

中国平安 PINGAN

专业·价值

专业 让生活更简单

证券研究报告

业绩承压、需求回暖，看好新质生产力 ——计算机行业2024年中期策略报告

计算机行业 强于大市（维持）

平安证券研究所 计算机团队

分析师： 闫磊 S1060519100002（证券投资咨询）

分析师： 黄韦涵 S1060523070003（证券投资咨询）

研究助理：王佳一 S1060123070023（一般证券从业资格）

2024年6月23日

请务必阅读正文后免责条款

平安证券

核心摘要

- **行业回顾及投资逻辑：**2024年一季度，行业上市公司业绩表现相比2023年持续承压。但从营收端来看，行业的下游需求已呈现回暖的趋势。市场表现来看，计算机指数表现不佳，涨幅跑输沪深300指数，且在31个申万一级行业排名倒数第二，受行业涨幅表现不佳的影响，行业的基金持仓呈现下降趋势。当前，行业估值处于历史相对较低水平。展望2024年，我们认为，随着超长期特别国债对科技创新的推动以及国内宏观经济回升向好，计算机行业的下游需求有望持续回暖，行业上市公司的业绩有望向好。我们看好AIGC、低空经济等新质生产力的未来发展，认为AIGC和低空经济主题将在二级市场获得更好的投资机会，维持对计算机行业的“强于大市”评级。
- **算力：算力是AIGC产业发展的底座，市场需求仍然旺盛。**当前，全球范围内大模型领域的竞争依然白热化。大模型的持续迭代升级将为AI芯片及服务器发展提供强劲动力。在供给端，全球AI芯片领军企业英伟达不断升级迭代高性能AI芯片，持续拔升全球大算力AI芯片的性能天花板；国内华为、寒武纪、海光信息等AI芯片企业持续迭代AI芯片产品，增强自身的市场竞争力。在需求端，根据我们的估算，大模型将为全球和中国AI服务器带来约985.82亿美元和439.18亿美元的市场空间，市场空间巨大。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。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升，在国内市场的市占率有望加速提升，我国国产AI芯片厂商将加快发展，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。
- **算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降。**我国大模型厂商持续迭代升级算法能力，在通用大模型领域，以阿里云通义千问2.5、智谱GLM-4等为代表的国产大模型能力已可对标GPT-4/GPT-4 Turbo。2024年以来，全球大模型价格呈现下降趋势。全球大模型厂商由单方面的性能角逐，逐渐转向性能与实用性并重的发展趋势，大模型纷纷实现提速降价，更加注重性价比，将有利于未来下游大模型部署成本的下降。
- **应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大。**下游大模型的部署成本将下降，将有望加速大模型在各场景的应用落地。手机厂商围绕AI影像、智能通话、智能搜索等功能已经提前布局AI应用，2024年起新一代AI手机渗透率将持续提升，与此同时，各品牌整机厂商纷纷加码AI PC产品，全球PC产业将稳步迈入AI时代，端侧智能发展前景广阔，成为AIGC应用进度较为靠前的景气赛道。另外，大模型已经开始在搜索、办公、传媒、医药、金融、教育、智能汽车等B端和C端应用场景持续落地。我们认为，随着我国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产业未来应用潜力巨大。
- **低空经济：政策持续催化，未来展翼腾飞。**低空经济是以eVTOL飞行器为重要牵引的一种综合性经济形态，无人机与通用航空产业为低空经济发展的核心产业。当前，政策推动低空经济进入快速发展期。根据赛迪顾问统计，2023年我国低空经济规模达到5059.5亿元，同比增长33.8%，预测到2026年我国低空经济规模将突破1万亿元。我们认为，在国家高度重视和地方积极布局的背景下，低空经济未来发展前景广阔。低空飞行器研发制造、低空基础设施、低空飞行保障服务等细分行业均将受益于低空经济的发展。
- **投资建议：**展望2024年下半年，我们认为，大模型的持续迭代升级将为AI算力带来巨大市场空间。同时，全球大模型价格呈现下降趋势，将有利于未来下游大模型部署成本的下降，将有望加速大模型在各场景的应用落地。参考我国数字经济的巨大体量，我国AIGC产业未来应用潜力巨大。我们坚定看好AIGC产业的发展。另外，政策推动低空经济进入快速发展期，在国家高度重视和地方积极布局的背景下，我们认为我国低空经济产业未来发展前景广阔。我们看好AIGC和低空经济产业发展。1) 算力方面，推荐浪潮信息、中科曙光、紫光股份、神州数码、海光信息、龙芯中科，建议关注寒武纪、景嘉微、拓维信息、软通动力；2) 算法方面，推荐科大讯飞；3) 应用场景方面，强烈推荐中科创达、恒生电子，推荐金山办公、顶点软件，建议关注万兴科技、同花顺、彩讯股份；4) 网络安全方面，强烈推荐启明星辰。低空经济产业方面，建议关注以下产业链相关标的：1) eVTOL整机厂商：亿航智能；2) 基础设施服务商：中科星图、莱斯信息、纳睿雷达；3) 无人机厂商：纵横股份；4) 通航运营：中信海直；5) 交通规划：深城交。
- **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) 低空经济产业发展不及预期。



目录CONTENTS

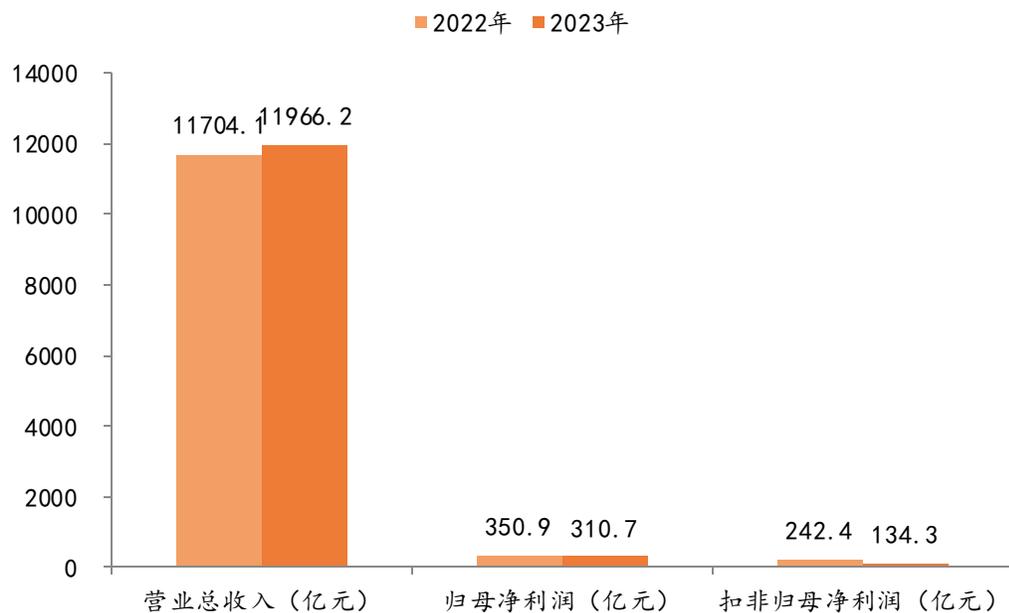
- 行业回顾：行业行情表现不佳，估值处于历史较低水平
- 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- 低空经济：政策持续催化，未来展翼腾飞
- 投资建议及风险提示



1.1 基本面：行业上市公司业绩承压，但需求呈现回暖趋势

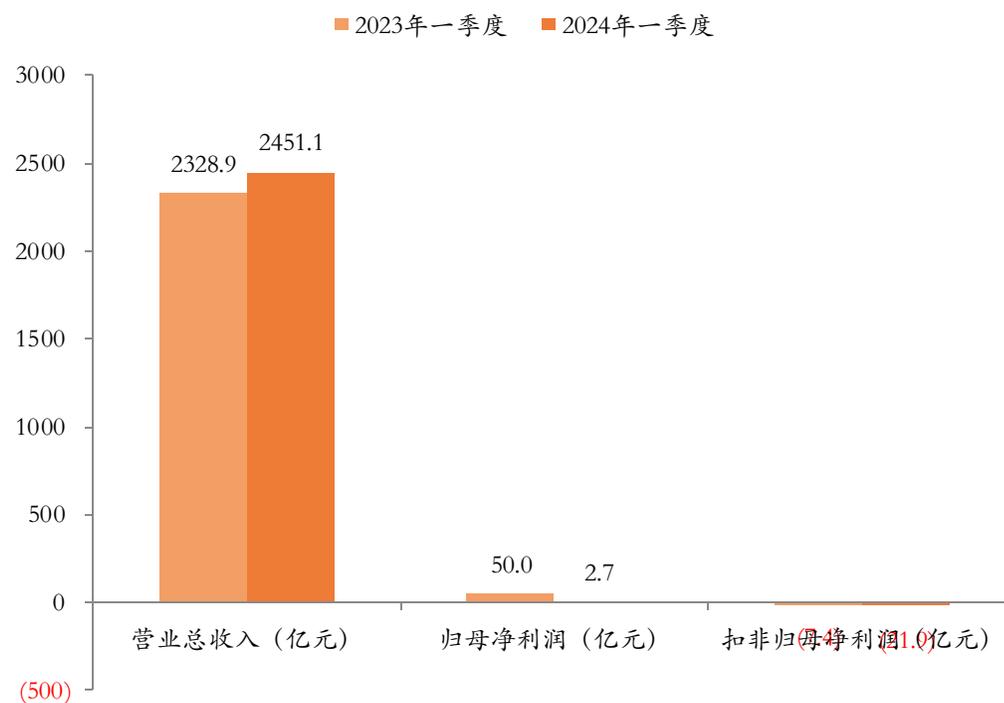
- 2023年，行业上市公司营业总收入合计实现11966.17亿元，同比增长2.24%；归母净利润合计实现310.71亿元，同比下降11.45%；扣非归母净利润合计实现134.33亿元，同比下降44.59%，行业上市公司整体业绩表现不佳。2024年一季度，行业上市公司营业总收入合计实现2451.08亿元，同比增长5.25%；归母净利润合计实现2.70亿元，同比下降94.60%；扣非归母净利润合计实现-21.90亿元，同比亏损进一步扩大，行业上市公司业绩表现持续承压。但从营收角度分析，2024年一季度的营收增速高于2023年全年的营收增速，也高于2023年一季度的营收增速（-1.13%），表明行业的下游需求呈现回暖的趋势。

2022年和2023年计算机行业上市公司业绩情况



数据来源：Wind，平安证券研究所

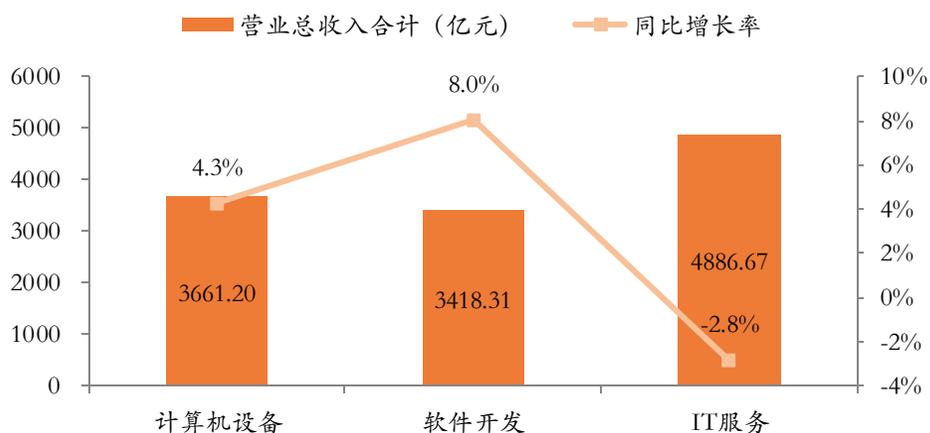
2023年一季度和2024年一季度计算机行业上市公司业绩情况



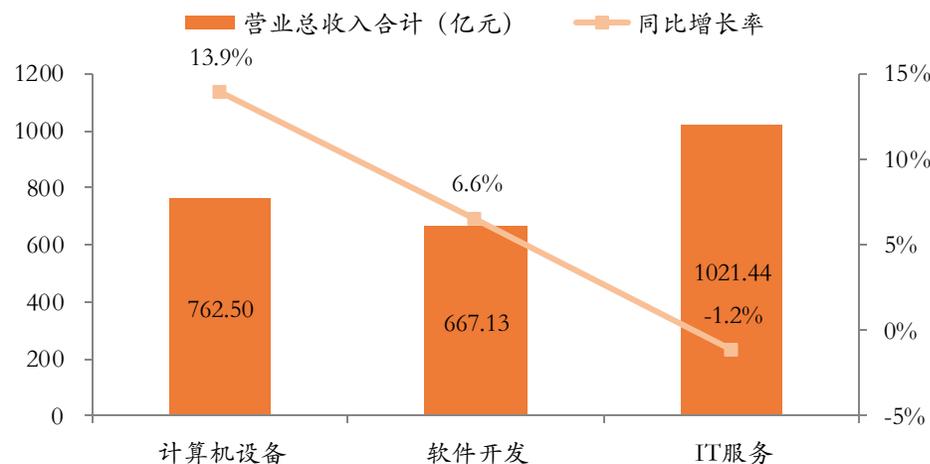
1.1 基本面：分板块看，计算机设备业绩表现好于软件开发和IT服务

- 按申万二级行业分类，计算机行业可分为计算机设备、软件开发和IT服务。分板块看，计算机设备业绩表现好于软件开发和IT服务，尤其是2024年一季度，计算机设备板块的营收和扣非归母净利润均实现同比两位数以上的增长。2023年，计算机设备板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为+4.29%、+1.86%、-21.68%，软件开发板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为+8.02%、-36.50%、由盈转亏，IT服务板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为-2.82%、-25.33%、-92.64%。
- 2024年一季度，计算机设备板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为+13.90%、+2.62%、+20.37%，软件开发板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为+6.55%、亏损扩大、亏损扩大，IT服务板块营业总收入合计、归母净利润合计、扣非归母净利润合计的增长情况分别为-1.15%、由盈转亏、由盈转亏。
- 展望2024年，我们认为，随着超长期特别国债对科技创新的推动以及国内宏观经济回升向好，计算机行业的下游需求有望持续回暖，行业上市公司的业绩有望向好。

2023年计算机行业各板块上市公司营收及增长情况



2024年一季度计算机行业各板块上市公司营收及增长情况





1.2 行情回顾：行业呈现明显震荡行情，整体处于下行态势

- 年初以来，计算机行业呈现明显的震荡行情，受业绩承压的影响，行业整体处于下行态势。年初至2月初，随着行业上市公司2023年业绩承压风险的持续释放，行业指数快速回调，到2月5日，回调到年初以来的最低点。之后，2024年春节期间，OpenAI在官网正式发布文生视频大模型Sora，全球视频大模型领域实现里程碑式进展，受Sora大模型发布的事件驱动，AIGC主题表现亮眼。另外，2024年3月，低空经济写入2024年国务院政府工作报告，低空经济主题表现强势。在AIGC、低空经济等主题的有力推动下，计算机行业指数快速反弹，至3月20日，反弹到年初以来的最高点。其后，3月下旬开始，行业上市公司2024年一季度业绩承压风险开始释放，行业指数从高点开始震荡下行。随着行业上市公司2024年一季报的陆续披露，当期业绩下滑幅度较大对行业指数的压力较大，行业行情自3月下旬之后整体处于下行态势。

年初以来计算机行业指数表现

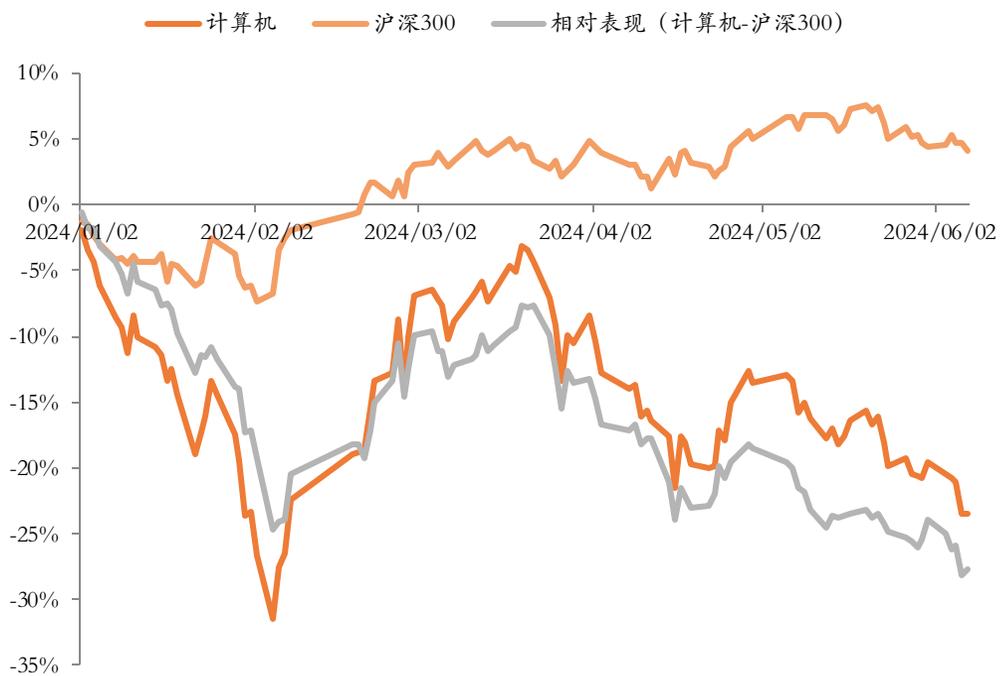


数据来源：Wind，平安证券研究所

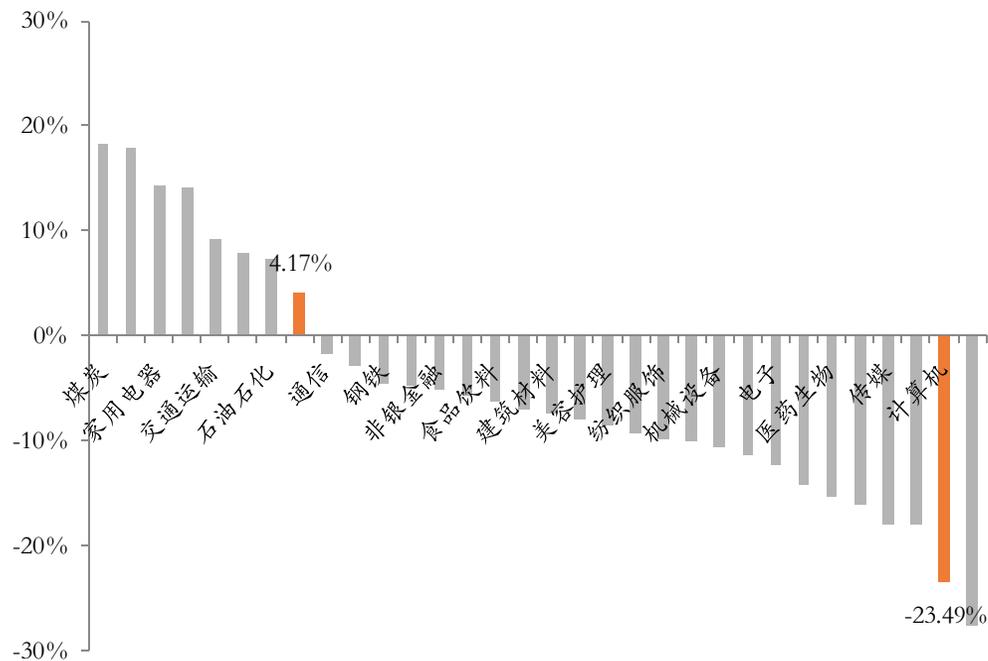
1.3 行情回顾：行业指数年累计涨幅跑输沪深300，位列倒数第2位

截至到2024年6月7日，申万计算机指数下跌了23.49%，跑输沪深300指数27.66个百分点，在31个申万一级行业中排名倒数第2位，排名靠后，表现不佳。

行业指数相比沪深300指数表现



行业指数相比沪深300指数表现



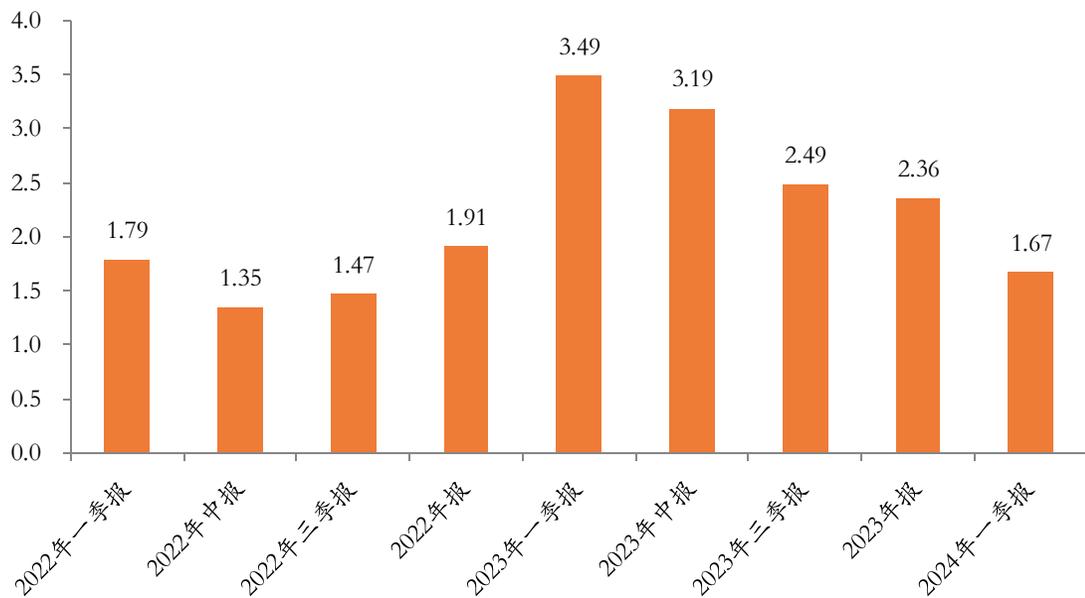
数据来源：Wind，平安证券研究所



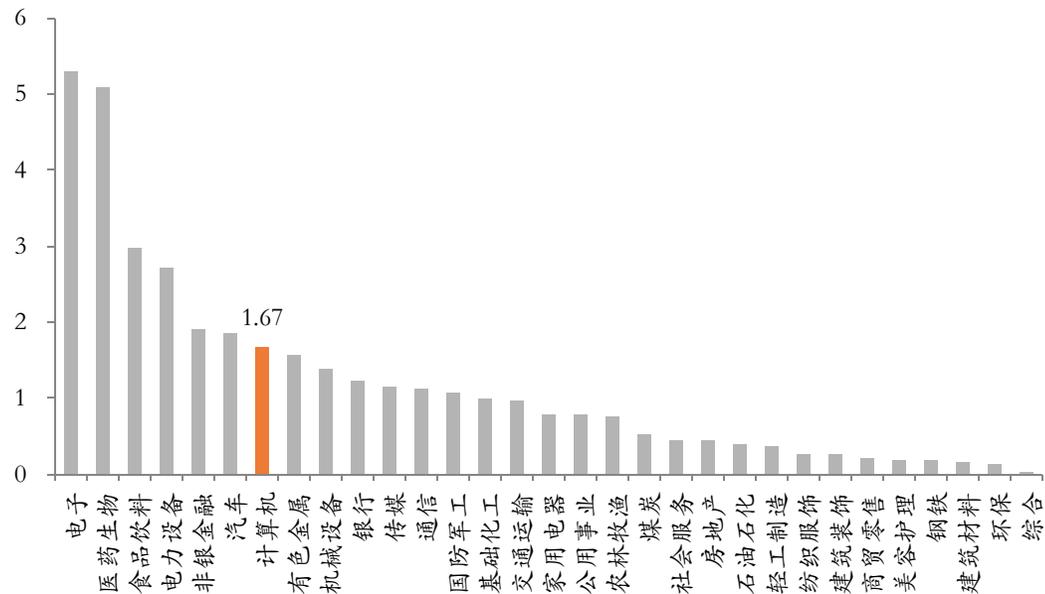
1.4 基金持仓：行业的基金重仓持股比例呈现下降趋势

➤ 根据公募基金2024年一季报数据，按申万行业分类，计算机行业2024年一季度的基金重仓持股比例为1.67%，相比2023年下降0.69个百分点。自2023年二季度以来，计算机行业的基金重仓持股比例已经连续四个季度下降。

2022年一季度以来计算机行业的基金重仓持股比例（单位：%）



计算机行业2024年一季度的基金重仓持股比例在31个申万一级行业排名第7（单位：%）





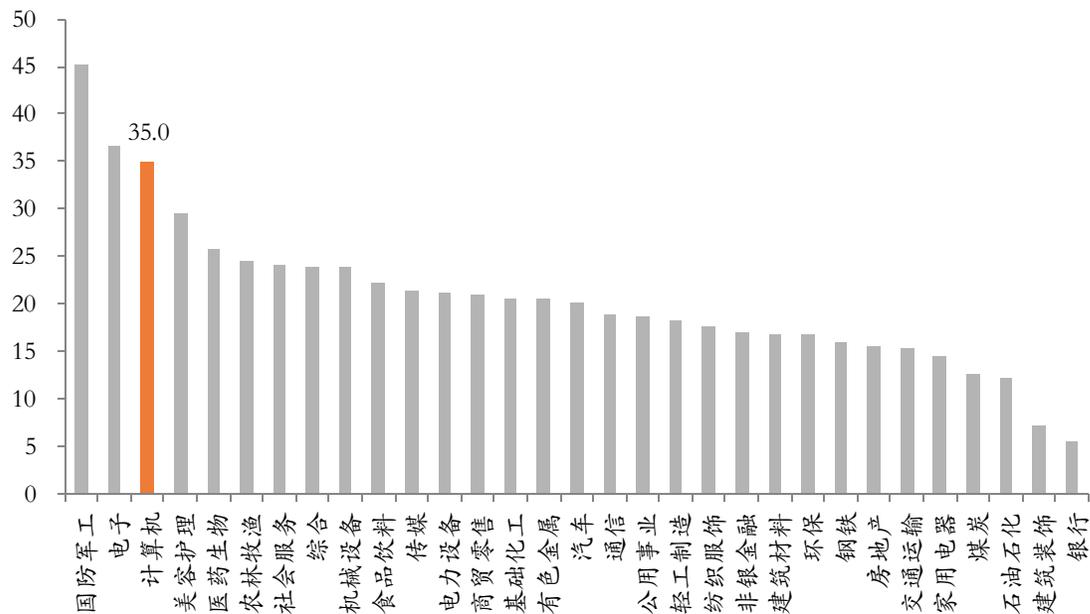
1.5 行业估值：估值处于历史相对较低水平

➤ 截至6月7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49.0倍。计算机行业6月7日市盈率（TTM，剔除负值）为35.0倍，在历史市盈率中位数水平之下。

● 计算机行业当前估值低于历史中位数水平



● 行业市盈率在31个申万一级行业排名第3位





1.6 行业展望：看好新质生产力的未来发展

- 2024年1月，习近平总书记在中共中央政治局第十一次集体学习时强调，加快发展新质生产力，扎实推进高质量发展。习近平总书记指出，科技创新能够催生新产业、新模式、新动能，是发展新质生产力的核心要素。
- 2024年3月，《政府工作报告》提出，大力推进现代化产业体系建设，加快发展新质生产力；同时指出，为系统解决强国建设、民族复兴进程中一些重大项目建设的资金问题，从今年开始拟连续几年发行超长期特别国债，专项用于国家重大战略实施和重点领域安全能力建设。国家发改委主任在十四届全国人大二次会议经济主题记者会上表示，初步考虑，超长期特别国债将重点支持科技创新、城乡融合发展、区域协调发展、粮食能源安全、人口高质量发展等领域建设。
- 展望2024，我们认为，在坚持高质量发展的新时代，包括AIGC（生成式人工智能）、低空经济等在内的新质生产力，将得到超长期特别国债等政策的大力支持，我们看好AIGC、低空经济等新质生产力的未来发展。在计算机行业整体需求处于回暖的趋势下，AIGC、低空经济等新质生产力在政策的大力推动下，将获得更好的发展机会。我们认为AIGC和低空经济主题将在二级市场获得更好的投资机会，我们维持对计算机行业“强于大市”的评级。



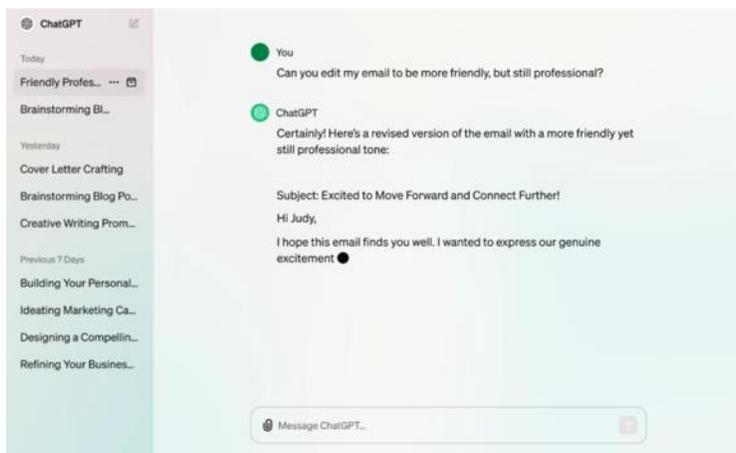
目录CONTENTS

- 行业回顾：行业行情表现不佳，估值处于历史较低水平
- 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- 低空经济：政策持续催化，未来展翼腾飞
- 投资建议及风险提示

ChatGPT火爆出圈，人工智能进入AIGC时代

- AIGC(AI Generated Content)即人工智能生成内容，也称为生成式AI，AIGC实现了从分析内容到创造生成新内容的跨越，而模型、数据集、算力、应用是催生AI技术新范式的重要因素。2022年11月，由OpenAI开发的大模型聊天机器人ChatGPT火爆出圈，短短5天，注册用户就超过100万，仅仅两个月月活用户数已经破亿。ChatGPT的火爆出圈，推动人工智能进入AIGC时代。2023年3月，OpenAI发布GPT-4。GPT-4可以接受图片作为输入，并且生成说明、分类和分析。GPT-4拥有超越ChatGPT的高级推理能力。
- 2024年2月，OpenAI发布文生视频大模型Sora，全球视频大模型领域实现里程碑式进展。作为文生视频大模型，Sora可以在保持视觉质量和遵循用户的文本提示的情况下，生成长达1分钟的视频。Sora呈现良好的多模态能力，不仅支持文本生成视频，还具备图像生成视频等能力。另外，Sora还有生成图像的能力。2024年5月，OpenAI在其春季发布会上，推出新旗舰模型GPT-4o。GPT-4o可接受文本、音频和图像的任意组合作为输入，并可生成文本、音频和图像的任意组合作为输出，多模态能力进一步提升。ChatGPT、GPT-4、Sora、GPT-4o等大模型爆款产品频现，打开了AIGC应用的想象空间。

◆ ChatGPT对话界面



◆ GPT-4可以接受图片作为输入



◆ Sora大模型生成的视频截图

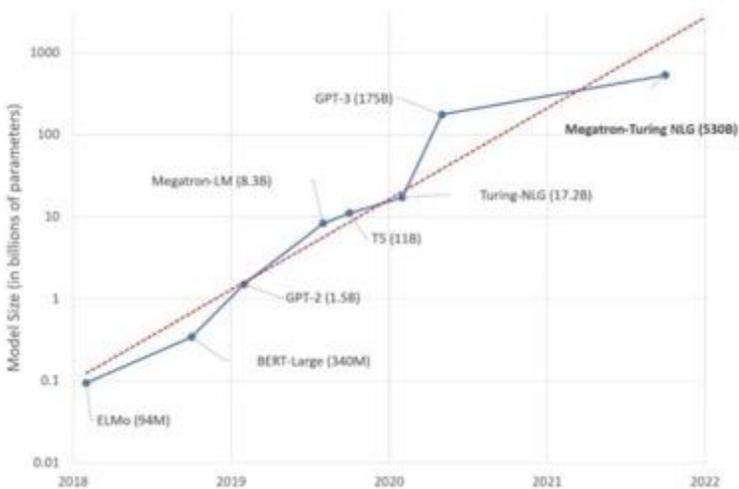




大模型的实现需要十分强大的算力来支持训练过程和推理过程

- 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。
- 根据《COMPUTE TRENDS ACROSS THREE ERAS OF MACHINE LEARNING》研究结果，大模型时代，算力翻倍的需求时间是9.9个月。AI芯片的性能提升成为决定大模型从理论实践到大规模应用的关键要素。

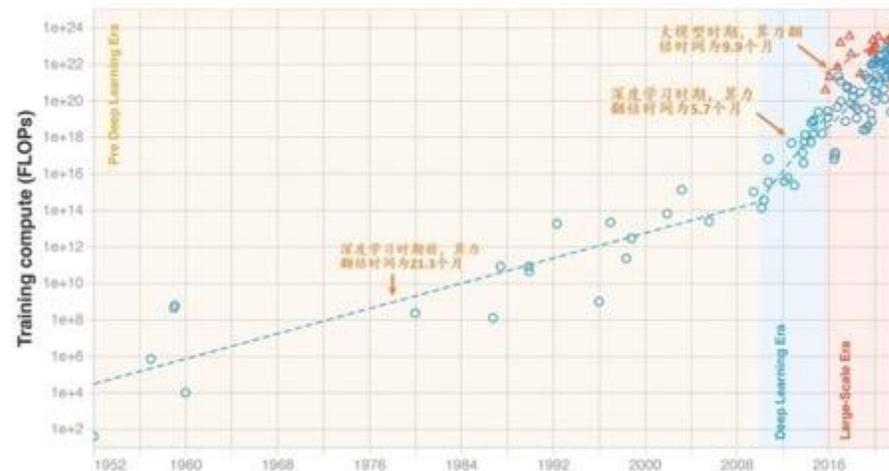
2018-2022年大模型参数增长变化趋势



模型计算所需算力、模型参数情况

Model	Total train compute (PF-days)	Total train compute (flops)	Params (M)
T5-Small	2.08E+00	1.80E+20	60
T5-Base	7.64E+00	6.60E+20	220
T5-Large	2.67E+01	2.31E+21	770
T5-3B	1.04E+02	9.00E+21	3,000
T5-11B	3.82E+02	3.30E+22	11,000
BERT-Base	1.89E+00	1.64E+20	109
BERT-Large	6.16E+00	5.33E+20	355
RoBERTa-Base	1.74E+01	1.50E+21	125
RoBERTa-Large	4.93E+01	4.26E+21	355
GPT-3 Small	2.60E+00	2.25E+20	125
GPT-3 Medium	7.42E+00	6.41E+20	356
GPT-3 Large	1.58E+01	1.37E+21	760
GPT-3 XL	2.75E+01	2.38E+21	1,320
GPT-3 2.7B	5.52E+01	4.77E+21	2,650
GPT-3 6.7B	1.39E+02	1.20E+22	6,660
GPT-3 13B	2.68E+02	2.31E+22	12,850
GPT-3 175B	3.64E+03	3.14E+23	174,600

人工智能不同时代对算力翻倍的需求时间

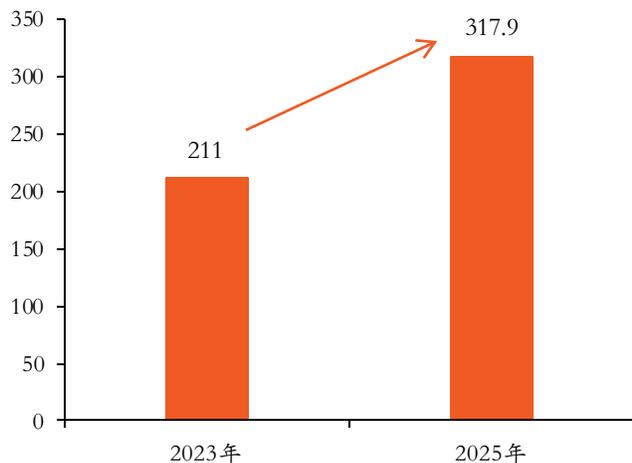




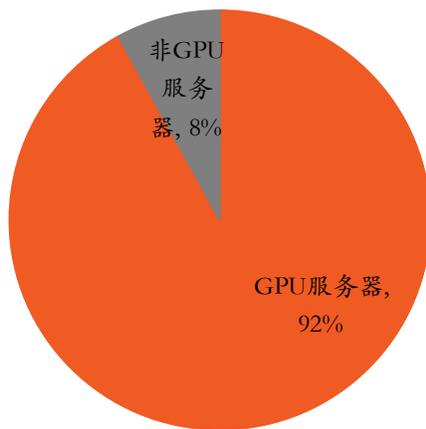
大模型发展助推算力需求释放，AI算力市场景气度持续向好

- 当前，全球范围内大模型领域的竞争依然白热化。2024年2月，OpenAI发布文生视频大模型Sora，全球视频大模型领域实现里程碑式进展。3月份，Anthropic推出Claude 3系列大模型，能力对标GPT-4。5月份，谷歌发布对标Sora的Veo。同时，国产大模型也在对标国际主流持续追赶。2024年6月，快手发布对标Sora的可灵大模型。科大讯飞将于2024年6月27日召开讯飞星火认知大模型V4.0发布会，大模型能力值得期待。
- 大模型需要大算力，大模型的持续迭代升级将为AI芯片及服务器发展提供强劲动力。1) 市场规模方面：据IDC统计，2023年，中国加速芯片的出货量达近140万张，其中国产AI芯片品牌的出货量超过20万张，国产化比例约为14%。2023年全球AI服务器市场规模预计为211亿美元，2025年有望达到317.9亿美元；2023年中国加速服务器市场规模为94亿美元，同比增长104%。其中GPU服务器以92%的市场份额占据主导。2) 算力结构方面：未来算力需求将逐渐向推理端倾斜。IDC预计，中国用于训练的AI服务器工作负载将由2023年的41.3%逐年提升至2027年的72.6%。

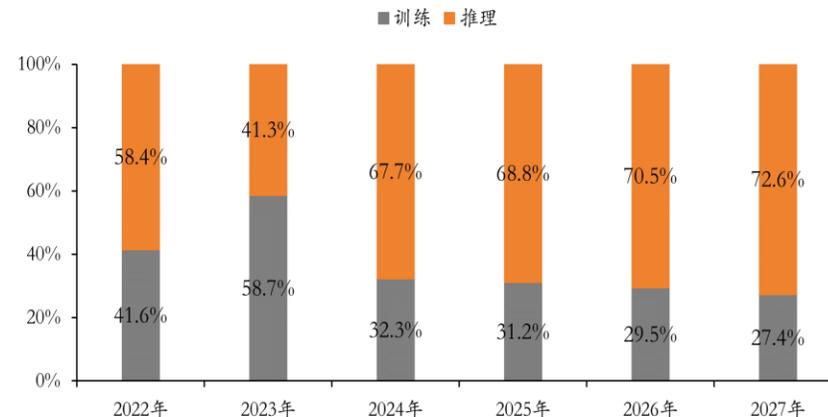
◆ 2023-2025年全球加速服务器市场规模 (亿美元)



◆ 2023年中国加速AI服务器市场结构



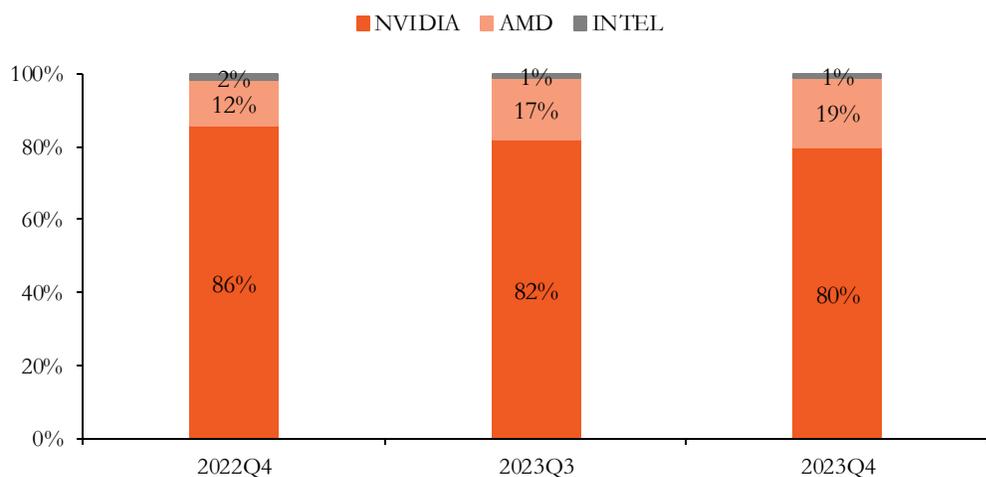
◆ 2022-2027年中国AI服务器工作负载预测



英伟达是全球GPU市场绝对龙头，GB200性能实现飞跃式提升

- 英伟达是全球人工智能计算领域的领导者，及全球GPU市场的绝对龙头。JPR统计数据显示，从出货量角度来看，2023年Q4，在全球独立显卡市场，英伟达以80%的市占率处于绝对领先地位。
- 英伟达推出全新Blackwell架构芯片，AI性能飞跃式提升。在2024年GTC大会上，英伟达推出全新Blackwell架构GPU芯片，该芯片基于台积电4nm工艺制造，并采用了双芯片设计，芯片间的互联速度高达10TB/s，总晶体管数量达2080亿个，同时搭配8颗HBM3e内存，内存容量达192GB，带宽达到8TB/s，相比Hopper架构平台提升明显。另外，公司还展示了GB200超级芯片，通过900GB/s超低功耗的NVLink连接，将两块Blackwell GPU与一块Grace CPU相连，可提供高达40 petaFLOPS的FP4算力，在标准1746亿参数的GPT-3基准测试中，GB200性能是H100的7倍，提供的训练算力是H100的4倍，推理能力是H100的30倍。GB200超级芯片的推出，将全球AI芯片性能提高到了一个新的高度，将能够有力支撑AIGC时代大模型的迭代。作为全球AI芯片领军企业，英伟达不断升级迭代高性能AI芯片，持续拔升全球大算力AI芯片的性能天花板。

全球独立GPU出货量市场份额



英伟达主要GPU产品性能参数对比

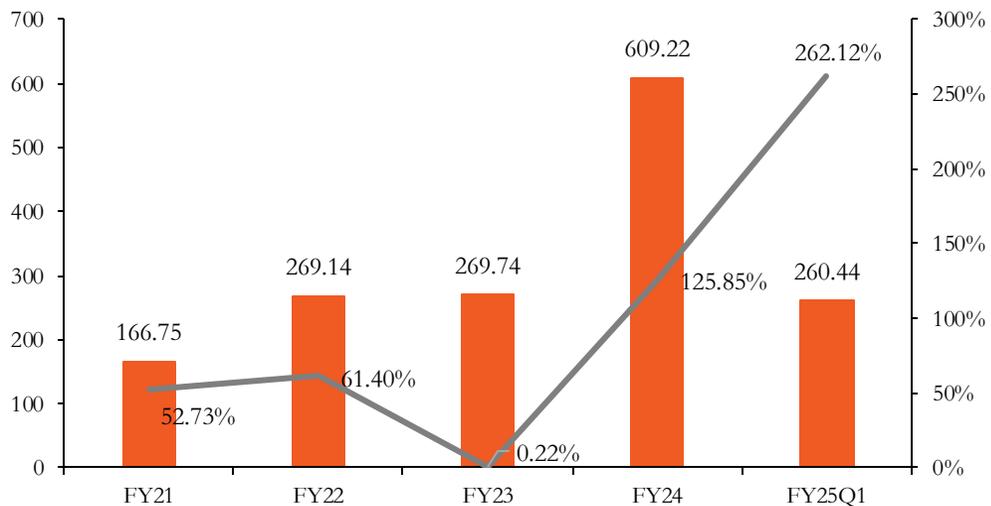
性能参数	H100 SXM	H200 SXM	GB200
FP16	1979 TFLOPS	1979 TFLOPS	10 PFLOPS
FP32	67 TFLOPS	67 TFLOPS	180 TFLOPS
FP64	34 TFLOPS	34 TFLOPS	90 TFLOPS
GPU 显存	80GB HBM3	141GB HBM3e	384 GB HBM3e
GPU 显存带宽	3.35TB/s	4.8TB/s	16 TB/s
最大热设计功耗 (TDP)	700W	700W	2700W

英伟达FY25Q1业绩超预期，彰显AI下游旺盛需求以及公司领先地位

- **英伟达FY25Q1业绩超预期。**英伟达发布FY25 Q1财报，FY25 Q1，公司实现营业收入260亿美元，同比增长 262%。公司收入端表现亮眼，主要受数据中心业务驱动。FY25Q1，公司数据中心业务实现收入226亿美元，创下历史新高，同比增长427%。FY25 Q1，公司实现归母净利润149亿美元，同比实现628%的高速增长。
- **英伟达FY25Q2 Blackwell开始出货，指引业绩继续上攻。**公司在FY25Q1业绩说明会上对FY25Q2的业绩进行了展望：2025财年二季度（对应2024年5月至7月），公司预计实现收入280亿美元，上下浮动2%，相较上财年同期增长107%，营收指引强劲，体现了AI领域下游旺盛的需求以及公司的领先地位。预计2025财年的GAAP和非GAAP运营费用将实现40%左右的增长。公司持续加码AI算力布局，为业绩持续释放提供强有力支撑。英伟达在其2025财年第一财季的电话会议上表示，当前公司Blackwell芯片产能正满负荷运转，计划节奏为2025财年第二财季开始量产，第三财季增产，第四财季会体现在数据中心业务中，预计在2025财年年报中会看到大量的Blackwell芯片收入。

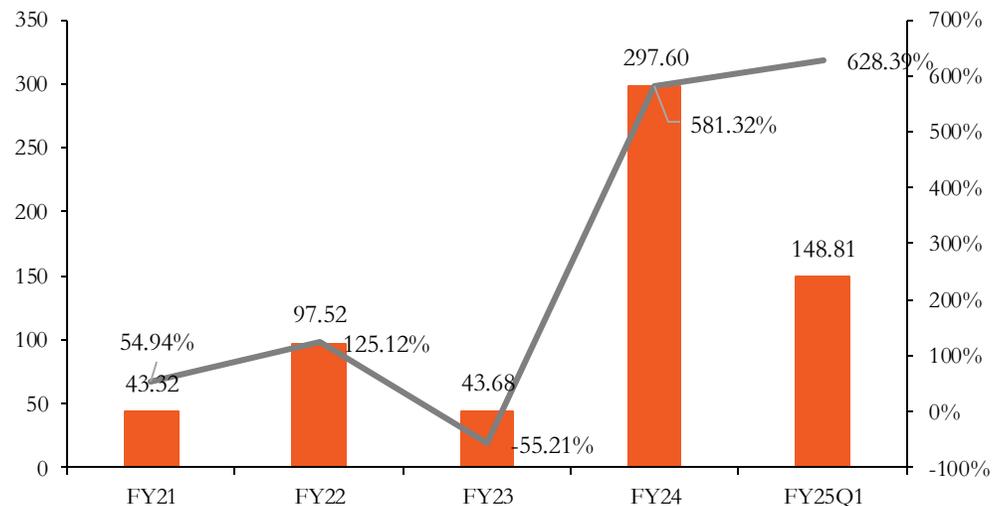
英伟达营业收入情况

营业收入 (亿美元) 同比增速



英伟达净利润情况

净利润 (亿美元) 同比增速



AMD更新产品路线图加速GPU迭代，市场份额有所提升

- AMD在全球GPU市场稳居第二，市场份额有所提升。JPR数据显示，2023年Q4，在全球独立显卡市场，AMD的出货量以19%的市占率位居第二，市场份额较上季度提升了2个百分点，市场份额环比提升。
- AMD更新加速器产品路线图，对标英伟达持续追赶。2024年6月，AMD在Computex 2024上更新了AMD Instinct加速器路线图，以满足对AI算力日益增长的需求。全新MI325X加速器将配备288GB HBM3E内存和6TB/秒的内存带宽，预计于2024年第四季度全面上市。MI325X将拥有业界领先的内存容量和带宽，分别是英伟达H200的2倍和1.3倍。MI350系列的首款产品MI350X加速器基于AMD CDNA 4架构，预计将于2025年上市。MI350X将采用先进的3nm工艺技术，支持FP4和FP6 AI数据类型，并具有高达288GB的HBM3E内存，与采用AMD CDNA 3架构的MI300系列相比，其AI推理性能将提高35倍。基于AMD CDNA “Next”架构的AMD Instinct MI400系列预计将于2026年上市。AMD产品对标英伟达持续追赶，全球AI芯片整体性能持续大幅提高。

◆ AMD全新加速器产品路线图



◆ AMD与英伟达部分产品性能对比

性能参数	H100 SXM	H200 SXM	MI250X	MI300X	MI325X
FP8	3958 TFLOPS	3958 TFLOPS	-	2.61 PFLOPS	2614.9 TFLOPS
FP16	1979 TFLOPS	1979 TFLOPS	362.1 TFLOPS	1.3 PFLOPS	1307.4 TFLOPS
BF16	1979 TFLOPS	1979 TFLOPS	362.1 TFLOPS	1.3 PFLOPS	1307.4 TFLOPS
GPU 显存	80GB HBM3	141GB HBM3e	128GB HBM2e	192 GB HBM3	288GB HBM3e
GPU 显存带宽	3.35TB/s	4.8TB/s	3.2TB/s	5.3TB/s	6TB/s
最大热设计功耗	700W	700W	560W	750W	-



华为对标国际主流持续追赶，平头哥芯片AI性能持续提升

- 华为的自研AI芯片昇腾910B采用了先进的达芬奇架构。2024年6月，在南京世界半导体大会上，华为表示，其最新的AI芯片910B在训练大型语言模型时表现出色，其效率可达到英伟达A100的80%，而在某些特定测试中，其性能甚至超越了A100达20%。910B目前已可实现千卡以上规模量级集群，根据科大讯飞公告，2023年10月，科大讯飞与华为联合发布我国首个全国产支持万亿参数大模型训练的万卡国产算力平台“飞星一号”，基于该平台训练完成的“讯飞星火V3.5”于2024年1月30日正式发布，“飞星一号”即采用了华为昇腾910B。华为即将推出的下一代AI芯片昇腾910C，目前仍处于内部测试阶段，性能或迎来又一次飞跃。
- 阿里旗下子公司平头哥于2023年8月发布了首个自研RISC-V AI平台，支持运行170余个主流AI模型，推动RISC-V进入高性能AI应用时代。同时，平头哥宣布玄铁处理器C920全新升级，C920执行GEMM计算较Vector方案可提速15倍。2023年11月，平头哥玄铁RISC-V上新了三款处理器：C907、C920、R910，将加速推动RISC-V在自动驾驶、人工智能、企业级SSD、网络通信等场景和领域的大规模商用落地。C920较上一代提升了最高3.9倍的AI性能，可以跑Transformer模型，适合机器学习、自动驾驶领域。

◆ 华为昇腾910B与英伟达A100性能对比

芯片	昇腾910B	英伟达A00 PCIe
FP32算力 (PFLOPS)	94T	19.5T
FP16算力 (PFLOPS)	376T	312T
内存容量 (BG)	64GB HBM2e	80B HBM2e
CPU-NPU PCIe规格	PCIe5.0×16 512GB	PCIe5.0×16 256GB
NPU-NPU带宽	392GB/s (HCCL)	600GB/s (NVLink)
芯片RRDMA出口带宽	200Gb/s RoCE芯片直出口	100Gb/s-200Gb/s IB需要通过IB网卡扩展出口

◆ 玄铁C907、C920、R910





海光信息DCU产品发展势头良好，已持续迭代至第三代

➤ 海光DCU（Deep Computing Unit 深度计算器）深算系列属于GPU的一种，采用“类CUDA”通用并行计算架构，产品性能稳步迭代。根据海光信息公司公告，公司深算二号已于2023年三季度发布，神算二号具有全精度浮点数据和各种常见整型数据计算能力，能够充分挖掘应用的并行性，发挥其大规模并行计算的能力，快速开发高能效的应用程序，性能相对于深算一号实现了翻倍增长，并已在大数据处理、人工智能、商业计算等领域实现商用。深算三号研发进展顺利。

海光信息深算系列DPU产品迭代节奏



海光信息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

海光信息DCU产品与行业典型可比产品参数对比

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 1.53Hz	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 × 16	PCIe Gen4 × 16	PCIe Gen4 × 16
GPU to GPU 互联	xGMI × 2, Up to 184GB/s	NVLink, Up to 600GB/s	Infinity Fabric × 3, Up to 276GB/s

寒武纪ASIC产品不断迭代，思元590实测训练性能实现大幅提升

- 寒武纪的第三代云端推训一体芯片思元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，公司包含思元370系列加速卡在內的云端产品已在互联网、运营商、金融、电力能源等多个行业及客户中进行了广泛的业务部署与落地。根据飞桨PaddlePaddle公众号消息，思元370系列产品与百度飞桨已完成II级兼容性测试，兼容性表现良好。思元590是寒武纪最新一代云端智能训练芯片，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，能够提供更大的内存容量和更高的内存带宽，IO和片间互连接口也较上代实现大幅升级。思元590尚未正式发布，已列示于百度飞桨官网的飞桨产品硬件支持表中。
- 国产AI芯片虽然性能较国外龙头厂商最新芯片仍有不小差距，但仍在大规模模型的时代持续迭代。国内华为、寒武纪、海光信息等AI芯片企业持续迭代AI芯片产品，增强自身的市场竞争力。国产高端AI芯片已经能够承接国内大规模模型厂商的大模型训练需求，如华为910B已经在支撑科大讯飞的大模型迭代且可实现千卡以上规模量级集群。

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	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

飞桨与寒武纪思元370系列产品兼容性测试证明





以浪潮信息为代表，我国国产AI服务器厂商具有全球竞争力

- ▶ **浪潮信息是全球AI服务器行业龙头企业。**从全球市场来看：根据公司2023年半年报，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。从中国市场来看：根据IDC数据，2023年，在中国加速服务器市场，从销售额来看，浪潮、新华三、宁畅位居前三，占据近70%的市场份额；从服务器出货量看，浪潮、坤前、新华三位居前三名，市场份额超过50%。
- ▶ **浪潮信息联合英特尔发布AI通用服务器，可运行千亿参数大模型。**2023年，浪潮信息发布全新一代G7算力平台，涵盖面向云计算、大数据、人工智能等应用场景的46款新品。最新一代融合架构的AI训练服务器NF5688G7，较上代平台大模型实测性能提升近7倍。2024年4月，浪潮信息与Intel联合发布AI通用服务器NF8260G7，NF8260G7在2U空间支持4颗英特尔至强处理器，具有AMX的AI加速功能，内存带宽1200GB/S，全链路UPI总线互连，传输速率高达16GT/s，能够更好满足千亿大模型低延时要求。

◆ NF5688技术规格

型号	NF5688-M7-A0-R0-00	NF5688-A7-A0-R0-00
高度	6U	
GPU	1块NVIDIA HGX-Hopper-8GPU模组	
处理器	2颗Intel®第四代至强®可扩展处理器，TDP 350W	2颗AMD第四代EPYC™处理器，最大cTDP 400W
内存	32条DDR5 DIMMs内存，速率最高支持4800MT/s	24条DDR5 DIMMs内存，速率最高支持4800MT/s
存储	最多支持24块2.5英寸SSD硬盘，其中最大支持16块NVMe	
PCIe插槽	支持10个PCIe 5.0 x16插槽（其中1个PCIe 5.0 x16插槽可替换为2个PCIe 5.0 x8速率的x16槽位） 可选支持Bluefield-3、CX7以及多种类型智能网卡	
RAID支持	可选支持RAID 0/1/10/5/50/6/60等，支持Cache超级电容保护	
前置I/O	1个USB 3.0端口，1个USB 2.0端口，1个VGA端口	
后置I/O	2个USB 3.0端口，1个MicroUSB端口，1个VGA端口，1个RJ45管理口	

◆ 浪潮信息服务器AI训推一体化方案架构

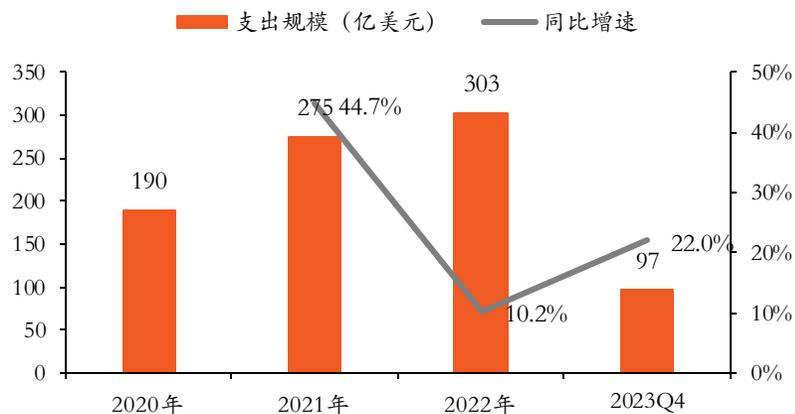




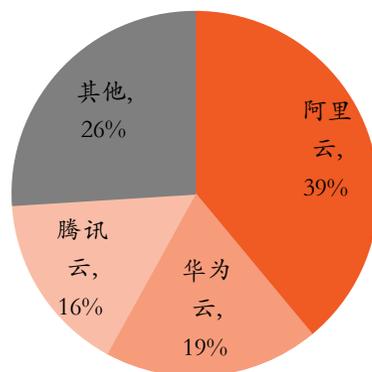
互联网厂商贡献AI服务器主要需求，运营商招标节奏加速

- **AI服务器市场的下游主要是以互联网厂商为主。**IDC统计数据显示，2023年，互联网依然是我国AI服务器最大的采购行业，市场份额近60%，此外金融、电信、交通和医疗健康等多数行业均有超过一倍以上的增涨。国内云计算支出势头不减，互联网厂商仍为主力。据Canalys统计，2023年第四季度，中国大陆云基础设施服务支出同比增长22%，达到97亿美元，占全球云支出的12%，其中阿里云、华为云和腾讯云继续保持主导地位，三者共同实现了28%的增长，占据了整个市场74%的份额。Canalys预计，2024年中国云基础设施服务支出的增速有望达到18%。我国云基础设施服务支出的快速增长将为我国AI芯片和服务器市场的持续高景气提供强力支撑。
- **三大运营商AI服务器招标进程加速。**2024年4月，中国移动采购与招标网发布公告：启动2024年至2025年新型智算中心采购招标，计划采购人工智能服务器7994台及配套产品、白盒交换机60台。此外，中国电信、中国联通此前也开启了AI服务器大规模采购。中国电信于2023年10月采购了4175台AI服务器；2024年3月，中国联通发标采购2503台AI服务器。三大运营商AI服务器招标进程加速，有望带动AI服务器需求持续释放。

2020年-2023Q4我国云基础设施服务支出及增速



2023Q4我国云基础设施服务支出市场份额



中国移动2024年至2025年新型智算中心采购项目（标包1）中标情况

中标公司	中标金额 (亿元)
昆仑技术	190.97
华坤振宇	190.97
宝德计算机	190.97
百信	191.36
长江计算	191.17
神州鲲泰	191.17
湘江鲲鹏	190.58

AI服务市场空间测算：大模型的训练和推理需要的算力成本高昂

- 大模型的训练和推理需要的算力成本高昂。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（A100GPU芯片）。根据英伟达网站信息，NVIDIA DGX A100服务器搭载8块A100芯片，估算需要1280台NVIDIA DGX A100服务器，每台NVIDIA DGX A100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25472万美元。若在云端训练，据LambdaLabs首席科学官Chuanli介绍，拥有1746亿个参数的GPT-3单次训练成本达到460万美元。

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（亿美元）
10240	8	1280	19.9	2.55

- 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NVIDIA DGX A100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。

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）	每天消耗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

AI服务市场空间测算：大模型将为AI算力市场带来巨大市场空间

- ▶ 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6.58（4.03+2.55）亿美元，按20家企业估算，算力需求为131.6亿美元。
- ▶ 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0.255亿美元，则合计训练成本为92.06亿美元，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DGX A100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307.58（92.06+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439.18亿美元的市场需求。以2023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估算（94/211≈44.55%），则将为全球AI服务器市场带来约985.82亿美元的市场空间。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。

◆ 大模型将为国内带来的AI服务器市场空间测算

	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	51亿美元（2.55亿美元*20）	80.6亿美元（4.03亿美元*20）	131.6
B端应用的企业	361（3614*10%）	92.06亿美元（0.255亿美元*361）	215.52亿美元（59.7万美元*100*361）	307.58
合计	381	143.06亿美元	296.12亿美元	439.18



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

- **美国对华半导体出口管制升级。**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，在2022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2024年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。受最终规则的影响，美国高端AI芯片的对华出口已受到非常严格的限制。
- 英伟达现已开发出针对中国区的最新改良版AI芯片，包括HGXH20、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，据IDC统计，2023年，中国加速芯片的出货量达近140万张，其中国产AI芯片品牌的出货量超过20万张，国产化比例约为14%，我国国产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。

◆ 英伟达 H20、L20、L2 性能相比国产芯片昇腾910B不占优势

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



目录CONTENTS

- ① 行业回顾：行业行情表现不佳，估值处于历史较低水平
- ② 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- ③ 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- ④ 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- ⑤ 低空经济：政策持续催化，未来展翼腾飞
- ⑥ 投资建议及风险提示



国产大模型能力追赶GPT-4，性能角逐同时注重性价比

➤ 如前所述，当前，全球范围内大模型领域的竞争依然白热化。我们认为，全球大模型领域竞争白热化将持续拔升大模型的整体能力水平，我国大模型厂商持续迭代升级算法能力，模型性能持续追赶国际领先厂商，将有利于AIGC在应用端的体验提升与场景拓展。同时，我们关注到，2024年以来，全球大模型厂商由单方面的性能角逐，逐渐转向性能与实用性并重的发展趋势，大模型纷纷实现提速降价，更加注重性价比，将有利于未来下游大模型部署成本的下降。

◆ 2024上半年全球大模型最新进展

1、全球范围内大模型领域的竞争依然白热化

- 2月，OpenAI发布文生视频大模型Sora，生成长达1分钟的视频，取得里程碑式进展。5月，OpenAI最新推出GPT-4o，实现性能与实用性双突破。
- 谷歌大模型全线升级，Gemini 1.5 Pro将上下文窗口长度从此前100万tokens进一步扩展至200万tokens。发布了视频生成模型Veo以及文生图模型Imagen 3。
- 开源大模型方面，Meta推出强大的开源人工智能模型Llama3，发布包括8B和70B参数的两个版本。其正在开发的400B版本模型能力对标GPT-4，未来有望赋能开源社区GPT-4的能力。

2、国内大模型厂商正在努力追赶GPT-4/GPT-4 Turbo/Sora的模型能力

- **通用大模型领域**，以百度文心4.0、阿里云通义千问2.5、智谱GLM-4、昆仑万维天工3.0、科大讯飞星火V3.5等最新版本为代表。
- **垂类大模型领域**，2024上半年月之暗面Kimi chat凭借突出的长文本处理能力，支持200万字超长无损上下文，访问量现象级激增，引发市场对于**对话大模型**的关注，5月字节火山引擎豆包大模型、腾讯元宝App等面向C端的AI助手陆续正式发布。**文生图/视频大模型领域**，美图奇想、万兴天幕、快手可灵等团队各自借助自身生态优势布局大模型，文生视频能力持续追赶Sora。

3、全球大模型价格呈现下降趋势，下游部署成本有望降低

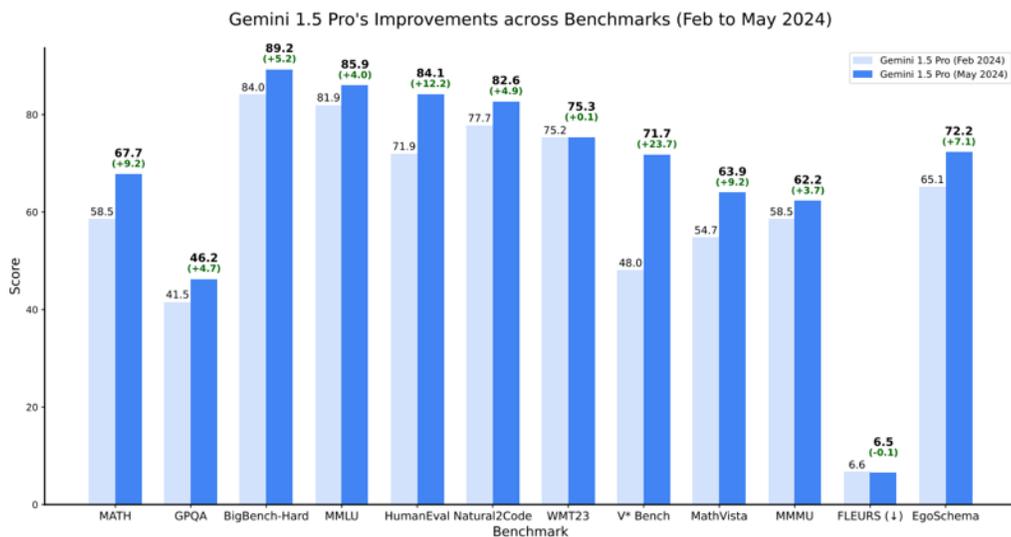
- **海外厂商来看**，2024年5月，OpenAI发布的GPT-4o支持免费试用，API价格比GPT-4-turbo降低了50%，输入为5美元/百万tokens。谷歌在2024年I/O开发者大会上发布谷歌Gemini 1.5 Flash，输入价格定为0.35美元/百万tokens（上下文小于128K）。
- **国内厂商来看**，5月6日，AI公司深度求索（DeepSeek）率先宣布降价，其发布的第二代MoE大模型DeepSeek-V2定为0.001元/千tokens的输入价格与0.002元/千tokens的输出价格，随后，智谱AI、火山引擎、阿里云、百度、科大讯飞、腾讯云等国内主要大模型厂商迅速跟进。

谷歌：Gemini系列全面升级，发布对标Sora的视频生成模型Veo

- **谷歌大模型全线升级。**继2023年12月谷歌推出其规模最大、功能最强的多模态大模型Gemini系列之后，2月15日，谷歌发布Gemini 1.5，在版本中首个登场的多模态通用模型Gemini 1.5 Pro，将稳定处理上下文的上限扩大至100万tokens。5月14日，在谷歌IO开发者大会上，谷歌大模型迎来多项升级，基础大模型方面，Gemini 1.5 Pro将上下文窗口长度从之前的100万tokens进一步扩展至200万tokens，谷歌首席执行官Pichai称这是目前市场处理上下文长度规模最大的基础大模型。
- **多模态大模型方面，**谷歌发布了视频生成模型Veo以及文生图模型Imagen3，Veo模型对标Sora，能够根据文字、图片和视频的提示生成长度超过1分钟、分辨率最高1080p的视频；Imagen3文生图模型是Imagen系列的升级版，从细节拟真度来看对标Midjourney v6。

◆ 5月Gemini 1.5 Pro性能相比2月份有明显提升

◆ 5月，谷歌发布视频生成模型Veo，推出Imagen最新产品Imagen 3



Meta: 推出最强开源大模型Llama3, 400B模型能力对标GPT-4

Meta发布最强开源大模型Llama3, 开源社区有望取得对标GPT-4的大模型能力。4月18日, Meta推出强大的开源人工智能模型Llama3, 发布包括8B和70B参数的两个版本, 为Llama2的重大升级。目前, Llama-3被誉为最强大的开源模型, 在一众榜单中取得开源SOTA (State-of-the-Art, 目前最高水平), Llama3 8B在MMLU、GPQA、HumanEval、GSM-8K等多项基准上超过谷歌Gemma 7B和Mistral 7B Instruct, Llama3 70B也在MMLU、HumanEval、GSM-8K基准上超越了闭源的谷歌Gemini Pro 1.5、Claude 3 Sonnet。Meta称, 目前正在开发的最新模型是400B参数, 将在未来几个月内推出。英伟达科学家Jim Fan认为, 即将推出的Llama3 400B将成为一个分水岭, 即社区将获得开源重量级的GPT-4模型, 它将改变许多研究工作和草根创业公司的计算方式。

◆ Llama 3在多项基准上超越对标的开闭源大模型

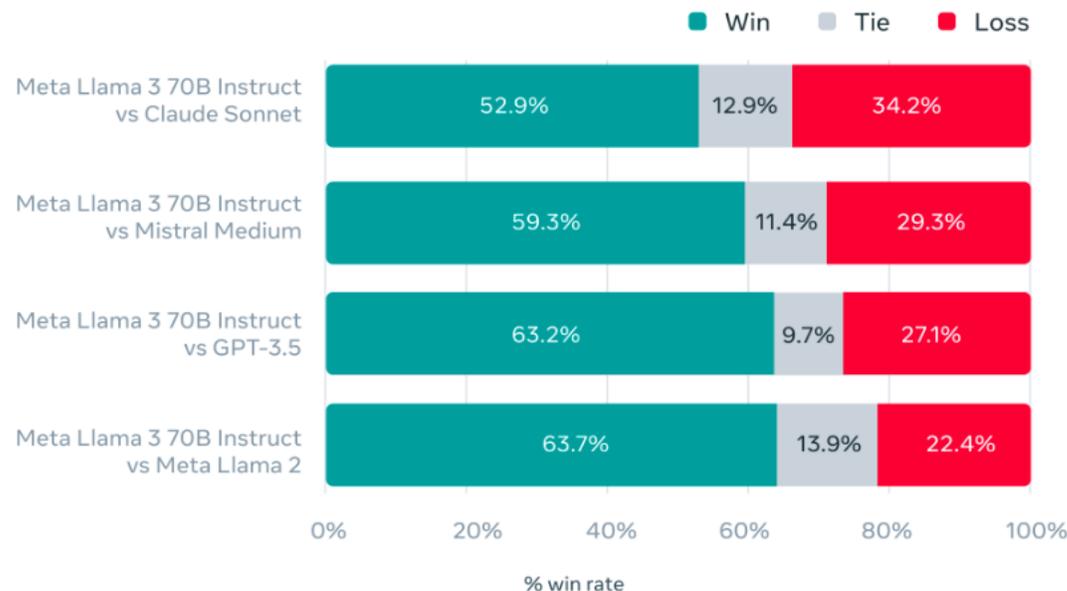
Meta Llama 3 Instruct model performance

	Meta Llama 3 8B	Gemma 7B - It Measured	Mistral 7B Instruct Measured
MMLU 5-shot	68.4	53.3	58.4
GPQA 0-shot	34.2	21.4	26.3
HumanEval 0-shot	62.2	30.5	36.6
GSM-8K 8-shot, CoT	79.6	30.6	39.9
MATH 4-shot, CoT	30.0	12.2	11.0

	Meta Llama 3 70B	Gemini Pro 1.5 Published	Claude 3 Sonnet Published
MMLU 5-shot	82.0	81.9	79.0
GPQA 0-shot	39.5	41.5 CoT	38.5 CoT
HumanEval 0-shot	81.7	71.9	73.0
GSM-8K 8-shot, CoT	93.0	91.7 11-shot	92.3 0-shot
MATH 4-shot, CoT	50.4	58.5 Minerva prompt	40.5

◆ 在与Claude Sonnet、Mistral Medium、GPT-3.5的对比中, Llama3表现更好

Meta Llama 3 Instruct Human evaluation (aggregated)



通用大模型——百度推出文心4.0工具版，阿里云发布通义千问2.5

- 1) **百度文心一言**：2023年3月，百度宣布正式推出大模型文心一言，10月发布对标GPT-4的4.0版本，2024年4月16日，Create 2024百度AI开发者大会上，李彦宏汇报了文心一言发布一年的成果：用户数突破2亿，API日均调用量超过两亿，服务用户数达到8.5万，千帆平台开发的AI原生应用超过19万。同时，发布文心大模型4.0工具版。针对MoE、小模型、智能体三个方向，百度提供了“开箱即用”的三大AI开发工具，包括智能体开发工具AgentBuilder、AI原生应用开发工具AppBuilder、各种尺寸的模型定制工具ModelBuilder。
- 2) **阿里云通义千问**：2023年4月，阿里云推出了超十万亿参数量的自研大语言模型通义千问。2024年5月9日，阿里云正式发布通义千问2.5大模型，其性能对标GPT-4 Turbo。相比上一代2.1版本，通义千问2.5在理解能力、逻辑推理、指令遵循、代码能力上分别提升9%、16%、19%、10%。Qwen1.5-110B参数开源模型在多个基准测评收获最佳成绩，超越Llama-3-70B。通义千问APP更名为“通义APP”，免费开放通义全栈能力，覆盖文生图、智能编码、文档解析、音视频理解、视觉生成等，主打“All in one”。

◆ 4月，文心大模型4.0工具版正式发布



◆ 通义千问2.5全面追赶GPT-4模型能力



智谱AI：GLM-4性能逼近GPT-4，多模态大模型CogVLM2追赶GPT-4V

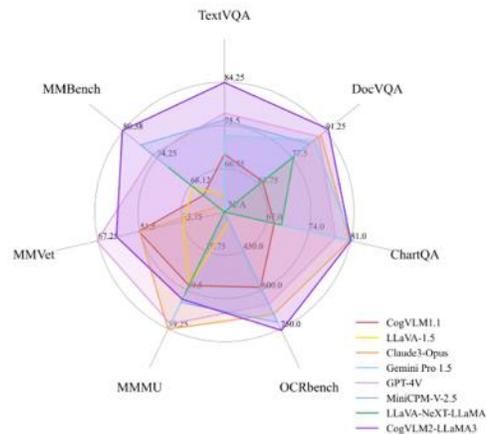
- 3) 智谱AI GLM系列：基础大模型方面，2022年8月，智谱AI推出千亿级开源大模型GLM-130B，2024年1月16日，智谱AI DevDay正式发布新一代基座大模型GLM-4，整体性能相比GLM3全面提升60%，逼近GPT-4；支持更长上下文；更强的多模态；支持更快推理速度，更多并发，大大降低推理成本；同时GLM-4增强了智能体能力。同时上线GLM-4 All Tools平台、GLMs智能体创建能力和GLMs商店，后续将推出开发者收入计划。全栈自主创新GLM-4全家桶，在多模态理解、复杂长文本、和自动解决复杂任务上都有较大提升，全方面对标GPT-4。
- 多模态大模型方面，继2023年5月18日和10月11日先后发布并开源 VisualGLM-6B 和 CogVLM之后，2024年5月22日，智谱AI宣布推出新一代多模态大模型CogVLM2。与上一代的 CogVLM模型相比，CogVLM2系列模型具有以下改进：1) 在不损失任何通用能力的前提下，在许多关键指标上有了显著提升，如在OCRbench基准上性能提升32%，在TextVQA基准上性能提升21.9%；2) 支持8K文本长度；3) 支持高达1344 * 1344的图像分辨率；4) 提供支持中英文双语的开源模型版本。CogVLM2的两个模型，尽管具有较小的模型尺寸，但在多个基准中取得SOTA性能；而在其他性能上，也能达到与闭源模型（例如GPT-4V、Gemini Pro等）接近的水平。

◆ GLM-4在多个数据集上达到接近GPT-4的水平

基础能力 (英文)						
	MMLU (5-shot)	GSM8K (5-shot)	MATH (4-shot)	BBH (3-shot)	HellaSwag (10-shot)	HumanEval (0-shot)
GPT-4	86.4	92.0	52.9	83.1	95.3	67.0
Gemini-Ultra	83.7	94.4	53.2	83.6	87.8	74.4
GLM-4	81.5	87.6	47.9	82.3	85.4	72.0
GLM-4 / GPT-4	94%	95%	91%	99%	90%	100%

(OpenAI, 2023; Gemini Team Google, 2023)

◆ CogVLM2的两个模型在多个基准中取得SOTA，性能对标闭源模型GPT-4V、Gemini Pro



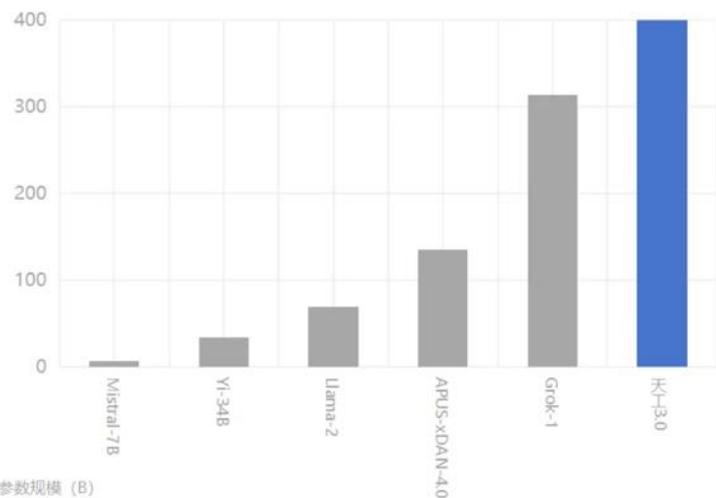
Model	Open Source	LLM Size	TextVQA	DocVQA	ChartQA	OCRbench	MMMU	MMVet	MMBench
LLaVA-1.5	✓	13B	61.3	-	-	337	37.0	35.4	67.7
Mini-Gemini	✓	34B	74.1	-	-	-	48.0	59.3	80.6
LLaVA-NeXT-LLaMA3	✓	8B	-	78.2	69.5	-	41.7	-	72.1
LLaVA-NeXT-110B	✓	110B	-	85.7	79.7	-	49.1	-	80.5
InternVL-1.5	✓	20B	80.6	90.9	83.8	720	46.8	55.4	82.3
QwenVL-Plus	✗	-	78.9	91.4	78.1	726	51.4	55.7	67.0
Claude3-Opus	✗	-	-	89.3	80.8	694	59.4	51.7	63.3
Gemini Pro 1.5	✗	-	73.5	86.5	81.3	-	58.5	-	-
GPT-4V	✗	-	78.0	88.4	78.5	656	56.8	67.7	75.0
CogVLM2-LLaMA3 (Ours)	✓	7B	69.7	-	68.3	590	37.3	52.0	65.8
CogVLM2-LLaMA3 (Ours)	✓	8B	84.2	92.3	81.0	756	44.3	60.4	80.5
CogVLM2-LLaMA3-Chinese (Ours)	✓	8B	85.0	88.4	74.7	780	42.8	60.5	78.9



昆仑万维：深耕AI搜索领域，发布天工3.0千亿参数大模型

- 4) 昆仑万维天工：2023年8月，昆仑万维推出了国内第一个AI搜索引擎——天工AI搜索。2024年2月6日，公司发布新版MoE大语言模型“天工2.0”与新版“天工AI智能助手”App，是国内首个搭载MoE架构并面向全体C端用户免费开放的千亿级参数大语言模型AI应用。
- 4月17日，昆仑万维推出“天工3.0”基座大模型。天工3.0拥有4000亿参数，超越了3140亿参数的Grok-1，是全球最大的开源MoE大模型，其在MMBench等多项权威多模态测评结果中超越GPT-4V。同时，天工3.0旗下的“天工SkyMusic”音乐大模型也面向全社会开放公测，采用自研大模型音乐音频生成技术路线，在与海外顶尖的AI音乐大模型Suno V3的横向测评中，“天工SkyMusic”在人声&BGM音质、人声自然度、发音可懂度等领域显著领先对手，并以6.65分的综合得分超越Suno V3。
- 6月3日，昆仑万维宣布开源2千亿稀疏大模型 Skywork-MoE，是首个完整将MoE Upcycling 技术应用并落地的开源千亿MoE大模型，也是首个支持用单台 4090 服务器推理的开源千亿MoE大模型。

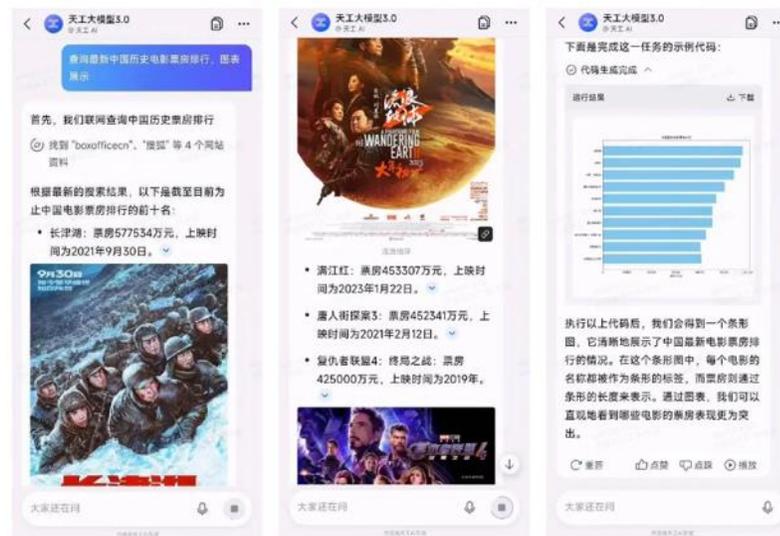
◆ 天工3.0模型参数超越Grok-1，成全球最大开源MoE专家混合大模型



◆ 天工SkyMusic综合性能超越Suno V3



◆ 天工3.0实时判断是否需要联网或调用工具，进行单轮或多轮的联网搜索、工具调用





科大讯飞：星火V3.5对标GPT-4，场景应用加速拓展

- 5) 科大讯飞星火大模型：2024年1月30日，科大讯飞召开讯飞星火认知大模型V3.5升级发布会。具体而言，文本生成提升7.3%，语言理解提升7.6%，知识问答提升4.7%，逻辑推理提升9.5%，数学能力提升9.8%，代码能力提升8.0%，多模态能力提升6.6%。对标GPT-4，星火大模型V3.5的语言理解能力、数学能力均已超过GPT-4 Turbo，代码达到GPT-4 Turbo的96%，多模态理解能力达到GPT-4V的91%。同时，星火语音大模型首次发布，在首批37个主流语种上已整体超越Whisper-large-v3，保持智能语音技术的领先水平。4月26日，讯飞星火V3.5春季上新，目前星火大模型通用长文本能力，包括长文档的信息抽取、知识问答、归纳总结、文本生成等，总体已经达到GPT-4 Turbo今年4月最新长文本版本的97%的水平，而多个垂直领域的知识问答任务上长文本总体水平超过GPT-4 Turbo。
- 5月底，讯飞星火上线All Tools功能，包括“联网搜索、代码解释器、文生图、天气、日期”等工具。星火能够根据用户指令，自己调用工具来完成复杂任务，不必用户在工具间手动切换。讯飞星火还上线了“数据分析助手”，支持上传Excel表格，快速完成数据分析，并生成分析表格与结论。

◆ 星火大模型总体已经达到GPT-4 Turbo今年4月最新长文本版本的97%的水平



◆ 讯飞星火用数据分析助手完成表格分析

数据分析助手
上传包含数据的Excel文件，我可以快速完成数据分析，生成分析表格与结论

```
product_sales_sorted = product_sales.sort_values(by='总销售额', ascending=False)
product_sales_sorted
```

根据您提供的数据，我对各产品的销售情况进行分析，并给出以下建议：

- 产品J**：总销售额最高，达到8800元，销售量也较高，为160个，平均单价为55元。这表明产品J在市场上受欢迎程度高，具有良好的销售表现。建议可以进一步加大产品J的库存和推广力度。
- 产品I**：总销售额为7000元，销售量为140个，平均单价为50元。产品I同样表现出较好的市场接受度，可以考虑维持当前的销售策略。
- 产品H、产品G、产品F**：这些产品的总销售额逐渐降低，分别为5850元、4400元和3150元，销售量和平均单价也有所不同。针对这些产品，建议密切关注市场反馈，调整销售策略以提升销量。
- 产品C、产品E、产品D、产品B和产品A**：这些产品的总销售额相对较低，特别是产品A，总销售额仅为1000元，平均单价最低。对于这些产品，建议分析其市场反响不佳的原因，可能需要调整定价策略、改进产品特性或加强营销推广。

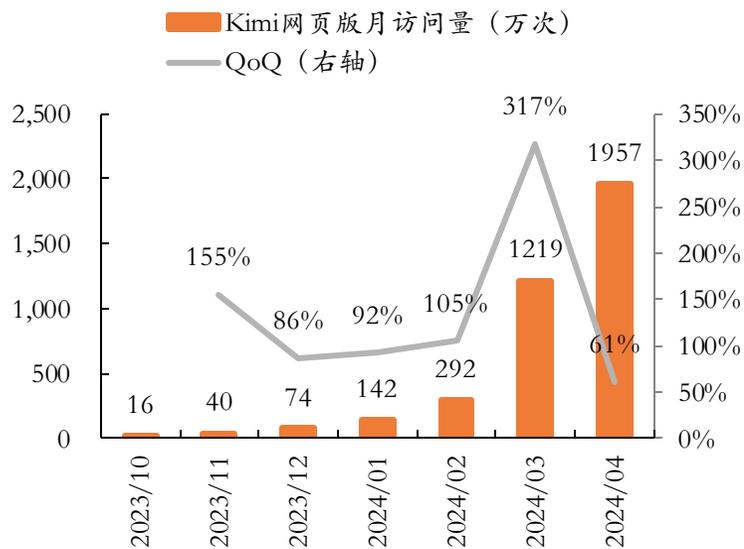
总体而言，对于销售额和销售量较高的产品应持续关注保持优势，而对于销售情况较差的产品则需要进一步分析原因，并考虑进行相应的调整以提高其市场竞争力。

退出助手 | 全新对话

对话大模型——月之暗面Kimi长文本处理能力突出，访问量现象级激增

- **对话大模型方面**，1) 月之暗面Kimi: 2023年10月10日，月之暗面发布首个支持输入20万汉字的智能助手产品 Kimi Chat。相较当时市面上以英文为基础训练的大模型服务，Kimi Chat具备较强的多语言能力。Kimi Chat在中文上具备显著优势，实际使用效果能够支持约20万汉字的上下文，2.5倍于Anthropic公司的Claude-100k（实测约8万字），8倍于OpenAI公司的GPT-4-32k（实测约2.5万字）。
- **Kimi长文本处理能力从20万字提升至200万字，3月访问量现象级激增，5月迎来商业化**。2024年3月18日，月之暗面宣布在大模型长上下文窗口技术上取得新的突破，其自研的Kimi智能助手已支持200万字超长无损上下文。Kimi Chat访问量的跃升，从推出时16万访问量，到2024年2月的292万，再到2024年3月的1219万，展示了Kimi Chat在长文本处理技术上的突破，使得Kimi Chat在信息整理领域迅速占据领先地位，吸引了大量用户的关注和使用。5月，Kimi智能助手上线“给Kimi加油”付费选项，可获得高峰期优先使用权益。

◆ Kimi网页版月访问量及环比增速



◆ 超长文本使用场景：用户上传教程后，让Kimi扮演德州扑克专家为自己提供出牌策略的指导



◆ Kimi增加“给Kimi加油”付费选项



对话大模型——角逐C端AI助手，字节豆包大模型、腾讯元宝App正式发布

- 2) **字节豆包**：5月15日，在2024春季火山引擎Force原动力大会上，字节跳动豆包大模型正式发布。据火山引擎总裁谭待介绍，经过一年时间的迭代和市场验证，豆包大模型正成为国内使用量最大、应用场景最丰富的大模型之一，目前日均处理1200亿Tokens文本，生成3000万张图片。基于豆包大模型，字节跳动打造了AI对话助手“豆包”、AI应用开发平台“扣子”、互动娱乐应用“猫箱”，以及星绘、即梦等AI创作工具，并把大模型接入抖音、番茄小说、飞书、巨量引擎等50余个业务，用以提升效率和优化产品体验。
- 3) **腾讯元宝**：5月30日，腾讯正式发布了基于混元大模型打造的App“腾讯元宝”，安卓、苹果等各大应用商店均可下载体验。据官方介绍，腾讯元宝提供AI搜索、AI总结、AI写作、AI绘画等特色功能，面向工作场景，它能够解析PDF、Word、txt等多种格式的文档，同时支持超长的上下文窗口，还能一次性解析多个网址、微信公众号链接，使用户处理工作中的复杂信息更为高效便捷；面向生活场景，腾讯元宝也提供了AI头像、角色扮演、口语陪练、同声传译等特色AI功能，用户还可以创建专属自己的个人智能体。

字节豆包场景示例



腾讯元宝提供AI助手功能



文生图/视频大模型——各凭优势布局，文生视频能力追赶Sora

- 1) **美图奇想**：1月2日，美图公司自研AI视觉大模型MiracleVision（奇想智能）通过《生成式人工智能服务管理暂行办法》备案，1月5日正式上线。奇想智能大模型于2023年6月内测，为美图秀秀、美颜相机、Wink、美图设计室、WHEE、美图云修等知名影像与设计产品提供AI模型能力的同时，也帮助美图公司搭建起由底层、中间层和应用层构建的人工智能产品生态。
- 2) **万兴天幕**：1月30日，万兴科技发布了国内首个音视频多媒体大模型万兴天幕。4月28日，万兴天幕大模型正式公测。目前，万兴天幕大模型的文生视频、视频生视频、文生音乐、视频配乐、文生音效相关能力已在万兴科技旗下Wondershare Filmora、Wondershare Virbo等产品上规模化商用，其中文生视频能力上已实现不同风格、丰富场景及主题的连贯性，一键生成时长支持60秒+。
- 3) **快手可灵**：6月6日，快手宣布，“可灵”视频生成大模型官网正式上线，快手介绍称，可灵大模型为快手AI团队自研，采用与Sora相似的技术路线，结合了多项自研技术创新。该模型支持生成长达2分钟、帧率30fps的视频，分辨率可高达1080p，并支持自由的宽高比。可灵大模型已上线快手旗下的快影App，面向公众开放邀测体验。

◆ 美图MiracleVision4.0



◆ 万兴“天幕”文生视频示例



◆ 快手对标Sora的“可灵”支持生成2分钟视频





大模型API价格呈现下降趋势，下游部署成本有望降低

- 海外来看，2024年5月，OpenAI发布GPT-4o，面向ChatGPT所有付费和免费用户发布，支持免费试用，API价格比GPT-4-turbo降低了50%，输入为5美元/百万tokens；谷歌在2024年I/O开发者大会上发布谷歌Gemini 1.5 Flash，输入价格定为0.35美元/百万tokens（上下文小于128K）。
- 国内来看，大模型价格亦呈现下降趋势。5月6日，AI公司深度求索（DeepSeek）率先宣布降价，其发布的第二代MoE大模型DeepSeek-V2定为0.001元/千tokens的输入价格与0.002元/千tokens的输出价格，随后，智谱AI、火山引擎、阿里云、百度、科大讯飞、腾讯云等国内主要大模型厂商迅速跟进（详见下列图表）。全球大模型逐渐呈现由单方面的性能角逐，转向性能与实用性并重的发展趋势，纷纷提速降价，更加注重性价比，有利于未来下游大模型部署成本的下降。

◆ 本轮国产大模型厂商降价统计

模型	推出/降价时间	API价格/百万Tokens					
		输入（元）			输出（元）		
		原价	现价	降价幅度	原价	现价	降价幅度
深度求索DeepSeek-V2	2024/5/6	-	1	-	-	2	-
智谱AI GLM-3-Turbo	2024/5/11	5	1	-80%	5	1	-80%
火山引擎 豆包通用模型pro-32k	2024/5/15	-	0.8	-	-	2	-
阿里云Qwen-Long（通义千问）	2024/5/21	20	0.5	-98%	20	2	-90%
百度ERNIE-Speed（文心一言）	2024/5/21	4	免费	-	8	免费	-
百度ERNIE-Lite（文心一言）	2024/5/21	3	免费	-	6	免费	-
科大讯飞 星火Lite	2024/5/22	18	免费	-	18	免费	-
腾讯云 混元-standard	2024/5/22	10	4.5	-55%	10	4.5	-55%
腾讯云 混元-lite	2024/5/22	8	免费	-	8	免费	-



目录CONTENTS

- 行业回顾：行业行情表现不佳，估值处于历史较低水平
- 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- 低空经济：政策持续催化，未来展翼腾飞
- 投资建议及风险提示

AIGC+PC/手机：终端成为重要载体，端侧智能驱动移动互联网产业发展

- ▶ 端侧（PC和手机）是AI发展的重要平台。手机厂商围绕AI影像、智能通话、智能搜索等功能已经提前布局AI应用，2024年起新一代AI手机渗透率将持续提升，与此同时，各品牌整机厂商纷纷加码AIPC产品，全球PC产业将稳步迈入AI时代，端侧智能发展前景广阔，成为AIGC应用进度较为靠前的景气赛道。

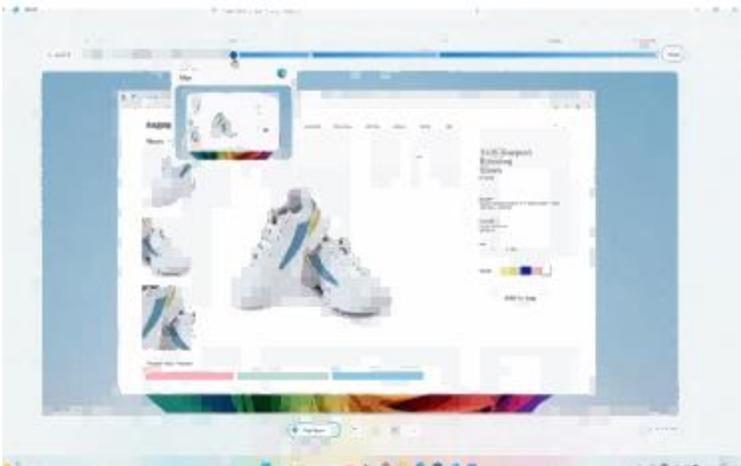
应用场景

- 个性化智能助手：日程管理、信息搜索、智能提醒、语言翻译
- 办公智能化：文档管理、会议记录、数据分析
- 创意辅助：设计建议、音乐创作辅助

具体案例

- 微软：最新推出“Copilot+PC”，增加了召回（Recall）功能，类似“时光倒流”，只要是屏幕显示过的内容，AI就可以找到。Windows 11 AI PC自6月18日起陆续上市。
- 苹果：最新推出“Apple Intelligence”，主要围绕文字、图像和交互三个模块。同时，用户将可以通过 Siri，直接调用 GPT-4o 的接口来进行对话。
- 中科创达：公司是操作系统和端侧智能的产品和技术厂商，公司将领先的端侧模型优化以及模型跨平台支持的能力，通过预集成的端侧编码助手Rubik Studio以及更加丰富的端侧大模型应用，助力OEM厂商快速推出功能全面、性能出众的AIPC。

通过Copilot+PC的Recall找到此前浏览过的心仪的鞋子



Apple Intelligence套件



中科创达在COMPUTEX 2024展示端侧智能产品魔方工具箱



AIGC+传媒：人机协同高效创作，多模态驱动传媒行业效能提升

➤ 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AI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。2024年2月，OpenAI发布文生视频大模型Sora，在全球视频大模型领域取得里程碑式进展。同月，苹果Vision Pro在美国上市，苹果CEO蒂姆·库克在WWDC2024上宣布，Vision Pro将于6月向中国内地、中国香港、日本和新加坡等地用户开售。

应用场景

- 文本方面：客服问答、新闻撰写、营销文本生成、文案生成、内容续写等
- 图片方面：虚拟人创作、平面设计、AI营销等
- 视频方面：动画片生成、广告生成等
- 游戏方面：剧情生成、地图生成等
- 电商方面：智能购物客服、智能推送等

具体案例

- 万兴科技：聚焦以视频创意为核心的数字创意软件业务，打造垂类音视频多媒体大模型“天幕”，并将系列算法能力应用于万兴喵影/Filmora、万兴PDF/PDFelement、亿图脑图/EdrawMind等原产品和万兴播爆/Virbo、TruMate、Kwicut等AIGC新品。
- 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。
- 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。2023年10月，公司宣布推出“中文逍遥”万字创作大模型，可实现一键生成万字小说、一张图写一部小说、一次读懂100万字小说等，今年5月已成功通过备案。

◆ 万兴科技：研发天幕多媒体大模型，提升AI原子能力

多媒体能力服务全球用户	应用	天幕多媒体大模型				
	组装配件	视频编辑框架	音乐创作引擎	特效生产引擎	软件智能助手	
	原子能力	大模型		基础算法池		
		视频大模型 文生视频生成 视频内容生成 文生视频生成 数字人播报	音乐大模型 AI编曲 文生音乐 虚拟人配音 语音生成	图片大模型 语言大模型	人脸识别 背景分割 人像分割 天空分割 降噪增强 视频超分 视频插帧 高光提取 视频降噪 视频去噪 音频分割 音频修复 节拍检测 音频去噪	
	基础设施	自研推理框架		数据生产平台		
		高性能算力平台		云原生容器平台		

◆ 蓝色光标：1) 数字人苏小妹，2) 协助吉林文旅打造“长白山下雪”首个冰雪文化虚拟数字人



AIGC+金融：金融科技深化发展，大模型可用于风险管理、投资研究等业务

▶ 金融机构数字化转型需求的不断提升，以及AI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。

应用场景

- 计算机视觉：智能安防、智能内控管理等
- 机器学习&知识图谱：风险识别、客户信用能力评估、知识检索等
- 智能语音&对话式AI：智能客服、智能营销、智能培训等
- 自然语言处理：文案生成、投研分析、智能风控、投顾助手等

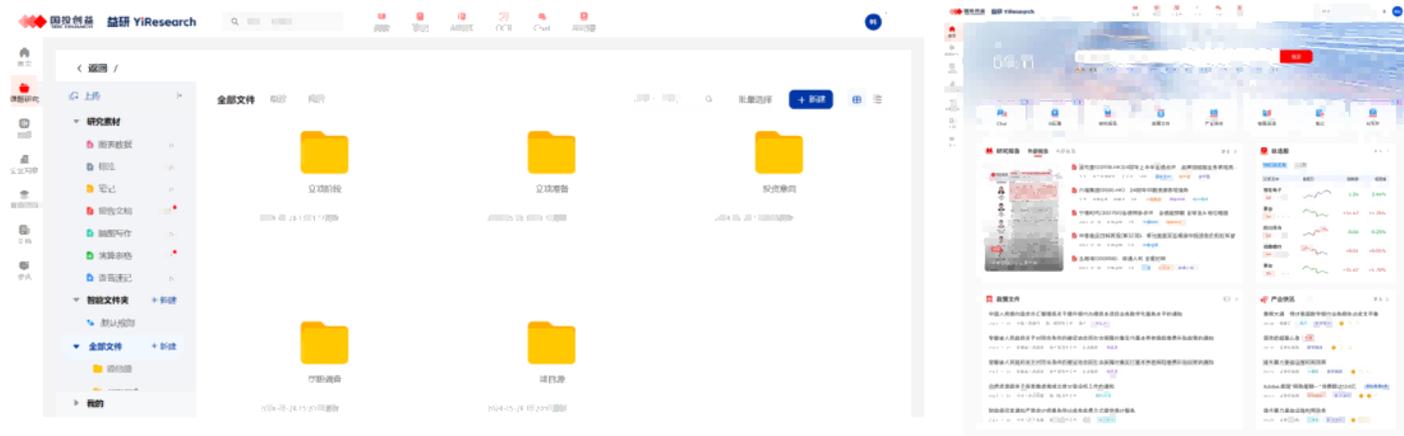
具体案例

- 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。2024年1月，公司自主研发的“妙想”金融大模型正式开启内测。
- 恒生电子：2023年6月，公司发布智能投研平台WarrenQ、金融智能助手光子以及金融行业大模型LightGPT，10月，公司发布金融大模型LightGPT最新能力升级成果以及应用产品。公司大模型应用产品与多家客户开展共建，今年4月，恒生LightGPT金融领域文本生成类算法备案通过。

同花顺i问财界面



恒生电子“投研+大模型”案例——“益研YiResearch”平台



AIGC+搜索：搜索为C端高频使用场景，AI驱动搜索更精准、更全面

➤ AI搜索带来更智能化的搜索体验，成为新的流量入口。AI搜索作为AI产业的催化剂，具体作用包括提升数据检索效率、优化个性化推荐、加强自然语言理解、促进跨模态交互，赋能各行各业如医疗诊断、金融分析、教育定制化服务，推动智能化转型。AI搜索工具方面，谷歌、百度利用AI增强搜索精度，Perplexity、秘塔AI等初创公司探索AI搜索新场景，共同推动信息检索技术的创新和应用。

应用场景

- 搜索引擎优化：实时资讯获取、跨模态检索、学术研究，提供更准确和个性化的搜索结果。
- AI助手：理解和响应用户的语音查询，并提供相关信息。
- 个性化推荐：利用AI搜索技术根据用户行为和偏好推荐商品。

具体案例

- Perplexity AI：基于GPT-3.5、Llama、Mistral等模型推出AI搜索产品，并推出信息来源编辑、Copilot（个性化搜索）等特色功能，产品和商业模式获得市场认可。
- 秘塔科技：自研大模型MetaLLM，产品包括秘塔AI搜索、秘塔写作猫，秘塔AI搜索3月访问量达到721万次，在“AI产品榜”国内总榜中仅次于百度文心一言、Moonshot Kimi。
- 昆仑万维：推出国内第一款融入大语言模型的搜索引擎天工AI搜索，目前，除了AI搜索、AI音乐、AI绘画、AI写作等功能，天工AI陆续推出了AI搜索研究/增强模式、AI PPT、AI智能体平台、AI数据分析等多项功能。

秘塔AI搜索主界面：简洁、深入、研究三个搜索选项，全网、学术两种模式



AI产品榜3月国内总榜

国内排名	产品名	分类	3月上榜访问量	3月上榜变化
1	百度文心一言	AI ChatBots	14.9M	48.14%
2	Kimi (Moonshot)	AI ChatBots	12.61M	321.58%
3	秘塔AI搜索	AI Search Engine	7.21M	551.36%
4	阿里通义千问	AI ChatBots	5.18M	43.92%
5	火山方舟	Model Training & Depk	3.26M	82.64%
6	清华智谱清言	AI ChatBots	3.02M	78.44%
7	讯飞星火	AI ChatBots	2.83M	46.47%
8	抖音豆包	AI ChatBots	2.74M	61.03%

昆仑万维天工AI搜索研究模式自动生成研究大纲、思维导图等



AIGC+办公：AI办公助手定价体系逐步落地，基于庞大客户群带来巨大想象空间

- 大模型在办公领域的应用主要以办公助手的形式体现，微软Copilot及金山办公WPS AI的定价体系陆续推出，使得大模型提升单客户价值量的逻辑进一步被验证，并且大模型算法的价值也得到进一步彰显。办公软件市场下游拥有庞大的客户群体，未来有望受益于AI功能的加持打开巨大想象空间。

应用场景

- 内容生成：生成会议纪要、对话问答、总结分析、智能改写等
- 理解能力：语言表达判断修改、文字润色、内容分析等
- 多模态：图片生成、语音生成、视频生成等
- 数理计算：数据分析、公式推理等

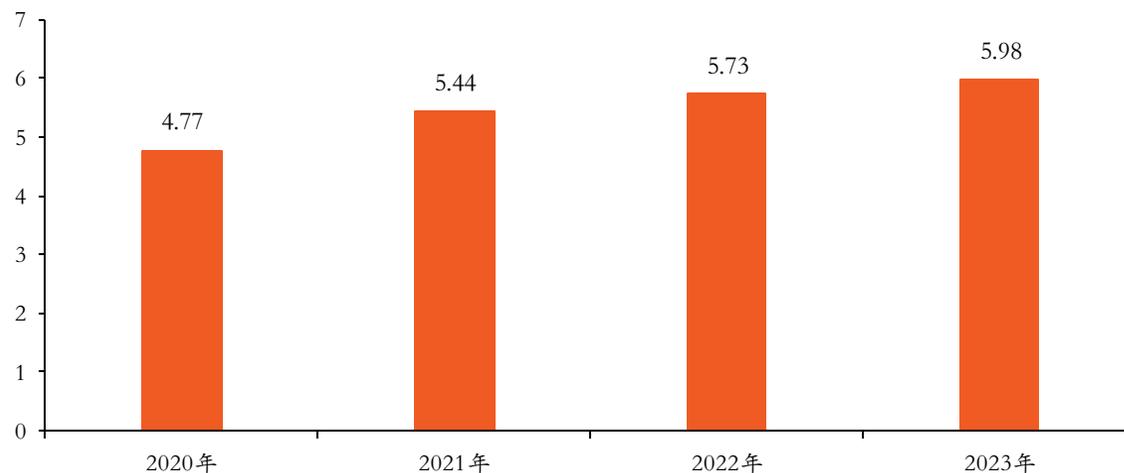
具体案例

- 微软：2023年3月，微软发布Office 365 Copilot作为AI助手工具为“Office全家桶”赋能，并于同年7月份公布了to B定价体系：以30美元/月的定价向Microsoft 365商业标准版、商业高级版、E3、E5的客户开放。当前，Microsoft 365商业标准版/商业高级版/E3/E5用户月费分别为12.5/22/36/57美元。
- 金山办公：2023年上半年，金山办公发布了基于大语言模型的智能办公助手WPS AI，并于2024年3月底将会员体系调整为WPS超级会员、WPS AI会员、WPS大会员，其中后两者包含AI功能，目前定价分别为连续包年198元/年、288元/年。

WPS会员体系

会员体系	会员权益	定价
WPS超级会员	WPS会员+稻壳会员+更多权益	128元/月
WPS AI会员	AI内容创作、AI生成PPT等10+项AI权益	198元/月（连续包年）
WPS大会员	WPS超级会员+WPS AI会员+1T云空间+更多权益	288元/月（连续包年）

金山办公主要产品月度活跃设备数（亿）



AIGC+教育：教育逐渐进入大模型时代，丰富落地场景提供广阔发展空间

➤ 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。Duolingo、科大讯飞、佳发教育等公司已将自研或其他主流大模型与自有业务或产品进行深度融合。大模型在教育领域有丰富的落地场景，未来具有广阔发展空间。

应用场景

- 精准教学：智慧体育、智能语言学习、早教、智能批改、VR\AR教学等
- 自动评阅：口语考评、试卷分析、智能阅卷等
- 科学管理：智慧校园、智能排课、智慧安保、智慧招考等
- 个性化学习：自适应学习、拍照搜题、教育机器人、AI学习助手等

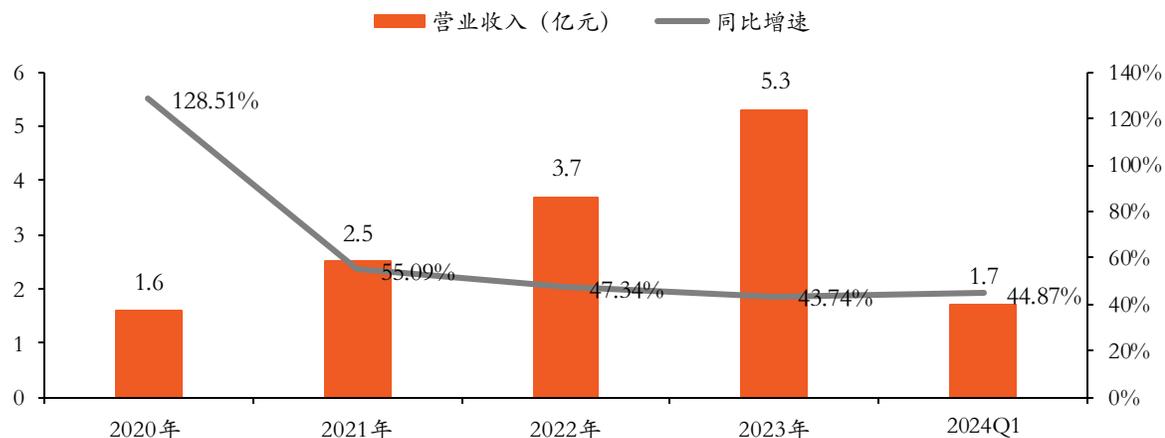
具体案例

➤ **Duolingo**：2023年3月，公司在 App 中引入了GPT-4。价格方面，由最初的 Duolingo Super 订阅套餐(12.99/月)提升到新套餐 Duolingo Max(14.99/月)。公司2020年推出了AI 模型Birdbrain为用户推荐练习题目，2023年6月，公司开始使用AI来创建题目。

➤ **科大讯飞**：公司自研讯飞星火大模型赋能教育业务成果显著，公司个性化学习手、讯飞 AI学习机、星火语伴APP、星火教师助手、AI心理伙伴等产品均已搭载讯飞星火大模型。其中，个性化学习手册英语学科产品率先深度融合星火认知大模型，2023年实现了单品销量同比增长40%。

➤ **佳发教育**：2023年，公司完成灵汭教育大模型及AI系列产品的发布，顺势推出“考试智能行为分析解决方案”、“理化生实验AI解决方案”、“英语听说AI解决方案”以及“体育教育AI解决方案”等一系列AI创新成果。目前灵汭具有百亿级的参数量，并且通过教育领域的专项数据，完成百亿级的Token训练数据的二次预训练，千万级Token的指令数据专项场景微调。

Duolingo营业收入情况



资料来源：WIND，各公司公告，佳发教育公众号，平安证券研究所

科大讯飞AI学习机产品



AIGC+医疗：医疗企业相继入局大模型赛道，有望赋能自有业务加速发展

- AIGC赋能诊疗全流程，与医学影像诊断、慢性病管理和生活方式指导、疾病排查和病理研究、药物开发等多个医学领域具有较高契合度。目前，医渡科技、嘉和美康、卫宁健康等公司纷纷推出医疗大模型产品，医疗企业相继入局大模型赛道，有望为自有业务深度赋能。

应用场景

- 医疗器械：AI心电图、AI影响、AI放疗器械、医疗机器人等
- 医疗服务：智能检验、智能病理判断、电子病历、AI医疗助理等
- 医药研发：AI制药、AI合同研发服务（CRO）等
- 医疗管理：智慧医院、健康管理、专家系统管理等

具体案例

- **医渡科技**：2023年，公司发布了自研医疗垂域大模型，在2024年5月份在面向中文医疗大语言模型的开放评测平台MedBench最新评测榜单中登顶榜首。作为YiduCore的核心算法部分，医渡科技大模型在为 用户赋能的同时，也不断反哺迭代YiduCore的算法。YiduCore已在近80个疾病领域建立疾病模型，覆盖超过1700家医院。
- **嘉和美康**：2023年，公司完成了嘉和医疗大模型的开发与训练，总训练Token数达到140 亿，目前，嘉和医疗大模型支持智能分诊、智能预问诊、病历自动生成、疾病诊断预测、诊疗方案推荐和智能随访等近20种医疗任务。
- **卫宁健康**：2023年，公司正式发布医疗垂直领域大模型WiNGPT，并以医护智能助手WiNEX Copilot的形态内置于全系列产品中。WiNEX Copilot已集成涵盖医院管理、医生增效以及患者端等100多个临床应用场景。2023年，WiNEX Copilot病历文书助手、影像报告助手等智能应用已在北京、上海等地的头部医院推进落地，在医疗领域大模型应用方面形成了比较优势。

医渡科技大模型在MedBench评测平台登顶榜首（评测名：HH-YIDU-Med）

模型名称	发布日期	类型	参数量	综合得分
HH-YIDU-Med	2023/12/21	闭源	70B	61.3
Inf-Med	2024/05/09	闭源	60B	56.4
UNIGPT-MED	2024/04/19	闭源	13B	54.7
jq_med_sft_v2	2024/05/07	闭源	6B	50.9
AntGLM-Med	2024/04/30	闭源	10B	49.3

WiNGPT通过Copilot辅助诊断模式与WiNEX产品全线深入融合



AIGC+汽车：FSD v12引领智驾进入端到端时代，国内大模型上车步伐加速

随着AIGC时代来临，借助于大模型的决策和推理能力，智能汽车的人机交互和用户体验将迎来新一轮变革。目前，各厂商通过AIGC技术努力提升智能汽车的智能座舱、自动驾驶等领域的产品体验，AIGC+汽车的应用场景空间广阔。

应用场景

- 智能座舱技术：语音交互、车联网、OTA
- 自动驾驶技术：图像识别与感知（雷达传感器、视觉传感器、定位及位姿传感器、车身传感器）；深度学习（无人驾驶软件系统）；信息共享（实时数据分享、实时数据处理）

具体案例

► **特斯拉**：2023年11月，特斯发布FSD V12第一个测试版本，仅向部分的员工开放，2024年1月开始正式向用户推送，目前，特斯拉正在努力将FSD V12引入中国。FSD v12是首个端到端AI自动驾驶系统，将城市街道的驾驶堆栈升级为端到端大模型架构，经过数百万个视频剪辑的训练，取代了超过 30 万行显式C++代码。FSD V12开始引领自动驾驶进入“端到端”时代。

► **国内**：1) **智能座舱**：百度的文心一言已有东风日产、红旗、长城以及吉利银河等多家车企宣布接入；2023年4月，商汤科技在上海车展期间发布了日日新SenseNova大模型，展示了其中文语言模型“商量SenseChat”以及“如影SenseAvatar”等与座舱的结合；理想汽车在2023年6月发布大模型MindGPT，其原有的座舱人工智能助手理想同学将基于大模型打造新一代的多模态人机交互技术体系。2) **智能驾驶**：2023年4月，毫末智行发布了自动驾驶生成式大模型DriveGPT，目标实现端到端自动驾驶；理想汽车自研大模型MindGPT，摆脱了对高清地图的依赖。

理想自研多模态认知大模型MindGPT



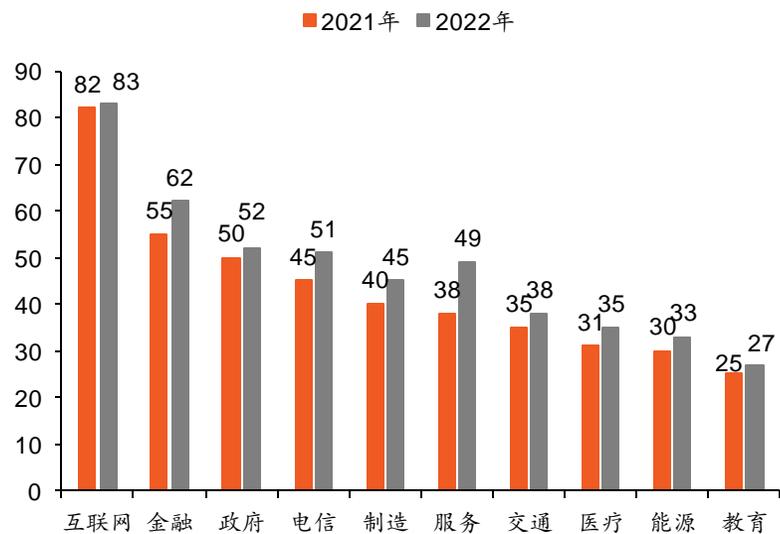
毫末智行DriveGPT

ChatGPT vs DriveGPT		HAOMO.AI
对比维度	ChatGPT	DriveGPT曹源·海若
建模方式	生成式预训练大模型 (GPT)	
系统流程	预训练 (Pre-training)、奖励模型 (Reward Model)、人类反馈强化学习 (RLHF)、提示语 (Prompt)	
数据形态	输入：自然语言文本串 输出：自然语言文本	输入：视觉传感器图片序列 输出：生成的未来2-3秒BEV序列，包含自行车和所有障碍物
产品定位	聊天机器人	自动驾驶机器人
应用场景	人机对话	自动驾驶
包含能力	<ul style="list-style-type: none">文本撰写知识问答公文写作数学计算代码编写考试应答	<ul style="list-style-type: none">场景理解场景识别场景重建感知数据生成认知数据生成驾驶辅助云端驾驶决策测评多智能体仿真 <p>基于多模态理解能力，具备对话式文理解能力</p> <p>实现对单帧图片、环视Bundle、4D Clips自动化标注</p> <p>实现对单帧或多帧结构的纯视觉NeRF三维重建</p> <p>实现不同场景的感知数据生成和障碍物、交通元素编辑</p> <p>基于DriveGPT生成各类接管场景的驾驶决策真值</p> <p>包含通用感知能力和驾驶决策能力，提升驾驶效果</p> <p>大模型通过“teacher-student”模式赋能车端小模型</p> <p>在仿真环境布置多个DriveGPT智能体，实现多车博弈仿真</p>

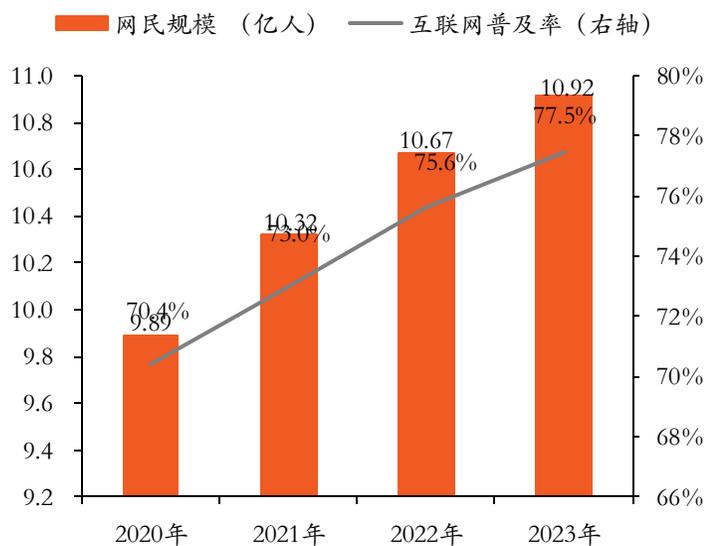
大模型赋能千行百业，AIGC未来发展前景广阔

- 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、服务、交通、医疗、能源、教育等行业的渗透率分别为83%、62%、52%、51%、45%、49%、38%、35%、33%、27%。随着大模型时代的到来，我国多家互联网企业的掌门人均表示，大模型时代，所有的产品都值得用大模型重做一次。根据中国互联网络信息中心(CNNIC)数据，截至2023年12月，我国网民规模达10.92亿，互联网普及率达77.5%。
- 我们认为，随着国产大模型的逐步成熟，在政策与技术的共振下，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。根据中国信通院数据，2022年，我国数字经济规模达到50.2万亿元，占GDP比重达到41.5%。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。

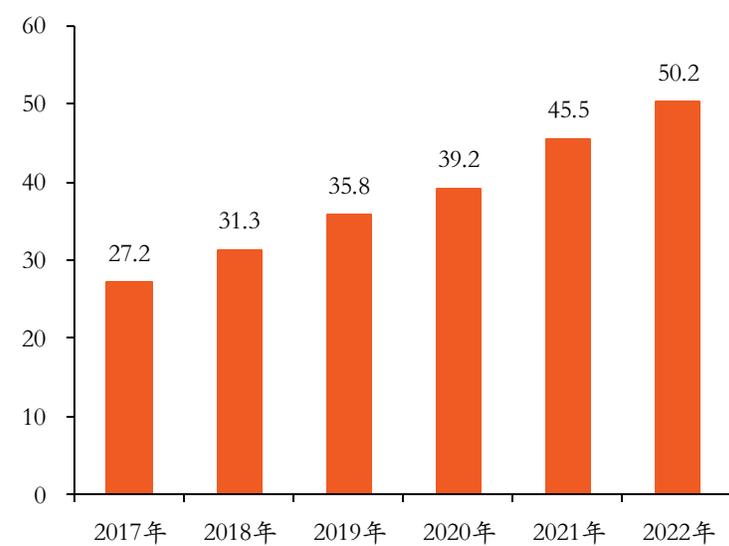
◆ 2021-2022年我国各行业人工智能渗透率 (%)



◆ 我国网民规模和互联网普及率



◆ 2017-2022年我国数字经济规模 (万亿元)





目录CONTENTS

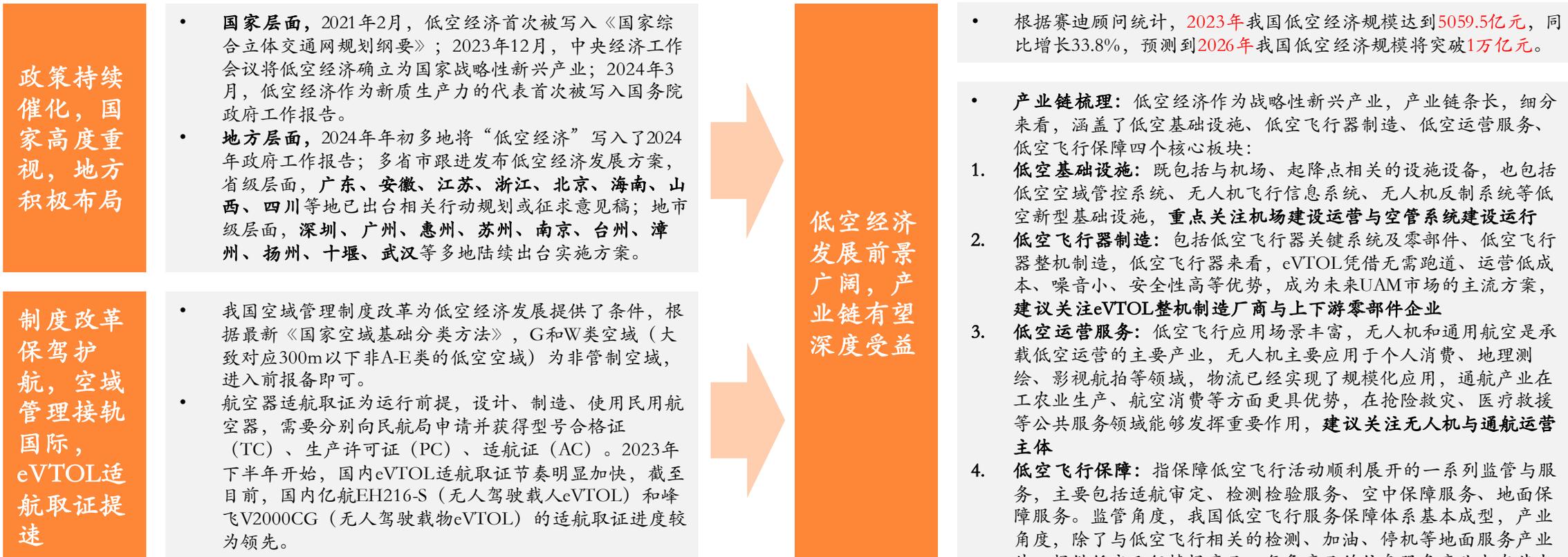
- 行业回顾：行业行情表现不佳，估值处于历史较低水平
- 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- 低空经济：政策持续催化，未来展翼腾飞
- 投资建议及风险提示



政策持续催化，低空经济产业链有望展翼腾飞

➤ 低空经济是以eVTOL飞行器为重要牵引的一种综合性经济形态，无人机与通用航空产业为低空经济发展的核心产业。当前，政策推动低空经济进入快速发展期。根据赛迪顾问统计，2023年我国低空经济规模达到5059.5亿元，同比增长33.8%，预测到2026年我国低空经济规模将突破1万亿元。我们认为，在国家高度重视和地方积极布局的背景下，低空经济未来发展前景广阔。具体来看，低空飞行器研发制造、低空基础设施、低空飞行保障服务等细分行业均将受益于低空经济的发展。

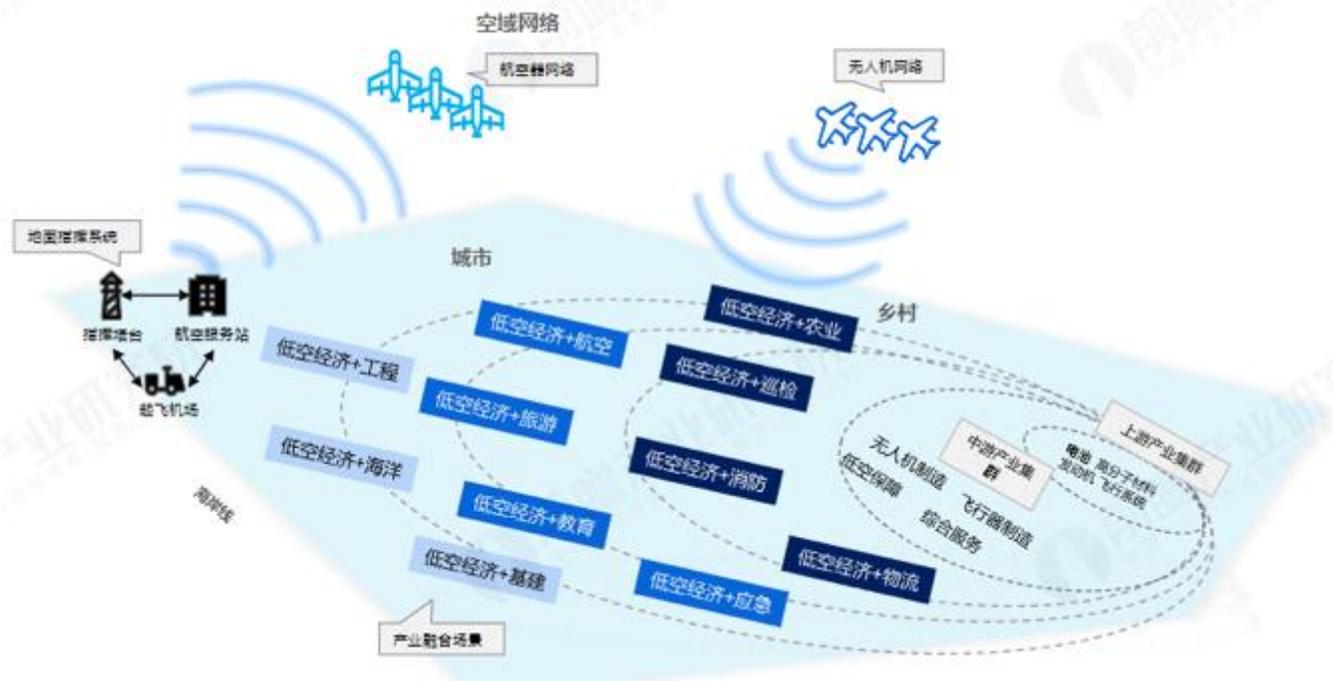
低空经济领域投资逻辑图



✍️ 低空经济首次写入国务院政府工作报告，被定义为新增长引擎

➤ 低空经济是以各种有人驾驶和无人驾驶航空器的各类低空飞行活动为牵引，辐射带动相关领域融合发展的综合性经济形态，具体包括距离地面垂直高度1000米以内、视地区特性和实际需求可延伸至3000米以内空域产生的一系列经济活动。早在2021年2月，低空经济就首次被写入《国家综合立体交通网规划纲要》，2023年12月中央经济工作会议将低空经济确立为国家战略性新兴产业，2024年3月，国务院政府工作报告指出，大力推进现代化产业体系建设，加快发展新质生产力，积极培育新兴产业和未来产业，积极打造生物制造、商业航天、低空经济等新增长引擎，此为低空经济首次被写入国务院政府工作报告。

◆ 低空经济是低空飞行活动与产业融合的新型经济形态



◆ 低空无人飞行器相关参数及应用

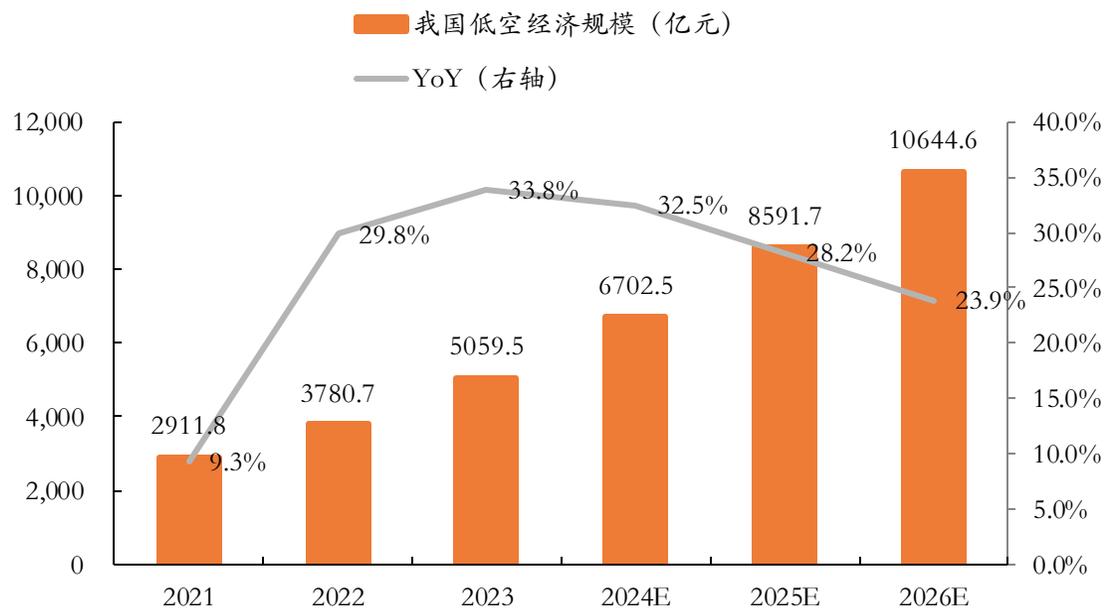
飞行器类型参数 (载人/无人、飞行安全风险)	飞行器应用
600-1000米 eVTOL	✓ 载人飞行器
300-600米 大型无人机/eVTOL • 最大起飞重量超过150千克	✓ 行业级无人机: 工业运输等
200-300米 中型无人机 • 最大起飞重量不超过150千克	✓ 行业级无人机: 消防、应急等
120-200米 小型无人机 • 空机重量不超过15千克 • 最大起飞重量不超过25千克	✓ 行业级无人机: 快递物流等
50-120米 轻型无人机 • 空机重量不超过4千克 • 最大起飞重量不超过7千克	✓ 行业级无人机: 即时物流配送, 城市管理
50米以下 微型无人机 • 空机重量小于0.25千克	✓ 消费级无人机为主



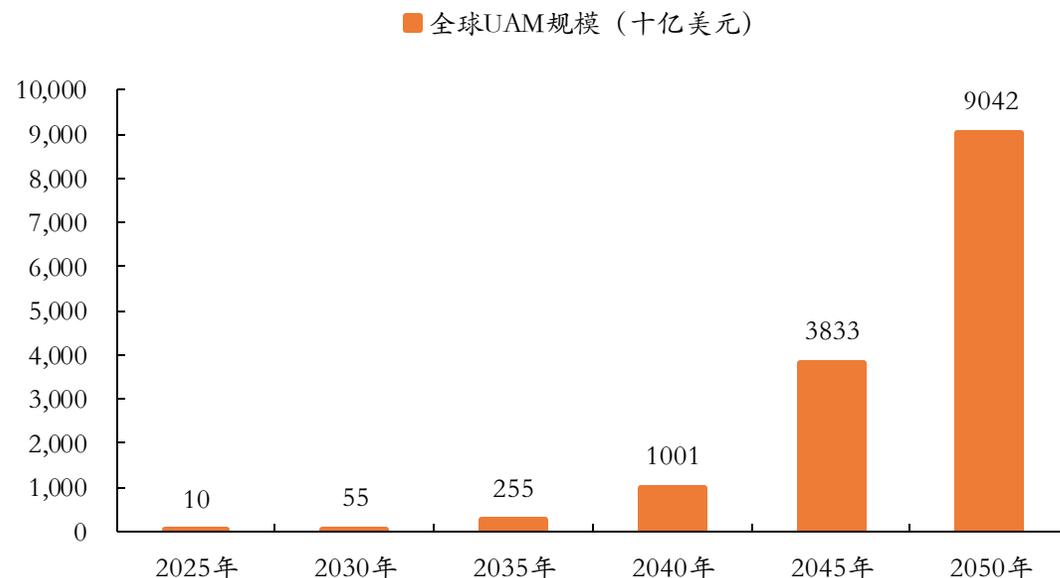
低空交通填补中高空与地面交通空白，发展前景广阔

传统的地面交通受限于平面交通路网，容量有限，而低空交通点对点可达，效率优化空间更大。低空经济从地面延伸至低空空域，从平面路网延伸至3维立体空间，在减轻地面交通压力的同时，也创造了一个新的产业空间。根据赛迪顾问统计，2023年我国低空经济规模达到5059.5亿元，同比增长33.8%，预测到2026年我国低空经济规模将突破1万亿元。全球市场来看，按照Morgan Stanley的预测，2030年全球城市空中交通（Urban Air Mobility, UAM）规模550亿美元，2040年达到1万亿美元，2050年达到9万亿美元。

工信部下赛迪顾问统计并预测我国低空经济规模



Morgan Stanley关于全球UAM市场规模的预测





国家与地方政策加快推进，低空经济步入快速发展期

➤ 国家层面，2010-2020年是低空概念提出以及低空经济初步发展的时期，2021年伴随首次写入国家规划，我国低空经济进入快速发展阶段。

➤ 为了适配我国低空经济的发展，可以看到“十四五”期间低空经济相关政策出台正在提速，2023年，无人驾驶航空器立法出台，空域管理条例等也在不断细化，2024年低空经济首次写入了政府工作报告。

2021至今
快速发展期

2010-2020
概念提出&
初步发展期

我国国家层面低空经济政策梳理

日期	文件/会议	主要内容
2024/03	工信部、科技部、财政部、民航局《通用航空装备创新应用实施方案（2024-2030年）》	到2027年，航空应急救援、物流配送实现规模化应用，城市空中交通实现商业运行，形成20个以上可复制、可推广的典型应用示范，打造一批低空经济应用示范基地，形成一批品牌产品。到2030年，以高端化、智能化、绿色化为特征的通用航空产业发展新模式基本建立，支撑和保障“短途运输+电动垂直起降”客运网络、“干-支-末”无人机配送网络、满足农作业需求的低空生产作业网络安全高效运行。
2024/03	第十四届全国人民代表大会第二次会议	“低空经济”首次写入政府工作报告，政府工作报告提出“积极打造生物制造、商业航天、低空经济等新增长引擎”。
2023/12	民航局《国家空域基础分类方法》	将我国空域划分为A、B、C、D、E、G、W等7类，其中：A-E类为管制空域，G、W类为非管制空域。
2023/12	中央经济工作会议	将低空经济纳入战略新兴产业范畴，提出“打造生物制造、商业航天、低空经济等若干战略性新兴产业”。
2023/11	国家空管委《中华人民共和国空域管理条例（征求意见稿）》	明确提出空域用户定义并提出空域用户的权利、义务规范，标志着我国空域放开有了实质性的突破。
2023/10	民航局《民用无人驾驶航空器系统物流运行通用要求第1部分：海岛场景》	规定了应用于海岛场景从事物流的民用无人驾驶航空器系统运行的通用要求。
2023/10	工信部等四部门《绿色航空制造业发展纲要（2023-2035年）》	提出发展目标，到2025年，使用可持续航空燃料的国产民用飞机实现示范应用，电动通航飞机投入商业应用，电动垂直起降航空器（eVTOL）实现试点运行。到2035年，建成具有完整性、先进性、安全性的绿色航空制造体系，新能源航空器成为发展主流，以无人化、电动化、智能化为技术特征的新型通用航空装备实现商业化、规模化应用。
2023/06	国务院、中央军委《无人驾驶航空器飞行管理暂行条例》	2024年1月1日起，《无人驾驶航空器飞行管理暂行条例》将正式施行，标志着我国无人机产业将进入“有法可依”的规范化发展新阶段。
2022/02	民航局《“十四五”通用航空发展专项规划》	提出“十四五”期间，发展规模实现新跃升，通用航空（含无人机）企业、飞行总量、航空器、执照等数量显著增加；保障能力取得新突破，力争低空空域改革取得实质性进展，推动低空空域分类划设，航路航线大幅拓展，在册通用机场布局合理。
2022/01	国务院《“十四五”旅游业发展规划》	提出完善邮轮游艇旅游、低空旅游等发展政策，选择一批符合条件的旅游景区、城镇开展多种形式的低空旅游，强化安全监管，推动通用航空旅游示范工程和航空飞行营地建设。
2022/01	国务院《“十四五”现代综合交通运输体系发展规划》	提出深化重点领域改革之一为：持续推进空管体制改革，完善军民航空管联合运行机制，实施空域资源分类精细化管理，优化全国航路航线网，深化低空空域管理改革。有序推进通用机场规划建设，构建区域短途运输网络，探索通用航空与低空旅游、应急救援、医疗救护、警务航空等融合发展。
2022/01	民航局、发改委、交通运输部《“十四五”民用航空发展规划》	提出构建运输航空和通用航空一体两翼、覆盖广泛、多元高效的航空服务体系。服务体系更加健全，货运网络更加完善，通用航空服务丰富多元，无人机业务创新发展。到“十四五”末，运输机场270个，市地级行政中心60分钟到运输机场覆盖率80%，通航国家数量70个。
2021/02	中共中央、国务院《国家综合立体交通网规划纲要》	提出发展交通运输平台经济、枢纽经济、通道经济、低空经济。这是“低空经济”首次被写入国家规划。
2019/05	民航局空管办《促进民用无人驾驶航空发展的指导意见（征求意见稿）》	提出以低空、隔离运行行为为起点，逐步积累实践经验和运行数据，不断提高面向国家、行业、社会及大众的航空服务能力。重点开展低空无人机公共航线划设和运行研究，组织开展垂直起降载人（VTOL）及物流无人机试运行，为制定适航、飞标、空管运行规则、标准提供依据。
2018/09	民航局《低空飞行服务保障体系建设总体方案》	明确了全国低空飞行服务保障体系由1个国家信息管理系统、7个区域信息处理系统以及一批飞行服务站组成。
2016/05	国务院办公厅《关于促进通用航空业发展的指导意见》	提出发展目标：到2020年，建成500个以上通用机场，基本实现地级以上城市拥有通用机场或兼顾通用航空服务的运输机场。通用航空器达到5000架以上，年飞行量200万小时以上，培育一批具有市场竞争力的通用航空企业。通用航空业经济规模超过1万亿元。
2014/07	《低空空域使用管理规定（试行）（征求意见稿）》	定义低空空域原则上是指全国范围内真高1000米（含）以下区域。提出低空空域按管制空域、监视空域和报告空域以及目视飞行航线进行分类。监视、报告空域的飞行计划，通航用户需向空军和民航局报备。
2010/11	国务院中央军委《关于深化我国低空空域管理改革的意见》	总体目标是，通过5至10年的全面建设和深化改革，在低空空域管理领域建立起科学的理论体系、法规标准体系、运行管理体系和服务保障体系，逐步形成一整套既有中国特色又符合低空空域管理规律的组织模式、制度安排和运作方式，充分开发和有效利用低空空域资源。具体实施分3个阶段，2011年前为试点阶段，2011年至2015年底为推广阶段，2016年至2020年为深化阶段。



国家与地方政策加快推进，低空经济步入快速发展期

➤ **地方层面**，2024年年初，多地区都将“低空经济”写入了2024年政府工作报告。据不完全统计，已有17个省（直辖市、自治区）将“低空经济”有关内容写入2024年政府工作报告，此外还有10个省（直辖市、自治区）的政府工作报告也有涉及低空经济的内容，以及我们看到，深圳、广州、成都、苏州、珠海、赣州等城市也将低空经济写入了政府工作报告。

➤ **各地政策持续催化**，2024年以来，多省市跟进发布低空经济发展方案，省级层面，广东、安徽、江苏、浙江、北京、海南、山西、四川等地已出台相关行动规划或征求意见稿；地市级层面，深圳、广州、惠州、苏州、南京、台州、漳州、扬州、十堰、武汉等多地陆续出台实施方案。

资料来源：各政府或各部门官网，平安证券研究所

多省市发布低空经济实施行动方案(5-6月)

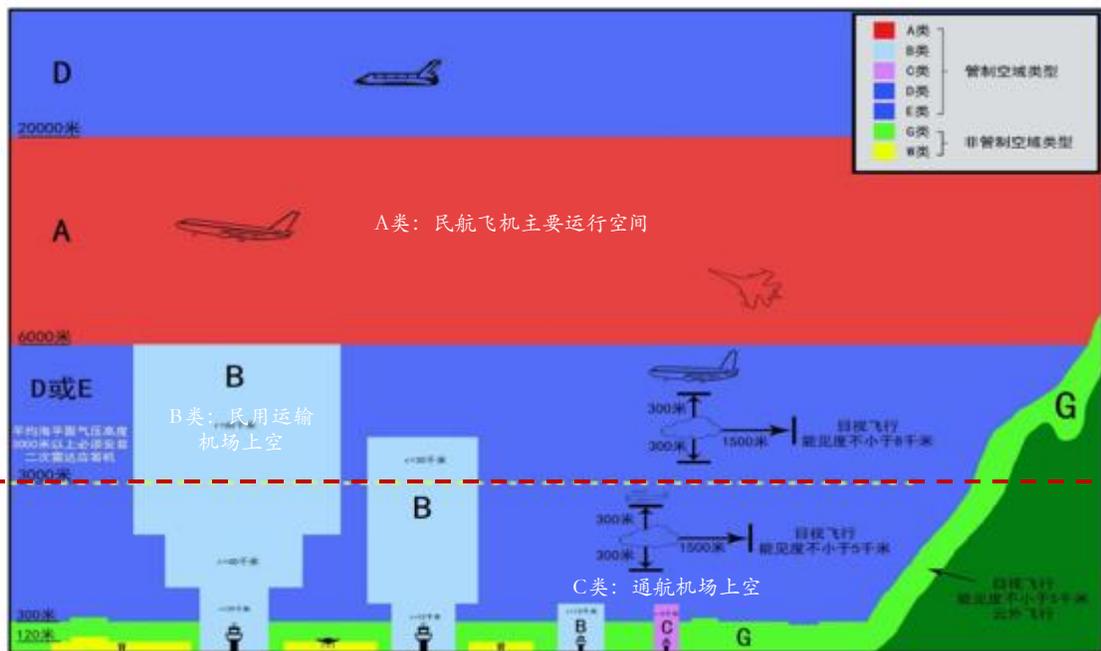
颁布日期	颁布部门/省市	文件/活动	内容
2024/5/7	成都市产业建圈强链工作领导小组	《成都市产业建圈强链2024年工作要点》	在低空经济方面，加大整机项目内培外引力度，精准补齐关键零部件短板，产业需求导向布局建设基础设施、公共平台、应用场景，加快打造西部低空经济中心。
2024/5/7	安徽省发展和改革委员会	《安徽省推动空天信息产业高质量发展行动方案（征求意见稿）》	构筑低空基础设施、空域资源利用、监管服务为重点的低空飞行保障体系，实施“低空+”应用场景工程，建设低空场景应用大市场，推进低空要素化、要素场景化、场景经济化发展。
2024/5/8	浙江省财政厅	《关于支持高水平建设民航强省 打造低空经济发展高地要素保障若干政策措施（征求意见稿）》	通过省市县三级联动，以市场需求为牵引、以科技创新为驱动，充分发挥要素政策杠杆作用，全面助力高水平建设民航强省、打造低空经济发展高地。
2024/5/10	苏州市	《苏州工业园区低空经济高质量发展行动计划（2024—2026年）》	以应用场景为抓手，以低空保障和特色服务为支撑，带动低空制造做大做强，全面提升低空经济体系化、数字化、融合化发展水平，努力打造全国低空经济示范区先导区，为园区加快形成新质生产力提供有力支撑。
2024/5/13	南京市	《南京市促进低空经济高质量发展实施方案（2024—2026年）》	通过三年时间的建设，在全市建成管理体制顺畅、运行机制高效、基础设施领先、示范效果显著、服务保障有力的低空飞行管理服务体系，支撑低空运行活动顺利开展；建成开放、创新、融合、协同的低空经济产业体系，成为全国低空经济产业高地；培育一批在全国有较高知名度的低空领域科技创新服务机构，成为全国低空经济主要科创策源地。
2024/5/15	南京市	《南京市关于支持低空经济高质量发展的若干措施（试行）》	《若干措施》主要涵盖持续扩大低空产业规模、建设低空飞行保障体系、拓展丰富低空应用场景、优化低空产业发展生态4个方面12条具体的政策措施。
2024/5/15	北京市经济和信息化局	《北京市促进低空经济产业高质量发展行动方案（2024—2027年）（征求意见稿）》	完善有条件地区低空空域环境，在延庆、平谷、密云等有条件的区构建数字低空一张网，完善航线网络体系建设；提升数字低空飞行服务保障能力，加强低空飞行服务保障系统建设。到2027年，本市将围绕应急救援、物流配送、空中摆渡、城际通勤、特色文旅等，新增10个以上应用场景，开通3条以上面向周边地区的低空航线，基本建成网络化的基础设施体系及低空应用生态。
2024/5/16	海南省	《海南省“低慢小”航空器活动区域管理办法（试行）》	涵盖海南省内、进出内地、进出境外等3个活动区域的“低慢小”航空器管理，并将通用机场货物、物品、人员、交通运输工具的管理纳入办法，即明确通过通用机场和临时起降点从事民用飞行活动的航空器，参照“低慢小”航空器管理。
2024/5/17	山西省	《山西省加快低空经济发展和通航示范省建设的若干措施》	大力拓展通航应用场景，构建通航公共服务网络；加快通航基础设施建设，提升运营服务保障能力；提高通航研发制造水平，强化创新平台驱动效能；激发通航消费市场活力，营造浓厚通航文化氛围；加大通航人才培养力度，夯实通航产业发展基础。
2024/5/20	台州市	《台州市低空经济高质量发展实施方案（2024—2028年）》	着眼发展新质生产力，以无人机、通用航空和商业卫星为主攻方向，以“源头创新—成果转化—研发制造—场景应用—技术培训—市场运维”为发展路径，全方位创新、全链条培育、全过程服务，培育低空制造产业集群（基地），打造成为全国领先的低空产业创新策源地、全国先进的无人机产业高地、长三角海联动低空经济服务示范区。
2024/5/20	漳州市	《漳州市促进低空经济产业高质量发展若干措施》	从生态培育、场景应用、产业环境配套等3方面提出13条措施，以深入推动漳州市低空经济产业高质量发展，加快低空经济产业战略布局，抢占技术新优势、发展新动能，打造低空经济创新发展高地。
2024/5/21	扬州市	《扬州市低空经济高质量发展实施意见》	将聚力打造低空经济体系实验室，并以航空航天（扬州）产业专项母基金为引导，撬动引领50亿元总规模的全社会资本投入扬州低空经济领域，更大力度鼓励社会资本参与低空经济发展。
2024/5/21	广东省	《广东省推动低空经济高质量发展行动方案（2024—2026年）》	广东省将依托广州、深圳、珠海三个低空经济核心城市强化引领支撑，发挥佛山、江门、惠州、东莞、中山等市制造业配套优势，推动粤东、粤西、粤北地区因地制宜打造低空应用场景，培育具有全球竞争力的低空经济产业集群。
2024/5/21	十堰市	《十堰市低空经济高质量发展实施方案（2024—2026年）》	《方案》明确——构建低空制造产业体系、提升低空制造创新能力、强化低空经济发展支撑等8个方面重点任务和加大制造产业招引力度、培育本地低空配套企业、打造“车机一体化”产业特色等32条措施。
2024/5/22	四川省	《关于促进低空经济发展的指导意见》	强调要充分发挥四川省研发、制造、空域、应用场景等优势，抢抓窗口期，在新赛道上加力加速冲刺，探索形成一批创新性引领性强的经验做法，加快研究出台配套政策措施，推动低空经济早“飞起来”，加快形成新质生产力。同时积极参与相关法规标准制定，构建全过程安全监管体系，确保空防安全、公共安全和飞行安全。
2024/5/22	武汉市	《关于支持武汉市低空经济高质量发展的实施意见》	意见涵盖支持低空经济重大项目落户；支持低空基础设施配套建设；支持低空经济标准规范制订；支持设立低空试点航线；支持企业提升科技创新能力；加强低空经济产业金融支持；支持低空经济人才引育；支持低空经济企业投资技改；促进低空经济产业集聚等九大低空经济发展措施。
2024/5/23	惠州市	《惠州市推动低空经济高质量发展行动方案（2024—2026）》	形成以低空应用创新为牵引、低空数字服务为动力、低空装备制造为支撑的低空经济产业体系，打造粤港澳大湾区低空经济产业高地；低空管理机制运转顺畅，“两网一平台”基础设施基本完备，应用场景加快拓展，低空航线成型成网、场景丰富多元、供给智能高效、监管安全规范，打造面向未来、立体协同、便捷智能的“天空智城”。
2024/5/31	广州市	《广州市人民政府办公厅关于印发广州市低空经济发展实施方案的通知》	到2027年，广州低空经济整体规模达到1500亿元左右。通航基础设施和飞行环境明显改善，以高端智能制造业为主导的产业体系初步形成，低空空域管理改革取得成效，低空飞行服务保障能力明显提升，低空领域技术创新水平全国领先。智能航空器销售方面，向全球生产销售“广州造”的首台飞行汽车，推动网联的载人航空器、飞行汽车、货运无人机、消费无人机、传统直升机等航空器制造业实现产值规模超1100亿元。城市先进空中交通商业运营方面，推动广州成为国内首个载人飞行商业化运营城市，低空经济跨境飞行、商务定制、短途客运、文旅消费、物流运输、应急医疗、会展服务等关键运营服务领域的市场规模达到300亿元。
2024/6/4	四川省	《四川省人民政府办公厅关于促进低空经济发展的指导意见》	到2027年，四川省将建成20个通用机场和100个以上垂直起降点，实现支线机场通航全覆盖，试点城市低空监管、服务、应用一体化信息平台建成投用，低空空域分类划设和协同管理取得突破性进展，在通航装备制造、低空飞行运营等领域各培育形成3-5家行业领军企业。开展低空经济试点建设，在全省范围内挑选发展基础较好的3-5个城市，联合省内外重点骨干企业，围绕通航短途运输、无人机物流配送、低空公共服务、低空消费、低空管理服务一体化信息平台等重点领域开展试点示范。省级财政安排2亿元资金专项支持低空经济发展。优先支持符合条件的项目申报地方政府专项债。鼓励金融机构针对低空飞行活动和基础设施建设创新金融产品。
2024/6/5	成都市	无人驾驶航空器飞行空域划设工作动员部署暨专题培训会	成都将启动无人驾驶航空器飞行空域划设工作，逐步将全市河道、天府绿道和超1500个各类公园等区域中的低风险区域上方真高120米以下空域开放为适飞空域。
2024/6/7	广东省	《广东省人民政府办公厅关于成立广东省推动低空经济高质量发展工作专班的通知》	广东省政府决定成立广东省推动低空经济高质量发展工作专班，由省府分管发展改革工作的副省长任召集人，省府协调发展改革工作的副秘书长、省发展改革委主要负责同志任副召集人，工作专班办公室设在省发展改革委。



低空空域管理改革促进军民融合，为低空经济发展提供条件

➤ 空域管理制度来看，自2000年以来我国低空空域管理经历多次改革，由军方主导逐步转向军民融合发展。美国在60年代开放了3000米以下的空域，eVTOL和直升机可以在无限制空域（一般为G级空域）自由飞行，低空空域管理基本趋向民用化管理。我国空域管理制度正在向国际接轨，根据中国民航局于2023年12月21日最新发布的《国家空域基础分类方法》，国家将空域划分为A、B、C、D、E、G、W七类，其中A-E类为管制空域，需空管部门进入许可；G和W类空域（大致对应300m以下非A-E类的低空空域）为非管制空域，进入前报备即可。eVTOL飞行范围通常涉及D、E、G、W类空域，无人驾驶类目前集中于W空域，为低空经济发展提供条件。

最新国家空域基础分类方法及要求：G、W类低空空域豁免飞行申请



低空经济对应 (3000m以下):
D/E类
G类
W类

空域类别	划设地域及范围	飞行类别	速度限制	通信要求	是/否需要ATC许可	监视设备
A类	标准气压高度6000米至20000米	A: 通常仅允许仪表飞行	A-B: 不适用	A-D: 持续双向	A-D: 是	A-C: 二次雷达应答机 (同等性能的监视设备)
B类	民用运输机场上空	B-G: 允许仪表和目视飞行	C: 仪表飞行不适用; 目视飞行 AMSL3000米以下, IAS不大于450千米/小时			
C类	建有塔台的通用航空机场上空		D-G: AMSL3000米以下, IAS不大于450千米/小时	E-G: 仪表飞行持续双向; 目视飞行保持守听	E: 仪表飞行需要; 目视飞行只需进入报告	
D类	标准气压高度高于20000米		E类			A、B、C、G类空域以外
E类	A、B、C、G类空域以外	G类	B、C类空域以外真高300米以下空域	W: 无	G-W: 否	E-E: AMSL3000米以上, 同A-C; 低于3000米安装可被监视的设备 G: 安装或携带可被监视的设备
G类	B、C类空域以外真高300米以下空域	W类	G类空域内真高120米以下的部分空域			

注: 1.ATC为空中交通管制、AMSL为平均海平面高度、IAS为指示空速的英文缩写。
2.当过渡高 (高度) 低于AMSL3000米时, 应当采用飞行高度层3000米代替AMSL3000米。



航空器适航取证为运行前提，国内eVTOL企业取证进度领先

- 我国航空器适航取证制度来看，根据《中华人民共和国民用航空法》和《中华人民共和国适航管理条例》，中国民用航空局（CAAC）是我国民用航空主管部门，任何单位或者个人设计民用航空器，应当向民航局申请并获得型号合格证（TC）；制造民用航空器，需要经生产许可审定并获得生产许可证（PC）；使用民用航空器，需要经单机适航检查并获得单机适航证（AC）。
- 2023年下半年以来国内适航取证节奏明显加快，截至目前，国内亿航EH216-S（无人驾驶载人eVTOL）和峰飞V2000CG（无人驾驶载物eVTOL）的适航取证进度较为领先。亿航于2020年12月提交EH216-S载人无人驾驶航空器系统型号合格证申请书，历时近三年，EH216-S于2023年10月获颁型号合格证（TC），成为全球首款取证的载人eVTOL，2023年12月取得标准适航证（AC），2024年4月7日取得生产许可证（PC）。2024年3月22日，峰飞自主研发的V2000CG无人驾驶航空器系统获颁TC，为全球首个通过型号合格认证的吨级以上eVTOL。国际适航取证进度来看，目前还没有一家eVTOL制造商获得FAA（美国联邦航空局）或EASA（欧洲航空安全局）的完全认证。

我国适航取证主要涉及的证书类型



型号合格证书(TC, Type Certificate)

设计民用航空器及其发动机、螺旋桨和民用航空器上设备时，应当向国务院民用航空主管部门申请领取型号合格证书



生产许可证书(PC, Production Certificate)

生产、维修民用航空器及其发动机、螺旋桨和民用航空器上设备时，应当向国务院民用航空主管部门申请领取生产许可证书、维修许可证书



适航证书(AC, Airworthiness Certificate)

具有中华人民共和国国籍的民用航空器，应当持有国务院民用航空主管部门颁发的适航证书，方可飞行

国内企业eVTOL适航认证情况

eVTOL型号	申请机构	申请日期	受理日期	批准审定计划CF	颁发TC证日期
广州亿航EH216-S	中国民航中南地区管理局	2020年12月	2021年1月	2022年9月	2023年10月12日
上海峰飞V2000CG 凯瑞鸥（载物版）	中国民航华东地区管理局	2022年9月27日	2022年9月29日	2023年5月	2024年3月22日
上海峰飞盛世龙（载人版）	中国民航华东地区管理局	2023年	2023年5月	-	-
成都沃飞长空 AE200	中国民航西南地区管理局	2022年11月22日	2022年11月23日	-	-
上海沃兰特VE25-100	中国民航华东地区管理局	已完成	2023年9月28日	-	-
上海时的科技	中国民航华东地区管理局	已完成	2023年10月27日	-	-
上海御风未来	中国民航华东地区管理局	2023年11月	-	-	-

✍️ 低空经济产业链：涵盖基础设施、飞行器制造、运营服务、飞行保障

➤ 低空经济作为战略性新兴产业，产业链条长，涵盖了低空基础设施、低空飞行器制造、低空运营服务、低空飞行保障四个核心板块；应用场景丰富，既包括传统通用航空业态，又融合了以无人机为支撑的低空生产服务方式。

◆ 低空经济全景图



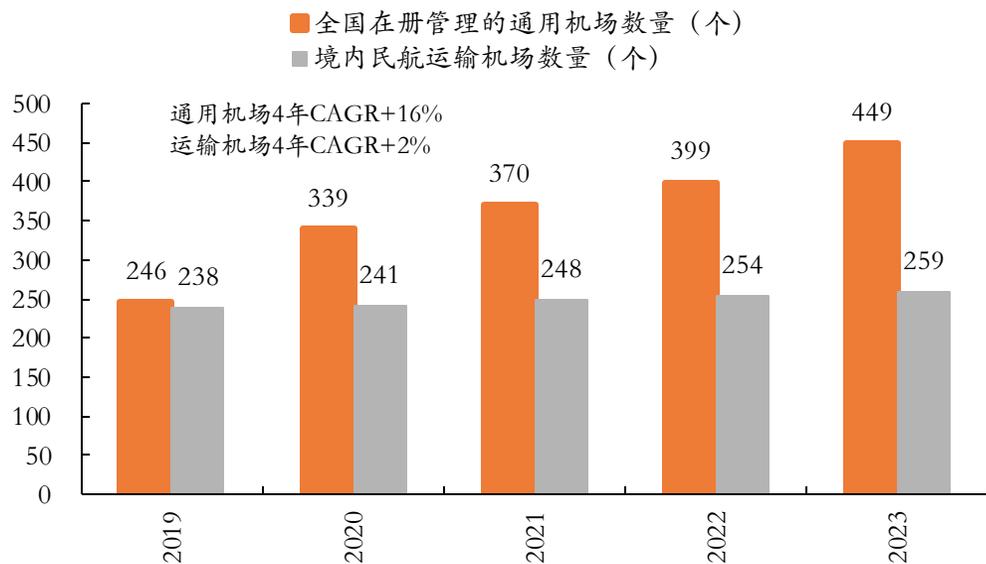
资料来源：赛迪顾问，平安证券研究所



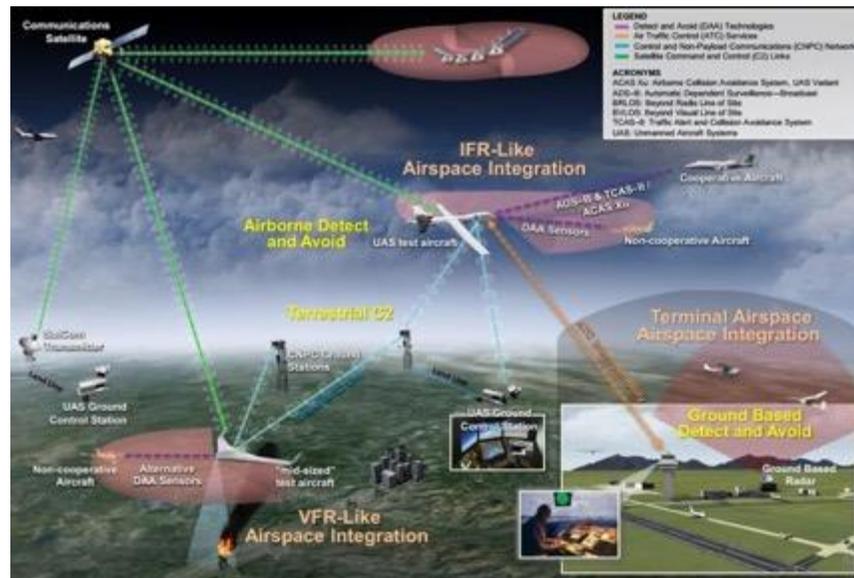
低空基础设施：重点关注机场建设运营与空管系统建设运行

- **低空基础设施**既包括与通航机场相关的设施设备，比如通用机场、直升机起降点、飞行营地、飞行服务站、维修基地、通讯、导航、气象、油库等基础设施建设和运营；也包括低空新型基础设施，比如低空空域管控系统、无人机飞行信息系统、无人机反制系统等建设与运行。根据民航局，截至2023年底，全国通用机场449个，4年复合增长率为16%（vs境内民航运输机场数4年CAGR+2%）。
- 新型基础设施方面，发展低空经济还需要解决空中交通管理、通信、导航和监视等问题，涉及一系列的软、硬件基础设施。以NASA的无人机空管系统UAS的关键技术为例，主要涉及：探测和躲避(DAA)技术涉及使用传感器(如雷达或照相机)来检测飞机是否离目标(如高楼或其他飞机)太近，并采取措施飞离潜在危险；指挥和控制(C2)技术确保无人驾驶飞机通过无线电与地面飞行员和空中交通管制保持安全联系；系统集成与运行(SIO)将通过利用集成的DAA、C2以及先进的飞行器技术，使无人机能够在NASA内稳健运行。

我国机场数量



NASA无人机空管系统 (UAS)





低空飞行器制造：eVTOL成为未来UAM市场的主流方案

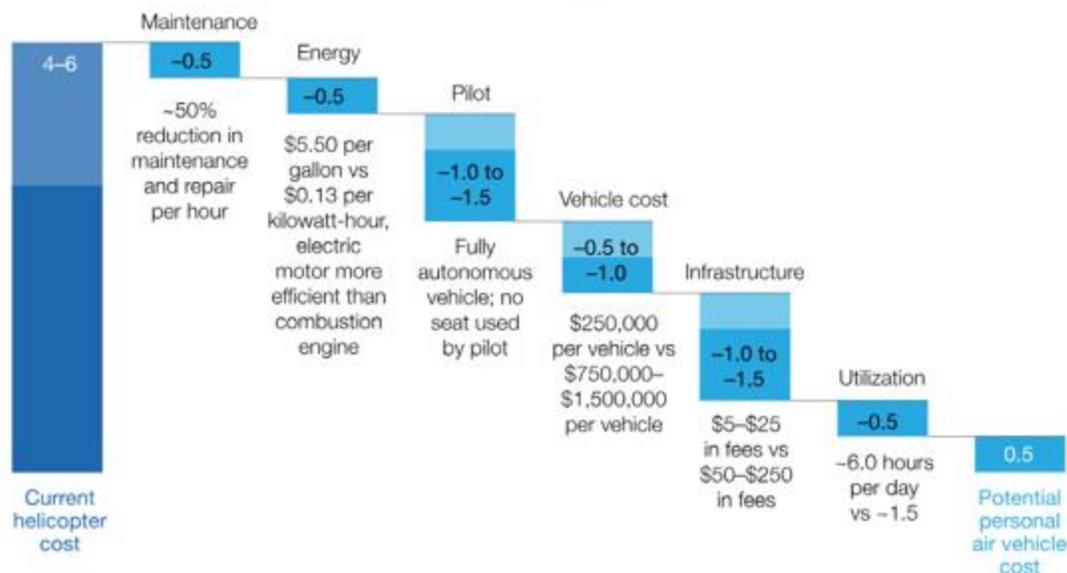
➤ 低空飞行器制造包括低空飞行器关键系统及零部件（动力系统、机载系统、飞控系统、抗干扰系统）、低空飞行器整机制造。低空飞行器中，由于eVTOL（electric Vertical Take-off and Landing，即电动垂直起降飞行器）飞行范围可达到50-200千米，巡航速度可达到50-200千米每小时，能以较高的性价比满足UAM的需求，吸引了市场最大的关注。相比其他类型飞行器，eVTOL有以下优势：1) 垂直起降，相比固定翼飞机，eVTOL不需要借助机场跑道可实现垂直起降场飞行；2) 使用电池作为能源，相比燃料动力更低碳环保，同时具有低成本优势，使用电力能源驱动低于使用传统燃料动力，运营成本大幅降低。根据麦肯锡测算，与传统直升机每可用座英里4-6美元的直接运营成本相比，eVTOL可以下降至0.5美元。3) eVTOL相比直升机噪音更低，尽可能飞到商业/社区中心，提高市内运行覆盖度。4) 电池和电机驱动技术简化机械设计结构，具备较高的安全可靠性和稳定性。

◆ eVTOL在城市和城际运行具有优势

◆ 麦肯锡测算相较燃油直升机，eVTOL可以将运营成本降低90%

机型种类	eVTOL	多旋翼	直升机	固定翼
任务范围	城市和城际	城市	郊区和城际	城际
概念图				
垂直起降	优	优	优	差
低噪音	优	优	较差	优
高安全	优	优	较差	良
能源效率	良	较差	中	优
低运营成本	良	较差	较差	较差
飞行范围(KM)	50-200	35-65	400+	1000+
巡航速度(KM/H)	50-200	50-110	200+	300+

Potential evolution for personal air vehicle, operating cost per seat-mile, \$

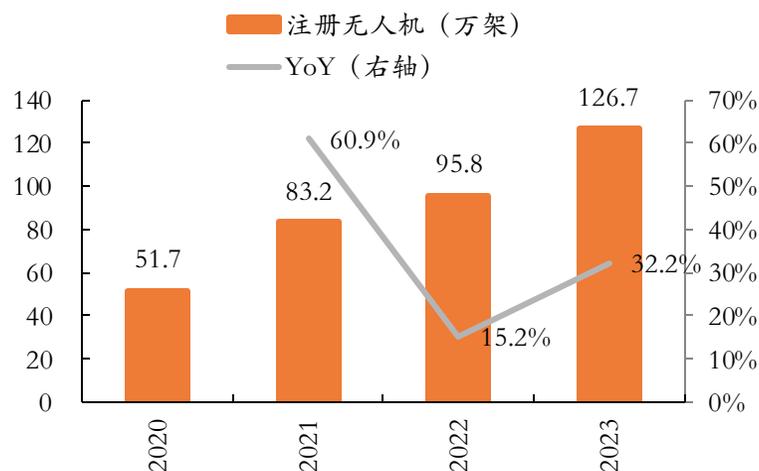


资料来源：中国民航报，McKinsey&Company，平安证券研究所

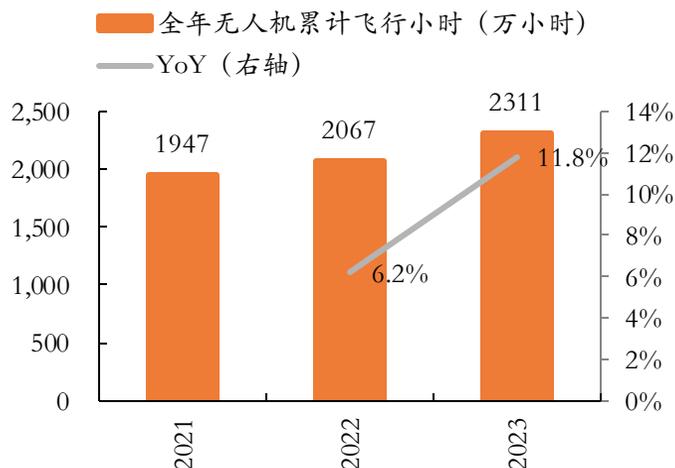
✍️ 低空运营服务：应用场景丰富，关注无人机与通用航空运营主体

- **低空运营服务**包括低空飞行各类场景（比如载人客运、载物货运、公共服务、警务安防等等）的运营，以及低空设施维护、低空飞行器维修、低空飞行培训等等。**无人机和通用航空是承载低空运营的主要产业。**
- **庞大的无人机产业是低空飞行发展的重要基础**，国内来看，无人机主要应用于个人消费、地理测绘、影视航拍等领域，在应急救援、通信中继、气象探测等应用场景也在不断被开发，同时，无人机物流已经实现了规模化应用。赛迪顾问统计2023年中国民用无人机产业规模达到1174.3亿元。根据民航局最新数据，截至2023年底，我国无人机设计制造单位大概有2000家，运营无人机的企业1.9万家；国内注册无人机126.7万架，同比增长32.2%；有统计的民用无人机全年飞行2311万小时，同比增长11.8%。
- **通用航空产业同样用途广泛、运营形态多元**，在工农业生产、航空消费等方面更具优势，在抢险救灾、医疗救援等公共服务领域能够发挥重要作用，也将整体受益于低空经济的发展。根据民航局最新数据，截至2023年底，我国通航企业（不含无人机）达到690家，在册通用航空器3303架，全年作业飞行137.1万小时，近三年年均增速达12%。

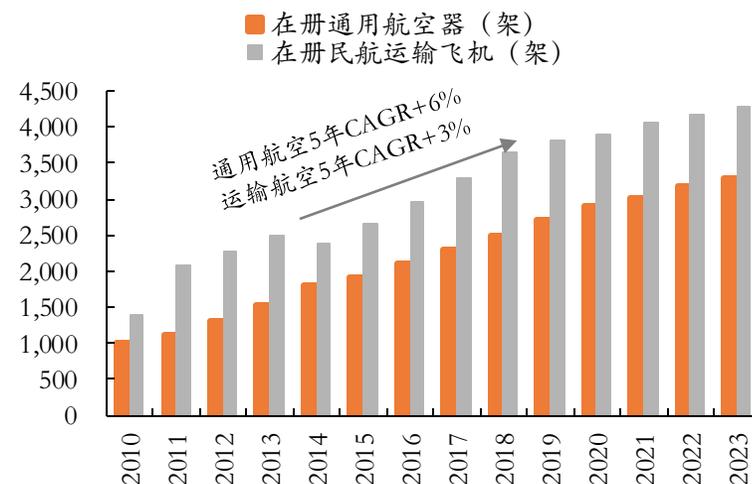
◆ 我国注册无人机数量



◆ 我国无人机累计飞行小时



◆ 我国民用飞机数量：通航&民航



资料来源：中国民航报，中国民航局，平安证券研究所

✍️ 低空飞行保障：监管体系基本成型，产业角度关注信息服务商

- **低空飞行保障**指保障低空飞行活动顺利展开的一系列监管与服务，主要包括适航审定、检测检验服务、空中保障服务、地面保障服务。
- 监管角度，按照任务分工，民航局承担低空飞行服务保障体系建设相关工作。目前，我国低空飞行服务保障体系基本成型，国家信息管理系统、区域信息处理系统与全国27个飞行服务站实现联网运行，日常通航飞行申请时限由飞行前一天15时提出、21时前批复，缩短为起飞前4小时提出、起飞前2小时批复，应急救援、抢险救灾等紧急特殊飞行任务随报随批。民用无人驾驶航空器综合管理平台上线，平台提供无人驾驶航空器实名登记、适飞空域查询、飞行活动申请等服务，实现低空飞行监管、服务全覆盖。
- 产业角度，除了与低空飞行相关的检测、加油、停机等地面服务产业外，提供低空飞行情报产品、气象产品的信息服务产业也在其中。

◆ 低空飞行服务各类系统的运行示意图





目录CONTENTS

- 行业回顾：行业行情表现不佳，估值处于历史较低水平
- 算力：算力是AIGC产业发展的底座，市场需求仍然旺盛
- 算法：国产大模型能力已可对标GPT-4，并且下游部署成本下降
- 应用：大模型应用落地有望加速，AIGC产业未来发展空间巨大
- 低空经济：政策持续催化，未来展翼腾飞
- 投资建议及风险提示

投资建议

- 展望2024年下半年，在全球范围内大模型领域的竞争依然白热化的背景下，我们认为，大模型的持续迭代升级将为AI算力带来巨大市场空间。同时，全球大模型价格呈现下降趋势，将有利于未来下游大模型部署成本的下降，将有望加速大模型在各场景的应用落地。参考我国数字经济的巨大体量，我国AIGC产业未来应用潜力巨大。我们坚定看好AIGC产业的发展。另外，政策推动低空经济进入快速发展期，在国家高度重视和地方积极布局的背景下，我们认为我国低空经济产业未来发展前景广阔。我们看好AIGC和低空经济产业发展。
- AIGC主题标的方面：1) 算力方面，推荐浪潮信息、中科曙光、紫光股份、神州数码、海光信息、龙芯中科，建议关注寒武纪、景嘉微、拓维信息、软通动力；2) 算法方面，推荐科大讯飞；3) 应用场景方面，强烈推荐中科创达、恒生电子，推荐金山办公、顶点软件，建议关注万兴科技、同花顺、彩讯股份；4) 网络安全方面，强烈推荐启明星辰。低空经济产业方面，建议关注以下产业链相关标的：1) eVTOL整机厂商：亿航智能；2) 基础设施服务商：中科星图、莱斯信息、纳睿雷达；3) 无人机厂商：纵横股份；4) 通航运营：中信海直；5) 交通规划：深城交。



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搜索、办公、传媒、医药、金融、教育、智能汽车等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) 低空经济产业发展不及预期。政策推动低空经济进入快速发展期，在国家高度重视和地方积极布局的背景下，我国低空经济进入快速发展期。但如果未来低空经济与产业融合的应用场景的落地进度低于预期，则我国低空经济产业将存在发展不及预期的风险。

附：重点公司预测与评级

股票简称	股票代码	6月7日	EPS (元)				PE (倍)				评级
		收盘价 (元)	2023A	2024E	2025E	2026E	2023A	2024E	2025E	2026E	
中科创达	300496.SZ	53.61	1.01	1.48	1.84	2.28	52.9	36.3	29.2	23.6	强烈推荐
恒生电子	600570.SH	19.26	0.75	0.91	1.11	1.32	25.6	21.1	17.4	14.6	强烈推荐
启明星辰	002439.SZ	17.64	0.61	0.79	0.99	1.25	29.0	22.3	17.8	14.1	强烈推荐
浪潮信息	000977.SZ	35.28	1.21	1.46	1.77	2.15	29.1	24.2	19.9	16.4	推荐
中科曙光	603019.SH	38.85	1.25	1.57	1.99	2.56	31.0	24.7	19.6	15.2	推荐
紫光股份	000938.SZ	22.79	0.74	0.87	1.05	1.28	31.0	26.3	21.7	17.8	推荐
神州数码	000034.SZ	28.40	1.75	2.13	2.62	3.23	16.2	13.3	10.8	8.8	推荐
海光信息	688041.SH	68.04	0.54	0.75	1.01	1.31	125.2	90.4	67.3	51.8	推荐
龙芯中科	688047.SH	95.90	-0.82	-0.26	0.11	0.39	-116.9	-362.8	874.0	248.1	推荐
科大讯飞	002230.SZ	41.51	0.28	0.31	0.37	0.45	146.1	135.9	112.6	93.1	推荐
金山办公	688111.SH	258.57	2.85	3.72	4.95	6.69	90.6	69.6	52.3	38.6	推荐
顶点软件	603383.SH	34.00	1.13	1.37	1.61	1.87	30.0	24.9	21.1	18.2	推荐
莱斯信息	688631.SH	59.44	0.81	1.03	1.32	1.70	73.4	57.7	45.0	35.0	-

注：莱斯信息2024-2026年EPS取自Wind一致预期。

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±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±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。

免责声明:

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2024版权所有。保留一切权利。