国联证券 GUOLIAN SECURITIES

证券研究报告

2024 年 03 月 31 日

电子 3 月周报（3.25—3.29）

大模型不断更新，算力条线仍为重点

投资建议： 强于大市（维持）

上次建议： 强于大市

相对大盘走势

作者

分析师：熊军

执业证书编号：S0590522040001

邮箱：xiongjun@glsc.com.cn

联系人：刘欢宇

邮箱：hyliu@glsc.com.cn

➢ **腾讯发布 BrushNet 图像修复模型：**

3 月 11 日，腾讯人工智能团队联合香港中文大学研发出了一种名为 BrushNet 的创新图像修复模型。BrushNet 是一个基于扩散模型的即插即用双分支模型，该模型的一个分支专注于提取遮罩图像的像素级特征，另一个分支负责图像生成，这种设计能够将像素级掩码图像特征嵌入任何预训练的扩散模型中，从而实现语义一致且质量增强的图像修复效果，可以实现高质量且自然连贯的重绘。

➢ **月之暗面升级 Kimi 智能助手**

3 月 18 日，月之暗面宣布，Kimi 智能助手已支持 200 万字超长无损上下文，并开启产品内测。Kimi 智能助手的智能搜索功能，可根据用户的问题，主动去互联网上搜索、分析和总结最相关的多个页面，生成更直接、更准确的答案。Kimi 诞生于 2023 年 10 月，是国内人工智能创业公司北京月之暗面科技有限公司推出的一款大模型应用。作为全球首个支持 200 万字上下文的中文大模型，Kimi 不仅提升了内容创作和整理的效率，还为小说、剧本创作等领域带来了深化和创新，同时在游戏互动、AI 陪伴和专业领域任务执行等方面开辟了新的应用场景。

➢ **AI21 发布 Jamba 突破性 AI 模型**

AI21 公司推出了名为 **Jamba** 的突破性 AI 模型，这是世界上第一个采用了 Mamba 结构状态空间模型(SSM)和 Transformer 混合架构的大规模生产级模型。Jamba 能够同时优化内存、吞吐量和性能表现。Mamba 结构由卡内基梅隆大学和普林斯顿大学的研究人员提出，主要解决 Transformer 内存占用大，随着上下文的增长推理速度变慢等问题，在 Jamba 推出之前，Mamba 用例更多停留在学术圈。

➢ **投资建议**

1）英伟达是全球 AI 龙头，其 GPU 技术有望持续引领市场，建议关注英伟达产业链快速发展带来的新机遇以及国内算力芯片在国产替代和自主可控逻辑下的渗透率提升。2）AI 需求兴起有望加速服务器平台向更大性能方向的产品替换。PCB 行业在这一发展过程中有望呈现产品价值量普遍提升的趋势。3）带宽由存储器决定，存力是限制 AI 芯片性能的瓶颈之一，建议关注存储产业链。4）高算力 AI 芯片导入有望加速服务器高功率密度演进趋势，而这对于散热效率提出了更高的要求。5）受益于全球数据量快速增长，光通信产业链建议重点关注。

风险提示：宏观经济增速不及预期，消费电子复苏和 IOT 产品出货量低于预期风险，AIGC 行业发展进程不及预期。

相关报告

1、《电子：GTC 2024 引领哪些硬件新方向?》2024.03.24

2、《电子：AI 手机有望驱动新一轮换机周期》2024.03.16

正文目录

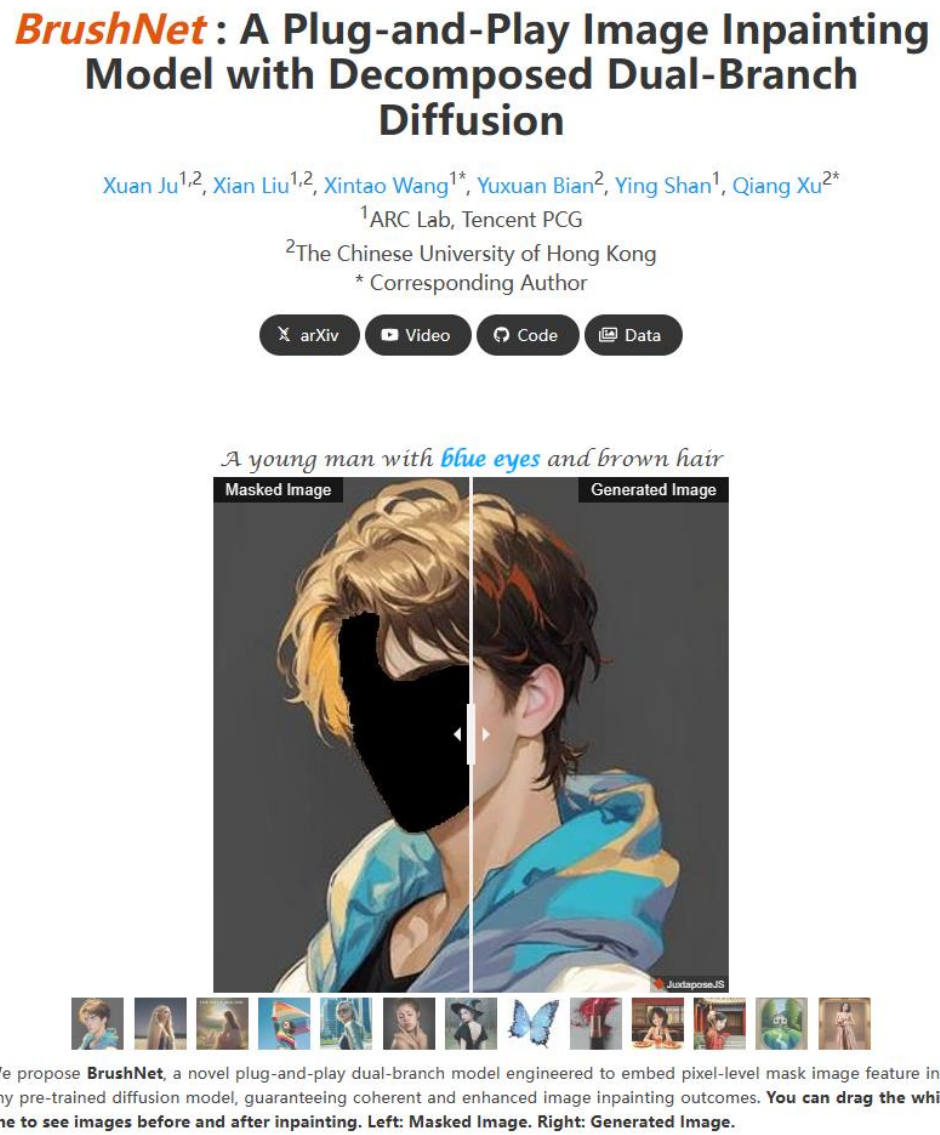
图表目录

1. 大模型不断更新

1.1 腾讯发布 BrushNet 图像修复模型

3 月 11 日，腾讯人工智能团队联合香港中文大学研发出了一种名为 BrushNet 的创新图像修复模型。BrushNet 是一个基于扩散模型的即插即用双分支模型，该模型的一个分支专注于提取遮罩图像的像素级特征，另一个分支负责图像生成，这种设计能够将像素级掩码图像特征嵌入任何预训练的扩散模型中，从而实现语义一致且质量增强的图像修复效果，可以实现高质量且自然连贯的重绘。

图表1：BrushNet 创新图像修复模型

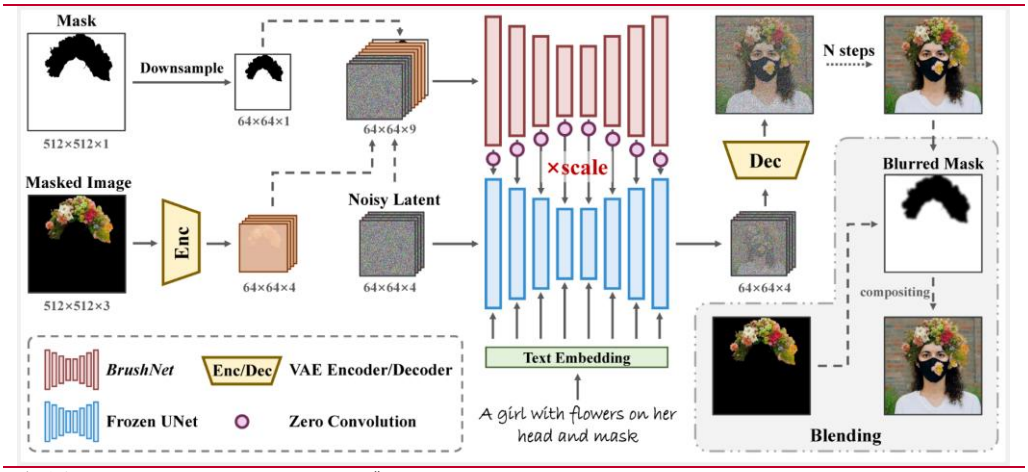


资料来源：Xuan Ju，Xian Liu，et al.《BrushNet：A Plug-and-Play Image Inpainting Model with Decomposed Dual-Branch Diffusion》，国联证券研究所

BrushNet 的核心创新点在于将掩码图像特征和噪声潜在向量分为两个独立分支进行处理。这种分支分离极大地减轻了模型的学习负担，使得模型能够以层级方式精细地融入掩码图像的关键信息。

在模型架构方面,BrushNet 首先对掩码进行下采样以匹配潜在向量大小,然后将掩码图像输入 VAE 编码器对潜在空间的分布进行校准。之后,噪声潜在向量、掩码图像潜在向量和下采样掩码被级联作为 BrushNet 的输入。从 BrushNet 提取的特征通过零卷积块层层叠加到预训练的 UNet 中。在去噪后,生成图像与掩码图像使用模糊掩码进行混合。

图表2:BrushNet 的工作原理



资料来源:Xuan Ju, Xian Liu, et al.《BrushNet:A Plug-and-Play Image Inpainting Model with Decomposed Dual-Branch Diffusion》,国联证券研究所

研究人员在多个基准测试上评估了 BrushNet,结果表明它在图像质量、掩码区域保留和文本一致性等 7 个关键指标上都显著优于现有模型。

图表3：BrushNet 性能表现优异

	Metrics	Image Quality			Masked Region Preservation			Text Align
	Models	IR$_{\times 10}$↑	HPS$_{\times 10^2}$↑	AS↑	PSNR↑	LPIPS$_{\times 10^3}$↓	MSE$_{\times 10^3}$↓	CLIP Sim↑
Inside	BLD [1]	9.78	25.87	6.17	21.33	9.76	49.26	26.15
	SDI [33]	11.72	27.06	6.50	21.52	13.87	48.39	26.17
	HDP [25]	11.68	26.90	6.42	22.61	9.95	43.50	26.37
	PP [56]	11.46	27.35	6.24	21.43	32.73	48.43	**26.48**
	CNI [51]	9.9	26.02	**6.53**	12.39	78.78	243.62	**26.47**
	CNI* [51]	11.21	26.92	6.39	**22.73**	24.58	**43.49**	26.22
	Ours	**12.36**	**27.40**	**6.53**	21.65	**9.31**	48.28	**26.48**
	Ours*	**12.64**	**27.78**	**6.51**	**31.94**	**0.80**	**18.67**	26.39
Outside	BLD [1]	7.81	26.77	6.23	15.85	35.86	21.40	26.73
	SDI [33]	10.27	27.99	**6.55**	18.04	19.87	15.13	27.21
	HDP [25]	9.66	27.79	6.46	18.03	22.99	15.22	26.96
	PP [56]	7.45	28.01	6.26	18.04	31.78	15.13	26.72
	CNI [51]	9.26	27.68	6.42	11.91	83.03	58.16	**27.29**
	CNI* [51]	9.57	27.76	6.28	17.50	37.72	19.95	26.92
	Ours	**10.82**	**28.02**	**6.64**	**18.06**	**22.86**	**15.08**	**27.33**
	Ours*	**10.88**	**28.09**	**6.64**	**27.82**	**2.25**	**4.63**	27.22

* with blending operation

Quantitative comparisons among BrushNet and other diffusion-based inpainting models in EditBench

Metrics	Image Quality			Masked Region Preservation			Text Align
Models	IR$_{\times 10}$↑	HPS$_{\times 10^2}$↑	AS↑	PSNR↑	LPIPS$_{\times 10^3}$↓	MSE$_{\times 10^3}$↓	CLIP Sim↑
BLD [1]	0.90	23.81	5.44	20.89	10.93	31.90	28.62
SDI [33]	1.86	24.24	5.69	23.25	6.94	24.30	28.00
HDP [25]	1.74	24.20	5.64	23.07	**6.70**	24.32	28.34
PP [56]	1.24	24.50	5.44	23.34	20.12	24.12	27.80
CNI [51]	1.49	24.46	**5.82**	12.71	69.42	159.71	28.16
CNI* [51]	0.90	23.79	5.46	22.61	35.93	26.14	27.74
Ours	**4.40**	**25.10**	**5.84**	**23.35**	6.81	**24.11**	**28.67**
Ours*	**4.46**	**25.24**	**5.82**	**33.66**	**0.63**	**10.12**	**28.87**

* with blending operation

资料来源：Xuan Ju, Xian Liu, et al.《BrushNet：A Plug-and-Play Image Inpainting Model with Decomposed Dual-Branch Diffusion》，国联证券研究所

与以前的图像修复方法相比(如 Blended Latent Diffusion、Stable Diffusion Inpainting、HD-Painter、PowerPaint 等)，BrushNet 的图像还原修复能力无论是在风格、内容，还是颜色和提示对齐等方面都表现出了优越的连贯性。

图表4：BrushNet 修复能力优异

资料来源：Xuan Ju, Xian Liu, et al.《BrushNet：A Plug-and-Play Image Inpainting Model with Decomposed Dual-Branch Diffusion》，国联证券研究所

1.2 月之暗面 Kimi 智能助手

2024 年 3 月 18 日，月之暗面宣布，Kimi 智能助手已支持 200 万字超长无损上下文，并开启产品内测。Kimi 智能助手的智能搜索功能，可根据用户的问题，主动

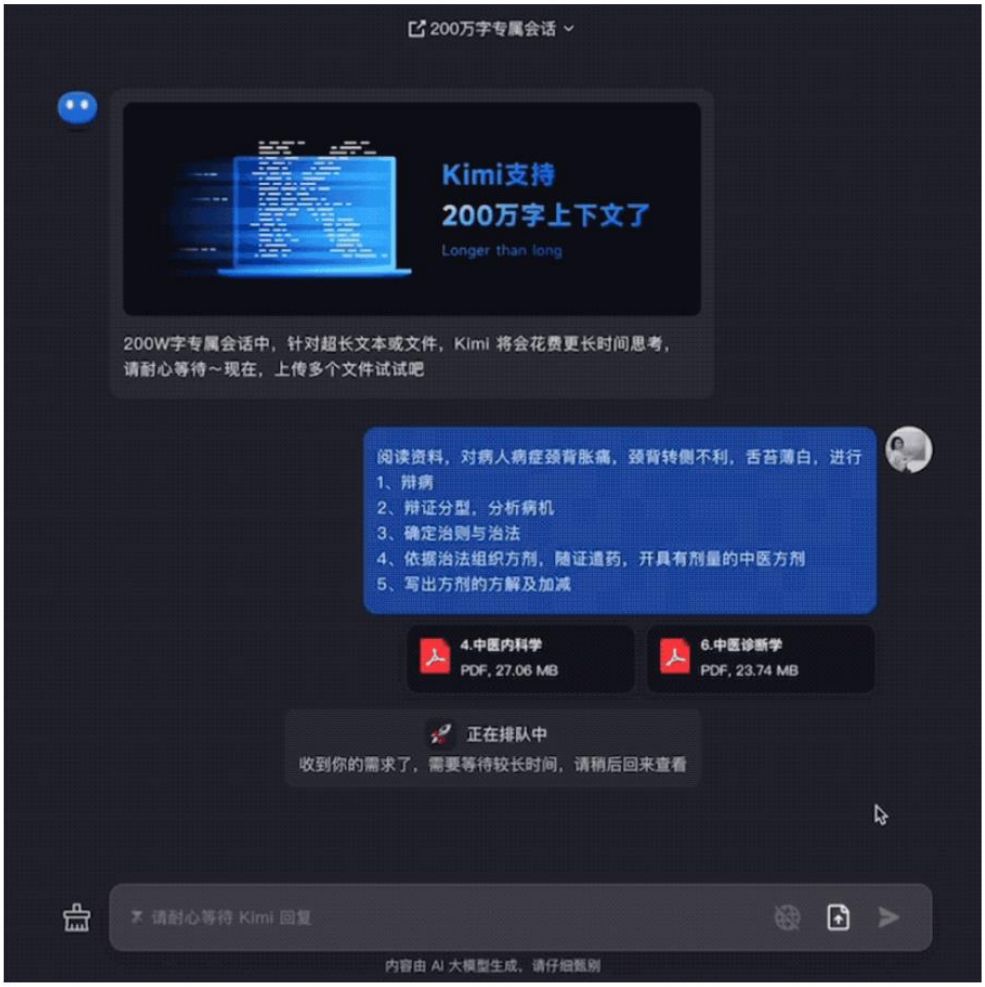
去互联网上搜索、分析和总结最相关的多个页面，生成更直接、更准确的答案。Kimi 诞生于 2023 年 10 月，是国内人工智能创业公司北京月之暗面科技有限公司推出的一款大模型应用。作为全球首个支持 200 万字上下文的中文大模型，Kimi 不仅提升了内容创作和整理的效率，还为小说、剧本创作等领域带来了深化和创新，同时在游戏互动、AI 陪伴和专业领域任务执行等方面开辟了新的应用场景。

图表5：支持 200 万字上下文的中文大模型 Kimi 已启动内测

资料来源：月之暗面公众号，国联证券研究所

过去要 10000 小时才能成为专家的领域，现在只需要 10 分钟，Kimi 就能接近任何一个新领域的初级专家水平。用户可以跟 Kimi 探讨这个领域的问题，让 Kimi 帮助自己练习专业技能，或者启发新的想法。有了支持 200 万字无损上下文的 Kimi，快速学习任何一个新领域都会变得更加轻松。例如上传一份完整的近百万字中医诊疗手册，Kimi 可以针对用户的问题给出诊疗建议；上传几十万字的经典德州扑克长篇教程后， Kimi 可以扮演德扑专家为自己提供出牌策略的指导；上传英伟达过去几年的完整财报， Kimi 可以成为英伟达财务研究专家，帮用户分析总结英伟达历史上的重要发展节点；上传一个代码仓库里的源代码，可以询问 Kimi 关于代码库的所有细节，即便是毫无注释的陈年老代码也能快速梳理清晰。

图表6：Kimi 可以针对用户的问题给出诊疗建议

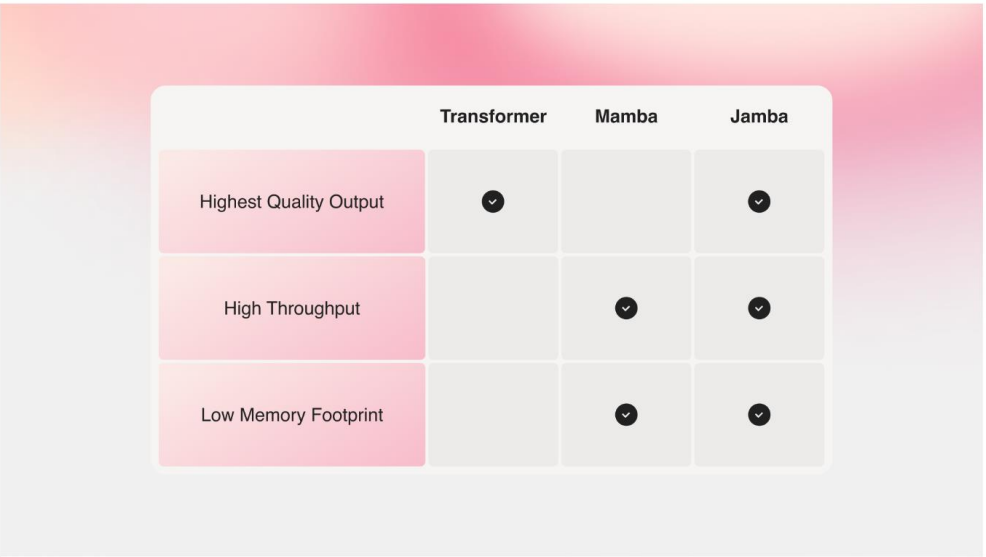


资料来源：月之暗面公众号，国联证券研究所

1.3 AI21 之 Jamba 突破性 AI 模型

AI21 公司推出了名为 **Jamba** 的突破性 AI 模型，这是世界上第一个采用了 Mamba 结构状态空间模型(SSM)和 Transformer 混合架构的大规模生产级模型。Jamba 能够同时优化内存、吞吐量和性能表现。Mamba 结构由卡内基梅隆大学和普林斯顿大学的研究人员提出，主要解决 Transformer 内存占用大，随着上下文的增长推理速度变慢等问题，在 Jamba 推出之前，Mamba 用例更多停留在学术圈。

图表7：Jamba 能够同时优化内存、吞吐量和性能表现



资料来源：AI21 官网，国联证券研究所

Jamba 的发布标志着 LLM 创新的两个重大里程碑：成功地将 Mamba 与 Transformer 架构结合在一起，并将混合 SSM-Transformer 模型提升到生产级规模和质量。生成式人工智能的起爆点是号称大模型“CPU”的 Transformer，但是 CPU 会一直升级换代，CPU 也不会只有一种，Jamba 是对 Transformer 这颗目前最热的大模型 CPU 的一个重大升级。Jamba 在各种基准测试中均优于或与同尺寸级别的其他最先进型号相媲美。

图表8：Jamba 在各种基准测试中表现优异

Common Benchmarks Comparison

	Reasoning				Aggregated assesment			
	HellaSwag	Arc Challenge	WinoGrande	PIQA	MMLU	BBH	TruthfulQA	GSM8K (CoT)
Lama2 13B	80.7%	59.4%	72.8%	80.5%	54.8%	39.4%	37.4%	34.7%
Lama2 70B	85.3%	67.3%	80.2%	82.8%	69.8%	51.2%	44.9%	55.3%
Gemma 7B	81.2%	53.2%	72.3%	81.2%	64.3%	55.1%	44.8%	44.5%
Mixtral 8x7B	86.7%	66.0%	81.2%	83.0%	70.6%	50.3%	46.8%	60.4%
Jamba	87.1%	64.4%	82.5%	83.2%	67.4%	45.4%	46.4%	59.9%

资料来源：AI21 官网，国联证券研究所

1.4 AIGC 需要大算力

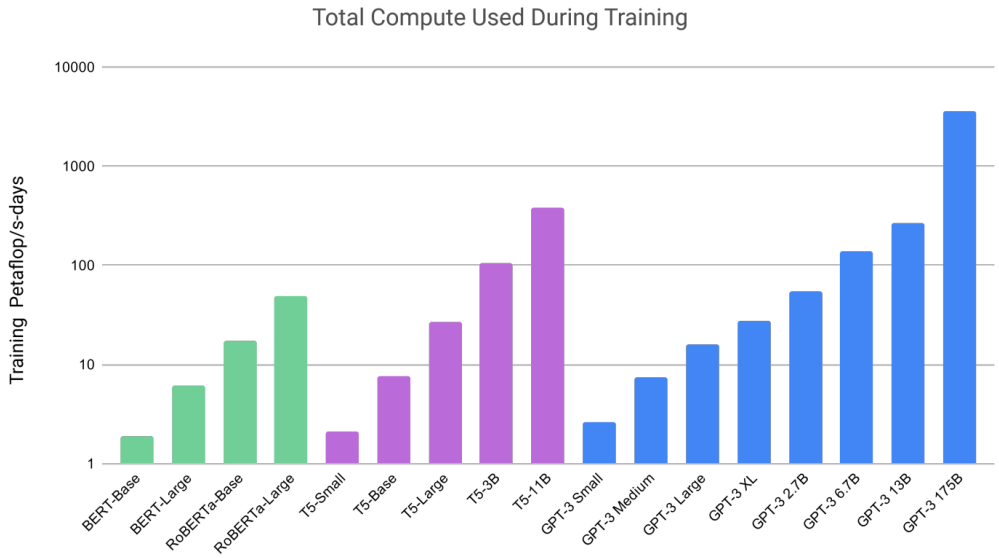
OpenAI 在 2018 年推出第一代 GPT 时，所采用的参数量为 1.17 亿个，此后 GPT 模型快速迭代，与之相对应的参数量也呈现指数增长，到 GPT3，参数量达 1750 亿，相比于初代 GPT 增长了近 1500 倍，预训练数据量更是从 5GB 提升到了 45TB。伴随着参数量和预训练数据量的提升，模型的性能实现了飞跃式提升。

图表9：Chat-GPT 模型与参数

模型	发布时间	参数量	预训练数据量
GPT	2018 年 6 月	1.17 亿	约 5GB
GPT-2	2019 年 2 月	15 亿	40GB
GPT-3	2020 年 5 月	1750 亿	45TB
GPT-4	2023 年 3 月	未公布	未公布

资料来源：中国社会科学网，国联证券研究所

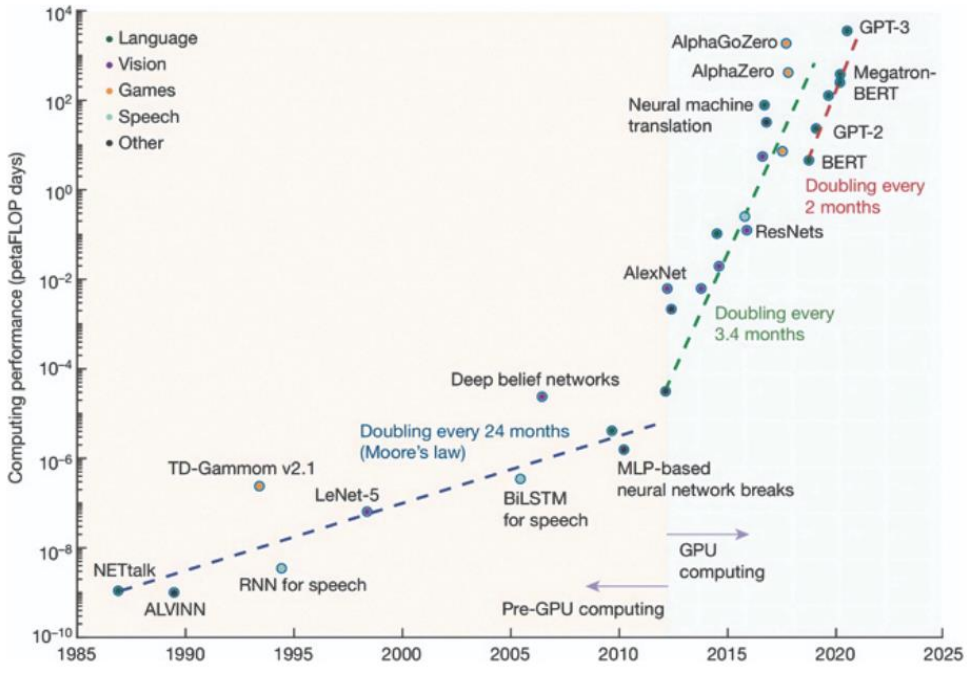
图表10：各语言模型训练算力需求对比



资料来源：Tom B. Brown, Benjamin Mann, et al.《Language Models are Few-Shot Learners》，国联证券研究所整理

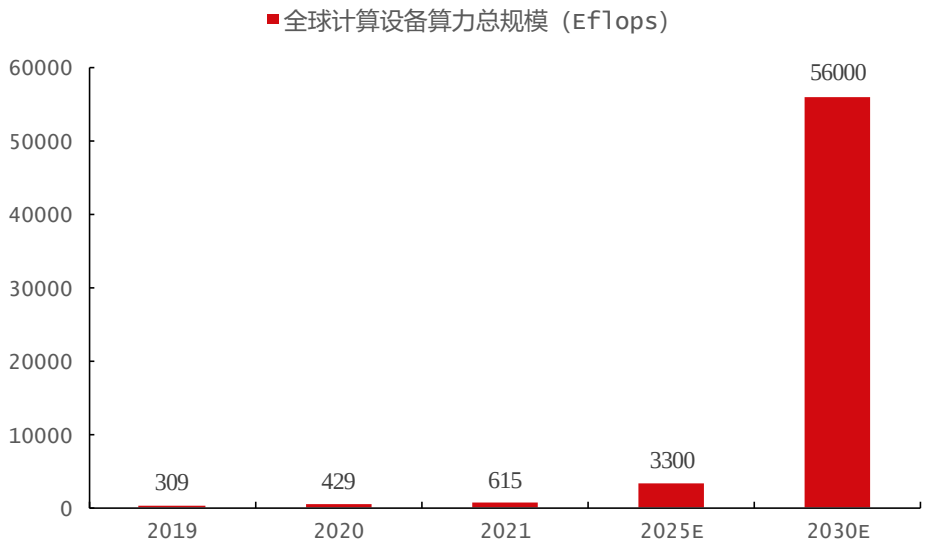
算力需求上升且增速变快。据中国信息通信研究院，2021 年全球计算设备算力总规模为 615EFlops，增速为 44%；预计到 2030 年，全球算力总规模将实现 56ZFlops，平均年增速将达到 65%。

图表11：算力需求上升且增速变快



资料来源：SHIQIANG ZHU, et al.《Intelligent Computing: The Latest Advances, Challenges, and Future》，国联证券研究所

图表12：全球计算设备算力总规模（Eflops）



资料来源：中国信息通信研究院，IDC，Gartner，国联证券研究所

2. 投资建议：算力条线仍为重点

2.1 算力芯片

英伟达是全球 AI 龙头，其 GPU 技术有望持续引领市场，建议关注英伟达产业链快速发展带来的新机遇以及国内算力芯片在国产替代和自主可控逻辑下的渗透率提

升。

2.2 PCB

AI 需求兴起有望加速服务器平台向更大性能方向的产品替换。PCB 行业在这一发展过程中将呈现产品价值量普遍提升的趋势。

2.3 存储芯片

AI 芯片需要处理大量并行数据，要求高算力和大带宽，算力越强、每秒处理数据的速度越快，而带宽越大、每秒可访问的数据越多，算力强弱主要由 AI 芯片决定，带宽由存储器决定，存力是限制 AI 芯片性能的瓶颈之一。AI 芯片需要高带宽、低能耗，同时在不占用面积的情况下可以扩展容量的存储器。

2.4 散热

高算力 AI 芯片导入有望加速服务器高功率密度演进趋势。大数据生成、业务模式变迁强调实时业务的重要性，导致高性能计算集群对于散热的要求提升。随着 ChatGPT 引爆新一轮人工智能应用的热情，海内外数据中心、云业务厂商纷纷开始推动 AI 基础设施建设，AI 服务器出货量在全部服务器中的占比逐渐提高，而这对于散热效率提出了更高的要求。

2.5 光芯片/光模块

受益于全球数据量快速增长，光通信逐渐崛起。在全球信息和数据互联快速成长的背景下，终端产生的数据量每隔几年就实现翻倍增长，当前的基础电子通讯架构渐渐无法满足海量数据的传输需求，光电信息技术逐步崛起。

3. 风险提示

- **宏观经济增速不及预期。**伴随全球半导体产业从产能不足、产能扩充到产能过剩的发展循环，半导体行业存在周期性波动。如果未来宏观经济形势发生剧烈波动，导致下游市场对各类芯片需求减少，半导体行业增长势头将逐渐放缓，行业内企业面临行业波动风险。
- **消费电子复苏和 IOT 产品出货量低于预期。**行业竞争加剧导致产品价格快速下滑的风险。全球通胀处于高位，如果售价过高或性能不达预期，可能抑制总体销量，从而导致供应链市场价格下降、行业利润缩减等状况。
- **AIGC 行业发展进程不及预期。**目前 AIGC 行业还处于起步阶段，未来发展空间以及发展速度还存在不确定性。国内公司由于起步较晚而无法与国际巨头竞争。国外领先公司如 OpenAI 已有领先行业的产品，并且形成成熟盈利模式，国内企业还处于探索阶段。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即：以报告发布日后的 6 到 12 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中:A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准；韩国市场以柯斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表指数涨幅 20%以上
		增持	相对同期相关证券市场代表指数涨幅介于 5%~20%之间
		持有	相对同期相关证券市场代表指数涨幅介于-10%~5%之间
		卖出	相对同期相关证券市场代表指数跌幅 10%以上
	行业评级	强于大市	相对同期相关证券市场代表指数涨幅 10%以上
		中性	相对同期相关证券市场代表指数涨幅介于-10%~10%之间
		弱于大市	相对同期相关证券市场代表指数跌幅 10%以上

一般声明

特别声明

版权声明

联系我们

北京：北京市东城区安定门外大街 208 号中粮置地广场 A 塔 4 楼

上海：上海浦东新区世纪大道 1198 号世纪汇一座 37 楼

无锡：江苏省无锡市金融一街 8 号国联金融大厦 12 楼

深圳：广东省深圳市福田区益田路 4068 号卓越时代广场 1 期 13 楼

电话：0510-85187583