版本（32K）的 4 倍。GPT-4 Turbo 将模型内部知识库更新至 2023 年 4 月，此前 GPT-4 知识库仅停留在 2021 年 9 月。当地时间 12 月 6 日，Google 宣布推出其规模最大、功能最强的多模态大模型 Gemini，支持 32k 的上下文长度，1.0 家族涵盖 Ultra、Pro、Nano 三类模型，Gemini Ultra 在 32 个广泛使用的文本领域学术基准中有 30 项超过 GPT-4，在图像、视频、音频多项基准测试中同样实现超越 GPT-4V。

国内大模型厂商情况来看，BAT、科大讯飞在今年 9 月-10 月陆续推出了其大模型的最新版本，模型能力向 GPT-3.5 和 GPT-4 看齐。9 月 7 日，在 2023 腾讯全球数字生态大会上，腾讯混元大模型正式亮相，并宣布通过腾讯云对外开放，模型中文能力超过 GPT3.5，英文能力目前不及 GPT3.5。10 月 17 日，百度世界大会 2023 在北京召开，李彦宏正式发布文心大模型 4.0，称 4.0 能力与 GPT4 相比毫不逊色。10 月 24 日，第六届世界声博会暨 2023 科大讯飞全球 1024 开发者节召开，讯飞星火 V3.0 重磅发布，根据公司介绍，讯飞星火 V3.0 全面对标 GPT3.5，中文能力客观评测超越 ChatGPT，英文能力对标 ChatGPT 48 项任务结果相当。10 月 31 日，阿里云栖大会上阿里云正式发布千亿级参数大模型通义千问 2.0，称在 MMLU、AGIEval、C-Eval 等 10 大主流测评集上综合性能超过 GPT3.5。

图表35 2023年下半年国内外厂商陆续发布大模型最新版本

公司	大模型名称	推出时间	大模型能力
腾讯	混元	2023年9月7日	中文能力超过GPT3.5，英文能力目前比 GPT3.5 差
百度	文心一言4.0	2023年10月17日	根据李彦宏在发布会上的发言，4.0能力可以对标GPT4
科大讯飞	讯飞星火V3.0	2023年10月24日	全面对标ChatGPT（GPT3.5），中文能力客观评测超越ChatGPT，英文能力对标 ChatGPT 48项任务结果相当
阿里云	通义千问2.0	2023年10月31日	在MMLU、AGIEval、C-Eval等10大主流测评集上综合性能超过GPT3.5
OpenAI	GPT4 Turbo	2023年11月7日	相较GPT4，GPT4 Turbo将上下文长度提升至128K，将模型内部知识库更新至2023年4月，整合了DALL·E 3、TextToSpeech、Whisper V3
Google	Gemini 1.0	2023年12月6日	Gemini Ultra在32个广泛使用的学术基准中有30项超过GPT4

资料来源：各公司官网，平安证券研究所

4.2 大模型赋能千行百业，AIGC 未来发展前景广阔

按模态划分，大模型可分为自然语言处理（NLP）大模型，视觉（CV）大模型、多模态大模型等。按应用场景划分，大模型可分为通用大模型和行业大模型。通用大模型是指可在不进行微调或少量微调的情况下完成多场景任务的大模型，具有强大的泛化能力，ChatGPT、华为盘古大模型以及最新发布的 Gemini 都是通用大模型。行业大模型则利用行业知识对大模型进行微调，以满足在金融、汽车、传媒等不同领域的需求，如金融领域的 BloombergGPT、汽车领域的毫末 DriveGPT、法律领域的 LawGPT_zh 等。

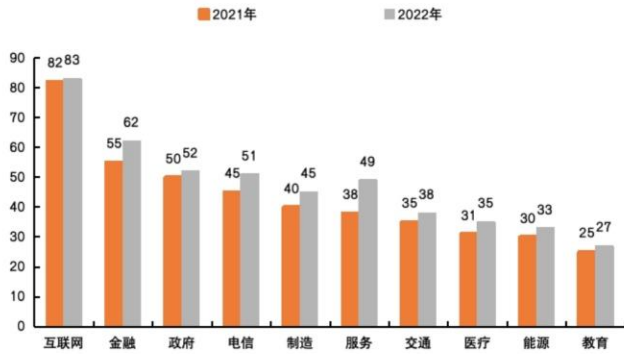
图表36 AI大模型按模态分类

分类	简介	代表
自然语言处理 (NLP) 大模型	NLP大模型通常用于处理文本数据和理解自然语言，主要特点是在大规模语料库上进行了训练，以学习自然语言的各种语法、语义和语境规则。	GPT系列（OpenAI）、Bard（Google）、文心一言（百度）
视觉 (CV) 大模型	CV大模型通常用于图像处理和分析，通过在大模型图像数据上进行训练，可以实现各种视觉任务，如图像分类、目标检测、图像分割、姿态估计、人脸识别等。	VIT系列（Google）、文心UFO（百度）、盘古CV（华为）、INTERN（商汤）
多模态大模型	多模态大模型是指能够处理文本、图像、音频等多种不同类型数据的大模型，这类模型结合了NLP和CV的能力，以实现多模态信息的综合理解和分析，从而能够更全面地理解和处理复杂的数据。	DingoDB多模向量数据库（九章云极 DataCanvas）、DALL-E（OpenAI）、悟空画画（华为）、midjourney、pika（Pika Labs）

资料来源：CSDN，平安证券研究所

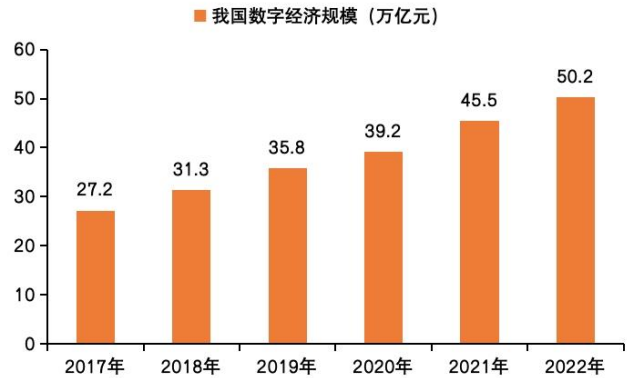
当前，人工智能在我国各行业已经得到广泛应用。根据 IDC 数据，2022 年，人工智能在我国互联网、金融、政府、电信、制造等行业的渗透率分别为 83%、62%、52%、51%、45%。随着国产大模型的逐步成熟，我国大模型产品面向我国庞大的互联网 C 端用户群和丰富的行业应用场景，将与产品和应用场景深度融合，赋能我国数字经济的发展。参考我国数字经济的巨大体量，我国 AIGC 产业未来发展前景广阔。

图表37 2021-2022 年我国各行业人工智能渗透率 (%)



资料来源：《2022-2023 中国人工智能算力发展评估报告》，平安证券研究所

图表38 2017-2022 年我国数字经济规模

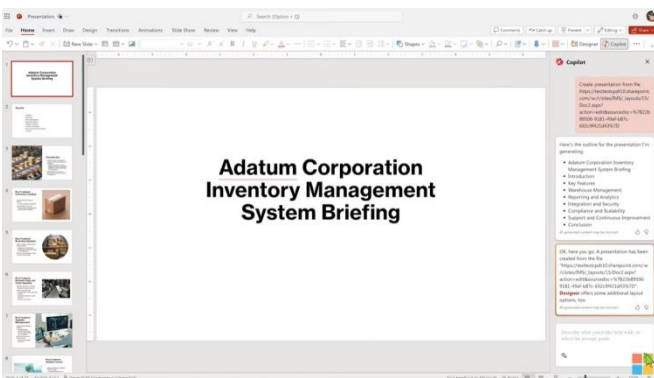


资料来源：中国信通院，平安证券研究所

■ AIGC+办公：微软 Copilot 将 AIGC 与办公软件结合，助力实现降本增效

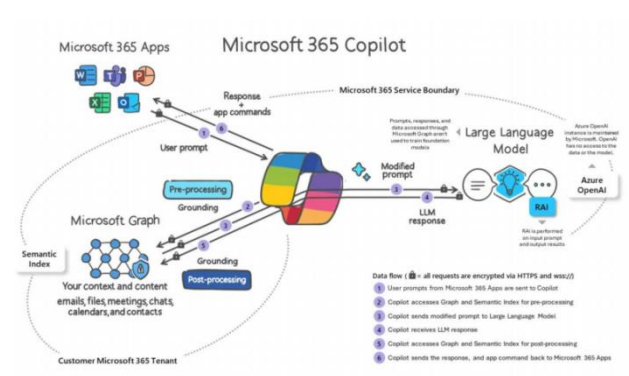
微软推出 Microsoft 365 Copilot 把 ChatGPT 引入 Office 办公软件，将 AIGC 与办公软件结合。AI 办公软件将大型语言模型的力量与业务数据和办公软件应用相结合，实现了发挥创造力、释放生产力、提高办公技能和改变生产方式等目的。2023 年，继 Microsoft 365 Copilot 于 11 月 1 日开启海外商用之后，12 月 6 日微软宣布 Copilot 将获得一些增强功能，从而生成更强大的文本和图像，包括获得对 GPT-4 Turbo 的支持、更新 DALL-E 3 模型、增添新的代码解释器功能和必应内部的深度搜索功能。我们认为 Copilot 有望带动各行业 AI 助手市场的发展，产生广泛的行业应用。

图表39 Microsoft 365 Copilot 生成 PPT



资料来源：微软公司官网，平安证券研究所

图表40 Microsoft 365 Copilot 全面注入 PC 端



资料来源：微软公司官网，平安证券研究所

金山办公：WPS AI 是具备大语言模型能力的生成式 AI 应用，搭载 WPS AI 后的组件可以实现智能问答、文本生成、文本分析、PPT 生成、PDF 翻译等功能。2023 年 11 月 16 日，WPS AI 开启公测，成为国内率先落地的 AI 办公应用工具。

万兴科技：公司的 PDFelement、万兴喵影（Filmora）和亿图图示（EdrawMax）等已经推出 AI 相关功能，通过赋能多模态创业业务场景，实现创意软件智能化的进一步提升。目前系列软件可提供文档智能处理、文案辅助创作、数字人营销、图片生成等功能。

福昕软件：公司在海外市场 Foxit PDF Editor Cloud 等产品集成 ChatGPT，在国内市场与百度“文心一言”达成合作。公司将 AI 赋能完整 PDF 解决方案全线产品，提供多模态智能识别、电子签名、文字翻译等功能。

图表41 WPS AI 成为国内率先落地的 AI 办公应用



资料来源：金山办公官网，中国经济网，平安证券研究所

图表42 万兴科技多项办公类产品推出 AI 功能



资料来源：万兴科技官网，平安证券研究所

■ AIGC+传媒：文生视频工具 Pika 出圈，多模态驱动传媒行业效能提升

AIGC 虽然处于发展早期，但成长曲线十分陡峭，随着越来越多由 AI 驱动的生产力工具被整合到工作流中，各产业生态势必将重塑。人工智能作为新的传媒行业发展引擎，未来将全面激活行业各业务，助推传媒行业业务发展进入新纪元。从应用方向上看，当前 AIGC 赋能工具主要集中在文本、代码、图像、音频等几大领域，与传媒行业密切相关，AIGC 将成为未来传媒行业发展的重要驱动力。2023 年 11 月，AI 初创公司 Pika labs 发布全新的文生视频产品 Pika 1.0，用户只需要输入一句话就可以生成想要的各种风格的视频，并通过简单的描述更改视频中的形象和风格，有望加速推动多模态 AIGC 应用在传媒行业的落地。

昆仑万维：公司坚定选择 AGI 和 AIGC 赛道，目前已推出天工巧绘 SkyPaint、天工乐府 SkyMusic、天工妙笔 SkyText、天工智码 SkyCode 等系列产品构建产品体系。公司还与奇点智源合作自研国产大语言模型，12 月 1 日，公司正式发布基于昆仑万维天工大模型所打造的天工 SkyAgents 平台，为国内领先的 AI Agent 开发平台。

蓝色光标：公司陆续推出了苏小妹、K、虚拟国风音乐女团吾音等虚拟 IP，MEME 数藏平台以及蓝宇宙等虚拟空间和虚拟产品。AIGC 技术产品方面，公司已发布包括分身有术、蓝标智播、创意画廊、销博特“创策图文”以及“萧助理”等多款产品，9 月 12 日，公司发布 Blue AI 营销行业模型，为介于通用大模型基础层和更具体的 AI 应用服务层之间的行业模型，聚焦内容生成、思维助手、体验创新三大主要垂直场景，提供数据分析、消费者洞察、营销策略、创意概念等功能。

中文在线：公司的海外产品 Chapters 和 My Escape 已在做接入 ChatGPT 测试，并推出 AI 主播、AI 绘画和 AI 文字辅助创作三款 AIGC 相关产品。10 月 13 日，公司宣布推出“中文逍遥”万字创作大模型，该模型可一键生成万字小说，能为作者提供全生命周期的 AI 辅助创作工具。

图表43 昆仑万维发布 AI Agent 平台天工 SkyAgents



资料来源：昆仑万维集团公众号，平安证券研究所

图表44 Blue AI 帮助出海品牌创作营销文案



资料来源：蓝色光标官网，平安证券研究所

■ AIGC+医药：赋能研诊疗全流程，提升健康管理和辅助决策效率

在医疗领域，从医学科研、药物研发到智慧诊疗的各阶段、医疗设备运维、医院管理等方面均有大模型产品涌现，参与主体主要是由高校及科研机构展开探索、企业推进商业化落地。2023 年 7 月 31 日，谷歌发布了首个全科医疗大模型——Med-PaLM M 多模态生成模型，此外，水木分子、华为、百度、科大讯飞等科研机构或企业陆续发布了医疗大模型。

水木分子：2023 年 9 月 21 日，水木分子团队发布新一代对话式药物研发助手 ChatDD 及全球首个千亿参数多模态生物医药对话大模型 ChatDD-FM 100B。ChatDD 化身“多场景助手”，通过 ChatDD-BI、ChatDD-Discovery 和 ChatDD-Trial 分别出现在立场场景、研发探索场景和临床试验场景中。

华为：2021 年发布的华为云盘古药物分子大模型，是由华为云联合中国科学院上海药物研究所共同训练而成的大模型，可以实现针对小分子药物全流程的人工智能辅助药物设计。实验验证结果表明，盘古药物分子大模型的成药性预测准确率比传统方式高 20%，进而提升研发效率，让先导药的研发周期从数年缩短至一个月，同时降低 70%的研发成本。

百度：百度的文心生物计算大模型于 2022 年 5 月推出，是将生物领域研究对象的特性融入模型，构建面向化合物分子、蛋白分子、基因组学信息的生物计算领域预训练大模型。2023 年 9 月，百度发布“产业级”医疗大模型——灵医大模型，发布会上，百度宣布面向大健康上下游产业开放灵医大模型测评、试用，推动医疗行业的数字化和智能化进程。

科大讯飞：2023 年 7 月，科大讯飞表示基于星火认知大模型，“讯飞医疗”全面升级医疗诊后康复管理平台，将“专业的诊后管理和康复指导延伸到院外”，该平台可以根据患者健康画像自动分析，为患者智能生成个性化康复计划，并督促患者按计划执行。10 月公司正式推出了讯飞星火医疗大模型，并首发医疗大模型加持下的“讯飞晓医”APP 及小程序，涵盖疾病预防、病症自查、报告解读、医疗信息查询、中医辨证等医疗场景，致力于辅助提升用户就医效率和就医体验。

图表45 ChatDD 给出分子亲和力计算排序结果

请调用本地小分子库，为我筛选与Survival motor neuron protein (SMN2) 亲和力较高的小分子

分子亲和力计算结果:

Target Name	Smiles	Affinity Type	Affinity Value
Survival motor neuron protein	CC(C)1cc(no1)C(=O)Nc1cccc(C)1	EC50	2.893
Survival motor neuron protein	CC(C)1cc(no1)C(=O)Nc1cccc1	EC50	4.023

资料来源：水木分子官网，中国日报网，平安证券研究所

图表46 百度发布国内首个产业级医疗大模型“灵医”



资料来源：百度灵医智慧，平安证券研究所

■ AIGC+金融：金融科技深化发展，大模型可用于风险管理、投资决策等业务

金融机构数字化转型需求的不断提升，以及 AIGC 技术对金融领域相关信息化产品的强大赋能，使得 AIGC+金融领域的应用发展十分迅速，BloombergGPT 的发布让人们看到未来金融行业 AIGC 应用的强大。目前在金融领域，大模型可以用于股票预测、信用评估、风险管理等业务，也可用于建设数字人系统、短视频生成平台、自动研报生成系统等。

同花顺：公司具有 AI 虚拟人、同花顺 AI 开放平台 i 问财等一系列人工智能产品，为金融行业持续提供数字虚拟人、智能客服机器人、智能质检机器人等 40 余项 AI 产品服务。

东方财富：公司陆续研发东方财富金融数据 AI 智能化生产平台、多媒体智能资讯及互动平台系统等多个人工智能相关项目，提供 AI 投资决策支持、AI 智能定投等业务服务。

恒生电子：2023 年 10 月，恒生金融科技大会上，公司发布了金融大模型 LightGPT 最新的能力升级成果，自 6 月底发布至今 LightGPT 在整体模型效果上提升 15%，安全合规性上提升 13%，推理速度上提升 50%，并面向金融机构实现 LightGPT-7B 的开源，推理和训练全面适配华为昇腾系列。同时在大模型应用方面，公司正式发布基于 LightGPT 打造的光子·文曲、光子·方圆、光子·善策和光子·慧营四款大模型应用，并宣布开放产品公测。

图表47 同花顺 i 问财界面



资料来源：同花顺官网，平安证券研究所

图表48 恒生光子·文曲为面向投顾内容创作的产品



资料来源：恒生电子 2023 恒生金融科技大会，平安证券研究所

■ AIGC+教育：教育数字化持续推进，大模型助力精准教学和个性化学习

AIGC 赋予教育材料新活力，将对传统教育的全流程环节范式产生一定影响。相对于阅读和讲座等传统方式，AIGC 为教育工作者提供了新的工具，使原本抽象、平面的课本具体化、立体化，以更加生动、更加令人信服的方式向学生传递知识。例如制作历史人物，为教学注入新的活力；合成逼真的虚拟教师，让数字教学更具互动性和趣味性。

科大讯飞：公司发布的讯飞星火认知大模型已嵌入科大讯飞 AI 学习机，使学习机获得充分结合学生刚需、充分考虑产品教育属性、过滤与学习无关的信息的类 ChatGPT 体验，8 月 15 日，科大讯飞正式推出了讯飞星火认知大模型 V2.0 版本，作为大模型在教育领域的典型应用，发布会上展示了星火认知大模型 V2.0 在科大讯飞 AI 学习机上新的应用——AI 一对一智能编程助手和 AI 一对一创意绘画伙伴。此外，公司还拥有智能评卷技术、个性化学习手册等产品。

好未来（学而思）：公司自研数学大模型 MathGPT，以数学领域的解題和讲題算法为核心打造而成，后续将推出基于该自研大模型的产品级应用。此外，公司还拥有“AI 老师监课系统”、“AI 老师记单词”、“AI 老师语言学习系统”以及“VR 沉浸式课堂”等产品，为教学核心环节赋能。按照《生成式人工智能服务管理暂行办法》规定，学而思大模型 MathGPT 完成了相关备案并已通过审核，正式成为首批通过备案的教育大模型。

图表49 科大讯飞 AI 学习机推出 1v1 智能编程助手



资料来源：科大讯飞官网，时代周报，平安证券研究所

图表50 好未来学而思学习机进行中文作文批改



资料来源：好未来官网，财联社，平安证券研究所

4.3 汽车驶向智能化时代，大模型提升自动驾驶泛化能力

大模型的引入加速自动驾驶算法向端到端的演进，助力高级别自动驾驶功能突破。自动驾驶算法系统分为端到端自动驾驶和模块化自动驾驶两类。模块化自动驾驶系统分为感知、决策和执行三层，其中 AI 在感知层和决策层发挥重要作用。端到端自动驾驶是指车辆将传感器采集到的信息，直接送入到一个统一的深度学习神经网络，神经网络经过处理之后直接输出自动驾驶汽车的驾驶命令。随着自动驾驶逐渐由 L2 迈入 L3 阶段，基于规则的感知算法和基于高精地图的定位方案从技术或商业化角度，都或已成为了高级别自动驾驶功能发展的瓶颈，端到端自动驾驶算法将成为实现高级别自动驾驶功能的突破口。

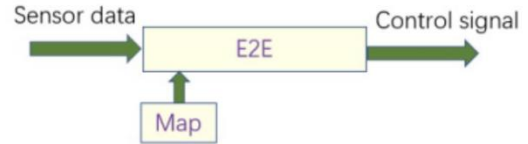
大模型强大的建模能力可以实现从原始传感器输入直接到驾驶决策输出，大模型的引入有利于自动驾驶技术从模块化阶段向端到端进化。

图表51 自动驾驶系统模块化结构



资料来源：焉知汽车公众号，平安证券研究所

图表52 自动驾驶系统端到端结构



资料来源：焉知汽车公众号，平安证券研究所

特斯拉自研算法，由BEV+Transformer到占用网络，引领自动驾驶算法迭代。特斯拉作为全球的自动驾驶技术指引，其自动驾驶感知算法的发展路径就是自动驾驶感知算法迭代历程的缩影。2014年至2016年期间，特斯拉自动驾驶系统采用了Mobileye的视觉算法技术。2016年，在告别了Mobileye后，特斯拉开始自研算法，采用鸟瞰视角(BEV)来建模自动驾驶系统的环境。2020年之后，特斯拉引入BEV+Transformer取代传统的2D+CNN算法，并采用特征级融合取代后融合，自动标注取代人工标注，自动驾驶开始进入大模型时代。2022年，特斯拉自动驾驶算法中引入时序网络，并将BEV升级为占用网络(Occupancy Network)。2023年8月，端到端AI自动驾驶系统FSD Beta V12首次公开亮相，完全依靠车载摄像头和神经网络来识别道路和交通情况，并做出相应的决策。

图表53 特斯拉自动驾驶算法迭代历程

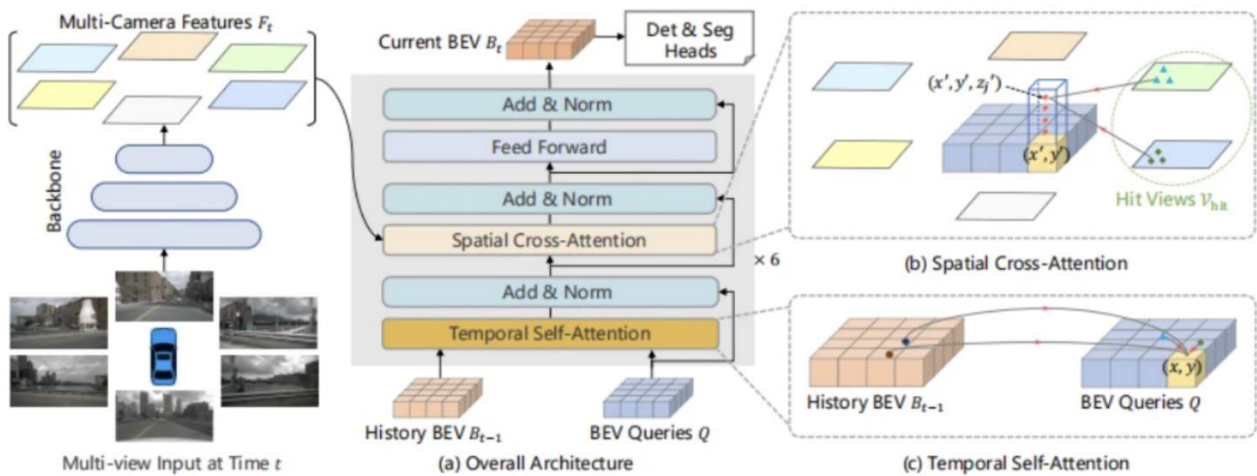


资料来源：汽车之心公众号，平安证券研究所

BEV+Transformer 助力城市 NOA 脱离高精度地图，是目前自动驾驶主流方案。2022 年之前，国内主机厂实现城市 NOA（导航辅助驾驶，Navigate on Autopilot）落地主要基于高精度地图+单车感知的方案，但在实施过程中发现高精度地图的弊端与限制逐渐暴露：1）无法实时更新且制作成本较高。高精地图内容极为精细，故更新速度缓慢，但自动驾驶需要地图至少达到日更以上水平，带来了极大的适配以及成本问题。2）政策制约。目前，高精地图包含内容过于详细，出于国家安全角度考量，其采集制作与应用限制颇多。

自动驾驶感知算法向 BEV+Transformer 架构升级，助力城市 NOA 脱离高精度地图。特斯拉推出的基于 BEV+Transformer 的自动驾驶感知路线，可将 2D 图像转化为 BEV 鸟瞰图视角下的 3D 图像，并基于更庞大、更复杂、参数更多的感知算法模块，利用 Transformer 的注意力机制，将各个传感器统一到同一个连续的四维（三维空间+一维时间）时空中。BEV+Transformer 架构逐渐成为自动驾驶主流方案。

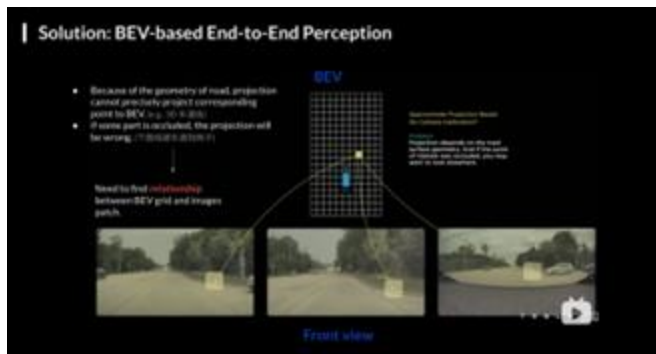
图表54 BEV+Transformer 架构



资料来源：佐思产研，平安证券研究所

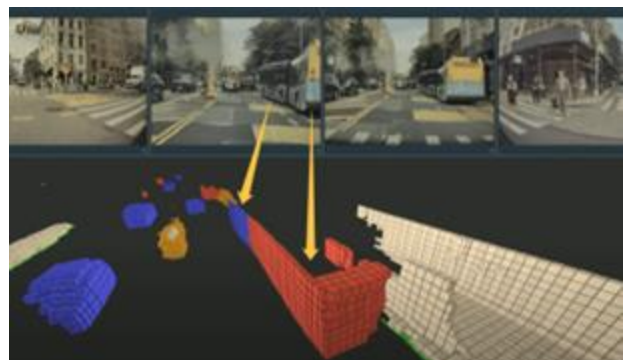
BEV 感知模型使得自动驾驶在极端天气等特定场景的安全性显著提升。BEV 感知模型的应用有效提升了自动驾驶在极端天气等特定场景下的安全性。在后融合模型中，当遇到雨雪等极端天气条件时，摄像头采集到的数据/视频流的清晰度会大幅下降，很难满足摄像头判断合格的标准，从而导致传递给后端进行规划控制的结果大幅降低。与后融合模型不同，从不同视角的摄像头采集到的图片在转换为 BEV 视角的过程中实现了特征级的融合，感知模型对数据的利用率明显提高。

图表55 多摄像头下的 BEV 感知



资料来源：智造前沿公众号，平安证券研究所

图表56 占用网络对起步中的公交车预测示意图



资料来源：智造前沿公众号，平安证券研究所

大模型提升自动驾驶泛化能力，助力城市 NOA 加速落地。高速领航辅助驾驶（高速 NOA）在特定路段下驾驶场景较为封闭，交通环境的标准化程度高，且交通参与者较为单一，不涉及到行人，驾驶状态的可预测性更强，因此高速 NOA 成为了自动驾驶率先落地的场景。而城市领航辅助驾驶（城市 NOA）在路况、交通参与者、场景异质性方面的复杂性均对自动驾驶模型的泛化能力提出了挑战。自动驾驶场景逐渐由高速拓展到城市。佐思汽研统计，2023 年 1-9 月，国内乘用车高速 NOA 渗透率为 6.7%，同比增加 2.5 个百分点；城市 NOA 渗透率为 4.8%，同比增加 2.0 个百分点。预计全年高速 NOA 渗透率将接近 10%，城市 NOA 超过 6%。AI 大模型的应用不仅通过提升了自动驾驶算法的泛化能力，同时有效降低了自动驾驶传感器硬件成本，有望加速城市 NOA 的商业化进程。

图表57 乘用车 NOA 正从高速向城市演进



资料来源：佐思汽研，平安证券研究所

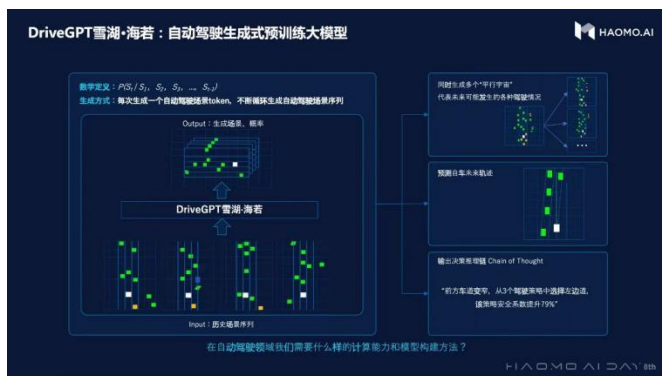
图表58 部分车企城市 NOA 落地计划

公司	功能	车型	日期	备注
小鹏	NGP	P5, P7L, G6	2022年9月	新增广州南沙 P5 车型无图城市 NGP (有高精地图)
			2023年9月	P5 逐步开放上海城市 NGP (有高精地图)
理想	NOA	L7, L8, L9	2023年上半年	落地北京、上海、广州、深圳，完成城市领航辅助驾驶功能，以及无图城市的城市 NGP 能力
			2023年下半年	落地 50 城，上线“AI 代驾模式”，大部分无图城市都称城市 NOA 能力
蔚来	NOA	ET5, ET7, ES6, ES7, ES8, EC6, EC7	2022年4月	落地 200 城，实现全场景（高速+城区+泊车）的领航辅助驾驶
			2023年4月	发布 ADMax 3.0 城市 NOA
蔚来	NAD	ET5, ET7, ES6, ES7, ES8, EC6, EC7	2023 Q2	城市 NOA (跟高精地图) 落地深圳，并在深圳多个无图城市 NOA 试驾
			2023 年底	城市 NOA (跟高精地图) 覆盖 100 座城市，推出城市 NOA 功能
华为 (问界、阿赛、智选)	NCA	问界 M5 智驾版、阿赛增程 11、智选鸿蒙 5 智驾版	2022年12月	发布 NCA
			2023 年初	推出全场景智慧领航功能 NAD (包含城市领航辅助)，进行 NAD NEB Beta 功能测试
蔚来	NOA	蔚来城市 NOA 智驾版	2022年9月	落地深圳并开通城市 NOA 功能
			2023年3月	落地北京、上海、广州并开通城市 NOA 功能
蔚来	NOA	蔚来城市 NOA 智驾版	2023 Q2	在 15 座城市实现城市 NOA (有高精地图)
			2023 Q3	在 15 座城市实现城市 NOA，并开放多个无图城市 NOA 试驾
蔚来	NOA	蔚来城市 NOA 智驾版	2023 Q4	在 45 座城市实现城市 NOA (有高精地图)
			2023 年底	在 45 座城市实现城市 NOA (有高精地图)
蔚来	NOA	蔚来城市 NOA 智驾版	2023 年上半年	落地北京、上海、广州并开通城市 NOA 功能
			2023 年下半年	落地北京、上海、广州并开通城市 NOA 功能
上汽智己	NOA	智己 L7, L7 Pro	2023 年初	落地北京、上海、广州并开通城市 NOA 功能
			2023 年底	落地北京、上海、广州并开通城市 NOA 功能

资料来源：浦东未来车产业公众号，平安证券研究所

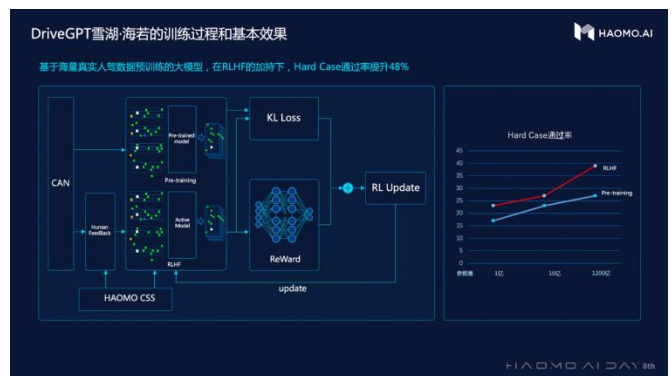
国内自动驾驶认知决策 AI 大模型开始落地，助力实现端到端方案。在自动驾驶认知决策层面，国内已有相关 AI 大模型落地。2023 年 4 月，毫末智行发布了业内首个自动驾驶生成式大模型 DriveGPT，中文名“雪湖·海若”。DriveGPT 通过引入驾驶数据建立 RLHF (人类反馈强化学习) 技术，对自动驾驶认知决策模型进行持续优化，最终目标是实现端到端自动驾驶参数规模达到 1200 亿，预训练阶段引入 4000 万公里量产车驾驶数据，RLHF 阶段引入 5 万段人工精选的困难场景接管 Clips。与 ChatGPT 使用自然语言进行输入与输出不同，DriveGPT 输入是感知融合后的文本序列，输出是自动驾驶场景文本序列，即将自动驾驶场景 Token 化，形成“Drive Language”，最终完成自车的决策规划、障碍物预测以及决策逻辑链的输出等任务。10 月，毫末智行发布了 3 款第二代智能辅助驾驶产品，摆脱对高精地图的依赖，同时成本高昂的激光雷达也成为了非必须可选项，3 款产品分别为：HP170、HP370、HP570，市场价格分别为：3000 元、5000 元、8000 元，定价相较此前市场上万元级别的智能驾驶软硬件价格而言具备优势，有望加速高级别自动驾驶大规模商用落地。

图表59 DriveGPT 自动驾驶生成式预训练大模型



资料来源：毫末智行公众号，平安证券研究所

图表60 DriveGPT 的训练过程和基本效果

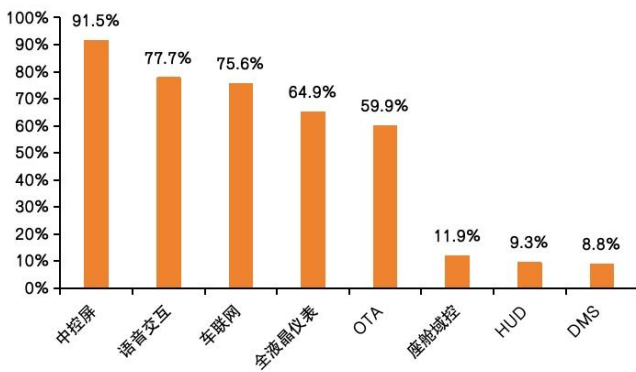


资料来源：毫末智行公众号，平安证券研究所

智能座舱已经进入规模化普及的新周期。根据盖世汽车研究院数据，2023 年上半年，中控大屏、车载语音交互和车联网的渗透率均超过 70%。自 2023 年以来，大模型上车的速度明显加快，智能座舱方面，车载语音助手成为 AI 大模型落地的重要途径。多模态大模型的应用可以通过提升车载助手多模态理解能力、感知能力等，重塑人车交互方式。

AI 大模型+座舱应用频繁落地。2023 年 6 月，理想旗下语音助手“理想同学”加入了其自研的认知 MindGPT，可以支持声纹识别、内容识别、AI 绘画、AI 计算等功能。10 月份，随着智己 LS6 上市，其自研的 GPT 大模型也正式上线。而蔚来也已申请 NIOGPT、NOMIGPT 等商标。此外，科大讯飞的星火大模型、百度的文心一言大模型、商汤的日日新大模型均已在智能座舱场景实现应用落地。

图表61 智能座舱核心产品/功能渗透率（标配）



资料来源：盖世汽车智能网联公众号，平安证券研究所

图表62 理想 MindGPT 上车



资料来源：全球智能汽车供应链公众号，平安证券研究所

4.4 大模型为人形机器人注入大脑，打开通用机器人可实现性

全球机器人市场规模持续扩大，中国市场增速领先。机器人指能够执行各种任务的自动化设备或系统，具备感知、决策和执行能力。受益于技术的不断进步、成本的下降以及对自动化解决方案的需求增加，全球机器人市场规模持续扩大，根据国际机器人联合会（IFR，International Federation of Robotics），2022 年全球机器人市场规模 513 亿美元，同比增长 19.6%，2017-2022 年 CAGR 为 14.0%。中国机器人市场增速保持领先，2022 年中国机器人市场规模 174 亿美元，同比增长 22.5%，2017-2022 年 CAGR 为 22.1%，增速领先于全球。

图表63 全球机器人市场规模及增速



资料来源：IFR，中商产业研究院，平安证券研究所

图表64 中国机器人市场规模及增速



资料来源：IFR，中商产业研究院，平安证券研究所

细分市场来看，IFR 将机器人分为工业机器人和服务机器人。1) **工业机器人**：主要用于工业生产和制造领域，通常具有高度精准的动作控制和重复性能力，用于自动化生产线上的各种任务，如装配、焊接、搬运、喷涂等。根据 IFR，2021 年全球工业机器人市场规模约 175 亿美元。值得注意的是，我国是全球最大的工业机器人市场，2022 年装机量达到 29.0 万台，占据了全球市场的 52%。2) **服务机器人**：设计用于为人类提供各种服务，可以在家庭、医疗、教育、零售、餐饮等领域执行特定的任务，如家庭清洁机器人、医疗辅助机器人、教育机器人、导购机器人等，IFR 统计 2021 年全球服务机器人市场规模约 172 亿美元。

图表65 全球工业机器人市场规模及增速



资料来源：IFR，中国电子学会，平安证券研究所

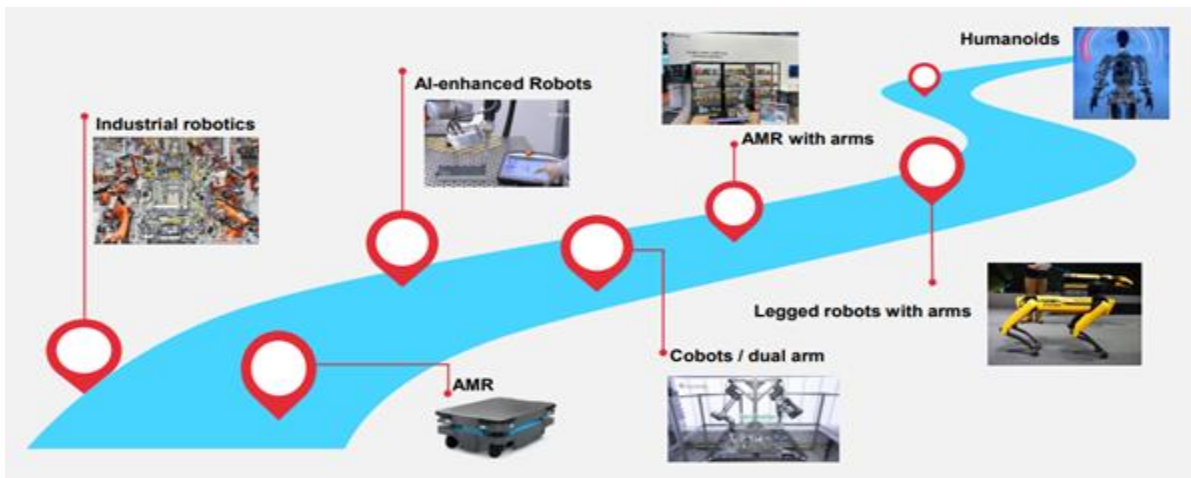
图表66 全球服务机器人市场规模及增速



资料来源：IFR，中国电子学会，平安证券研究所

机器人智能化迭代升级，人形机器人具备“拟人”优势。机器人的发展历史可以追溯到 20 世纪 40 年代，其智能化程度经历了不断的迭代升级，从最初的工业机器人到现代的服务机器人、协作机器人等等。如今人形机器人是机器人领域的重要分支，之所以能够成为当前研究领域的热点，是因为其“拟人”优势所赋予机器人功能的多样性与普适性，从而带来应用场景与领域的广泛性。人形机器人近似于人的机体和动能设计，意味着能够像人一样行动，完成人在制造业、农业、服务业中从事的各种重复、繁琐工作，更易成为人类社会活动的重要参与者。

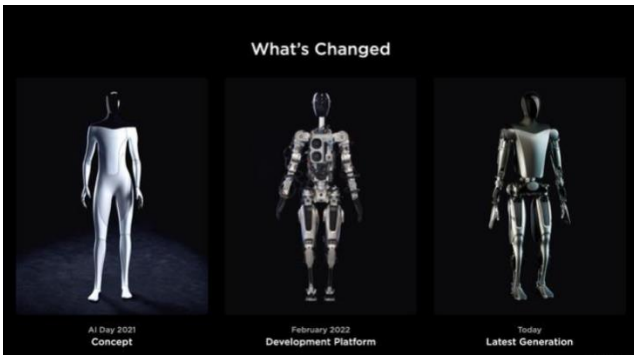
图表67 人形机器人是机器人领域最新的发展趋势



资料来源：IFR，平安证券研究所

特斯拉高调入局，掀起人形机器人产业热潮。2022年9月30日，特斯拉AI Day 上人形机器人“Optimus”（擎天柱）首次亮相，海内外市场为之振奋。2023年5月特斯拉股东大会上视频展示了迭代后的Optimus，在电机扭矩和力度控制等方面更精确，预计量产后单价有望低于20000美元。2023年9月，特斯拉机器人X官方账号发布视频，Optimus已可以自主对物品分类，同时能够自动校准它的手臂和腿，仅通过视觉和关节位置编码器，即可定位其肢体位置。12月13日，马斯克在社交平台X展示了将于12月发布的Optimus第二代机器人(Gen 2)步行速度提升30%，平衡感和身体控制能力有所改善。根据特斯拉透露的未来方向，Optimus的应用场景未来可能会涉及工业生产，也就是由人形机器人全程接管汽车生产领域，实现真正的无人生产。

图表68 2022年特斯拉Optimus外形发展历程



资料来源：特斯拉官网，平安证券研究所

图表69 2023年12月新一代Optimus做深蹲



资料来源：特斯拉官网，平安证券研究所

人形机器人产业百花齐放，但尚未形成大规模商业化应用。人形机器人代表企业包括 Boston Dynamics、优必选、Agility Robotics 等公司，但当前绝大多数人形机器人都未实现商业化落地。2022年特斯拉高调入局掀起产业热潮，我们看到海内外企业纷纷加紧布局人形机器人赛道。从国内企业来看，2022年8月小米发布了人形机器人CyberOne，2023年以来国内厂商人形机器人产品发布节奏进一步加快，7月傅利叶发布人形机器人GR-1，8月宇树科技发布其首款人形机器人H1，达闼紧随其后发布了最新版旗舰人形服务机器人XR-4，10月小鹏首次展示了自研的人形机器人PX5，在多家科技企业大力投入的背景下，我们认为我国人形机器人商业化进程有望加速。

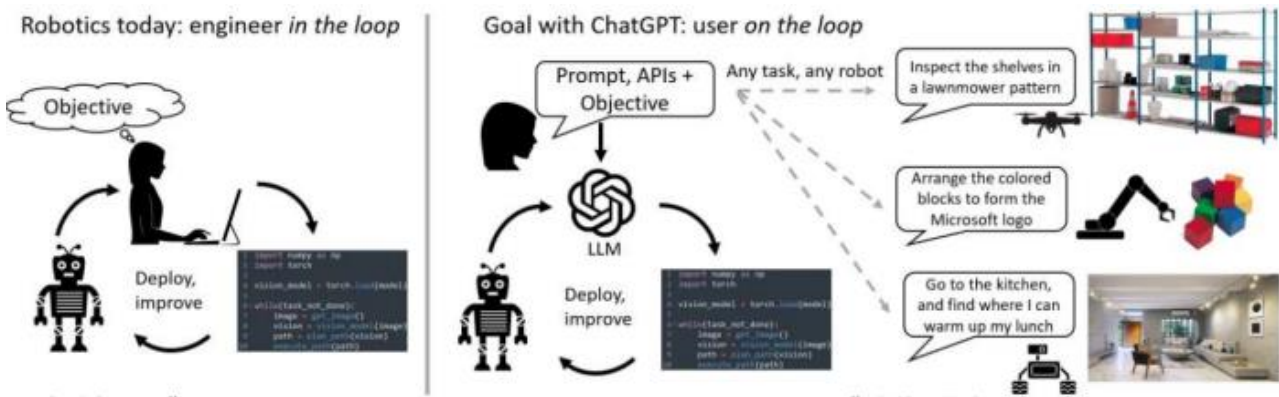
图表70 全球具有代表性的人形机器人梳理

公司	产品名称	发布时间	产品类型	身高体重	速度	自由度	是否销售
Boston Dynamics	Atlas	2013	工业应用、救援任务	175cm 80kg	9千米/小时	28	否
优必选	Walker X	2018	商用服务、科研开发	130cm 63kg	3千米/小时	41	是
Agility Robotics	Digit	2019	物流配送	155cm 45kg	5.4千米/小时	16	是
特斯拉	Optimus	2022	汽车驾驶	173cm 56kg	8千米/小时	28+11	否
小米	Cyberone	2022	人机交互服务	177cm 52kg	3.6千米/小时	21	否
Engineered Arts	Ameca	2022	娱乐	187cm 49kg	-	-	是
追觅	追觅人形机器人	2023	智能生活服务	178cm 56kg	-	44	否
傅利叶	GR-1	2023	康复陪护	165cm 55kg	5千米/小时	40	是
宇树	Unitree H1	2023	人形作业替代	180cm 47kg	大于5.4千米/小时	19	否
达闼	XR-4	2023	人机交互服务	168cm 65kg	5千米/小时	60+	否
智元	远征A1	2023	工业应用、家庭管家	175cm 53kg	7千米/小时	49+	否

资料来源：各公司官网，平安证券研究所

大模型为人形机器人注入大脑，打开通用机器人可实现性。我们认为通用性受限是掣肘人形机器人商业化落地的原因，过去人形机器人动作普遍源自提前创建的行为库，限制了其在更泛化应用场景上的推广。人形机器人需要具备更强大的智能化和自主决策能力，而大模型具有庞大的先验知识库与强大的通识理解能力，可以满足人形机器人对于通用性的场景和技能的要求，实现复杂的运动控制、多模态感知和决策、自主学习和优化等。当人形机器人的使用者从工程师变成了一般用户，使用门槛降低将助力产业走向更大规模的爆发。

图表71 大模型赋能任务级交互：将自然语言指令快速转换为代码



资料来源：微软《ChatGPT for Robotics: Design Principles and Model Abilities》，平安证券研究所

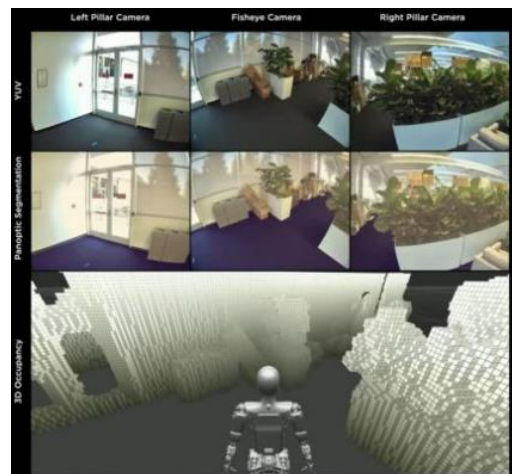
特斯拉 Optimus：移植智驾系统 FSD 端到端架构，实现高运动控制水平。最新特斯拉 Optimus 自主对物体进行分类的能力来自于端到端训练的神经网络，与特斯拉自动驾驶 FSD 的逻辑架构较为接近，视频作为输入，控制作为输出，并由此来控制各个部件和关节的移动。特斯拉通过打通自动驾驶和机器人的底层模型，仅靠视觉和编码器就能够完成非常高的运动控制水平，同时特斯拉 Dojo 平台也有望进一步增强其人形机器人的 AI 能力。

图表72 9月特斯拉 Optimus 展示按颜色分拣



资料来源：特斯拉官网，平安证券研究所

图表73 特斯拉人形机器人系统移植 FSD 智驾系统



资料来源：特斯拉官网，平安证券研究所

谷歌：机器人大大模型起源于 LLM，保持高速迭代。谷歌机器人大大模型由大语言模型演化而来，逐步由基础认知理解向底层操作控制发展，迭代速度极快。从 2022 年 4 月推出 SayCan 模型，初次引入大模型用于做任务理解和拆分，5 月推出 Gato 输出多模态感知能力，12 月推出 RT-1 使用传统神经网络的方法来执行 SayCan 任务，再到 2023 年 3 月推出 PaLM-E 结合自主可靠决策和多模态感知两种能力，6 月推出能够自主生成新数据集的 RoboCat，7 月 RT-2 将 VLM 大模型与 RT-1 的机器人执行数据集共同微调训练，10 月谷歌创建 Open X 数据集，融合目前最全的机器人场景与任务数据。

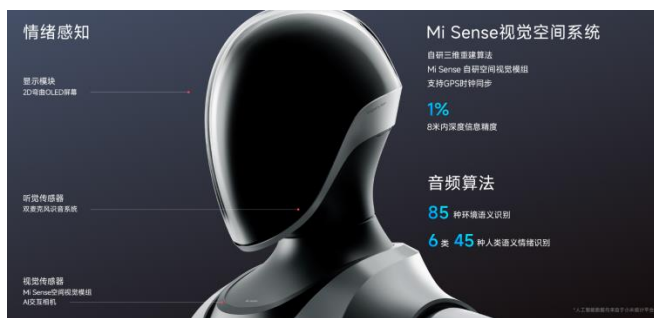
图表74 谷歌机器人大大模型梳理

大模型	发布时间	原理	创新意义
SayCan	2022年4月	Say主要由LLM组成，负责理解高层级任务的含义；Can由预训练得到的Affordance函数构成	首次引入LLM帮助机器人理解任务
Gato	2022年5月	将多模态信息token化后输入transformer架构	帮助机器人实现多模态感知
RT-1	2022年12月	将机器人轨迹数据输入transformer架构，可得到离散化动作token	收集了大量的机器人真实数据
PaLM-E	2023年3月	由谷歌大语言模型PaLM与拥有220亿个参数的最大视觉模型ViT-22B结合而成	让机器人能够接收持续的多模态的输入，本质为具有解码器的大语言模型
RoboCat	2023年6月	将多模态模型Gato与包含各种机器人手臂图像和动作的训练数据集相结合	可以利用少量人工收集的真实机器人数据集，自主生成新的训练数据集
RT-2	2023年7月	直接采用已有的视觉语言模型（VLM）作为主模型，再结合RT-1的视觉/语言/机器人动作数据集与互联网级别数据共同微调	兼具视觉语言模型和机器人动作的优势
RT-X	2023年10月	开放训练数据集Open X-Embodiment，并用此数据集训练前述的机器人模型RT-1和RT-2得到新的模型RT-1-X与RT-2-X	创造了一个仍然在持续增长的共享与开源的数据集

资料来源：谷歌《Do as I can, not as I say: Grounding language in robotic affordances》，谷歌《A Generalist Agent》，谷歌《RT-1: Robotics Transformer for Real-World Control at Scale》，谷歌《PaLM-E: An Embodied Multimodal Language Model》，谷歌《RoboCat: A Self-Improving Foundation Agent for Robotic Manipulation》，谷歌《RT-2: Vision-Language-Action Models》，谷歌《Open X-Embodiment: Robotic Learning Datasets and RT-X Models》，平安证券研究所

小米 CyberOne：家用机器人定位，AI 赋能情绪感知。2022 年 8 月，小米首款全尺寸人形仿生机器人 CyberOne 亮相，身高 177cm，体重 52kg，全身 21 个自由度，13 个关节位，运动时速 3.6km/h。人形机器人最主要的部分关节，以及其中的电机，最大模组峰值扭矩为 300N·m。雷军提到，CyberOne 以人工智能为内核，标准人形为载体，是小米对未来科技生态的一次探索，也是小米多元融合技术体系的新成果。CyberOne 在视觉和感知方面有较强的能力，不仅采用了弯曲 OLED 屏幕，还增加了双麦克风系统和 Mi Sense 空间视觉模组，以及 AI 交互系统，能够识别 85 种语义环境、45 种人类语义情绪。CyberOne 的最终定位是在娱乐陪伴、教育益智、管家、消费服务等服务机器人领域。

图表75 小米 CyberOne 在视觉和情绪感知方面的能力



资料来源：小米官网，平安证券研究所

图表76 小米 CyberOne 展示打架子鼓



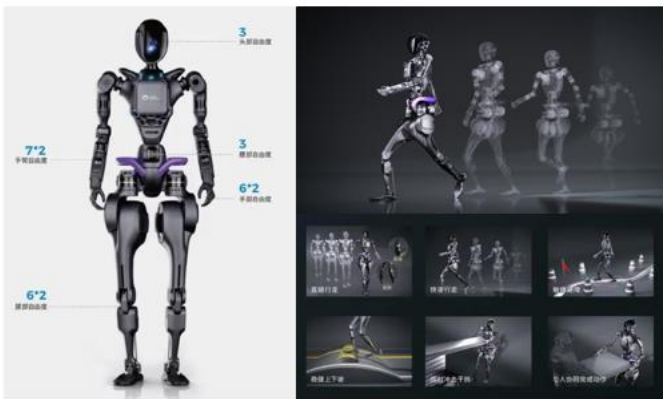
资料来源：小米官网，平安证券研究所

傅利叶 GR-1：支持大模型二次开发，商业化落地进度领先。傅利叶智能通用机器人 GR-1 身高 1.65 米，体重约 55 公斤，拥有高度仿生的躯干构型和拟人化的运动控制，搭载自主研发的 FSA 一体化执行器，全身总关节执行器数量达 44 个，最大关节峰值扭矩达 230N·m，从芯片到零部件其国产化率已超过 90%。GR-1 可在不稳定、无辅助、强干扰环境下，快速

行走（最大速度超过 5km/h）、后退行走、站走切换、原地转弯、扭腰、转头、跑步、避障、越障、上下坡等多种运动能力，根据计划 GR-1 会率先在康复陪护等民生场景中应用。9 月 26 日，傅利叶智能在公众号宣布 GR-1 开启对外预售，11 月 3 日有关负责人表示目前已开始陆续交付，商业化落地进度全面领先于同业。

海外人形机器人产业整体发展领先，当前以特斯拉为代表的汽车企业、以波士顿动力为代表的机器人企业、以谷歌为代表的人工智能企业均在发力人形机器人赛道。国内以小米、傅利叶等为代表的企业也正在切入，展望未来，伴随我国人形机器人厂商的积极布局和国产大模型能力的提升，我国人形机器人产业有望借助大模型实现通用性能提升，加快大规模商业化落地进程。

图表77 傅利叶 GR-1 具备快速行走等运动能力



资料来源：傅利叶智能官网，平安证券研究所

图表78 傅利叶 GR-1 支持底层功能 SDK、APP 源代码和基础预训练大模型开发，以及硬件扩展



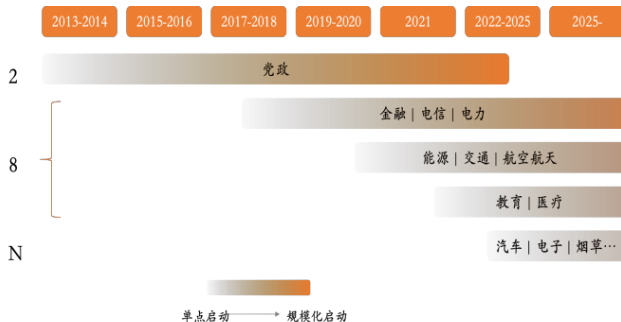
资料来源：傅利叶智能官网，平安证券研究所

五、信创工程市场空间巨大，华为将成为信创产业发展重要推动力

5.1 党政信创先行，金融、电信等八大行业紧随其后

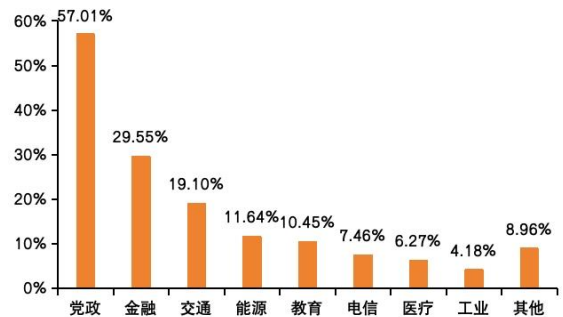
党政信创先行，行业信创大幕开启。我国信创产业的发展将按照 2+8+N 的落地体系逐步展开：党政是信创产业落地的起步领域，2013 年即开始在公文系统领域实行国产化替代，金融、电信等八大行业紧随其后，N 个行业的信创国产化替代也有望逐步启动。根据零壹智库数据，截止 2021 年 12 月，党政、金融的落地实践率分别为 57.0%和 29.6%，信创产业正从党政领域快速向其他行业延伸拓展。

图表79 2+8+N 信创发展进程



资料来源：亿欧智库，平安证券研究所

图表80 各领域信创应用落地进程



资料来源：工信部及下属机构，地方经信委，零壹智库，平安证券研究所

在党政信创领域，根据龙芯中科、中国长城、太极股份、中国软件等信创核心企业 2023 年半年报和季报的财报信息，我们判断，2023 年党政信创市场阶段调整。考虑到我国信创产业的发展进程，我们判断，2024 年，党政信创有望加速。在行业信创领域，以金融信创和电信信创来看，2023 年，信创订单在金融、电信等行业持续落地。我们判断，2024 年，金融、电信等八大行业信创将陆续进入常态化发展阶段。信创产业的发展，将为我国信创相关厂商带来良好的市场机遇。

金融信创招标进程加速，进入应用落地爆发期。金融业作为行业信创的试验田，其国产替代步伐逐渐从办公系统延伸至核心业务软件系统，已经进入规模化量产落地阶段。根据艾瑞咨询数据，2022 年试点单位已经扩充至 5000 家，并逐步向中小金融机构渗透。2023 年以来金融信创招标进度加快。根据 C114 通信网数据，近期，中信银行 65 亿信创大单落地，含包括服务器在内的 9 类设备。国内金融 IT 相关企业也在信创领域持续取得突破。根据顶点软件公司公告，顶点软件新交易体系 A5 信创版已在东吴证券、东海证券全面上线，华宝证券、华鑫证券等 3 家券商也在切换过程中。根据恒生电子公司公告，恒生电子截至 2023 年 6 月末，33 个主产品信创项目竣工覆盖率达到 29%，信创合同覆盖率提升到 58%。上半年还通过了数据库国产化认证工作，信创生态建设持续推进。金融行业对自主可控能力、技术供应链的韧性等安全稳定性的要求逐渐攀升，有望释放巨大存量替换需求。

图表81 2020-2022 年金融信创试点单位数量 (个)



资料来源：艾瑞咨询，平安证券研究所

图表82 中信银行 65 亿信创订单招标细则

分类	子分类	预估采购金额 (万元)
服务器	ARM芯片服务器	340094
	C86芯片服务器	102020
网络设备	数据中心级交换路由器	100000
	非数据中心级交换路由器	11850
	防火墙设备	9850
	波分设备	17000
存储设备	集中式ARM高端存储	39348
	分布式ARM对象存储	23400
	中高端存储光纤交换机	8972
合计		652534

资料来源：C114 通信网，平安证券研究所

电信信创硬件采购加速，AI 算力布局进一步打开信创市场空间。电信运营商作为国家通信安全的保障，早在 2019 年已开始有信创采购需求并不断加码投入。当前我国电信行业的信创建设仍以基础硬件为主，尤其信创服务器已实现规模化应用。2022 年 10 月，移动公示 2021-2022 年 PC 服务器集中采购第二批次 (标包 7-9) 结果，综合此前标包 1-6 与第一批次及补采，中国移动 PC 服务器集采中，海光芯片服务器 59982 台，鲲鹏芯片服务器 58901 台，整体信创芯片服务器占比达 41.43%。2023 年 2 月，联通公示《2022 年通用服务器集中采购项目》，合计采购服务器 55352 台，共计 46 亿元，信创 CPU 服务器金额占比 41.4%。海光芯片服务器 10300 台，鲲鹏芯片服务器 12108 台。

图表83 2020-2023 年我国电信行业信创建设进展

时间	2020年	2021年	2022年	2023年
项目招标	<ul style="list-style-type: none"> 三大运营商启动200亿集采大单，采购信创服务器92353台 信创CPU服务器占比超过20%，中国联通信创服务器占比超30% 	<ul style="list-style-type: none"> 3月，电信招标信创服务器，采购数量24823台 10月，移动启动2021-2022年第一批PC服务器采购项目，海光芯片服务器17164台，鲲鹏芯片服务器27063台 11月，电信2021-2022年服务器集采项目，鲲鹏、海光、飞腾芯片服务器53401台 	<ul style="list-style-type: none"> 1月，联通2022年云服务器集中采购招标，服务器44818台，预估规模327767万元；7月，联通软研院发布应用安全信创数据库购置招标，一为OLTP分布式信创数据库，预算836万元，二为OLTP集中式数据库，预算626万元 10月，移动公示2021-2022年PC服务器集中采购第二批次(标包7-9)结果，综合此前标包1-6与第一批次及补采，中国移动PC服务器集采中，海光芯片服务器59982台，鲲鹏芯片服务器58901台，整体信创芯片服务器占比达41.43% 	<ul style="list-style-type: none"> 1月，阿里云中标2022-2023年中国联通软研院数据中台计算能力大数据平台信创采购项目(商用大数据平台)，预算金额2351万 2月联通公示《2022年通用服务器集中采购项目》，合计采购服务器55352台，共计46亿元，信创CPU服务器金额占比41.4%，海光芯片服务器10300台，鲲鹏芯片服务器12108台
实验室建设及标准适配推进	<ul style="list-style-type: none"> 联通建立信息技术应用创新联合实验室，并于21年在广东、浙江、青海等7省建立信创攻关分基地 移动与华为构建的IT联合生态实验室运行；山东移动联合华为推出鲲鹏生态开放实验室 	<ul style="list-style-type: none"> 浙江移动信创适配实验室通过相关部门认证审查；中国移动成立信创实验室 电信浙江信创适配实验室通过相关部门认证审查；“中国电信上海信创基地”揭牌 	<ul style="list-style-type: none"> 电信举办信创合作联盟及云适配研讨沙龙 移动部署PAAS平台与飞腾服务器CPU(腾云S2500处理器)完成产品兼容性测试与认证 	<ul style="list-style-type: none"> 中国移动、华为、清华大学、中国电信、中国联通等11家初创成员共同发起的多样性算力产业及标准推进委员会正式成立

资料来源：艾瑞咨询，平安证券研究所

5.2 我国信创基础软硬件产品性能已实现从可用到好用，信创工程市场空间巨大

当前，我国信创基础软硬件产品性能提升，已实现从可用到好用。以芯片为例，龙芯中科和海光信息等公司均已发布新一代 CPU 芯片，芯片性能大幅提升，性能可对标国际主流桌面/服务器 CPU 芯片。

龙芯中科：坚持架构自研，具备较强生态支持能力。龙芯中科是国产 CPU 的引领者，其产品矩阵包括：龙芯 1 号、2 号、3 号系列处理器及配套桥片。龙芯 1 号为嵌入式专用处理器，面向物联网终端、仪器设备、数据采集等工控领域；龙芯 2 号为单芯片 SOC，面向网络设备、行业终端、智能制造等工控领域；龙芯 3 号为高性能通用处理器，面向桌面和服务器等信息化领域，配套桥片主要与龙芯 3 号系列处理器配套使用。

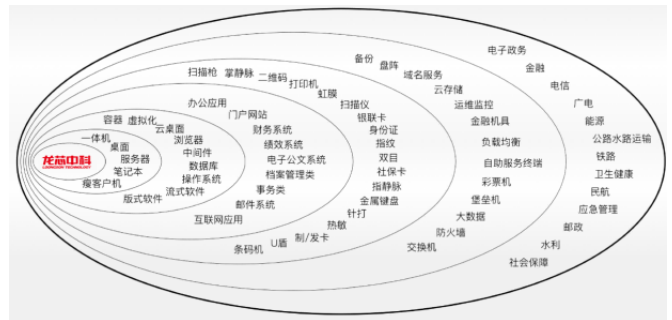
龙芯中科秉承独立自主和开放合作的运营模式，从指令集/IP 核授权、到芯片级/主板级开发以及系统内核应用等方面对生态伙伴进行全方位的开放支持。目前，龙芯中科的信息化合作伙伴上千家，可以提供从端到云的完整解决方案，相关软硬件开发人员数万人，已经形成较强的产业链与生态支撑能力。11 月 28 日，在龙芯中科 2023 新品发布会上，龙芯中科发布新一代国产桌面 CPU 龙芯 3A6000，龙芯 3A6000 处理器总体性能与 Intel 公司 2020 年上市的第 10 代酷睿四核处理器（Intel 酷睿 i3）相当。在发布会现场，基于龙芯 3A6000 处理器的众多整机产品发布。其中包括同方计算机、联想开天、超越科技、升腾资讯、攀升、国光信息、北方自控、视睿、海尔雷神、宝德网安、百信、黄河信产、大众电子、方正数码等 50 余家品牌企业，共同发布基于龙芯 3A6000 的桌面计算机、笔记本、板卡、存储产品、网络安全设备、工业控制计算机等产品。另外，在本次发布会上，龙芯中科还推出了打印机主控芯片龙芯 2P0500。

图表84 龙芯 3A6000 处理器产品性能参数

龙芯3A6000是龙芯第四代微架构首款处理器，采用自主龙芯指令集（LoongArch™），基于全新研制的LA664处理器核，龙芯3A6000处理器性能在龙芯3A5000处理器基础上实现大幅提升，单核定/浮点性能分别提升60%和90%，多核定/浮点性能分别提升100%和90%。龙芯3A6000处理器 SPEC CPU 2006 Base 单线程定/浮点分值分别达到43.1/54.6分，达到国际市场主流水平			
主频	2.0GHz-2.5GHz	峰值运算速度	240GFlops
物理核数	4	逻辑核数	8
处理器核	64位超量处理器核 LA664；支持 LoongArch™指令系统	典型功耗	38W@2.5GHz

资料来源：公司官网，平安证券研究所

图表85 龙芯生态布局



资料来源：公司官网，平安证券研究所

海光信息：兼容 X86 生态，对标国际主流高端处理器。公司高端处理器分为 CPU 系列和 DPU 系列，其中 CPU 为通用处理器，包括海光 3000、海光 5000、海光 7000 系列产品，内置多个处理器核心，集成通用高性能外设接口，适用于数据处理和事务处理类通用型应用。海光信息处理器产品兼容 X86 指令集及“类 CUDA”环境，软硬件生态丰富，能够很好的支持国内外主流的操作系统、云计算、数据库、大数据、人工智能、商用计算软件等，生态兼容性良好。

公司产品性能基本达到国际上同类型主流高端服务器处理器的水平，在国内处于相对领先地位。对比 Intel 在 2020 年(与海光 7285 同期)发布的 6 款至强铂金系列产品(能够反映 Intel2020 年发布的主流 CPU 产品的性能水平)，海光 7285 产品 CPU 的 SPEC CPU2017 的实测性能与 Intel 同期发布的主流处理器产品的实测性能总体相当。

图表86 公司与 Intel 服务器 CPU 的参数对比

产品名称	发布时间	4路测试结果		双路测试结果		性能差异(Intel数据/海光数据-1)	
		Speccpu_INT	Speccpu_FP	Speccpu_INT	Speccpu_FP	Speccpu_INT	Speccpu_FP
Intel8380HL(铂金)	2020Q2	784	657	392	329	12.64%	6.66%
Intel8380H(铂金)	2020Q2	784	653	392	327	12.64%	6.01%
Intel8376HL(铂金)	2020Q2	765	641	383	321	9.91%	4.06%
Intel8376H(铂金)	2020Q2	756	643	378	322	8.62%	4.38%
海光7285	2020Q1	-	-	348	308	-	-
Intel8360HL(铂金)	2020Q3	690	599	345	300	-0.86%	-2.76%
Intel8360H(铂金)	2020Q3	688	597	344	344	-1.15%	-3.08%

资料来源：海光信息公司招股说明书，平安证券研究所

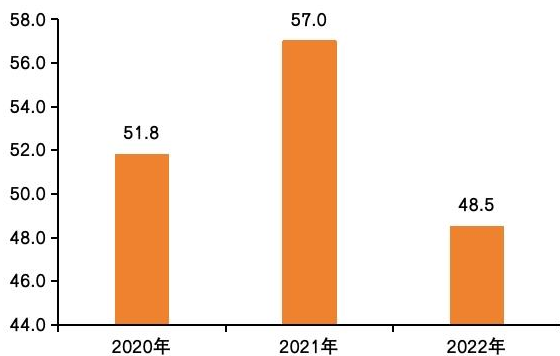
我国信创工程市场空间巨大。PC 和服务器合计出货量超过 3000 万台。信创产业链包括 PC 和服务器整机，以及芯片、操作系统、中间件、数据库、办公软件等核心组件。需求量基本假设和测算过程如下：

1) 信创 PC：根据国家统计局数据，党政机关人数约 2000 万，我们假设党政人均 PC0.7 台，且已经有 600 万台 PC 完成替代，剩余 PC 假设分 6 年替换完成；八大行业人数约 3000 万，假设人均 PC0.6 台，分 6 年替换完成。按照 2023-2028 年 10%、20%、20%、20%、20%、10%的比例估算，2023-2028 年信创 PC 的市场容量分别为 260 万台、520 万台、520 万台、520 万台、520 万台、260 万台。

2) 信创服务器：根据 IDC 数据，2021 年我国服务器总出货量为 391.1 万台，以此为基数，假设未来服务器出货量年均复合增长率为 5%，假设 2023 年信创服务器出货量能占到服务器总出货量的 12%，并自 2024 年起占比每年同比提高 3 个百分点。以此估算，2023-2028 年，信创服务器市场容量分别为 52 万台、68 万台、86 万台、105 万台、126 万台、149 万台。

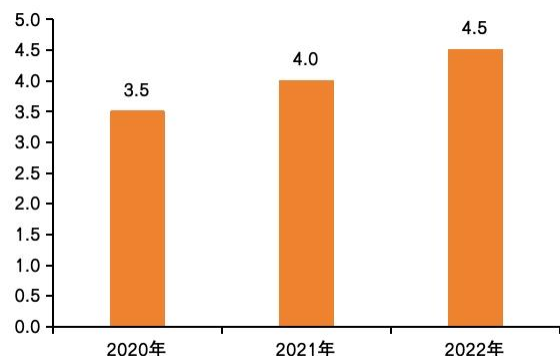
根据我们的测算，预计 2023-2028 年，信创 PC 的市场容量分别为 260 万台、520 万台、520 万台、520 万台、520 万台、260 万台，合计为 2600 万台；服务器市场容量分别为 52 万台、68 万台、86 万台、105 万台、126 万台、149 万台，合计为 586 万台。2023-2028 年，PC 和服务器合计出货量为 3186 万台。

图表87 2020-2022 年我国 PC 出货量（百万台）



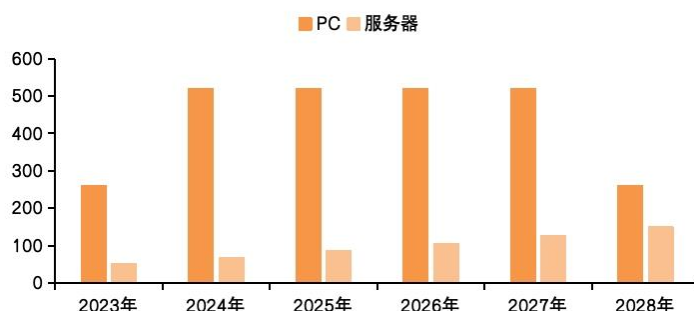
资料来源：Canalys，平安证券研究所

图表88 2020-2022 年我国服务器销售量（百万台）



资料来源：赛迪顾问，平安证券研究所

图表89 信创产业 PC、服务器市场容量预测（万台）



资料来源：平安证券研究所

PC 和服务器合计市场规模超过 4000 亿元。信创市场规模测算如下：

1) PC 市场：结合市场报价以及招标信息，我们假设每台 PC 单价为 5500 元/台，并保持稳定，则 2023-2028 年，信创 PC 市场规模分别为 143 亿元、286 亿元、286 亿元、286 亿元、286 亿元、143 亿元，6 年市场规模合计 1430 亿元。每台 PC 对应 1 块处理器芯片、1 套操作系统、1 套办公软件，结合市场均价测算（均为 500 元左右），假设 PC 处理器芯片、操作系统、办公软件单价均为 500 元，则 2023-2028 年，PC 处理器芯片、操作系统、办公软件市场规模如下所示。

图表90 信创 PC 市场规模预测（亿元）

年份	2023年	2024年	2025年	2026年	2027年	2028年
整机	143	286	286	286	286	143
芯片	13	26	26	26	26	13
操作系统	13	26	26	26	26	13
办公软件	13	26	26	26	26	13

资料来源：平安证券研究所

2) 信创服务器市场：结合市场报价，我们假设每台服务器单价为 5 万元/台，并保持稳定，则 2023-2028 年，信创服务器市场规模分别为 260 亿元、340 亿元、430 亿元、525 亿元、630 亿元、745 亿元，6 年市场规模合计 2930 亿元。PC 和服务器 6 年合计市场规模 4360 亿元。每台服务器对应 1 套操作系统，假设每台服务器平均有 2 个处理器芯片、每 3 台服务器中有 1 台安装数据库软件、每 3 台服务器中有 1 台安装中间件，服务器芯片、操作系统、数据库、中间件的单价假设分别为 5000 元/个、3000 元/套、6 万元/套、4 万元/套，则 2023-2028 年，服务器芯片、操作系统、数据库、中间件市场规模如下所示。

图表91 信创服务器市场规模预测（亿元）

年份	2023年	2024年	2025年	2026年	2027年	2028年
整机	260	340	430	525	630	745
芯片	52	68	86	105	126	149
操作系统	16	20	26	32	38	45
数据库	102	138	174	210	252	300
中间件	68	92	116	140	168	200

资料来源：平安证券研究所

5.3 华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力

在 CPU 芯片方面，华为具有鲲鹏 920 系列产品和鲲鹏 916 系列产品，其中鲲鹏 920 系列产品单处理器整型计算性能，相比上一代提升 2.9 倍；在整机方面，华为有华鲲振宇、新华三、神州数码、宝德、清华同方等 10 多家合作伙伴；在服务器操作系统方面，公司欧拉开源操作系统（openEuler，简称“欧拉”）长生命周期版本更新到 openEuler 22.03 LTS SP2 版本，创新版本更新到 openEuler 23.09 创新版本；在数据库方面，华为 openGauss 的长期支持版本（LTS）更新到 openGauss 5.0.0(LTS)，社区创新版本（Preview）更新到 openGauss 5.1.0(Preview)。当前，华为已经建立了涵盖芯片、整机、操作系统、数据库等基础软硬件产品的较为完整的信创产品体系，我们认为，华为将是我国信创产业发展的重要推动力。

图表92 华为信创产品体系（示意图）



资料来源：华为鲲鹏社区网站，平安证券研究所

■ 鲲鹏芯片：计算产业生态完备，打造国产替代算力底座

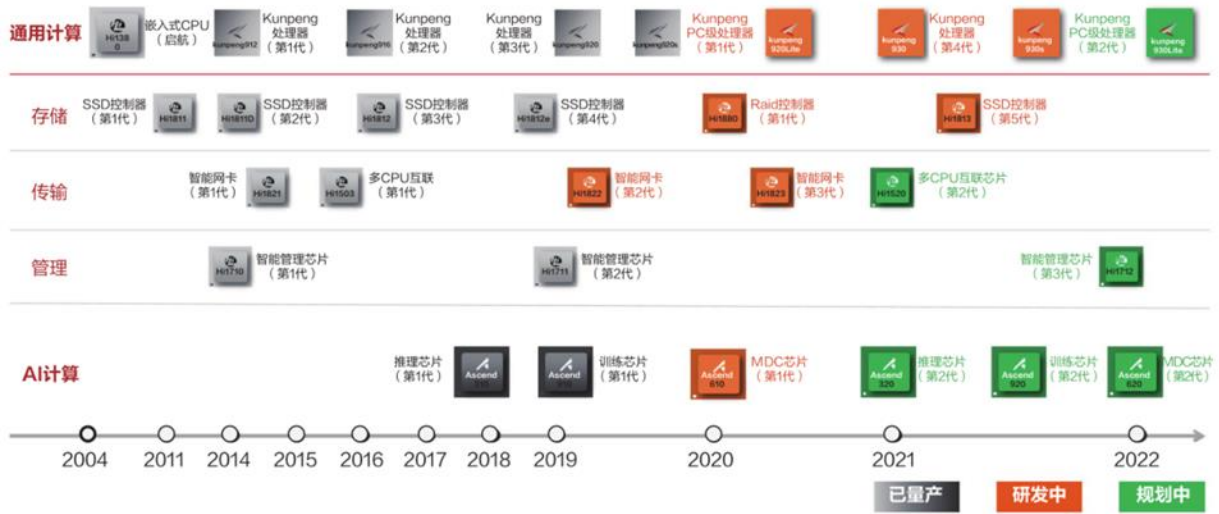
华为持续投入鲲鹏处理器，以量产一代、研发一代、规划一代的节奏稳步推进，并围绕鲲鹏处理器打造了“算、存、传、管、智”五个子系统芯片族。鲲鹏 920 是目前业界领先的 ARM 处理器，SPECint Benchmark 评分显示，对标 intel 旗舰级处理器 Xeon Platinum 8180，48 核版鲲鹏 920 与其性能相当，功耗低 20%；64 核版鲲鹏 920 性能超 intel 约 33%。作为低功耗、高性能的 ARM 处理器，为鲲鹏服务器主板及整机产品提供芯片支撑，是鲲鹏计算产业的底座。此外，华为基于鲲鹏处理器打造了 TaiShan 服务器，赋能整个鲲鹏计算产业链、构建完整生态。鲲鹏计算产业链与 ARM 共享全球生态，协同加速发展。

图表93 鲲鹏计算产业定义



资料来源：华为云官网，平安证券研究所

图表94 鲲鹏“算、存、传、管、智”芯片族



资料来源:《鲲鹏计算产业发展白皮书》, 平安证券研究所

■ 鲲鹏整机: 协同合作伙伴拓展市场, 华为整机发展势头迅猛

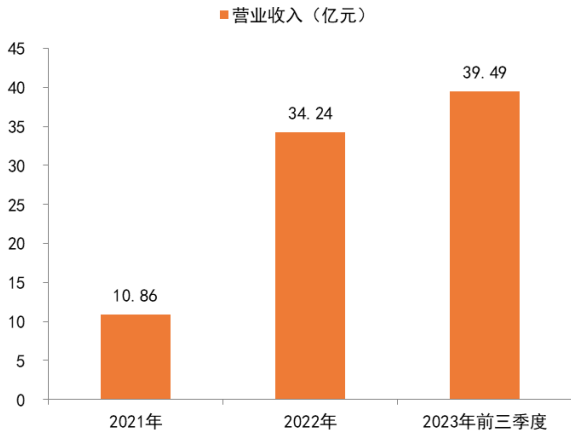
华为与合作伙伴协同发展, 在整机信创市场发展势头迅猛。除了自研泰山 (TaiShan) 服务器, 华为与合作伙伴协同发展, 广泛依靠合作伙伴拓展整机信创市场。根据鲲鹏社区网站信息, 华为目前有华鲲振宇、新华三、神州数码、宝德、清华同方等 10 多家合作伙伴。以华鲲振宇和神州数码为例, 华为整机在信创市场发展势头迅猛。根据高新发展公司公告, 2021 年、2022 年、2023 年前三季度, 华鲲振宇营收分别为 10.86 亿元、34.24 亿元、39.49 亿元, 营收跳跃式增长。根据神州数码公司公告, 2020-2022 年, 神州数码自主品牌业务收入分别为 7.28 亿元、16.50 亿元、25.70 亿元, 2023 年前三季度, 神州数码信创业务 (神州鲲泰服务器) 收入超 20 亿, 同比增长 69%。神州数码信创业务已经实现了行业客户的全面突破, 其中运营商三大客户实现全覆盖。

图表95 鲲鹏合作伙伴整机产品



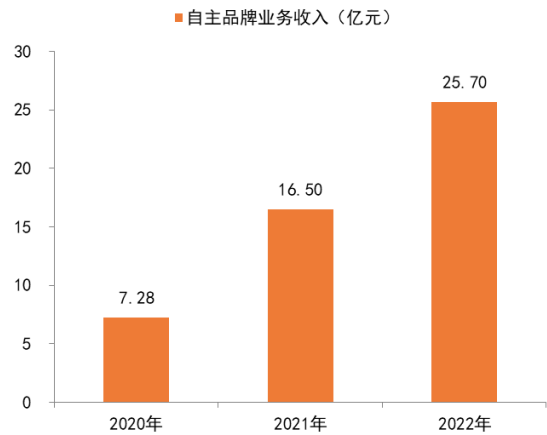
资料来源: 鲲鹏社区网站, 平安证券研究所

图表96 华鲲振宇 2021年-2023Q3 营收情况



资料来源：高新发展公司公告，Wind，平安证券研究所

图表97 神州数码 2020-2022年自主品牌业务营收情况

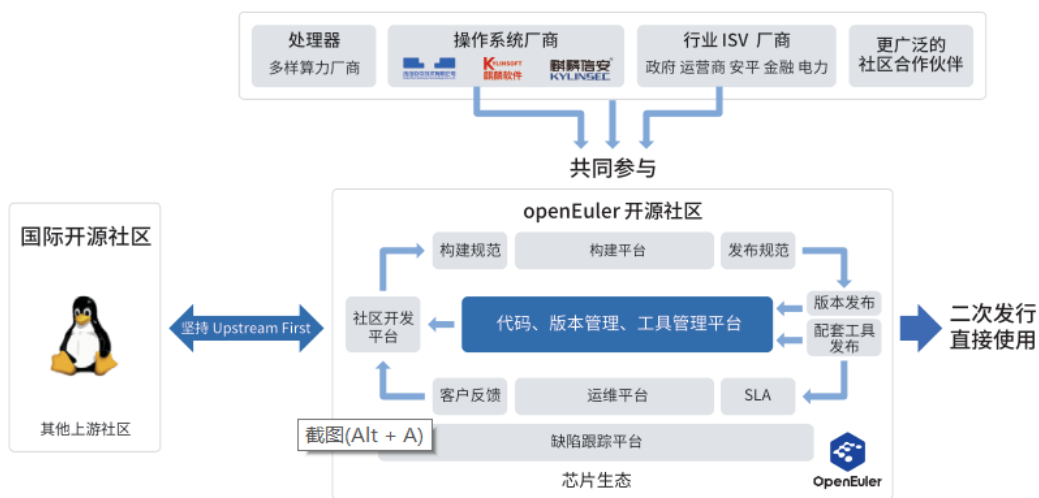


资料来源：神州数码公司公告，Wind，平安证券研究所

■ 欧拉：openEuler 开源共建，长生命周期版本和创新版本交叉迭代

欧拉是面向数字基础设施的操作系统，可广泛部署于服务器、云计算、边缘计算、嵌入式等各种形态设备，应用场景覆盖 IT、CT 和 OT。2019 年底华为决定将欧拉操作系统正式开源，并命名为 openEuler。openEuler 的版本管理包括两种版本：长生命周期版本（Long Term Support，简称为 LTS）和创新版本。其中，长生命周期版本每两年发布一次，在创新版本基础上提供长生命周期管理，维护性能、可靠性和兼容性，该版本为企业级用户提供一个安全稳定可靠的操作系统；创新版本在长生命周期版本之间每半年发布一次，集成 openEuler 以及其他社区最新版本技术进展，目前 openEuler 长生命周期版本更新到 openEuler 22.03 LTS SP2，创新版本更新到 openEuler 23.09。当前，与 openEuler 合作的操作系统厂商较多，推动了 openEuler 装机量的迅速提升。根据 openEuler 网站信息，当前国内外已经有超过 20 家厂商发布 openEuler 路线的操作系统商业发行版。

图表98 openEuler 平台框架



资料来源：openEuler 社区网站《openEuler 23.09 技术白皮书》，平安证券研究所

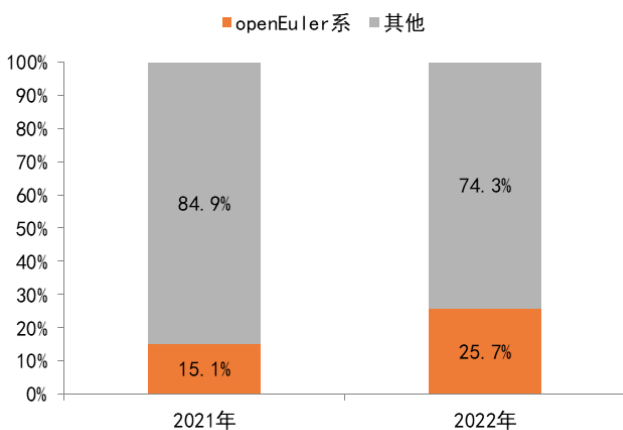
图表99 国内外 20 多家厂商发布 openEuler 路线的操作系统商业发行版

发布 openEuler 商业发行版的厂商			
 宝德	 FUSION	 烽火星空	 恒生
 Intewell	 凝思	 普华基础软件	 麒麟软件
 麒麟信安	 润和软件	 SUSE	 turbolinux
	 依信软件技术有限公司	 HBC	 ISCAS
 iSOFTSTONE	 CS&S	 ThunderSoft	 中科方德
 中科红旗	 SEAWAY		

资料来源: openEuler 社区网站, 平安证券研究所

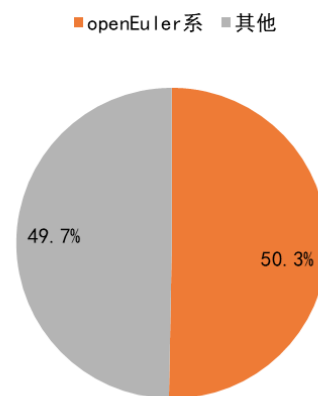
openEuler 装机量市占率已跨越操作系统的生态拐点, 在政府、电信、金融领域表现突出。根据沙利文《2022 年中国服务器操作系统行业市场研究报告》数据, 2022 年, 中国服务器操作系统装机量达到 401.2 万套, 同比增长 13.9%。其中, openEuler 系服务器操作系统市场份额达到 25.7%, 市场份额已经跨过 16% 的生态拐点。在政府、电信、金融领域, openEuler 系服务器操作系统表现突出, 2022 年装机量市场份额分别为 50.3%、40.6%、30.6%, 市场份额均排名第一。随着信创产业的持续推进, 我们判断, openEuler 系服务器操作系统装机量市场份额将持续提升。

图表100 2021-2022 年 openEuler 系服务器操作系统装机量市场份额



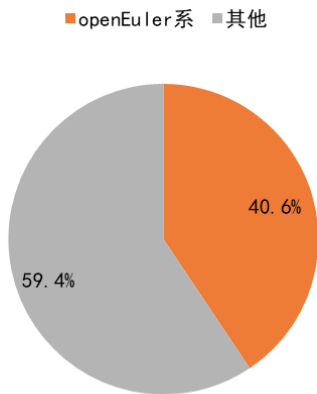
资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

图表101 openEuler 系服务器操作系统政府领域装机量市场份额



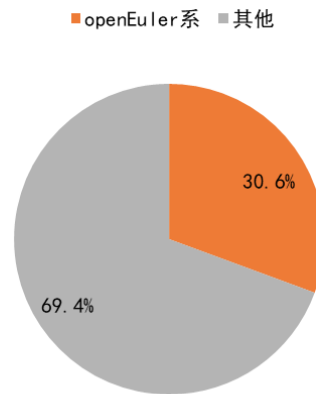
资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

图表102 openEuler 系服务器操作系统电信领域装机量市场份额



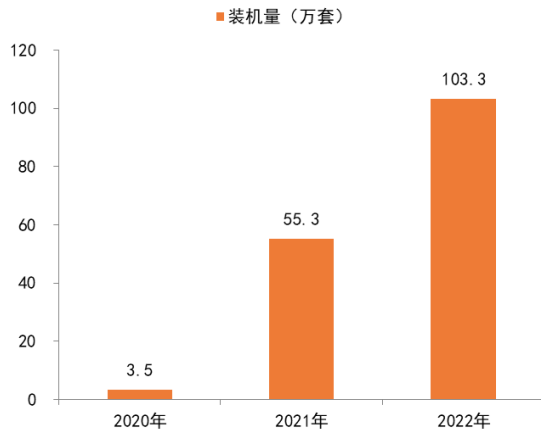
资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

图表103 openEuler 系服务器操作系统金融领域装机量市场份额



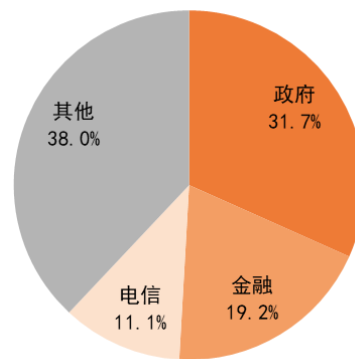
资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

图表104 2020-2022 年中国 openEuler 系服务器操作系统装机量



资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

图表105 2022 年中国 openEuler 系服务器操作系统装机量结构



资料来源: openEuler 社区网站, 沙利文《2022 年中国服务器操作系统行业市场研究报告》, 平安证券研究所

■ 华为数据库: 从内部自用 to 共建生态, 已经历二十多年的发展

华为数据库从内部自用发展到共建生态, 已有二十多年的发展历程。根据 openGauss 网站信息, 华为数据库产品的发展经历了四个发展阶段: 1) 内部自用, 2001-2011 年, 华为数据库是华为内部自用的企业级内存数据库; 2) 产品化, 2011-2019 年, 华为数据库支撑公司内部 40+ 主力产品, 在全球 70+ 运营商规模商用 3 万+套, 服务全球 20+ 亿人口; 3) 云&开源, 2019-2020 年, 2019 年 5 月 15 日, 华为 GaussDB 正式全球发布, 兼容行业主流生态, 完成金融等行业对接, 2020 年 6 月 30 日, 华为正式宣布开源数据库能力, 开放 openGauss 数据库源代码, 并成立 openGauss 开源社区; 4) 生态构建, 2021 年至今, openGauss 分享企业级数据管理能力, 引领生态建设。

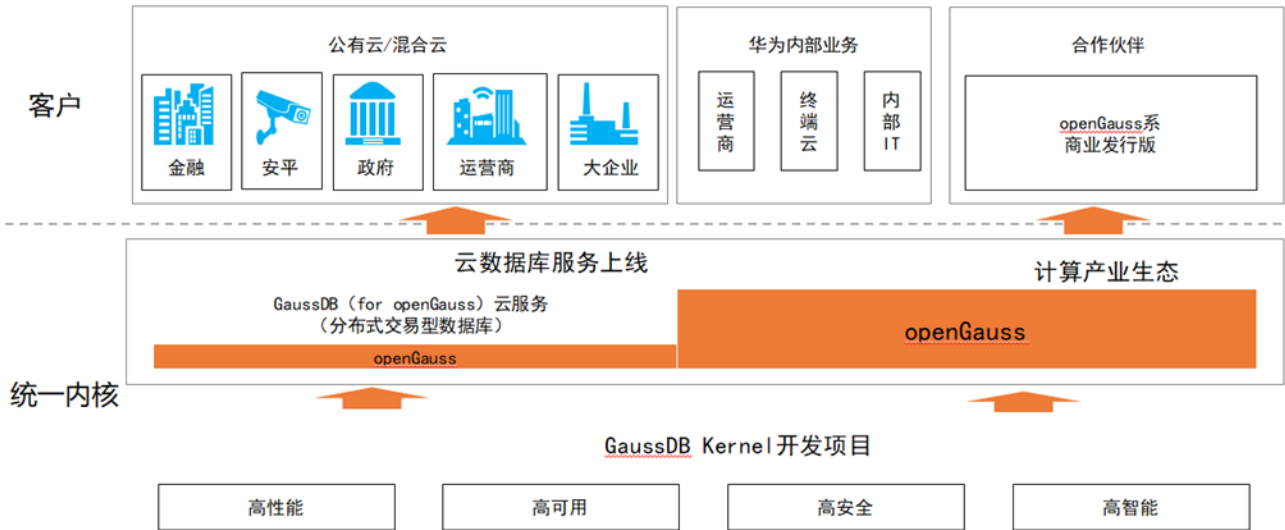
图表106 华为数据库发展历程

发展阶段	时间	相关内容
内部自用	2001-2011年	<ul style="list-style-type: none"> 企业级内存数据库
产品化	2011-2019年	<ul style="list-style-type: none"> G行核心数据仓库、DWS华为云商用 Z行核心业务系统替换商业数据库 支撑公司内部40+主力产品，在全球70+运营商规模商用3万+套，服务全球20+亿人口
云&开源	2019-2020年	<ul style="list-style-type: none"> 2019. 5. 15 GaussDB全球发布 构筑合作伙伴生态 兼容行业主流生态，完成金融等行业对接 2020. 6. 30 openGauss开源
生态构建	2021年至今	<ul style="list-style-type: none"> 分享企业级数据管理能力 引领生态建设 促进数据库教育事业发展

资料来源： openGauss 网站《openGauss 技术架构》，平安证券研究所

华为数据库的公司内部配套、公有云的 GaussDB、开源 openGauss 共享代码基线，内核将长期演进。华为数据库的发展是商用+自用+开源相结合，公司内部配套、公有云的 GaussDB、开源 openGauss 共享代码基线，商用、自用、开源是同一个单机内核，内核持续迭代演进，持续增强华为数据库的市场竞争力。openGauss 社区版本分为长期支持版本(LTS)和创新版本(Preview)，其中长期支持版本为规模上线使用，发布间隔周期为 1 年，提供 3 年社区支持，当前最新版本更新到 openGauss 5.0.0 (LTS)；创新版本为联创测试使用，发布间隔周期为 1 年，提供 6 个月社区支持，当前最新版本更新到 openGauss 5.1.0 (Preview)。根据 openGauss 网站信息，截至目前，有 20 款基于 openGauss 开发的商业发行版通过了 openGauss 的社区认证。

图表107 华为数据库的公司内部配套、公有云的 GaussDB、开源 openGauss 共享代码基线



资料来源： openGauss 网站《openGauss 技术架构》，平安证券研究所

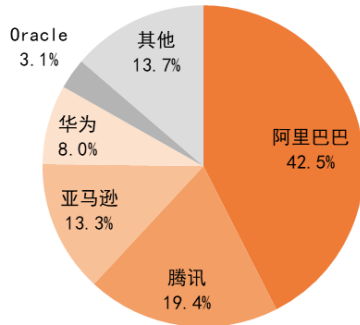
图表108 基于 openGauss 开发的商业发行版

合作伙伴公司名称	认证的商业发行版	合作伙伴公司名称	认证的商业发行版
北京沐融信息科技股份有限公司	MuDB 1.0.0	苏大信创（青岛）信息科技有限公司	【全真数据库平台】V5.0.0
北京海量数据技术股份有限公司	【Vastbase G100】V2.2	中软国际科技服务有限公司	CSIDB V1.0.0
云和恩墨（北京）信息技术有限公司	MogDB 2.0.1	宏数通信息科技（苏州）有限公司	宏数通数据库系统 V5.0.0
北京太阳塔信息科技有限公司	LNxDB-RDS V3.4	北京华宇信息技术有限公司	ArteryBase数据库管理系统 V3.6
成都虚谷伟业科技有限公司	有容数据库 V1.0	云和恩墨（北京）信息技术有限公司	MogDB 3.0.0
天津神舟通用数据技术有限公司	神通数据库管理系统 V7.0.21	软通动力信息技术（集团）股份有限公司	天鹤数据库 ISSEDB 5.0.0
天津南大通用数据技术股份有限公司	GBase 8c V5	云和恩墨（北京）信息技术有限公司	MogDB 5.0.0
超聚变数字技术有限公司	FusionDB 22	京东科技信息技术有限公司	StarDB 5.0.0
天津南大通用数据技术股份有限公司	GBase 8s V8.8.5	南京烽火星空通信发展有限公司	FitRDB数据库 V1.0.0
北京海量数据技术股份有限公司	【Vastbase G100】V2.2.10	首杭数字科技（山东）有限公司	AScoDb数据库V1.0.0

资料来源：openGauss 网站，平安证券研究所

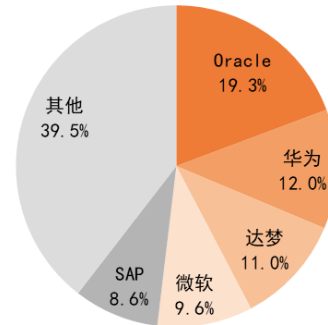
华为高斯数据库 GaussDB 是华为自研的自用+商用数据库版本，广泛应用于金融等行业，在我国关系型数据库市场份额领先。2023 年 6 月，在华为全球智慧金融峰会 2023 上，华为正式发布新一代分布式数据库 GaussDB。根据峰会的信息，GaussDB 已在华为内部 IT 系统和多个行业核心业务系统得到应用，并在金融、能源等行业的核心业务系统得到广泛应用，支撑了中国工商银行、中国邮政储蓄银行、中国建设银行、中国农业银行等国有大行的核心业务，此外，也服务了陕西省财政厅、中海油、国家管网、国网陕西电力、首都公路发展集团等关键行业企业。华为数据库在我国关系数据库市场份额领先。根据 IDC 数据，2021 年下半年，华为在我国关系型数据库软件市场（公有云模式）厂商份额排名第四，在我国关系型数据库软件市场（本地部署模式）厂商份额排名第二。我们认为，随着信创产业的发展以及 GaussDB 在下游行业的持续推广，华为在我国关系型数据库软件市场的厂商份额有望持续提升。

图表109 2021年下半年我国关系型数据库软件市场
(公有云模式)厂商份额



资料来源: IDC, 墨天轮, 平安证券研究所

图表110 2021年下半年我国关系型数据库软件市场
(本地部署模式)厂商份额



资料来源: IDC, 墨天轮, 平安证券研究所

六、投资建议及风险提示

6.1 投资建议：看好 AIGC 和信创主题的投资机会

投资逻辑：展望 2024 年，在政策、市场、技术等多重因素的驱动下，我们看好 AIGC 产业和信创产业的发展，看好两条投资主线：（一）当前，全球范围内大模型领域的竞争依然白热化，这将持续拔升大模型的整体能力水平，提升大模型的多模态能力。大模型需要大算力，大模型算法的迭代升级将为全球和中国 AI 算力市场的增长提供强劲动力。在美国对华半导体出口管制升级的背景下，我国国产 AI 芯片产业链将加快成熟，相关 AI 芯片和服务器厂商将深度受益。另外，在应用端，我国国产大模型能力持续升级，AIGC 产业未来发展前景广阔。（二）信创产业的发展。我国信创产业的发展将按照 2+8+N 的落地体系逐步展开，我们判断，2024 年，党政信创将有望加速，金融、电信等八大行业信创将陆续进入常态化。华为拥有较为完整的信创产品体系，将是我国信创产业发展的重要推动力。信创华为为产业链值得重点关注。我们看好 AIGC 和信创主题的投资机会，维持对计算机行业的“强于大市”评级。

投资标的：AIGC 主题，1) 在算力方面，推荐浪潮信息、中科曙光、紫光股份、海光信息，建议关注工业富联、寒武纪、景嘉微；2) 在算法方面，推荐科大讯飞；2) 在应用场景方面，强烈推荐中科创达，推荐金山办公、恒生电子，建议关注拓尔思、同花顺、彩讯股份、航天宏图、盛视科技；4) 在网络安全方面，强烈推荐启明星辰。信创主题，推荐龙芯中科、太极股份、顶点软件，建议关注神州数码、高新发展、中国软件、麒麟信安、诚迈科技、中国长城、软通动力、海量数据、普联软件、远光软件。

6.2 风险提示

1) AI 算力供应链风险上升。美国对华半导体出口管制升级，将倒逼我国国产 AI 芯片产业链加快成熟。但如果我国国产 AI 芯片的迭代速度不达预期，将影响我国 AI 算力的发展。

2) 大模型产品的应用落地低于预期。当前，我国国产大模型已经开始在教育、医疗、汽车、办公、工业、智能硬件等 B 端和 C 端应用场景持续落地，如果产品的市场拓展不及预期，则我国大模型产品的应用落地将存在低于预期的风险。

3) 信创产业发展不及预期。当前，我国信创基础软硬件已经实现从可用到好用，但如果信创基础软硬件与行业客户业务系统的适配进度低于预期，则将影响信创向业务系统的拓展，我国信创产业将存在发展不及预期的风险。

平安证券研究所投资评级：

股票投资评级：

- 强烈推荐（预计 6 个月内，股价表现强于市场表现 20% 以上）
- 推 荐（预计 6 个月内，股价表现强于市场表现 10% 至 20% 之间）
- 中 性（预计 6 个月内，股价表现相对市场表现在 $\pm 10\%$ 之间）
- 回 避（预计 6 个月内，股价表现弱于市场表现 10% 以上）

行业投资评级：

- 强于大市（预计 6 个月内，行业指数表现强于市场表现 5% 以上）
- 中 性（预计 6 个月内，行业指数表现相对市场表现在 $\pm 5\%$ 之间）
- 弱于大市（预计 6 个月内，行业指数表现弱于市场表现 5% 以上）

公司声明及风险提示：

负责撰写此报告的分析师(一人或多人)就本研究报告确认：本人具有中国证券业协会授予的证券投资咨询执业资格。

平安证券股份有限公司具备证券投资咨询业务资格。本公司研究报告是针对与公司签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本公司研究报告仅提供给上述特定客户，并不面向公众发布。未经书面授权刊载或者转发的，本公司将采取维权措施追究其侵权责任。

证券市场是一个风险无时不在的市场。您在进行证券交易时存在赢利的可能，也存在亏损的风险。请您务必对此有清醒的认识，认真考虑是否进行证券交易。

市场有风险，投资需谨慎。

免责条款：

此报告旨在发给平安证券股份有限公司（以下简称“平安证券”）的特定客户及其他专业人士。未经平安证券事先书面明文批准，不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其他人。

此报告所载资料的来源及观点的出处皆被平安证券认为可靠，但平安证券不能担保其准确性或完整性，报告中的信息或所表达观点不构成所述证券买卖的出价或询价，报告内容仅供参考。平安证券不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能仅依靠此报告而取代行使独立判断。

平安证券可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的判断，可随时更改。此报告所指的证券价格、价值及收入可跌可升。为免生疑问，此报告所载观点并不代表平安证券的立场。

平安证券在法律许可的情况下可能参与此报告所提及的发行商的投资银行业务或投资其发行的证券。

平安证券股份有限公司 2023 版权所有。保留一切权利。

平安证券

平安证券研究所

电话：4008866338

深圳

深圳市福田区益田路 5023 号平安金融
融中心 B 座 25 层

上海

上海市陆家嘴环路 1333 号平安金融
大厦 26 楼

北京

北京市丰台区金泽西路 4 号院 1 号楼
丽泽平安金融中心 B 座 25 层