

国联期货研究所

研究所

交易咨询业务资格编号
证监许可[2011]1773号

分析师:

林菁

从业资格证号: F03109650

投资咨询号: Z0018461

联系人: 丁家吉

从业资格证号: F03117223

相关研究报告:

《芳烃产业链专题报告:
下游行业库存周期的演绎
与展望》

《芳烃与调油系列一》

《芳烃与调油系列二——
风口上的乙苯调油》

探索算力对芳烃下游材料的需求

➤ 纯苯-苯酚的下游化工树脂是芯片基材的原料

算力性能主要由高速服务器中的GPU、CPU的主板和载板, NVSwitch, 存储芯片, 及硬盘等零部件提供。这些零件中都存在印刷电路板(PCB)的身影。PCB又由覆铜板(CCL)制作而成, 覆铜板原材料与芳烃下游的化工合成树脂息息相关。电子级的酚醛树脂和环氧树脂均属于高端产品, 它们拥有阻燃、耐热、介电性能好、具备绝缘能力、稳定性高等特性, 在半导体领域应用于覆铜板、光刻胶和塑封料(封装)等核心技术领域。

➤ AI大模型带来算力需求爆发式增长

根据MIC及Trendforce测算, 2023年全球AI服务器出货量逾125万台, 同比增长超过47%。麦肯锡2022年发布的《人工智能现状报告》指出, 机器学习的训练计算量在2016年后每10个月翻一番。毕马威发布的《普惠算力开启新计算时代》指出, 全球头部AI模型训练算力需求每3-4个月翻一番。

➤ 芳烃下游增量需求展望

经我们测算, 保守估计半导体领域(芯片、光刻胶、封装)对酚醛和环氧树脂的需求量约达5.14-8.48万吨/年和7.78-11.9万吨/年。该领域需求量相当于消耗苯酚15.53万吨(占其2023年产量约3.68%)。电子酚醛树脂供需在理论上国内还存在缺口, 特种环氧树脂我国的设计产能或可以覆盖需求。

整体来看, 半导体用树脂的需求增长对苯酚基本面的边际影响尚不重要, 未来两年相对PC(聚碳酸酯)的规模增量更小, 仍需观察后续算力需求的变化。

目录

一、从一滴油到一颗芯：	- 4 -
1.1 芳烃下游与算力产业的关系	- 4 -
1.2 酚醛树脂、环氧树脂的基本面	- 5 -
二、AI 产业链分析：	- 7 -
2.1 上游硬件端	- 7 -
2.2 中游及下游软件	- 9 -
2.3 模型发展的瓶颈	- 10 -
三、算力需求的规模测算：	- 11 -
3.1 算力的基础概念	- 11 -
3.2 训练侧算力需求测算	- 12 -
3.3 推理侧算力需求测算	- 15 -
四、原材料需求量有限但增速高	- 16 -

图表目录

图 1: 从算力服务器到上游原料的逻辑链条	- 4 -
图 2: 覆铜板原材料组成示意图	- 5 -
表 1: 英伟达 A100 和 H100 性能数据	- 5 -
图 4: 酚醛树脂产能利用率 (%)	- 6 -
图 5: 酚醛树脂行业毛利 (元/吨)	- 6 -
图 6: 环氧树脂 E-12 产能利用率 (%)	- 7 -
图 7: 环氧树脂 E-12 生产毛利 (元/吨)	- 7 -
图 8: 环氧树脂 E-51 产能利用率 (%)	- 7 -
图 9: 环氧树脂 E-51 生产毛利 (元/吨)	- 7 -
图 10: 人工智能产业链梳理	- 8 -
图 11: 英伟达 DGX A100 服务器拆分图	- 9 -
图 12: 算力训练对全球语料、视频数据储备消耗	- 11 -
表 2: 英伟达 A100 和 H100 性能数据	- 12 -
图 13: SORA 生成原理示意图	- 13 -
图 14: 千亿参数规模以上的大模型 (不完全统计)	- 14 -
表 3: 英伟达 A100 服务器零部件规格、质量	- 16 -
表 4: 英伟达 H100 服务器零部件规格、质量	- 17 -
表 5: AI 算力对芳烃下游增量需求的测算	- 18 -
表 6: 计入电子树脂需求后的苯酚平衡表	- 19 -

一、从一滴油到一颗芯：

1.1 芳烃下游与算力产业的关系

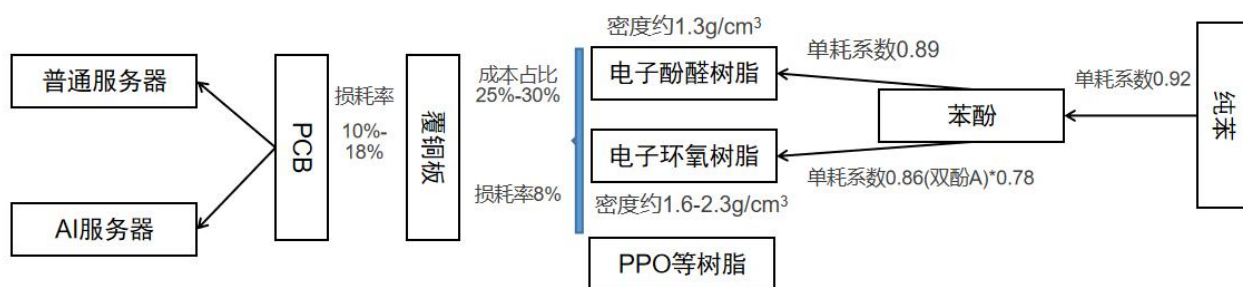
2023年初OPENAI发布首个千亿参数模型ChatGPT 3.5后，全球语言类模型和生成类模型实现爆发式增长。根据Statista的预测，2023年至2030年，全球人工智能市场规模年复合增速超过30%，2030年或超1.8万亿美元。根据彭博智库及Market.U.s等机构的预测，2023年至2032年，全球生成式AI市场规模年复合增速预期在30%-50%，2032年或达10000亿美元。

全球生成式AI遍地开花，带来算力需求翻倍的增长。算力性能主要由高速服务器中的GPU、CPU的主板和载板，NVSwitch，存储芯片，及硬盘等零部件提供。这些零件中都有印刷电路板（PCB）的身影。PCB又由覆铜板（CCL）制作而成，覆铜板原材料与芳烃下游的化工合成树脂息息相关。

本文切入算力芯片，研究当今增速最高的终端需求，对芳烃下游材料产生的消耗，能达到多少量级。

下图1从左至右展示了终端产品（服务器）到PCB覆铜板原件，再到树脂材料，最后追溯至原料苯酚和纯苯的逻辑链条。

图1：从算力服务器到上游原料的逻辑链条

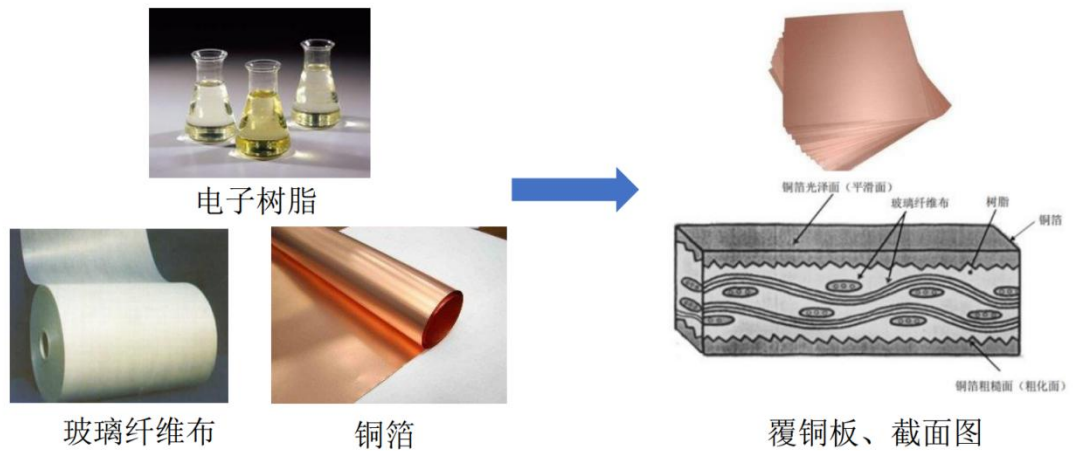


来源：国联期货研究所卓创资讯百川盈孚公开信息等

分解来看，PCB在电子设备中起到导电连接、支撑、信号传输、电路布局等重要功能。PCB由核心层（CORE）和PP片构成，PP片是多层PCB的内层粘合材料及绝缘材料。

CCL（覆铜板）是PCB的基材，即核心层，其导电性、物理强度、耐热度等品质影响PCB的性能。覆铜板的上下两面分别覆有铜箔，作导电用途；两个导电层之间一般用玻璃纤维布和固态树脂填充。一般电子树脂占覆铜板成本的比重约25%-30%。

图 2：覆铜板原材料组成示意图



来源：国联期货研究所《同宇新材招股说明书》

半导体行业通常用介质损耗因子 (dielectric loss factor) 对覆铜板的性能进行分类。Df 值越小，说明该材料在交流电场中的能耗越低，信息传输速率越高，是高频高速覆铜板的关键指标。高速覆铜板的性能、应用场景和对应的树脂材料划分如下表所示：

表 1：英伟达 A100 和 H100 性能数据

应用场景	损耗分类	性能等级	适用树脂
5G/6G 传输基站	超低损耗	Df = 0.002-0.005	聚苯醚树脂(PPO)；碳氢树脂 (Hydrocarbon chemical) 等
AI 高速服务器	低损耗	Df = 0.005-0.008	聚苯醚树脂(PPO)；碳氢树脂 (Hydrocarbon chemical)；特种环氧树脂；双马来酸酐酰亚胺树脂(BMI)；苯并噁嗪树脂 (Benzoxazine)等
普通服务器	中损耗	Df = 0.008-0.010	特种环氧树脂；双马来酸酐酰亚胺树脂(BMI)；苯乙烯马来酸酐共聚物(SMA)等
手机、电脑、平板	常规损耗	Df > 0.010	改性环氧树脂；多功能环氧树脂；线性酚醛树脂等

来源：国联期货研究所公开信息整理

算力芯片使用的主流树脂材料一般为 PPO、BMI、HC 和特种环氧树脂等，而苯并噁嗪树脂多为固结剂。

1.2 酚醛树脂、环氧树脂的基本面

电子级的酚醛树脂和环氧树脂均属于高端产品，它们拥有阻燃、耐热、介电性能好、具备绝缘能力、稳定性高等特性，在半导体领域中应用于覆铜板、光刻胶和塑封料（封装）等核心技术领域。

酚醛树脂中低端产品主要制成模塑料、耐火铸材、建筑材料等，涉及地产、钢铁冶炼等行业，占总需求的九成以上。受地产和钢铁行业景气下行影响，酚醛树脂传统需求衰减。

根据百川盈孚，2023年国内消费量仅94.63万吨，同比降约29%；行业年产能约214.04万吨，近三年变化较小；平均开工率从2020年的80%-90%，到2024年下降至50%以下。

行业前三大企业占比分别为圣泉集团（18.69%）、杭摩集团（11.68%）和宇世巨化工（9.34%），集中度较低。小规模产线较多，可统计的5万吨以下产能共占比约19.08%，未详尽统计的产能共占比约17.94%。小装置停车成本低、开工灵活，大装置停车成本较高。

2020-2023年，我国酚醛树脂净进口量约0.25万吨、0.73万吨、-2.02万吨及-3.89万吨，转为净出口国。根据海关数据，2023年酚醛树脂平均进口价较出口价高1637美元/吨，进口多为高端品（电子级）。中低端品产能过剩转向出口，国内表需逐年下降的同时，也存在缺乏高端品的结构性问题。

图4：酚醛树脂产能利用率（%）

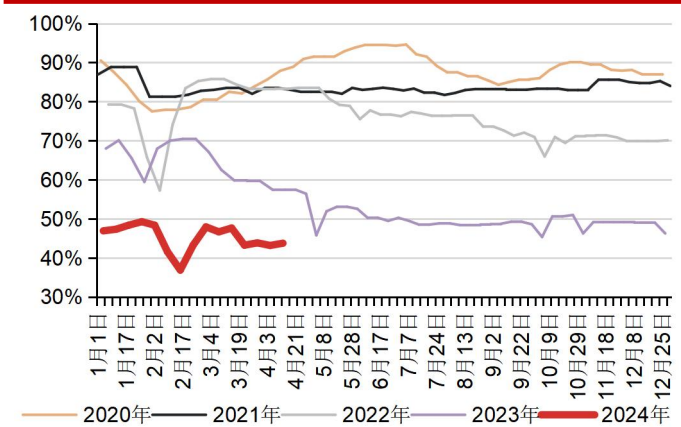
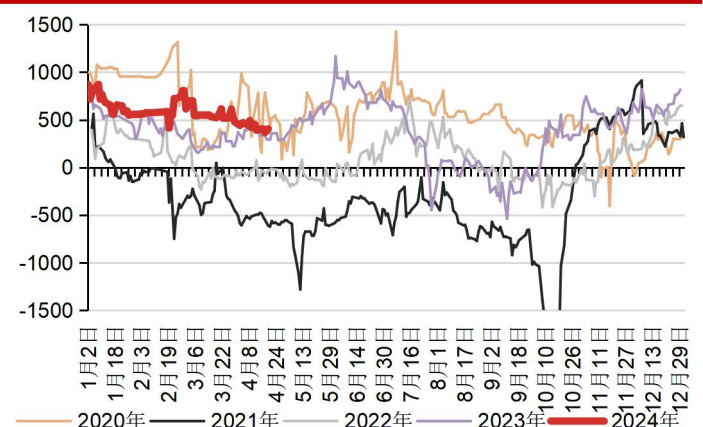


图5：酚醛树脂行业毛利（元/吨）



来源：国联期货研究所 百川盈孚

根据卓创资讯，环氧树脂传统需求分布于涂料（41%）、电子电器（28%）和复合材料（20%）等行业。2023年国内环氧树脂需求量约180万吨，同比增长约2.06%。

行业产能2023年达约327万吨，五年复合增速约9.05%；产量2023年达182.84万吨，五年复合增速约9.44%，增速较快。行业集中度较分散，前三大企业分别占比16.07%、7.66%和5.56%，10万吨及以下的产线合计占比41.5%。

随着近三年以15-20%高增速投产，环氧树脂行业毛利水平逐年回落，小产线因经济效益不佳关停的时间较长，近五年开工率在五至六成区间内波动。

进出口方面，2023年我国转为环氧树脂净出口国，近四年净进口量分别为35.76万吨、21.43万吨、9.58万吨和-1.29万吨。与酚醛树脂相似的是，我国存在缺乏高端料、中低端料过剩的结构性问题，2023年平均进口价与出口价之差达约1937美元/吨。

图6: 环氧树脂E-12产能利用率 (%)

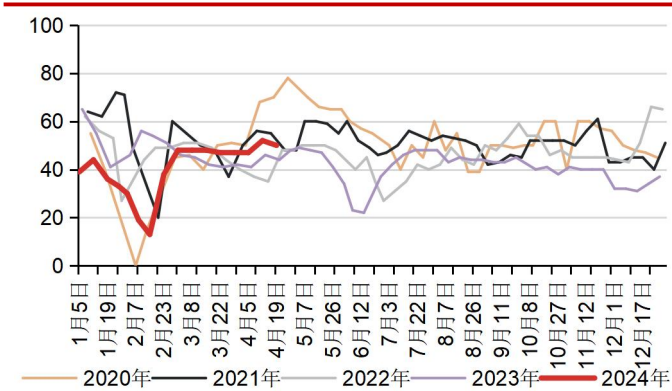
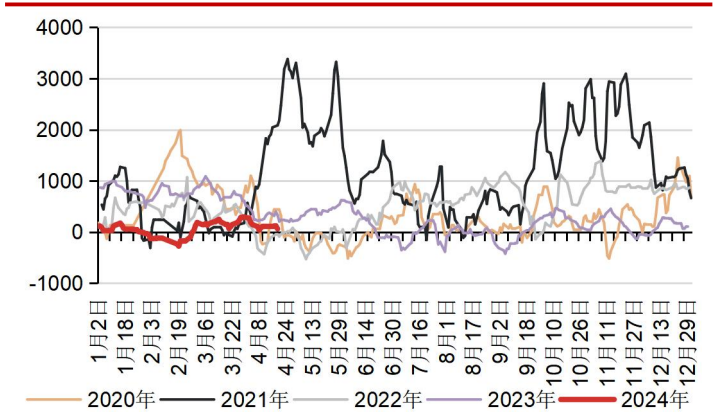


图7: 环氧树脂E-12生产毛利 (元/吨)



来源: 国联期货研究所 卓创资讯 红桃3

图8: 环氧树脂E-51产能利用率 (%)

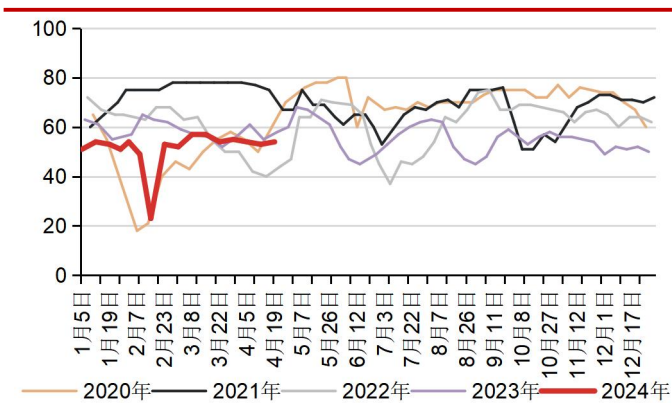
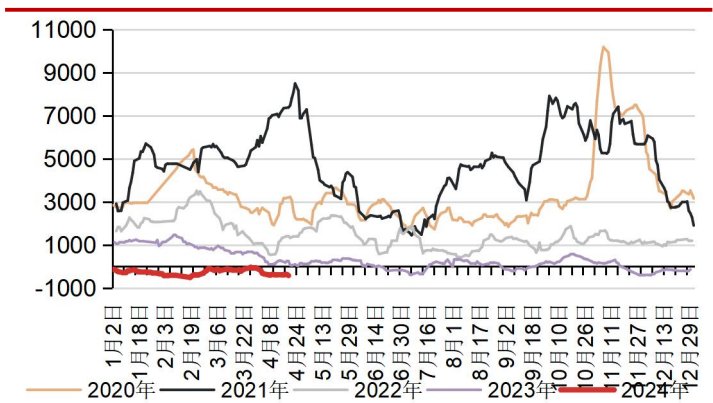


图9: 环氧树脂E-51生产毛利 (元/吨)



来源: 国联期货研究所 卓创资讯 红桃3

二、AI 产业链分析:

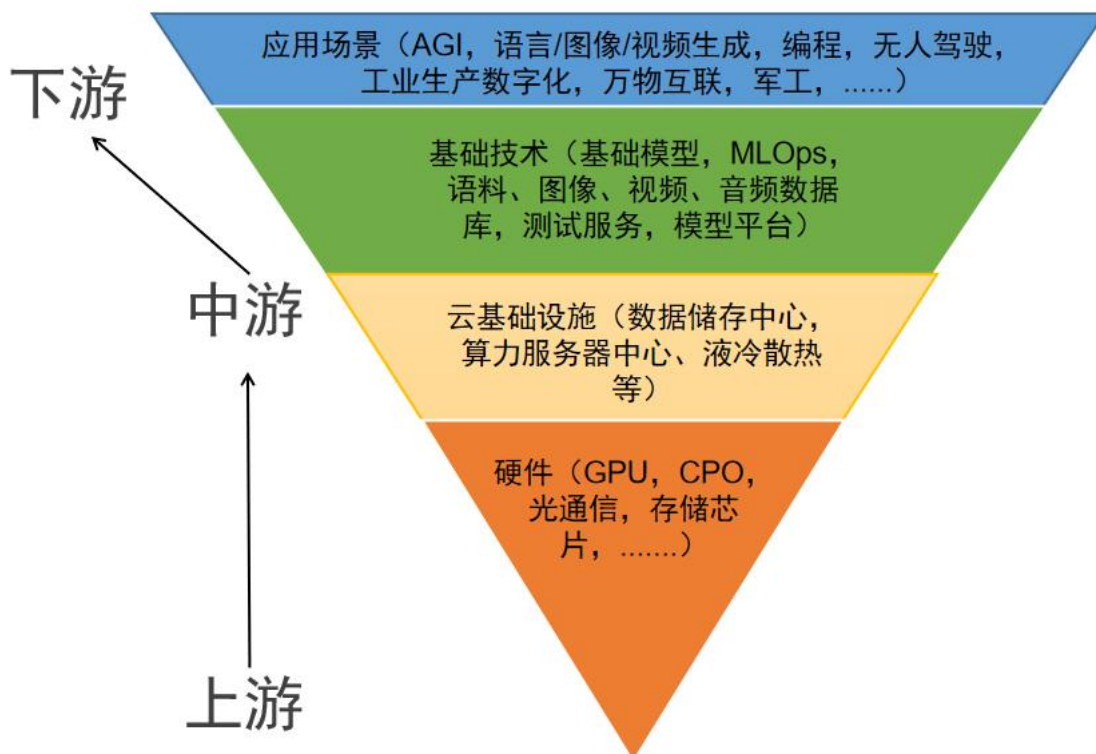
2.1 上游硬件端

国外如 META、谷歌、微软、苹果，国内如阿里、腾讯等国际互联网大厂均涌入了人工智能领域的投资。AI 应用端的创新一如智能手机时代般，深入百行百业，融入日常工作和生活娱乐之中。AI 硬件端因算力芯片的稀缺性，引发各国政府及商业资本展开了算力军备竞赛，推动芯片、服务器等基础设施的产能及性能加速升级。在新一轮全要素生产率提升的周期下，人工智能市场为全球制造业带来了新增长点，被市场誉为“第四次工业革命”。

市场对 AI 产业链一般按照六分法进行梳理，包括“电脑硬件、云平台、基础模型、模型开发及平台、应用、服务”，可以简单归纳为“硬件、云基础设施、基础技术（模

型等)和应用场景”。在当前AI行业发展的过程中,越处于底层(上游),参与竞争的企业就越少,例如垄断算力GPU的英伟达,获得了整个产业链最大部分的利润。

图 10: 人工智能产业链梳理



来源: 国联期货研究所

产业链的上游——硬件,是本文重点研究的领域。AI 半导体支持大模型的训练和运行,集成于算力服务器之中。GPU 架构拥有大量的小型、高效的核心,可以同时处理大规模多线程任务,适合应对并行计算。而 CPU 的核心数量较少,但其能处理复杂的逻辑判断任务,更适合任务调度和串行处理。大模型的训练需要通过神经网络中的深度学习方法,将数据计算任务分解为成千上万个计算集,这被称为“高度并行”。因此,算力服务器的计算工作以 GPU 为主、CPU 为辅。

综合市场调查信息,全球数据中心用 GPU 的市占率中英伟达占约 92%,其竞争对手超威半导体 (AMD) 仅占 3%,处于绝对垄断地位。我们以英伟达的 DGX A100 算力服务器为例,该产品中包含了 8 块 A100 GPU 主板,1 块 CPU 主板,6 块 NVSwitch (负责 GPU 之间通信的模块)和若干网卡、固态硬盘、存储芯片等配件构成。

图 11: 英伟达 DGX A100 服务器拆分图



来源：英伟达 DGX 系列白皮书 国联期货研究所

算力芯片自 2023 年以来一直处于供不应求状态。戴尔公司透露，GPU 的交付周期在紧张时期长达 40-52 周，2024 年随着供应产能的提升，才缩短到 8-12 周。但考虑到未来全球 AI 模型的个数和参数量是指数级增长的，目前芯片的整体性能和产能难以满足需求。同时，欧美国家限制我国进口高端算力芯片及光刻机等设备，也将刺激我国加速推进算力芯片研发制造的投入。

从规模来看，根据北京研精毕智，2022 年全球 AI 服务器行业市场出货量达到 85 万台，同比增长约 11%。根据 MIC 及 Trendforce 测算，2023 年全球 AI 服务器出货量逾 125 万台，同比增长超过 47%。另根据财报信息，英伟达在 2023 年 A100 GPU 芯片产量达 125-150 万张，H100 GPU 芯片产量达 50-70 万张，该公司在 2024 年计划将两款芯片的产量翻倍。

2.2 中游及下游软件

产业链中游的基础设施包括数据运算中心和数据存储中心，集中存放算力服务器和液冷系统。数据中心五到六成的成本开支在数据运算和存储上，四到五成开支在电力、散热、建筑、地产等方面。数据中心运营商的业务模式主要为客服直接提供算力租赁服务，或为客户提供场地以托管服务器和网络设备。

基础技术也是产业链中游的重要环节，包括 MLOps（机器学习运营）、基础模型、

训练数据服务（语料库）等。由于AI开发成本较高，普通资本很难覆盖从模型开发到终端应用的全链条，部分云平台企业提供了MLOps的服务，帮助应用端企业进行模型的设计、训练、部署和监控等内容。

语料，图像，视、音频等数据是模型训练的必需品，全球流媒体企业向各类多模态模型的开发商提供数据服务。

终端应用场景方面，人工智能不局限于聊天机器人和影视制作，在工业生产、军事活动、无人驾驶、药物研发、材料学实验、虚拟建模等高精尖领域都有AI的身影。AI能够帮助生产企业降低产品研发的沉没成本，提高工作效率，解决人力短缺的问题。

AI项目开发社区GitHub 2023年度报告显示，当年生成式AI项目数新增6.5万，同比增长约248%，推动GitHub总项目数年增速达27%。

根据AI产品榜发布的“3月AI应用全球总榜”来看，排名前百的软件中，聊天或检索机器人占约24%，文/图/音/视频生成工具占约29%，办公/生活辅助占约14%，市场/金融分析占约5%。面向C端的AI应用仍在全球市场中使用频率和时长最高。

2.3 模型发展的瓶颈

AI模型在追求参数量更大、更智能的同时，也面临着理论发展瓶颈。

训练数据是决定模型性能的主要因素，《Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning》论文中估计，全球高质量语料数据或在2026年被消耗殆尽，低质量语料数据或在2030-2040年间被消耗殆尽，视觉数据或到2050-2060年被消耗殆尽。高质量数据荒的时间节点很近，或对提升人工智能造成困难。

此外，DeepMind在《Training Compute-Optimal Large Language Models》论文中论证了语料数据的规模和模型效果（最佳损失）呈现的是凸函数关系。凸函数意味着模型的分差和语料数据之间存在一个最优点——二者量级均不大。即语料数据规模无穷大，**损失函数¹**也不会再被优化，模型效果反而变差。

因此，即使未来数据被极大丰富，大模型训练也不一定能“数”尽其用。数据撞墙问题或制约人工智能的发展速度。

¹机器学习中的损失函数，用来衡量模型的预测值与真实值的差异。训练模型一般通过函数取极小值来做参数估计，以找到最优的语料库和模型方差的比值，从而能够更精准地学习到数据的规律和特征。

图 12: 算力训练对全球语料、视频数据储备消耗

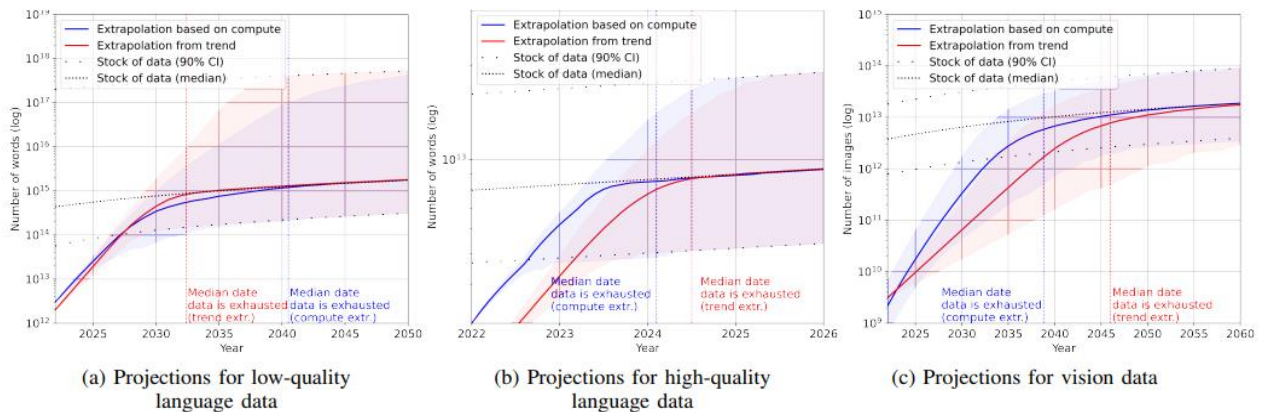


Fig. 1: Projections of data usage. Each graph shows two extrapolations of data usage, one from past trends and one from compute availability estimations plus scaling laws. Both projections are constrained to be lower than the estimated data stock. In all three cases, this constraint causes a slowdown in data usage growth.

来源: Pablo Villalobos et al., "Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning", 2022

三、算力需求的规模测算:

3.1 算力的基础概念

本文的研究重心在于探讨 AI 行业的算力规模提升, 对芳烃下游的增量需求。因此, 我们将针对“硬件端”的量级预期进行分析。

我们在第一章中分析到 AI 行业最靠近上游的进入壁垒越高, 高参数量的模型对芯片质量和运行架构的要求极高, 市场预期一到两年内, 很难出现英伟达的挑战对手。该细分行业是求大于供的市场, 产能领先于需求投放的难度较大。我们将测算需求规模增速的预期, 作为供给增速的指引。

前文提到, 提供 AI 模型的训练、推理能力的最终产品为算力服务器, 目前市场主流的英伟达 A100 和 H100 服务器均搭载了 8 块 GPU 芯片。单位 GPU 对应的算力性能一般用 FLOPS (每秒执行浮点运算的次数) 来衡量, 一单位 TFLOPS (Tera Floating Point Operations Per Second) 相当于每秒执行一万亿次的浮点运算。

英伟达 A100 和 H100 的重要性能数据如下表所示:

表 2：英伟达 A100 和 H100 性能数据

型号	H100 80GB SXM5	A100 80GB SXM4
GPU 架构	Hopper	Ampere
GPU 核心版本	GH100	GA100
单精度浮点核心(CUDA Core)	16896	6912
显存容量	80GB HBM3	80GB HBM2e
显存带宽	3.35TB/S	2039 GB/s
NVLink	NVLink 4.0 NVSwitch 900GB/s	NVLink 3.0 NVSwitch 600 GB/s
性能 (FP16 Tensor Core)	1979 TFLOPS	312 TFLOPS
最大功耗	700W	400W

来源：国联期货研究所 英伟达官网

由于多 GPU 通信速度不足、并行效率瓶颈、内存带宽限制及模型设计等问题，会导致在实际模型训练中，GPU 的算力利用率下降。根据英伟达统计的历史数据，目前 A100 和 H100 的算力利用率约在 40%-60% 的区间内。

本文在后续的算力测算中，假设 A100@FP16 和 H100@FP16 的算力利用率为 50%，对应算力性能为 156 TFLOPS 和 990 TFLOPS。

接下来，我们将分析大模型从训练。到运营后产生的推理，所需的算力规模。

3.2 训练侧算力需求测算

英伟达在《Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM》论文中列出了模型训练所需时间的公式：

$$\text{Training time} \approx \frac{8 \times T \times P}{n \times X}$$

其中“P”代表模型的总参数量，英伟达在案例里代入了 OpenAI 的“GPT-3”的数据，约 1750 亿参数量。“T”代表实际用于训练的数据量，GPT-3 约使用了 3000 万个 tokens。

“n”代表训练所需的 GPU 个数，该案例假设使用 1024 块。“X”代表单个 GPU 的算力，该案例假设使用 A100@FP16 芯片（40% 算力利用率）约 140 TFLOPS。经测算，1024 块 A100 芯片训练参数量为 175 billion 的模型，大约需要 34 天。

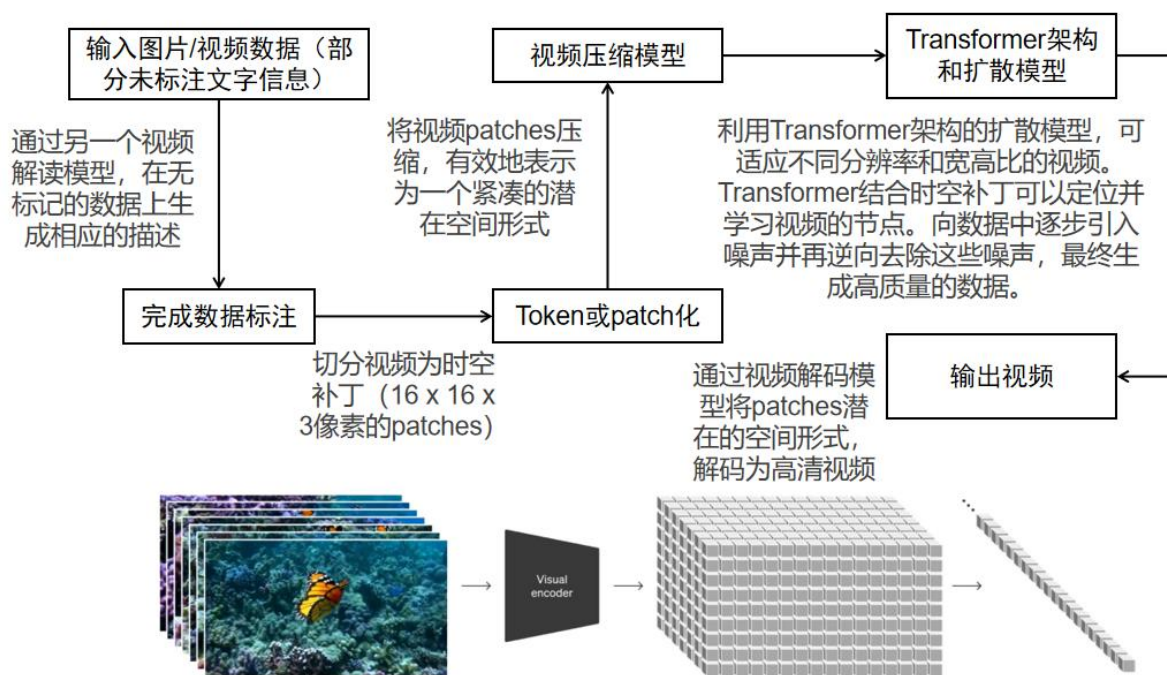
该公式显示，参数量相近且使用同一款芯片的模型，训练的要求时间越长，所需的芯片数量越少。

举一反三，与 OpenAI “GPT-4” 规模相近（参数量约 1800 billion、训练集约 5000 billion tokens）的模型，假设训练的时间要求为一个月，经测算得到需要使用约 28058 块 H100 芯片。以八卡为一台的标准，即需要 3507.25 万台的 H100 服务器。以 H100 和 A100

的算力性能之比换算，约需要2.17万台 A100 服务器。

语言类生成模型的算力需求，相对音视频生成模型而言，是小巫见大巫。2024年2月16日，OpenAI 发布文生视频模型“SORA”。通过输入一段文字，SORA 可以生成长达60秒的动画或真人视频。其原理基于扩散模型（Stable Diffusion）、Transformer 架构和时空补丁（Spacetime Patches）等技术，示意图如下：

图 13: SORA 生成原理示意图



来源：国联期货研究所 SORA 官网

由于 OpenAI 未公开 SORA 模型的参数值和训练集的数据量，市场的测算方式各不相同。综合多家券商机构的测算结果，假设使用 H100 芯片且训练要求时间为 3 个月，训练一个文生视频模型所需的算力最大估计数为 52.32 万张，中位估计数为 12.74 万张，最小估计数为 0.66 万张。考虑到可比性，假设训练时间缩短至一个月，那么训练 SORA 所需的算力约为大参数语言类模型的 15-60 倍。





















在预测 AI 模型训练侧算力需求规模之前，我们需要回顾当下的大模型供给格局。

中、美走在大模型开发领域的前沿，两国 AI 模型总数约占全球九成。根据《中国人工智能大模型地图研究报告》，截至 2023 年 5 月，国内 10 亿参数以上的大模型数量约 79 个，同期美国 10 亿参数规模以上的大模型已超 100 个。根据《经济学人》相关报道，美国 2022 年大模型投资额达 474 亿美元，约为中国的 3.5 倍，且 2023 年仍保持高速增长。

根据《北京市人工智能行业大模型创新应用白皮书（2023年）》，截至2023年10月，国内10亿参数规模以上的大模型数量约254家，较2022年末激增200余家，预计2024年一季度国内或超过300家。

另据DataLearner不完全统计，截至2024年3月，全球千亿参数规模以上的大模型至少约30家以上。

图 14：千亿参数规模以上的大模型（不完全统计）

	WizardLM-2 8x22B 基础大模型	1760 亿参数		Mixtral-8x22B- MoE 基础大模型	1760 亿参数		C4AI Command R+ 基础大模型	1040 亿参数		DBRX Base 基础大模型	1320 亿参数
	Grok-1 基础大模型	3140 亿参数		DeepSeekMoE 145B Chat 聊天大模型	1446 亿参数		PaLM 2 基础大模型	3400 亿参数		StableLM 基础大模型	1750 亿参数
	GPT-4 基础大模型	18000 亿参数		GPT-3.5 聊天大模型	1750 亿参数		HunYuan-NLP 1T 基础大模型	10000 亿参数		OPT 基础大模型	1750 亿参数
	Galactica 基础大模型	1200 亿参数		PaLM-Coder 编程大模型	5400 亿参数		GLM-130B 基础大模型	1300 亿参数		GLaM 基础大模型	12000 亿参数
	Gopher 基础大模型	2800 亿参数		PanGu-Alpha 基础大模型	2000 亿参数		BLOOM 基础大模型	1760 亿参数		MT-NLG 基础大模型	5400 亿参数

来源：国联期货研究所 Datalearner

综合来看，截至2024年3月，全球大规模级文生视频模型约在5家以内（SORA、PIKA、Runway GEN-2、Pixverse等），万亿参数规模以上的模型（参考GPT-4）大约在10家以内（已知有谷歌、OpenAI、META、英特尔、悟道、华为、阿里、腾讯等），千亿参数规模以上的模型（参考GPT-3）大约在30-50家左右，十亿参数规模以上的模型（假设中、美共占九成）大约在500-600家左右。

下文的规模测算，我们做了以下假设：（1）国内受限制只能使用性能接近A100的算力芯片；（2）大模型的增量规模基于上一段落的存量估计；（3）10-1000亿参数量的模型假设均使用A100的芯片。

1.保守预期2024年全球新增3家与SORA等规模的视频模型，平均训练时间为1个月，那么将带来约9.5万台H100服务器的海外增量需求（取市场估计的中位数）和29.7万台A100服务器的国内增量需求。

2.保守预期2024年全球新增约10家（海外6家、国内4家）万亿参数的大模型。假设平均训练时间为1个月，那么将带来约2.1万台H100服务器海外增量需求和8.68万台A100服务器的国内增量需求。

3.保守预期2024年全球新增约100家千亿参数的大模型（以GPT-3为标准）。假设平

均训练时间为1个月，那么将带来约1.28万台A100服务器的增量需求。

4.保守预期2024年全球新增约500家十亿参数以上的大模型（平均500亿参数量，1000亿训练集）。假设平均训练时间为1个月，那么将带来约1万台A100服务器的增量需求。

5.如果2025年OpenAI发布GPT-5（预估参数量大约为10万亿参数量，训练集约20万亿），假设训练时间为90天，那么需要约25980台H100服务器。

3.3 推理侧算力需求测算

除了训练大模型需要算力以外，在模型完成训练后，用户的访问将使大模型生成大量的tokens，这也需要服务器运算的支持。

生成文字或图片对应的推理运算消耗不同。OpenAI官方数据显示，单位token的推理成本约为 $2N$ FLOPS，其中“N”是模型的参数量。1个英文单词约等于1.33个token，1个汉字约等于2个token。

我们以GPT-3模型（参数量1750亿）为例，测算其单日所需的算力。根据similarweb的数据，CHATGPT于2024年2月日访问量约0.58亿次。假设单位用户每次对话约1000个汉字，则单日合计产生约1160亿个token，换算得到的算力约 4.06×10^{22} FLOPS。这相当于需要性能为156 TFLOPS的A100芯片（单日算力约 1.35×10^{19} FLOPS）约3007块。考虑到峰值需求（80%的访问量集中在20%的访问时间里），则需要储备约12029块A100，即1503台服务器。

另根据OpenAI官方收费标准，生成图片每包含1000个token收费0.004美元，512*512像素的图片收费0.018美元，这意味着单张512*512图片的token约4500个。1024*1024像素的图片收费0.02美元，即单张1024*1024图片的token约5000个。

我们以GPT-4模型（参数量1.8万亿）为例，测算其单日生成图片所需算力。OpenAI在2023年世界人工智能大会上公布GPT-4日活跃人数达0.14亿次。假设每人每天生成一张图片，单日约700亿个tokens，即 25.2×10^{22} FLOPS。这相当于需要18667块A100芯片，考虑峰值需求约74667块A100芯片，即9333台服务器。

综合上述的信息和公式，我们沿用“2.2”章节的假设条件，测算大模型推理测的算力需求增量。

1.保守预期2024年全球新增约10家（海外6家、国内4家）万亿参数的大模型，假

设日均访问量1000万次（该估计偏保守，以苹果SIRI接入AI模型为例，SIRI全球活跃用户数约5亿，每月最少用一次，相当于0.17亿的模型日调用量），假设每人每天分别生成一张1024像素图片和1000个汉字且考虑峰值影响。经测算将带来约0.89万台H100服务器的海外需求和3.74万台A100服务器的国内需求。

2.保守预期2024年全球新增约100家千亿参数的大模型（以GPT-3为标准），假设日均访问量达到1000万次（百度单日最高被调用曾高达5000万次），每次1000个汉字量及一张1024像素图片且考虑峰值影响。经测算将带来约9.07万台A100服务器需求。

3.保守预期2024年新增约500家十亿参数以上的大模型（平均500亿参数量，1000亿训练集）。参考similarweb发布的2024年2月国内AI产品访问量数据，假设日均访问量达到100万次，每次1000个汉字量及一张1024像素图片且考虑峰值影响。经测算将带来约1.29万台A100服务器需求。

综上所述，保守估计2025年全球音/视频及语言类模型的增量，较2024年的增量再增长一倍。那么2024-2025年AI训练侧及推理侧的算力增量需求，分别约67.25万台和137.09万台服务器。

四、原材料需求量有限但增速高

由于国内对芯片材料单耗的研究文献较少，我们引用了国金证券研究所报告《AI服务器中到底需要多少PCB》中，对英伟达服务器零部件的尺寸规格测算结果。并结合市场上通用PCB产品的厚度，整理了不同零部件的三维尺寸数据。

表3：英伟达A100服务器零部件规格、质量

组成部分	整机个数	PCB厚度 ² (mm)	整机PCB重量(g/台)	所需材料
GPU载板	8	2.4	189.28	电子酚醛
NVSwitch	6	2.5	227.50	电子环氧
GPU加速卡(OAM)	8	3	1820.02	电子环氧
GPU模组板(UBB)	1	3.9	2957.53	电子环氧
CPU载板	2	2.4	47.32	电子酚醛
CPU主板	1	2.5	2401.42	电子环氧
DIMM系统内存	32	3	630.94	电子酚醛

² 该列数据基于我们对公开数据的整理，部分零部件通过对PCB的PP片层数测算厚度，部分零部件参考通用产品的厚度。

网卡	10	3.9	768.96	电子酚醛
拓展卡	1	3.9	64.08	电子酚醛
存储操作系统驱动板	1	3.9	76.90	电子酚醛
电源	6	1.5	280.97	电子酚醛
硬盘	8	1.5	157.74	电子酚醛
前控制台板	1	1.5	24.65	电子酚醛
合计			9647.3	

来源：国联期货研究所 公开信息整理

表 4：英伟达 H100 服务器零部件规格、质量

组成部分	整机个数	PCB 厚度(mm)	整机 PCB 重量 (g/台)	所需材料
GPU 载板	8	2.4	189.28	电子酚醛
NVSwitch	4	2.5	151.67	电子环氧
GPU 加速卡 (OAM)	8	3	1820.02	电子环氧
GPU 模组板 (UBB)	1	3.9	2957.53	电子环氧
CPU 载板	2	2.4	47.32	电子酚醛
CPU 主板	1	2.4	2305.36	电子环氧
DIMM 系统内存	32	2.4	504.75	电子酚醛
网卡	6	2.4	283.92	电子酚醛
拓展卡	2	2.4	94.64	电子酚醛
存储操作系统驱动板	1	2.4	47.32	电子酚醛
电源	6	1.5	280.97	电子酚醛
硬盘	8	1.5	157.74	电子酚醛
前控制台板	1	1.5	24.65	电子酚醛
合计			8865.17	

来源：国联期货研究所 公开信息整理

上表 3 和 4 中的“整机 PCB 重量”，参考国金证券报告《AI 服务器中到底需要多少 PCB》中测算的英伟达服务器零件面积，乘以我们整理的零件厚度，测算得到三维体积。

AI 服务器中的 GPU 的加速卡 (OAM)、模组版 (UBB)、NVSwitch 和 CPU 主板，这些核心部件因承担高速信息传输、数据运算等核心工作，对介质损耗具有低乃至超低损耗的要求。我们假设上述四大芯片均采用特种环氧树脂，服务器的其他零部件均采用与电脑、手机等电子器件要求相近的电子级酚醛树脂。

根据经验值，使用覆铜板制作 PCB 折损率约 10%-18%，使用树脂材料制作覆铜板的损耗率大概在 8% 左右，酚醛树脂的密度大约为 1.3 g/cm³，环氧树脂的密度大约为 1.6-2.3 g/cm³（取 2 g/cm³）。通过扣除折损率以及原材料密度的转换，我们分别测算得到单台 A100 服务器和 H100 服务器对应的原材料重量。

制作覆铜板所需要的电子级酚醛树脂和环氧树脂，均为苯酚的下游材料。根据卓创资讯，0.855 吨的苯酚和 0.279 吨的丙酮通过离子交换树脂法，可以制成 1 吨的双酚 A。0.78 万吨的双酚 A 和环氧氯丙烷等材料可以制成 1 吨的环氧树脂。根据百川盈孚，0.89 万吨的苯酚和甲醛等材料可以制成 1 吨的酚醛树脂。

麦肯锡 2022 年发布的《人工智能现状报告》指出，机器学习的训练计算量在 2016 年后每 10 个月翻一番。毕马威发布的《普惠算力开启新计算时代》指出，全球头部 AI 模型训练算力需求每 3-4 个月翻一番。根据机构的统计结果，我们在需求测算中将乐观情景，保守假设为算力需求每 12 个月翻一番。中性情景使用了本文第二章中训练侧和推理侧的规模测算。在两种情景下分别预测未来两年 AI 算力对电子树脂、苯酚和纯苯的需求增量。

表 5: AI 算力对芳烃下游增量需求的测算

乐观情景（算力每隔 12 个月翻一番）	2023 年预测值	2024 年预测值	2025 年预测值
AI 服务器出货量（万台）	125	250	500
电子环氧树脂消耗量（吨）	9258.09	18516.18	37032.36
电子酚醛树脂消耗量（吨）	2801.03	5602.06	11204.12
双酚 A 增量需求（万吨）	0.72	1.44	2.89
苯酚增量需求（万吨）	0.87	1.74	3.48
纯苯增量需求（万吨）	0.80	1.60	3.20
算力 PCB 占酚醛树脂需求比重（%）	0.31%	0.62%	1.23%
算力 PCB 占环氧树脂需求比重（%）	0.51%	1.03%	2.06%
中性情景（依下游 AI 模型需求推理）			
	2023E	2024E	2025E
AI 服务器出货量（万台）	125	192.25	329.35
电子环氧树脂消耗量（吨）	9258.09	14218.20	24325.03
电子酚醛树脂消耗量（吨）	2801.03	4231.93	7135.67
双酚 A 增量需求（万吨）	0.72	1.11	1.90
苯酚增量需求（万吨）	0.87	1.33	2.27
纯苯增量需求（万吨）	0.80	1.22	2.09
算力 PCB 占酚醛树脂需求比重（%）	0.31%	0.47%	0.79%
算力 PCB 占环氧树脂需求比重（%）	0.51%	0.79%	1.35%
需求增速	47.06%	52.86%	70.38%

来源：国联期货研究所百川盈孚卓创资讯

中性情景的 AI 服务器出货量，参考本文 2.2 和 2.3 小节的训练侧和推理侧算力规模的预测。上表结果显示，到 2025 年由于 AI 算力产生的环氧树脂需求量占 2023 年国内产量约 1.35%-2.06%，酚醛树脂需求量占 2023 年国内产量约 0.79%-1.23%。

根据华强电子网，2009 年我国覆铜板制造业对酚醛树脂的需求量约 1 万吨。另据我国覆铜板协会统计，2009-2022 年我国覆铜板产量约增长 2.58 倍，粗略估计 2022 年产量达 2.58 万吨。由此可见，虽然算力难以成为酚醛树脂主要需求，但相对原覆铜板需求的低基

数，算力增量的拉动较为明显（增速超27%以上）。追溯到上游来看，算力芯片拉动的苯酚总需求（2023年表需）增速约0.75%左右。

根据公开信息，我国目前在产及在建的电子级酚醛树脂产能有：圣泉集团（2万吨）、东材科技（1.5万吨）、彤程新材（0.5万吨）和同宇新材（1.4万吨）等。我国目前在产及在建的特种环氧树脂产能有：圣泉集团（3万吨）、东材科技（1.5万吨）、同宇新材（拟投5.7万吨）、宏昌电子（拟投8万吨）和德源环氧电子科技（5万吨）等。

2022年我国环氧塑封料产能超14万吨，占全球约35%；相当于每年国内消耗1.82万吨环氧树脂和0.98万吨酚醛树脂，全球消耗5.2万吨环氧树脂和2.8万吨酚醛树脂。2022年我国光刻胶产量约19万吨，光刻胶用树脂用量占比约10%-40%，相当于每年消耗1.9-7.6万吨的环氧树脂和酚醛树脂。

综上所述，2025年我国高端酚醛和环氧树脂的供给或超5万吨/年和23万吨/年，保守估计半导体领域（芯片、光刻胶、封装）对酚醛和环氧树脂的需求量约达5.14-8.48万吨/年和7.78-11.9万吨/年。该领域的需求量相当于消耗苯酚15.53万吨（占其2023年产量约3.68%）。电子酚醛树脂供需在理论上国内还存在缺口，特种环氧树脂我国的设计产能或可以覆盖需求。

将半导体用树脂的需求计入苯酚平衡表后，假设苯酚及下游的开工率保持不变，2024-2025年，苯酚基本面或由过剩转为出现缺口。随着苯酚投产放缓及下游投产增速加快，苯酚未来两年有望在纯苯下游中一改弱势，支撑对纯苯的需求增长。

但整体来看，半导体用树脂的需求增长对苯酚基本面的边际影响尚不重要，未来两年相对PC（聚碳酸酯）的规模增量更小，仍需观察后续算力需求的变化。

表6：计入电子树脂需求后的苯酚平衡表

单位（万吨）	2023	2024E	2025E
苯酚产能	635	657	713
苯酚产量	422	427.05	463.45
苯酚净进口	33.83	30	27
酚醛树脂产量	94.63	100	100
双酚A产能	480.5	522.5	594.5
双酚A产量	330	365.75	416.15
环氧树脂产量	182.84	195.57	206.37
PC产量	248.6	273.65	273.65
双酚A理论供需差	-3	-18.08	21.90
电子树脂额外需求	12.92	13.79	15.53
苯酚理论供需差	35.83	-3.09	-19.31

来源：国联期货研究所百川盈孚卓创资讯

联系方式

国联期货研究所无锡总部

地址：无锡市金融一街8号国联金融大厦6楼(214121)

电话：0510-82758631

传真：0510-82757630

国联期货研究所上海总部

地址：上海市浦东新区滨江大道999号高维大厦9楼（200135）

电话：021-60201600

传真：021-60201600

免责声明

本报告中信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。

报告中的内容和意见仅供参考，并不构成对所述期货操作的依据。由于报告在撰写时融入了研究员个人的观点和见解以及分析方法，如与国联期货发布的其他信息有不一致及有不同的结论，未免发生疑问，本报告所载的观点并不代表国联期货公司的立场，所以请谨慎参考。我公司及其研究员对使用本报告及其内容所引发的任何直接或间接损失概不负责。

本报告所提供资料、分析及预测只是反映国联期货公司在本报告所载明日期的判断，可随时修改，毋需提前通知。

本报告版权归国联期货所有。未经书面许可，任何机构和个人不得进行任何形式的复制和发布。如遵循原文本意的引用，需注明引自“国联期货公司”，并保留我公司的一切权利。

期市有风险 投资需谨慎