

应用端百花齐放，硬件侧本固枝荣

2023 年 04 月 11 日

► **ChatGPT 引入插件 Plugin，开启 AI 生态。**在 ChatGPT 联网前，用户仅能查询到 2021 年 9 月之前的消息，但随着 OpenAI 开放网络浏览器和代码解释器，用户不仅能够查询到最新新闻。从这一点看，我们认为 ChatGPT 未来最大的进步在于从“有监督式”学习转向“无监督式学习”。在联网前，ChatGPT 都是从已有固话训练集中获取学习素材，需要认为调整训练参数才能提升学习精度；联网后，ChatGPT 能够根据实时更新数据素材进行“自我迭代”，搜索精准度、理解准确度都将迎来新的提升。

► **算力仍是重中之重，带动产业链成长。**我们认为 ChatGPT 体现了当前 AI 训练速度、训练精度的提升，背后的根本逻辑在于庞大训练数据集的支撑，而对于硬件层面来说，计算与传输是保证 ChatGPT 平稳运行的核心。根据贝壳财经咨询，北京时间 2 月 9 日，由于训练量超负荷，ChatGPT 官网无法登录，页面显示“超负荷”，侧面印证了当前的网络资源/计算能力无法保证未来大量涌入的算力需求。ChatGPT 的总算力消耗约为 3640PF-days，从摩尔定律角度看，以 ChatGPT 为代表的 AI 发展速度呈指数级上涨，约每隔 3~4 个月翻倍（摩尔定律翻倍周期约 18 个月）。

► **从 ChatGPT 所需要 GPU 数量测算，未来所需要高功率机柜近 7.5 万个。**根据钛媒体，GPT 3.5 完全训练需要 A100 芯片近 3 万颗，对应英伟达 DGX A100 服务器近 3750 台。考虑到英伟达 DGX A100 服务器单台最大功率为 6.5kw，假设平均功率为 4kw/台，则 10kw 功率机柜可搭载 2.5 台 AI 服务器。为满足 3750 台 AI 服务器，需要建设机柜 1500 个，假设算力需求扩至 50 倍，共需要高功率机柜约 7.5 万个。

► **AI 大趋势下，数据中心耗电成为“阿喀琉斯之踵”。**数据中心制冷主要采取风冷+液冷，液冷有望成为大数据中心主流方案。根据 IDC 圈，液冷的冷却能力较风冷而言高 1000~3000 倍，且液冷对环境要求更低，适应性更强。根据 Dell 'Oro 统计预测，2022 年-2027 年数据中心基础设施（DCPI）收入 CAGR 有望保持 8%，超 350 亿美元。且随着英特尔、英伟达、AMD 等处理器与加速器性能的提升，数据中心机架功率密度会再度提升。

► **投资建议：**AI 模型及相关应用带动算力增长，建议关注第三方头部 IDC 运营厂商**奥飞数据、数据港、宝信软件（计算机组覆盖）、光环新网、润泽科技（计算机组覆盖）**；同时我们认为随着未来高密度数据中心兴起，叠加 ESG 要求，IDC 储能温控环境有望核心受益，建议关注**科华数据、申菱环境、英维克、依米康**。

► **风险提示：**市场竞争加剧，技术更新风险，电价波动风险。

重点公司盈利预测、估值与评级

代码	简称	股价 (元)	EPS (元)			PE (倍)			评级
			2021A	2022E	2023E	2021A	2022E	2023E	
300738.SZ	奥飞数据	14.51	0.38	0.25	0.30	39	58	48	/
300442.SZ	润泽科技	67.8	-0.01	1.28	1.90	/	53	36	推荐
300383.SZ	光环新网	12.81	0.47	-0.50	0.33	27	/	39	推荐
603881.SH	数据港	36.36	0.34	0.35	0.49	107	104	74	推荐
600845.SH	宝信软件	59.36	0.92	1.15	1.46	65	52	41	推荐
301018.SZ	申菱环境	44.37	0.58	0.96	1.45	77	46	31	/
002837.SZ	英维克	34.84	0.61	0.64	0.84	57	54	41	推荐

资料来源：，民生证券研究院预测；注：股价为 2023 年 04 月 10 日收盘价；未覆盖公司数据采用 wind 一致预期，英维克、数据港 2022 年为真实数据

推荐

维持评级



分析师 马天诣

执业证书：S0100521100003

电话：021-80508466

邮箱：matianyi@mszq.com

研究助理 谢致远

执业证书：S0100122060027

邮箱：xiezhiyuan@mszq.com

相关研究

- 1.通信行业点评：算力集群有望成为 IDC 大趋势-2023/04/02
- 2.通信行业点评：算力，AI“星辰大海”的边界-2023/03/27
- 3.通信行业点评：微软发布 Copilot，GPT 应用侧和算力侧谁最先受益？-2023/03/17
- 4.IDC 行业点评：东风送暖，拨云见日-2023/03/10
- 5.卫星互联网行业点评：星网牵头，五大运营商联手推进卫星互联网进程-2023/03/09

目录

1 应用：ChatGPT 引入插件 Plugin，AI 生态伊始	3
1.1 看海外：ChatGPT 插件或开启苹果 App Store 模式	3
1.2 析国内：各家大厂纷纷发布大模型	4
2 硬件：算力仍是重中之重，带动产业链成长	10
2.1 超越摩尔定律，未来 AI 算力有望几何式增长	10
2.2 从 CPU 到 GPU，AI 模型推动行业结构性变化	12
2.3 数据中心走出新生态，打造算力池	15
3 温控：解决高功率算力池下的“阿喀琉斯之踵”	21
4 重点公司	23
4.1 奥飞数据	23
4.2 光环新网	23
4.3 数据港	24
4.4 英维克	25
4.5 申菱环境	26
4.6 依米康	26
5 风险提示	28
插图目录	29
表格目录	29

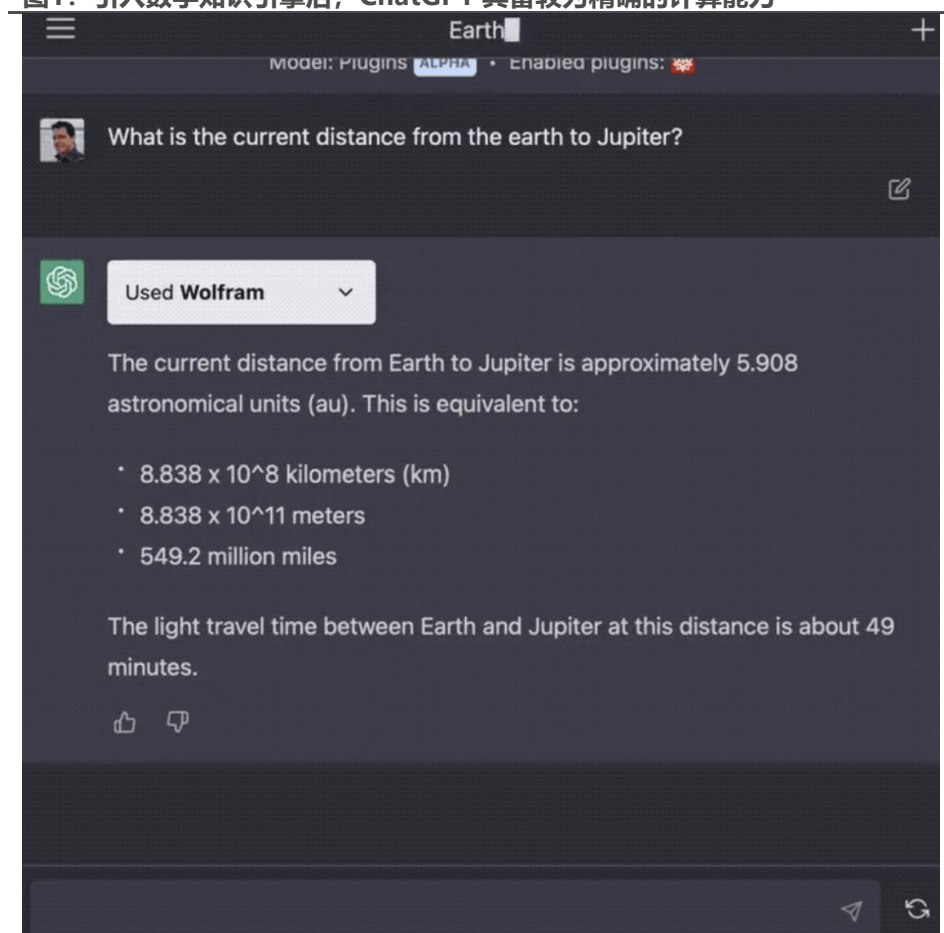
1 应用：ChatGPT 引入插件 Plugin，AI 生态伊始

1.1 看海外：ChatGPT 插件或开启苹果 App Store 模式

插件支持功能促进 ChatGPT 更好接入第三方应用。北京时间 3 月 24 日，人工智能公司 OpenAI 宣布为 ChatGPT 添加插件支持，未来可访问第三方应用程序。目前 OpenAI 针对开发人员及 ChatGPT 用户，将插件集成至其使用产品的 API 中，通过“开发—使用”闭环增加用户访问量，从而打造人机交互范式社区。目前已有多家巨头接入使用插件。

ChatGPT 通过插件功能已实现类似 Python 的效果，赋能百业。根据 OpenAI 官方演示，ChatGPT 接入数学知识引擎 Wolfram Alpha 后，数值计算精确度得到较大提升。

图1：引入数学知识引擎后，ChatGPT 具备较为精确的计算能力



资料来源：OpenAI 官网，民生证券研究院

联网打开 ChatGPT 更多想象空间。在 ChatGPT 联网前，用户仅能查询到 2021 年 9 月之前的消息，但随着 OpenAI 开放**网络浏览器**和**代码解释器**，用户不仅能够查询到最新新闻，同时能够将自己的私人数据如**个人文档、笔记、邮件**等输入 ChatGPT 供其训练。从这一点看，我们认为 ChatGPT 未来最大的进步在

于从“有监督式”学习转向“无监督式学习”。在联网前，ChatGPT 都是从已有固话训练集中获取学习素材，需要人为调整训练参数才能提升学习精度；联网后，ChatGPT 能够根据实时更新数据素材进行“自我迭代”，搜索精准度、理解准确度都将迎来新的提升。

图2：ChatGPT 可线上链接多个第三方库资源

 Expedia Bring your trip plans to life—get there, stay there, find things to see and do.	 FiscalNote Provides and enables access to select market-leading, real-time data sets for legal, political, and regulatory data and information.	 Instacart Order from your favorite local grocery stores.	 KAYAK Search for flights, stays and rental cars. Get recommendations for all the places you can go within your budget.
 Kiarna Shopping Search and compare prices from thousands of online shops.	 Milo Family AI Giving parents superpowers to turn the manic to magic, 20 minutes each day. Ask: Hey Milo, what's magic today?	 OpenTable Provides restaurant recommendations, with a direct link to book.	 Shop Search for millions of products from the world's greatest brands.
 Speak Learn how to say anything in another language with Speak, your AI-powered language tutor.	 Wolfram Access computation, math, curated knowledge & real-time data through Wolfram Alpha and Wolfram Language.	 Zapier Interact with over 5,000+ apps like Google Sheets, Trello, Gmail, HubSpot, Salesforce, and more.	

资料来源：量子位，民生证券研究院

打造类 App Store 商业模式，多家知名品牌已经接入，AI 生态圈正处于起步阶段。目前 OpenAI 邀请的第一批开发伙伴主要包含 Shopify、Slack 以及旅游平台 Expedia、生鲜电商 Instacart、旅游搜索引擎服务商 Kayak、在线餐厅预订服务平台 OpenTable、自动化平台 Zapier 等。如用户可通过 Instacart+ChatGPT 插件，发送预定商品指令，ChatGPT 处理指令后电商平台 Instacart 可直接下单并将货品送至客户地址。

我们认为 ChatGPT 已经成为 OpenAI 的生态平台底座，开发人员可在 OpenAI 平台上创建并分享自己基于 ChatGPT 底层开发的插件。类似于苹果的 App Store 允许第三方应用程序在其设备上广泛接入。

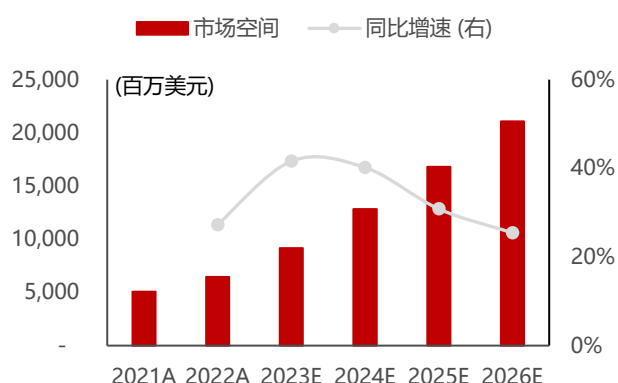
1.2 析国内：各家大厂纷纷发布大模型

供需两侧增长较快，国内环境给予大模型良好生长土壤。从需求端看，我国 AI 软件及应用市场增长较快。根据 IDC 统计，2021 年中国人工智能软件及应用市场规模为 51 亿美元，预计 2026 年达 211 亿美元。我们认为当前人工智能已进入大规模应用的关键时期，但过去受高开发门槛、复杂多样应用场景、数据标注等问题桎梏，AI 应用并未形成规模化效应。从**供给侧**看，信息化的发展使得数据量井喷式增长，带来数据“宝藏”的同时也对技术提出更多挑战，但不可否认的是，数据既是 AI 发展的驱动力，同样也可成为其发展的瓶颈，数据的井喷式增

长也带来了算力复杂度提升。根据 IDC 统计，我国数据规模有望从 2021 年的 18.51ZB 增长至 2026 年的 56.16ZB，CAGR 24.9%，增速居于全球第一。

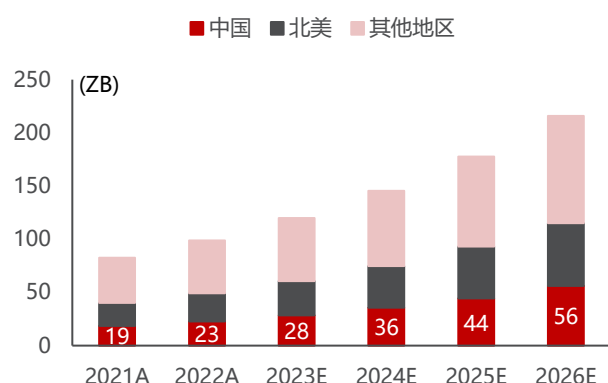
“大模型”是打通人工智能技术通用性“任督二脉”的关键。过去在分散化的模型研发模式下，单一的 AI 应用场景下多个任务需要由多个模型共同支撑完成，每一个模型建设都需要算法开发、数据处理、模型训练与调优过程。预训练大模型增强了人工智能的通用性、泛化性，基于大模型通过零样本或小样本精调，就可实现在多种任务上的较好效果。大模型“预训练+精调”等模式带来了新的标准化 AI 研发范式，实现 AI 模型在更统一、简单的方式下规模化生产。

图3：我国 AI 软件及应用市场规模



资料来源：IDC 咨询，民生证券研究院

图4：我国数据量规模全球领先



资料来源：IDC 咨询，民生证券研究院

百度：文心大模型已拥有模型+平台+产品构成的全套体系。百度文心已经构建了文心大模型层、工具平台层、产品与社区三层体系。在产品能力上，文心大模型通过与百度飞桨 Paddle-Paddle 深度学习平台组合，可以满足市场大规模落地需求；在应用能力方面，百度已在金融、能源、制造、城市、传媒、互联网等行业拥有实际落地的标杆案例；在生态能力方面，以社区用户为基础，文心大模型可以实现与开发者、行业用户、上下游产业的正向互动。

百度文心大模型参数量达到 2600 亿，在市场格局中处于第一梯队。文心大模型构建了“基础+任务+行业”的三级模型体系，基础大模型作为整体的底座来支撑任务与行业大模型的建设，任务和行业大模型则通过结合真实场景与数据反哺基础大模型优化。基础大模型旨在技术挑战、通用性、泛化性探索；任务大模型由对话、搜索、信息抽取、生物计算等多个典型任务构成；最后所得出的行业大模型已经形成了包含了与来自 8 个行业的头部企业或机构共建的 11 个行业大模型。

行业大模型是百度文心大模型深入产业落地的重要措施与载体。百度与行业头部企业、机构联合研发的融合行业数据、知识以及专家经验的大模型，在各行业的技术效果突破、产品创新、生产流程变革、降本增效等维度产生价值。如国家电网与百度联合发布知识增强的电力行业大模型；浦发银行与百度在行业数据、大算力和 AI 算法上优势互补，联合研发面向金融行业的大模型；百度与人民网合

作，引入舆情数据中心沉淀的行业知识来更好训练知识增强的传媒行业大模型实现更少的标注数据下大幅提升传媒行业自然语言处理任务效果等。

图5：百度文心大模型全景图

产品与社区	文心一格 AI艺术和创意辅助平台			文心百中 大模型驱动的产业级搜索系统			畅谷社区 大模型创意与探索社区						
工具与平台	EasyDL-大模型 零门槛 AI 开发平台			BML-大模型 全功能 AI 开发平台			大模型 API						
	大模型套件												
	数据标注与处理		大模型精调		大模型压缩		高性能部署		场景化工具				
文心大模型	行业大模型												
	国网-百度文心		浦发-百度文心		航天-百度文心		人民网-百度文心		冰城-百度文心		电影频道-百度文心		
	深燃-百度文心		吉利-百度文心		泰康-百度文心		TCL-百度文心		辞海-百度文心				
	NLP 大模型				CV 大模型			跨模态大模型			生物计算大模型		
	医疗 ERNIE-Health		金融 ERNIE-Finance		商品图文搜索表征学习 VIMER-UMS			文图生成 ERNIE-ViLG		文档智能 ERNIE-Layout		化合物表征学习 HelixGEM	
	对话 PLATO	搜索 ERNIE-Search	信息抽取 ERNIE-UIE		OCR图像表征学习 VIMER-StrucText			视觉-语言 ERNIE-ViL	语音-语言 ERNIE-SAT	地理-语言 ERNIE-GeoL	蛋白质结构预测 HelixFold		
	跨语言 ERNIE-M	代码 ERNIE-Code	图网络 ERNIE-Sage		多任务视觉表征学习 VIMER-UFO								
	语言理解与生成				视觉处理 多任务学习 VIMER-TCIR		自监督视觉表征学习 VIMER-CAE		单序列蛋白质结构预测 HelixFold-Single				
	ERNIE 3.0 Tiny (轻量级)	ERNIE 3.0 (百亿级)	鹏城-百度文心 (千亿级)		ERNIE 3.0 Zeus (任务知识增强)								

资料来源：文心大模型官网，民生证券研究院

百度文心围绕大模型产业应用的不同研发环节，面向不同开发者或用户，打造系列工具平台与场景化产品。对于开发者拥有面向 NLP 工程师的大模型套件 ERNIEKit;对于下游应用，文心开放了 API 接口供其使用；对于用户则推出了基于大模型驱动的新一代产业级搜索系统文心百中，以及 AI 艺术与创意辅助平台文心一格。

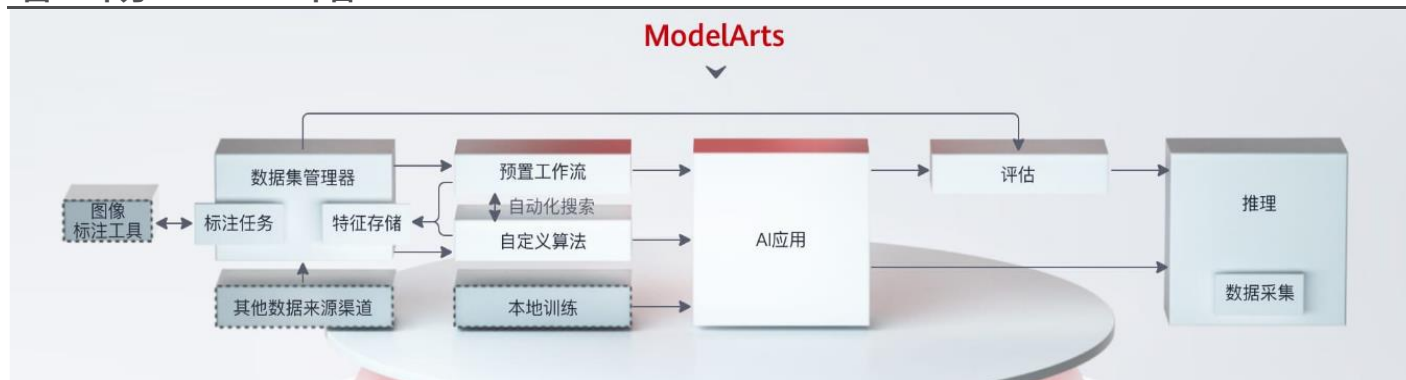
华为：盘古大模型已深耕行业，发掘应用。华为的盘古 AI 大模型项目于 2020 年立项，并于 2021 年 4 月发布“盘古大模型”。依托华为自身的全栈式 AI 解决方案，将大模型与 ModelArts 平台深度结合。当前盘古大模型已经发展出包括基础大模型（L0）、行业大模型（L1）、行业细分场景模型（L2）三大阶段的成熟体系。

盘古大模型首个落地应用“矿山 AI 大模型”。在智能矿山领域，华为旨在将 AI 应用门槛降低，将应用场景串联，将 IT、CT、OT 系统相互联系。为此，华为拥有以“矿山 AI 大模型”为代表的 IT 技术；以 5G、F5G 为代表的 CT 技术；以“矿鸿”操作系统为代表 OT 技术。“IT+CT+OT”的“3T 融合”将矿山内部众多的工作“微场景”联接成为“中场景”，将“中场景”联接成为“巨场景”，再将“巨场景”联接成为“智能矿山”。进而推动集团型能源企业的数字化转型。

华为 ModelArts 一站式 AI 平台为开发者提供平台支持。ModelArts 平台拥有为机器学习与深度学习提供海量数据预处理及交互式智能标注、大规模分布式训练、自动化模型生成，及端-边-云模型按需部署的能力，并具有数据处理、

算法开发、模型训练、AI 应用管理和部署等功能，可以帮助用户快速创建和部署模型，管理全周期 AI workflow。

图6：华为 ModelArts 平台



资料来源：华为云官网，民生证券研究院

腾讯：混元大模型已成为深入产业落地的重要措施。腾讯在 2022 年 4 月首次对外披露了混元大模型，该大模型作为协同了腾讯预训练研发力量的成果，完整的覆盖了 NLP 大模型、CV 大模型、多模态大模型以及众多行业领域的任务模型。

腾讯太极机器学习平台为混元大模型提供底层支持。腾讯自主研发的一站式机器学习生态服务平台——太极机器学习平台，为 AI 工程师提供了从数据预处理、模型训练、模型评估到模型服务的全流程高效开发工具，已成为大模型训练的有效保障。

图7：太极-HCF ToolKit 加速大模型落地



资料来源：量子位公众号，民生证券研究院

混元 AI 大模型已在腾讯多个业务场景中落地。混元大模型先后支持了包括微信、QQ、游戏、腾讯广告、腾讯云等多项产品和业务，通过 NLP、CV、跨模态等 AI 大模型，在创造了增量价值的同时降低使用成本。其中在广告内容理解、行业特征挖掘、文案创意生成等方面的应用，在为腾讯广告带来大幅 GMV 提升

的同时，也初步验证了大模型的商业化潜力。

图8：太极-HCF ToolKit 加速大模型落地



资料来源：量子位公众号，民生证券研究院

阿里巴巴：通义大模型于 2022 年 9 月发布并宣布相关核心模型开源开放。

在 2022 年 9 月 2 日，阿里发布 "通义" 大模型系列，核心模型通过 "魔搭" 社区向全球开发者开源开放。"通义" 系列大模型以统一学习范式和模块化设计理念统一架构、模态、训练、应用等方面，使用开源社区 "魔搭" 进行模型服务共享，推出 "飞天智算平台" 提升 AI 训练效率。目前阿里 "通义" 大模型已广泛用于电商、设计、医疗等领域，助力其降本增效。

图9：阿里巴巴通义大模型落地



资料来源：量子位公众号，民生证券研究院

2022 年 11 月，阿里推出 AI 开源社区 "魔搭" (ModelScope)。自 2021 年起，阿里达摩院先后发布多个版本的多模态及语言大模型，在超大模型、低碳训练技术、平台化服务、落地应用等方面实现突破，引领了中文大模型的发展。

"魔搭"旨在打造下一代开源的模型即服务共享平台，致力降低 AI 应用门槛。
2023 年 3 月，阿里在"魔搭"上线了"文本到视频生成扩散模型"，实现视频生成功能。

2 硬件：算力仍是重中之重，带动产业链成长

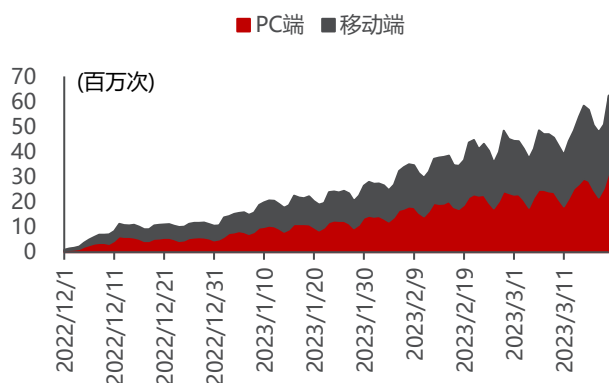
2.1 超越摩尔定律，未来 AI 算力有望几何式增长

ChatGPT如火如荼，开启大算力时代。我们认为ChatGPT体现了当前AI训练速度、训练精度的提升，背后的根本逻辑在于庞大训练数据集的支撑，而对于硬件层面来说，计算与传输是保证ChatGPT平稳运行的核心。根据贝壳财经咨询，北京时间2月9日，由于训练量超负荷，ChatGPT官网无法登录，页面显示“超负荷”，侧面印证了当前的网络资源/计算能力无法保证未来大量涌入的算力需求。ChatGPT的总算力消耗约为3640PF-days，从摩尔定律角度看，以ChatGPT为代表的AI发展速度呈指数级上涨，约每隔3~4个月翻倍（摩尔定律翻倍周期约18个月）。

ChatGPT大模型需要海量算力给予支持。根据OpenAI团队发表于2020年的论文《Language Models are Few-Shot Learners》，训练1次1746亿参数的GPT-3模型需要的算力约为3640 Petaflop/s-day（假设模型1PetaFLOP/s效率运行，需要训练3640天）。根据TrendForce统计，GPT-3训练时大概用到了20000个英伟达A100 GPU芯片。

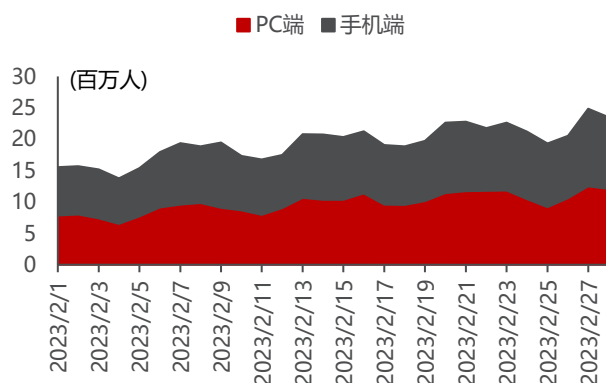
从OpenAI官网访问量看，根据Similarweb数据统计，2月ChatGPT应用PC端+手机端平均日活1939万人，假设每人平均提1000字（假设均为英文单词）左右的问题，则合计产生约194亿单词，根据OpenAI，平均每4个字母对应1个token，我们假设每个单词包含5个字母，则共需要消耗242.5亿个token。假设24小时平均分配任务，平均每天需要的总算力= $242.5\text{亿} \times 1000\text{亿} / (24\text{小时} \times 3600\text{秒}) \times 100 = 280.67\text{ Petaflop/s}$ ，考虑OpenAI访问存在峰值，假定访问峰值是一天均值的2倍。目前英伟达A100芯片峰值算力约19.5TFLOP/s，则需要A100 GPU芯片约 $280.67 \times 2 / 19.5 \times 1000 = 28786\text{片}$ 。

图10：OpenAI 官网访问量



资料来源：Similarweb，民生证券研究院

图11：ChatGPT 日活



资料来源：Similarweb，民生证券研究院

表1: NLP 模型参数规模及训练所需算力

日期	模型	Token大小	参数规模 (M)	特征规模 (M)	训练算力
2014/9/10	Seq2Seq				11,000
2017/6/12	Transformer	512	65	77	23,000
2018/2/15	ELMo		94		3,300
2018/10/11	BERT Large	512	330	230	250,000
2018/11/6	GPT-1	512	110	85	57,000
2019/2/14	GPT-2	1024	1,500	2,000	
2019/7/26	RoBERTa Large	512	1,500	2,000	4,300,000
2019/8/17	Megatron	1024	8,300	4,700	8,100,000
2019/9/26	ALBERT xtl	512	235	2,500	31,000,000
2020/2/13	Microsoft TNLG	1024	1,700	5,700	28,000,000
2020/2/23	ELECTRA Large	128	330	38	3,100,000
2020/5/28	GPT-3	2048	175,000	63,000	310,000,000
2020/6/30	Gshard		600,000		
2020/6/20	Baidu ResSys-C	N/A	2,000,000	N/A	N/A
2020/6/20	Baidu ResSys-E	N/A	10,000,000	N/A	N/A

资料来源: GitHub, 民生证券研究院

表2: CV 模型参数规模及训练所需算力

日期	模型	输入分辨率	参数大小 (M)	训练算力
2012/6/1	AlexNet	227×227	61	460
2014/9/4	VGG-19	224×224	138	11,000
2015/12/2	InceptionV3	299×299	24	100,000
2015/12/10	ResNet152	224×224	55	11,000
2016/2/26	InceptionV4	299×299	82	
2016/10/7	Xception	299×299	23	450,000
2016/11/16	ResNeXt101 (64×4)	224×224	83	12,000
2016/12/3	DenseNet201	224×224	20	2,800

资料来源: GitHub, 民生证券研究院

AI 大模型运算律超越摩尔定律。根据 GitHub 社区文章《AI and Memory Wall》, CV, NLP 和语义模型的模型运算量平均每 2 年翻 15 倍; Transformer 模型的运算量平均每 2 年翻 750 倍。而摩尔定律下内存硬件大小的增长平均每 2 年仅翻 2 倍。

GPU 成为 AI 模型训练基石。GPT-3.5 训练过程中使用了微软专门建设的 AI 计算系统, 其为 1 万个英伟达 V100 GPU 组成的高性能网络算力群。总算力消耗约 3640 Petaflop/s-day。

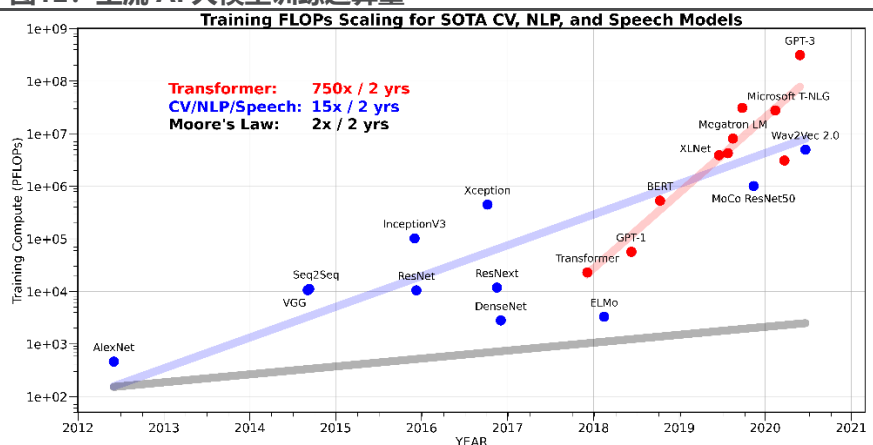
对于人工智能计算架构来说, CPU 配合加速芯片的模式是典型的 AI 部署方案: CPU 提供算力, 加速芯片提升算力并助推算法的产生。常见的 AI 加速芯片按照技术路线可以分为 GPU、FPGA、ASIC 三类。其中, 应用于图形、图像处理领域的 GPU 可以并行处理大量数据, 非常适合深度学习的高并行、高本地化数据场景, 是目前主流的人工智能计算架构。

表3: GPU、ASIC、FPGA 对比

技术路线	GPU	FPGA	ASIC
芯片架构	晶体管大部分构建计算单元 运算复杂度低, 适合大规模并行计算	通过冗余晶体管与连线实现逻辑可编程, 更接近 底层IO, 计算效率高	晶体管根据算法定制, 无冗余 功耗低, 计算性能高
定制化程度	通用	半定制化	全定制化
灵活性	高	高	低
成本	高	较高	低
功耗	高	较高	低
优点	并行运算能力强, 产品成熟	计算效率较GPU与CPU高	体积小, 功耗低, 计算效率高
缺点	价格贵, 功耗散热高	编程难度大, 峰值性能不如ASIC, 量产成本高	算法固定不可编辑, 研发周期长
擅长领域	图像处理, 粗粒度并行计算	算法更新频繁或市场较小专用领域	市场需求较大专用领域
主要玩家	  	 	
市场格局	传统GPU市场, 三足鼎立, 英伟达领先	双雄争霸, 赛灵思领先	市场较为分散, 竞争激烈

资料来源: IDC, 民生证券研究院

图12: 主流 AI 大模型训练运算量



资料来源: GitHub, 民生证券研究院

2.2 从 CPU 到 GPU, AI 模型推动行业结构性变化

偶然中的必然, AI与GPU的不解之缘。2011年, Dan等研究人员研究深度学习时, 曾利用英伟达GTX 480和GTX 580的GPU与英特尔Core i7-920的CPU分别对神经网络模型进行计算。结果显示GPU计算加速效果为远超CPU: 针对网络规模小的模型加速效果GPU为CPU的10倍; 而针对大规模网络模型GPU效果更是高于CPU的60倍。2016年AlphaGo战胜李世石标致着AI浪潮的兴起, 更带动了全球AI产业发展, GPU也逐渐成为AI计算的“标准配置”。

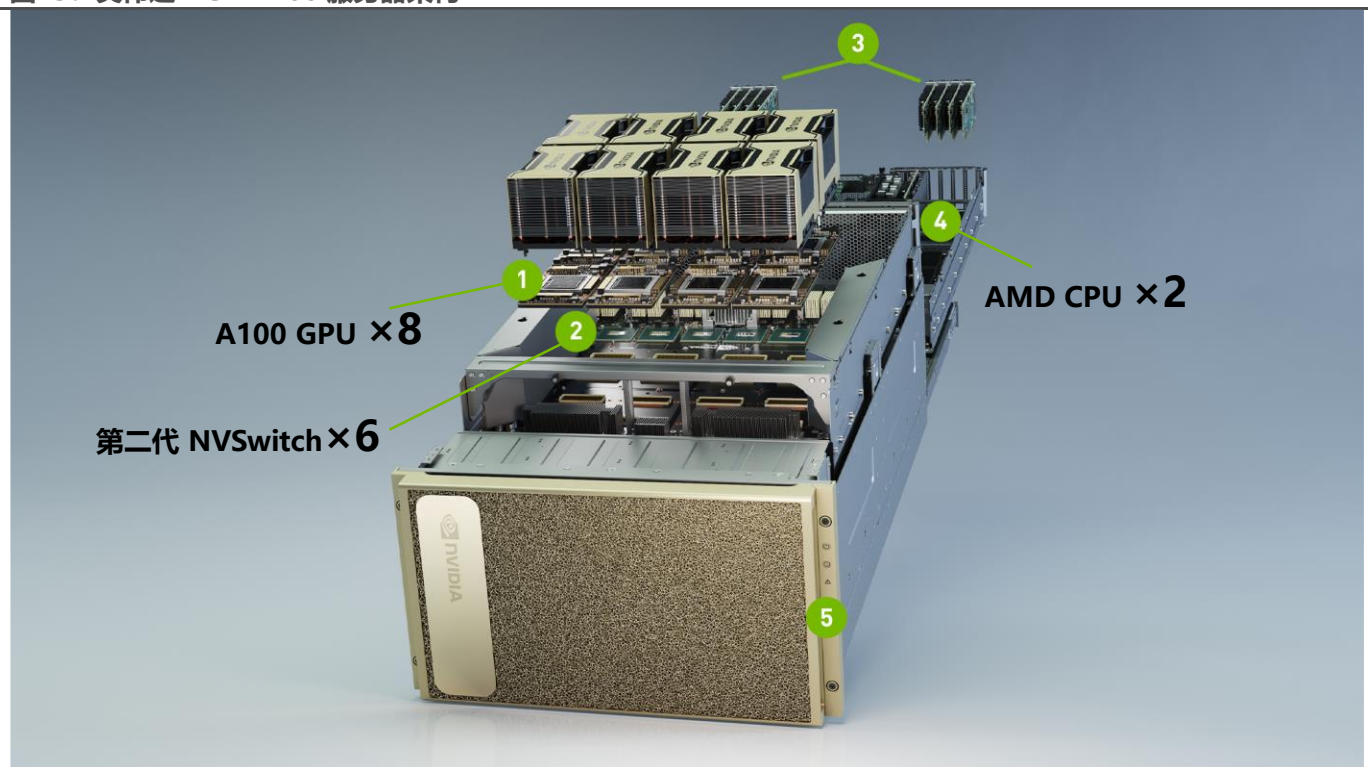
CPU与GPU相互辅助, GPU负责更为复杂计算。不同于CPU核心, GPU专注于某些特定计算任务, 可以将GPU比作“大学教授”, CPU比作“学生”, 虽然教授(GPU)知识结构/计算能力强于学生(CPU), 但学生(CPU)数量庞大, 在处理简单重复计算时效率高于GPU。

从CPU与GPU结构出发, CPU主要从主存中读写数据, 并通过总线与GPU交互。GPU不仅拥有较多的计算核心, 更有独立存储功能。单台服务器上可以安装

多块GPU卡，但**GPU卡的发热量较大**，普通的空调系统难以给大量GPU卡降温，所以大型数据中心通常使用水冷散热，并且选址在温度较低处。

GPU计算特性同AI服务器较为匹配。训练AI模型的过程需要同时对所有数据样本执行几乎相同的操作，而GPU的架构设计具有快速同时处理多个任务所需的并行处理能力，因为针对AI模型的AI服务器大多采用GPU芯片架构。

图13：英伟达 DGX A100 服务器架构

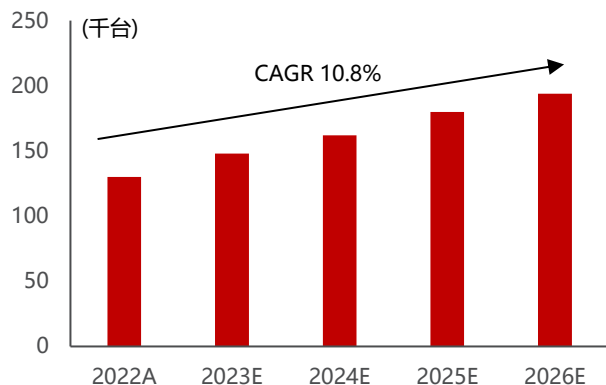


资料来源：英伟达官网，民生证券研究院

AI服务器全球出货量迅速增长，侧面印证GPU需求紧俏度。根据TrendForce统计预测，2022年搭载GPCPU（General Purpose GPU）的AI服务器年出货量占整体服务器比重近1%，2023年出货量同比增长约8%，2022-2026年复合增长率将达10.8%。2022年AI服务器采购中，北美四大云端厂商谷歌、亚马逊 AWS、Meta、微软合计占比66.2%。国内市场方面，字节跳动采购力道最为显著，年采购占比达6.2%，紧随其后的是腾讯（2.3%）、阿里巴巴（1.5%）、百度（1.5%）。

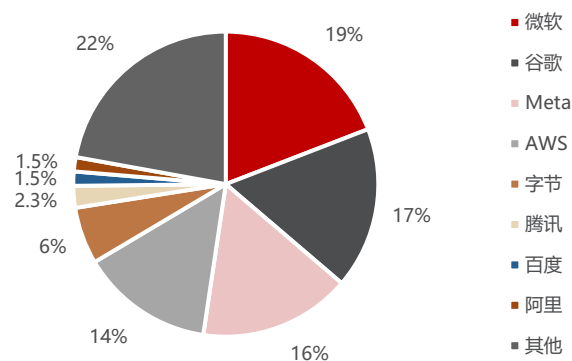
根据AI服务器及GPU搭载比例，模型训练阶段约需要使用近三千台AI服务器。根据JPR统计，截至4Q22，英特尔GPU市场份额达71%，英伟达以17%位居第二，AMD份额保持在12%左右。

图14: AI 服务器出货量稳步增长



资料来源: TrendForce, 民生证券研究院

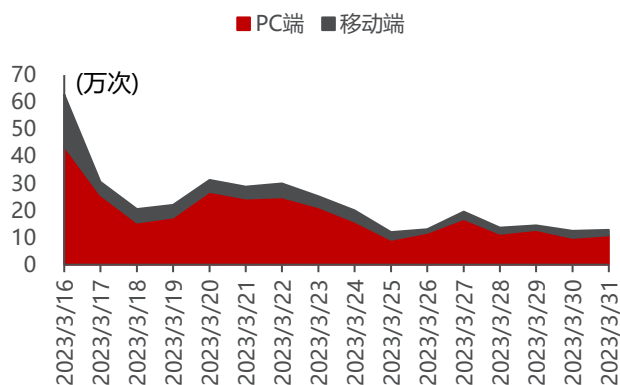
图15: 全球 A 服务器采购量占比



资料来源: TrendForce, 民生证券研究院

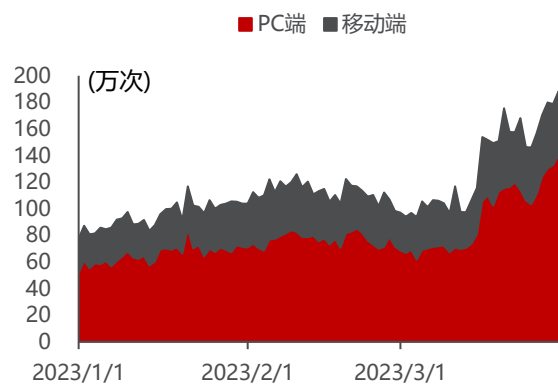
从算力需求计算上来看, 我们认为总算力同未来用户访问次数相关。根据我们上一节对 GPT 算力需求测算, 仅 OpenAI 预训练阶段就需要近 3 万颗英伟达 A100 芯片; 按 1 台 AI 服务器搭载 8 颗 AI 芯片比例测算, 大约需要 3598 台 AI 服务器。我们以 Midjourney (文字生成 AI 图片应用) 访问量为例, 2023 年 3 月平均访问次数 133 万次, 2 月 PC 端+手机端平均日活 1440 万人, 接近 OpenAI 访问量。

图16: 百度文心一言访问频次



资料来源: Similarweb, 民生证券研究院

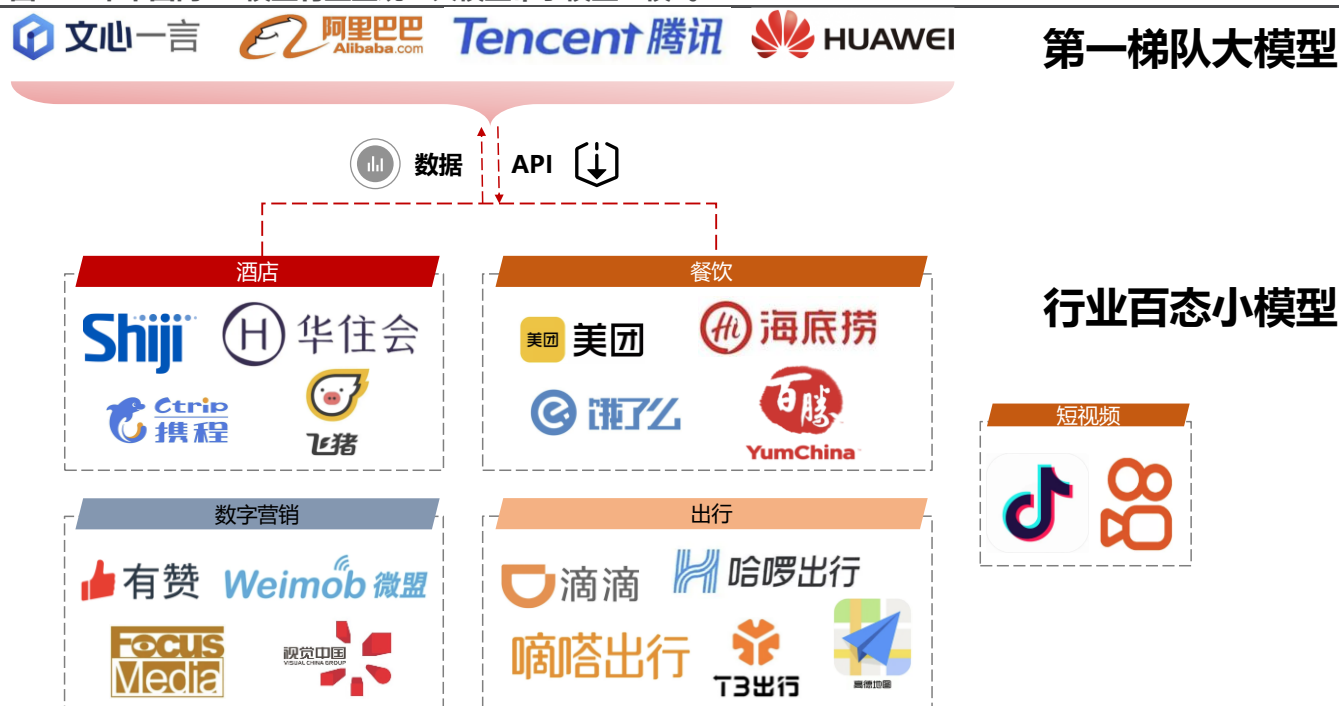
图17: Midjourney 访问频次



资料来源: Similarweb, 民生证券研究院

从国内大模型视角出发, 考虑我国繁多的应用, 算力需求将进一步拉大。虽然我们认为国内 AI 大模型同海外相比仍待发展, 但正因为国内模型较海外存在差距, 更需要算力及数据训练支撑提升模型精度, GPU、服务器需求有望快速增长。我们认为国内在 AI 模型上会逐渐延伸为: **个别大模型+若干小模型**的情况。即几家互联网巨头分别推出自家 AI 大模型, 二线互联网厂商基于自己应用场景推出垂直“小模型”, 接入大模型训练。我们认为国内大模型也将类似于 OpenAI 开放 API 接口, 垂直小模型将通过 API 接口调用大模型运算能力/训练准确度。百度推出文心一言 API+飞桨接口也同样印证了我们上述想法。从上述维度看, 国内 AI 模型也将走出 MaaS (Model as a service), 百度、腾讯、阿里等云厂商将大模型接入云后, 为垂直行业提供数据训练服务。

图18：未来国内 AI 模型有望呈现“大模型带小模型”模式

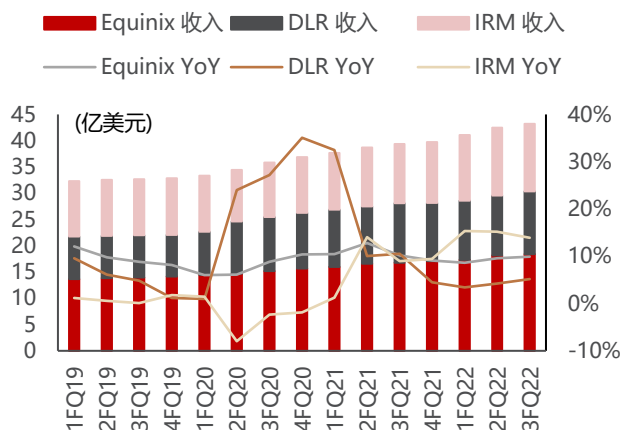


资料来源：民生证券研究院整理

2.3 数据中心走出新生态，打造算力池

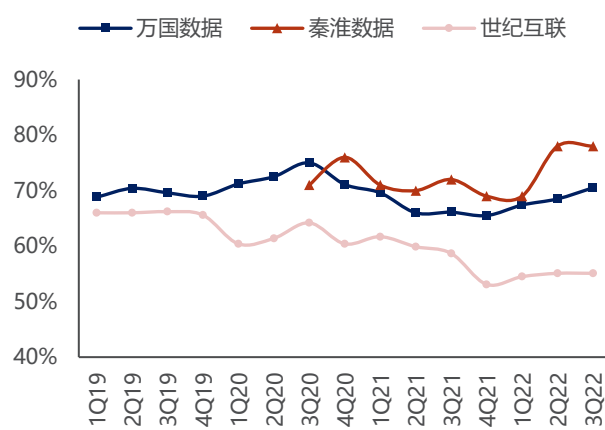
全球维度看，以 EQIX、DLR、IRM 等 IDC 龙头公司收入均呈稳步增长，3FQ22 三家公司收入分别同比增长 9.9%/5.2%/13.9%。我们认为海外 IDC 企业收入稳步增长主要受益于下游客户加大对数字化转型投入正相关。此外，海外区域分布式混合云也加大了流量算力需求。回顾国内，自 3Q21 电价市场改革以来，IDC 公司总成本承压。结合 REITs、ESG 等新兴投融资方式，我们认为“绿色化能力较差”的 IDC 公司将面临融资成本上升，PUE 改造升级等内外部压力，预计未来行业竞争格局出现改善。

图19：海外 IDC 公司收入及增速



资料来源：各公司公告，民生证券研究院

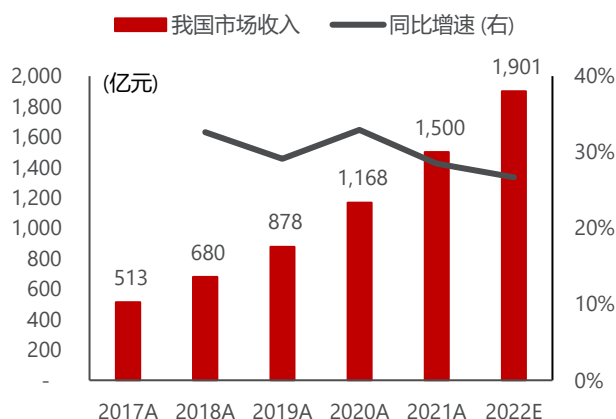
图20：万国数据、秦淮数据、世纪互联上架率回升



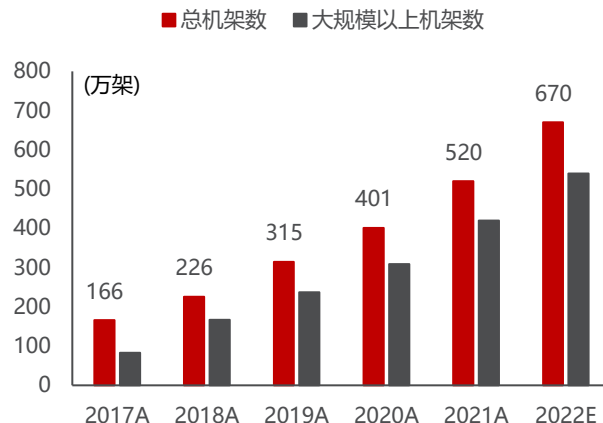
资料来源：各公司公告，民生证券研究院

AI+数字经济拉动 IDC 需求稳定增长；需求拉动，数据中心加速向大型化转型。根据中国信通院统计，2021 年我国数据中心行业市场收入达到 1,500 亿元左右，近三年年均复合增长率达到 30.7%。AI+数字经济加速云产业发展，国内互联网行业受人口红利减弱及反垄断政策的影响，互联网厂商近年来资本开支呈现疲软态势。我们认为，AI 的激增算力将加大对算力提升的需求，加之数字经济政策进一步推动企业上云，云计算产业有望加速发展。

我国数据中心机架规模稳步增长。按照标准机架 2.5kW 统计，2021 年，我国在用数据中心机架规模达到 520 万架，近 5 年 CAGR 超过 30%。其中，大型以上数据中心机架规模增长更为迅速，按照标准机架 2.5kW 统计，机架规模 420 万架，占比达到 80%。

图21：我国数据中心市场规模


资料来源：中国信通院，民生证券研究院

图22：我国数据中心机架规模


资料来源：中国信通院，民生证券研究院

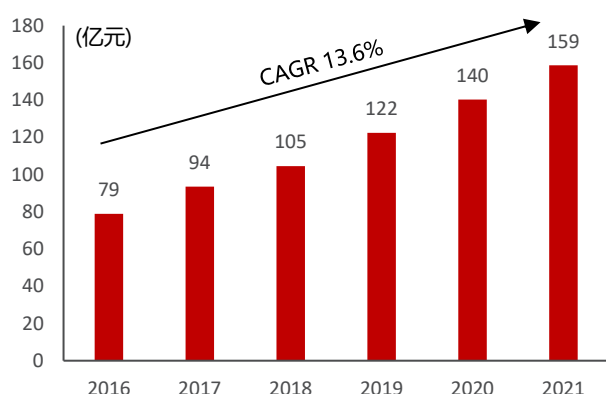
IDC 基础设施有明显的地域性特征。IDC 的地域集中性和中国网络拓扑结构及客户需求密切相关，因为 IDC 需要保障客户数据对外的网络连接顺畅。用户选择 IDC 资源时，网络时延是重要的考虑因素；对时延要求敏感的应用如远程服务、在线娱乐、在线支付等需求快速增长，且向头部城市靠拢，导致一线城市 IDC 需求强劲增长。国内网络结构中，核心节点为北京、上海、广州、沈阳、南京、武汉、成都和西安八个城市，其中北京、上海、广州为三个中心，与其他核心节点互联并负责与国际 Internet 互联。

复盘 IDC，社交媒介改变/数据井喷有望带来需求高增。我们认为上海数字经济发展位于全国前列，沪上 IDC 市场空间增长一定程度代表了数字经济背景下国内 IDC 行业的发展。根据科智咨询统计，2021 上海地区传统 IDC 业务市场规模达到 159 亿元，实现平稳增长。供给侧和需求侧的双重影响，上海地区潜在有效供给持续增加，需求平稳增长，以公有云为代表的泛互联网行业仍是上海新增需求的主要驱动力。未来几年，基于上海本地互联网企业云化进程加速、全国互联网企业在华东区域的规模化布局的这大背景下，5G、AI 等技术有望带动消费

互联网进一步创新，同时随着传统产业持续数字化转型，上海市数据中心产业仍有望保持快速增长。

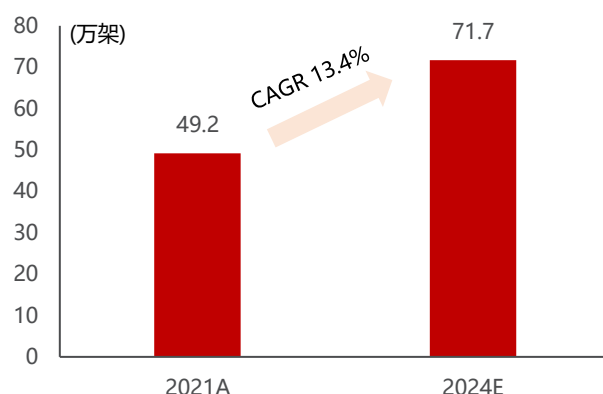
2021 年北京地区数据中心增长同样印证了前文逻辑。2021 年，北京及周边地区 IDC 市场最主要的需求来源为视频、电商、游戏等互联网行业，互联网行业 IDC 需求占比达到 65.2%。我们认为北京地区需求侧的增长同样来自于社交媒介的改变。2020-2021 年短视频的广泛传播一方面加大了用户流量；另一方面短视频的数据格式相较文字更为复杂，多维度计算同样加大了算力需求。

图23：上海传统 IDC 业务市场空间



资料来源：科智咨询，民生证券研究院

图24：北京及周边地区 IDC 机柜供给快速增长



资料来源：中国工信产业网，民生证券研究院

模型训练成本趋呈上升趋势。过去计算机行业假设为“数字处理会变得越来越便宜”，但考虑到不断上升的研究复杂性和竞争性，斯坦福人工智能研究所副所长克里斯托弗曼宁表示：最前沿模型的训练成本还在不断上升。

训练计算环境成本与模型大小成正比。根据马萨诸塞大学阿默斯特校区研究人员公布的研究论文显示，大模型底座以神经网络为主。而在调整神经网络体系结构以尽可能完成详尽的试验，并优化模型的过程，相关成本较高。单 BERT 模型的碳足迹约为 1400 磅二氧化碳。

我们认为，AI 计算系统正在面临计算平台优化设计、复杂异构环境下计算效率、计算框架的高度并行与扩展、AI 应用计算性能等挑战。算力的发展对整个计算需求所造成的挑战会变得更大，提高整个 AI 计算系统的效率迫在眉睫。

智算中心大势所趋。根据 IDC 发布《2022-2023 中国人工智能计算力发展评估报告》，中国人工智能计算力继续保持快速增长，2022 年智能算力规模达到 268 百亿亿次/秒(EFLOPS)，已超过通用算力规模。

智算中心建设对软硬件均有一定要求。目前我国的智算中心基本采用了高标准建设，100P 算力是起步目标，该算力大约相当于 5 万台高性能电脑。我们认为如果说过去传统的数据中心是“快捷酒店”，那么智算中心可以理解为“高端

型酒店”，主要针对拥有 AI 模型计算的“高净值客户”提供数据训练支撑。而对于第三方数据中心厂家而言，智算中心的投资模式也不同于传统 IDC 机柜。对于传统 IDC 机柜，内部服务器、CPU 等大多有下游客户自行购买，且应用场景也主要为云计算、简易推荐算法等。而智算中心主要针对 AI 大模型训练，对于 BAT 等大型互联网厂商而言，自行采购 GPU、AI 服务器带来的成本足以被模型带来的收益所覆盖；而对于**二三线互联网厂商**而言，要想训练出自己的大模型不可避免需要投资如英伟达 A100/A800 等高性能芯片，高昂的成本或使上述小厂“望而却步”，但如通过第三方数据中心公司采购，则成本可被“嫁接”至数据中心厂商的资本开支，会随逐年折旧摊薄，二三线互联网“小模型”接入训练后同时能够带动第三方数据中心上架率的提升。

图25：中国科大瀚海 20 超级计算系统配置

序号	类别	规格型号/主要特征	主要参数	数量	备注
1	管理/登录节点	华为FusionServer 2288H V5	2*Intel Xeon Scale 6248(20核, 2.5GHz), 192GB DDR4 2933MHz内存, 2*1TB NVMe硬盘	4	
2	普通双路CPU计算节点	华为FusionServer X6000(XH3 21L V5节点)	2*Intel Xeon Scale 6248(20核, 2.5GHz), 192GB DDR4 2933MHz内存, 240GB SSD硬盘, 板级液冷技术	720	适合一般CPU作业
3	双V100 GPU计算节点	华为FusionServer G5500(G530 V5节点)	2*Intel Xeon Scale 6248(20核, 2.5GHz), 2*NVIDIA Tesla V100 GPU, 384GB DDR4 2933MHz内存, 1TB NVMe硬盘	10	支持GPU作业
4	八A100 GPU计算节点	融科联创RW-4124GO-NART	2*AMD Rome 7742 (2.25GHz, 128核), 1TB DDR4 3200MHz, 8*NVIDIA A100 Tensor Core GPU,NVLink, 2*1.92TB+4*3.84TB NVME	2	支持大GPU作业
5	2TB AEP大内存计算节点	华为FusionServer 2288H V5	2*Intel Xeon Scale 6248(20核, 2.5GHz), 384GB DDR4 2933MHz内存, 2TB Intel AEP内存, 1TB NVMe硬盘	8	适合大内存作业
6	国产ARM CPU计算节点	华为Taishan 2280V2	2*海思Hi1620 CPU (48核,2.6GHz), 256GB DDR4 2666MHz内存, 300GB SAS硬盘	20	10台各含6块Atlas 300 AI卡
7	存储系统	长虹GS7990(OEM DDN)	1*GS7990主机箱+2*SS9012磁盘扩展机箱, 248*8TB NL-SAS硬盘, 1.5PB可用容量	1	可用容量1.5PB, 持续IO聚合带宽>20GB/s
8	并行文件系统	长虹GRIDScaler(OEM DDN)	基于IBM GRIDScaler	1	
9	高速计算网络	Mellanox HDR100		1	ARM节点为25Gbps以太网

资料来源：中科大官网，民生证券研究院

超算中心涉及多个核心 GPU 芯片，有望拉动板块整体投资。以中科大瀚海 20 超级计算系统为例，该系统共计 752 个节点，含 30640 颗 CPU 核心、20 块 Nvidia Tesla V100GPU 卡、16 块 NVIDIA A100 Tensor Core GPU 卡及 60 块华为 Atlas AI 卡，理论峰值双精度浮点计算能力达 2.69Petaflop/s。

根据我们测算，但超算中心有望拉动投资额近 500 亿元。中科大瀚海 20 超级计算系统总功率约 39kw，相较普通机柜功率高 2~4 倍，属高密度机柜。假设当前第三方 IDC 公司所建高功率机柜以 10kw 功率为主，则投资额约为 227 亿元。

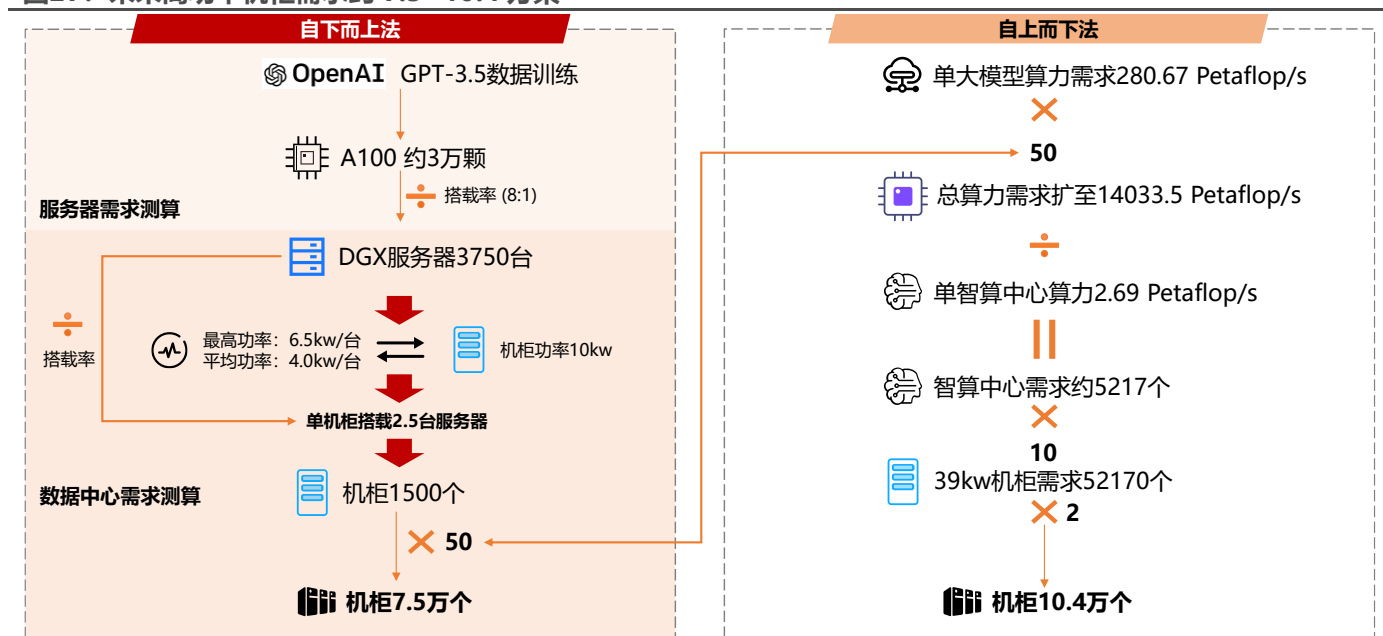
图26：智算中心投资成本

设备	数量	详细信息	单价 (万元)	总成本 (万元)
服务器				
GPU	8	英伟达A100	12	96
CPU	2	华为Taishan 2280V2	3	7
存储	1	长虹GS7990(OEM DDN)	4	4
网络网关				
高速计算网络	1	Mellanox HDR100	8	8
IB交换机	1	200G Infiniband高速网络	12	12
核心交换机	0.25	800G端口	20	5
路由等设备				10
光模块	20	200G光模块	0.4	8
IDC建设成本	39KW	密度为一般机柜的2-4倍	8	312
总计				462

资料来源：中科大官网，民生证券研究院

智算中心有望拉动投资额百亿元，高功率机柜 10 万余个。单智算中心算力约 2.69Petaflop/s，根据我们上文测算，单大模型训练算力需求约 280.67 Petaflop/s。我们认为未来“大模型带小模型”将成为趋势，假设算力需求放大 50 倍（多家二三线互联网小厂接入大模型 API 训练），则总算力需求扩至 14033.5 Petaflop/s，此时需要智算中心约 5217 个（合 52170 个机柜）。但由于上述计算以 39kw 功率为基础计算，密度通常为普通高功率机柜 2~4 倍，取放大系数为 2，则需要机柜 10.4 万个。

图27：未来高功率机柜需求约 7.5~10.4 万架



资料来源：中科大官网，英伟达官网，钛媒体，民生证券研究院测算

从 ChatGPT 所需要 GPU 数量测算，未来所需要高功率机柜近 7.5 万个。根据钛媒体，GPT 3.5 完全训练需要 A100 芯片近 3 万颗，对应英伟达 DGX A100 服务器近 3750 台。考虑到英伟达 DGX A100 服务器单台最大功率为 6.5kw，假设平均功率为 4kw/台，则 10kw 功率机柜可搭载 2.5 台 AI 服务器。为满足 3750

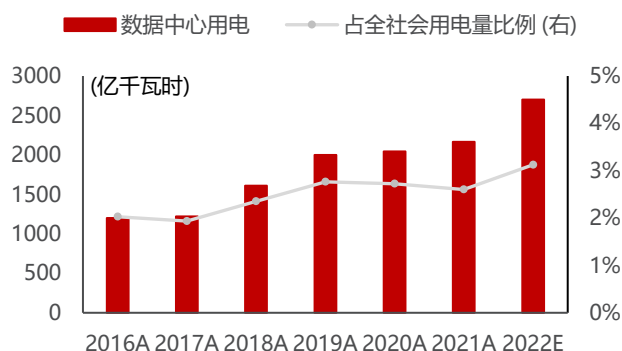
台 AI 服务器，需要建设机柜 1500 个，假设算力需求扩至 50 倍，共需要高功率机柜约 7.5 万个。

3 温控：解决高功率算力池下的“阿喀琉斯之踵”

AI 大趋势下，数据中心耗电成为“阿喀琉斯之踵”。以电量=功率*时间计算，智算中心的运行取决于 IT 设备提供服务的重要程度。大中型数据中心大多配有双路市电+后备柴发作为保障，IT 设备的运行时不间断。根据英伟达，其 DGX-1 服务器功率约 3.5kw，全年功耗=3.5*360*24=30240kwh (1kWh=1 度电)。根据国家能源局统计，2022 年城乡居民生活用电量约 13366 亿 kwh，年末全国总人口约 14.1 亿人，平均每人每年用电 948kwh，换算**每台 AI 服务器每年接近 32 个居民全年耗电量**。根据我们前文测算，大模型训练需要近 3750 台 AI 服务器，而数据中心的总能耗不单单来自于 IT 设备，为 IT 设备提供支撑的电力、制冷等也会消耗电力。

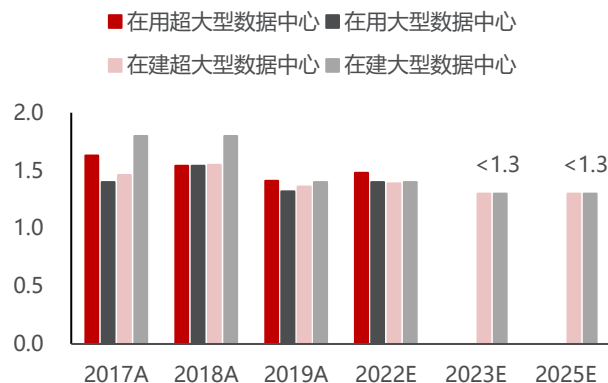
短期维度看，数据中心耗电量快速增长同 PUE 指标是主要矛盾。根据国家能源局统计预测，2022 年我国数据中心总耗电量约 2700 亿千瓦时，同比增长约 24.7%，占全社会用电量比重近 3.1% (较 2022 年提升 0.5pct)。在能耗变高的同时，在用大型数据中心 PUE 也存在较大提升空间。根据《东数西算下新型算力基础设施发展白皮书》统计预测，2022 年我国在建超大型数据中心平均 PUE 需 <1.4，23 年需 <1.3。

图28：我国数据中心电量占全社会电量比重持续上升



资料来源：国家能源局，民生证券研究院

图29：我国数据中心 PUE 情况



资料来源：中国智能计算产业联盟，民生证券研究院

“10℃法则”——数据中心温度与平均无故障运行时间 (MTBF) 的关系。对于现代电子元器件而言，环境温度每提高 10℃，元器件寿命就会降低 30%~50%，影响小的也基本都在 10% 以上。数据中心的计算机服务器、交换机、存储器等 IT 设备的集成度越来越高，精密性也越来越高，使数据中心机房的空调负荷特点更加显著，表现为热负荷大、湿负荷小、单位体积发热量越来越大。

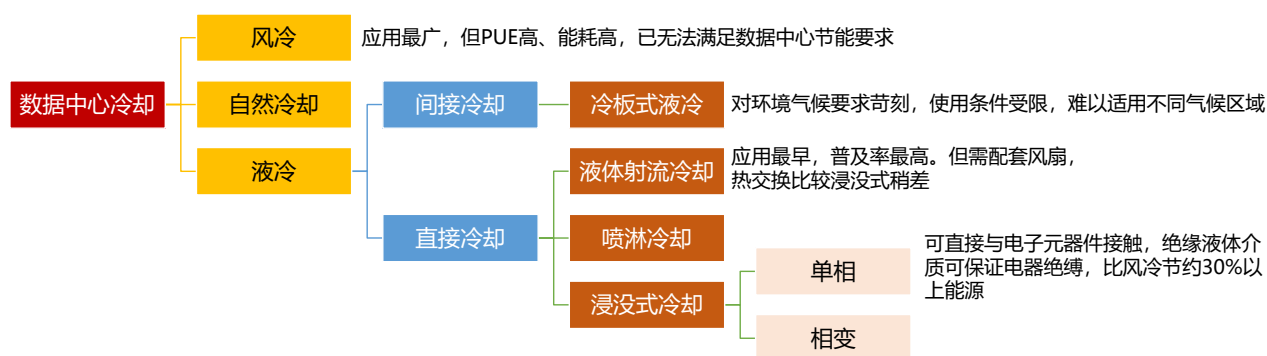
根据《新基建：数据中心创新之路》¹，数据中心计算设备热负荷可根据公式 $Q = P \times \eta_1 \eta_2 \eta_3 (kw)$ 计算，其中 η_1 为同时使用系数，可按主机 $\eta_1 = 1.0$ ； η_2 为利用系数，通常取值 0.8~1.0； η_3 为负荷工作均匀系数，取值 0.8~1.0。根据上文测

¹ 郭亮. 新基建：数据中心创新之路 [M]. 北京：人民邮电出版社，2020.

算，训练大模型生态远期所需机柜约 10.4 万个，合计 A100 GPU 约 83.2 万个，计算设备热负荷 $Q_1 = 83.2 \times 4\text{kw/台} = 332.8 \times 10^4\text{kw}$ ，约等于总制冷量需求。

数据中心制冷主要采取风冷+液冷，液冷有望成为大数据中心主流方案。根据 IDC 圈，液冷的冷却能力较风冷而言高 1000~3000 倍，且液冷对环境要求更低，适应性更强。根据 Dell 'Oro 统计预测，2022 年-2027 年数据中心基础设施 (DCPI) 收入 CAGR 有望保持 8%，超 350 亿美元。且随着英特尔、英伟达、AMD 等处理器与加速器性能的提升，数据中心机架功率密度会再度提升。根据 Dell 'Oro 预测，直接液冷和浸入式冷却(单相和两相)预计将大幅增长，到 2027 年将达到 17 亿美元，占热管理收入的 24%。

图30：数据中心冷却技术分类



资料来源：CCDC，民生证券研究院

表4：单相浸没式液冷技术具有较好的应用前景

	传统风冷	冷板式液冷	单相浸没式液冷	相变浸没式液冷
散热能力	★★	★★★	★★★★	★★★★★
年均PUE值	1.5~1.9	1.2~1.3	1.08~1.12	1.08~1.12
噪音	噪音大	风扇减小，少量发热源仍需风冷散热	液冷无噪音	液冷无噪音
承重要求	8KN/m ² ~10KN/m ²	≥12KN/m ² ~	≥15KN/m ² ~	≥15KN/m ² ~
技术成熟性	★★★★	★★★	★★★	★★
建设成本	★★	★★★	★★★	★★★★★
可维护性	★★★★★	★★★★	★★★	★★
单机柜高密度支持	<10kw; >15kw后成本上升	15kw~100kw	50kw~100kw+	50kw~100kw+

资料来源：CDCC，民生证券研究院

4 重点公司

4.1 奥飞数据

下游服务客户覆盖广，提供全方位数据中心解决方案。公司已经成为国内较有影响力的专业 IDC 服务商，通过多年的精细化运营和稳步发展，现已接入中国电信、中国联通、中国移动、教育网等全国性骨干网络，搭建了一个安全、稳定、高效的智能网络平台，服务客户覆盖短视频、网络游戏、直播、门户网站等多个行业，积累了良好的商业信誉，获得了市场的广泛认可。公司拥有较为完善的基础网络设施，服务种类齐全，涵盖 IDC 服务以及网络接入、数据同步、网络数据分析、网络入侵检测、网络安全防护、智能 DNS、数据存储和备份等增值服务，能够为客户提供全方位的互联网数据中心解决方案。

国内核心区域已有规模优势，抓紧海外市场资源布局。截至 2022 年 6 月末，公司在北京、广州、深圳、海口、南宁、廊坊拥有 12 个自建自营的数据中心，可用机柜约为 22,000 个，同时还有多个数据中心项目正在扩容和建设中，在核心区域已经具有一定的规模优势；同时公司数据中心布局主要在一线城市和一线周边城市，以及部分战略节点城市，对于优质客户有较强的吸引力，同时利用公司原有的网络资源和组网能力，更可满足客户全国布点需求。同时公司继续抓紧海外市场资源布局，成效显著。奥飞国际网络以香港为核心，通过搭建海缆资源网络，互联了公司在东南亚、欧洲和美洲节点，与全球各互联网热点区域主流运营商实现 IP 互联和资源覆盖，用户国际业务提供稳定可靠安全的 IDC 服务，业务范围也进一步扩大。

4.2 光环新网

行业深耕二十载，品牌知名度高。公司在北京、上海等一线城市及周边拥有丰富的品质数据中心资源，凭借领先的技术水平、优质的产品服务和市场影响力，为客户提供安全高效、标准规范、绿色可持续发展的互联网基础设施服务。同时大力发展云计算业务，利用亚马逊云科技国际领先的云服务技术，依托光环有云、无双科技等子公司，实现为用户提供从 IAAS 到 PAAS、SAAS 多层级的多云生态服务。

持续优化技术方案，节能减排降本增效。公司继续践行建设绿色数据中心理念，持续优化节能减排工作。公司数据中心已陆续完成与政府能耗在线监测平台对接，积极配合相关部门进行碳排放情况核查，同时根据掌握的碳排放数据，设定科学的减排目标、配额履约、管理碳资源。另一方面，公司持续部署新能源设施，挖掘节能潜力，从数据中心运营角度设计减排方案，对机房制冷系统进行升级，采用高效节能设备提高制冷效率，降低 PUE。酒仙桥机房、亦庄中金云网机房通过改造制冷系统、更换冷却塔、冷机变频改造等方式降低 PUE 水平。公司积

极与上下游合作伙伴共同研究探讨安全、可靠的储能技术方案应用，寻求在波峰波谷电价差较大的地区施行储能运营方案降低电价成本，未来在满足自身数据中心的储能需求外，实现向工业园区内的其他企业进行供能的可行性方案。

投资建议：我们看好“东数西算”+AI 大参数模型背景下，下游企业需求复苏。我们认为随着公司多地数据中心扩建平稳落地，多元的客户结构也有望增强公司收入规模化效应，并带来更平稳的经营性现金流。预计 22 年~24 年公司收入 72.4/76.9/80.6 亿元，EBITDA 13.4/15.9/19.2 亿元，对应 4 月 10 日收盘价 EV/EBITDA 19/16/13x，归母净利润-8.9/6.0/7.7 亿元，23 年~24 年归母净利润对应 4 月 10 日收盘价 P/E 38/30x。维持“推荐”评级。

风险提示：IDC 市场竞争加剧；云计算需求不及预期。

表5：光环新网盈利预测与财务指标

项目/年度	2021A	2022E	2023E	2024E
营业收入（百万元）	7,700	7,239	7,689	8,060
增长率（%）	3.0	-6.0	6.2	4.8
归属母公司股东净利润（百万元）	836	-893	598	766
增长率（%）	-8.4	/	扭亏为盈	28.0
每股收益（元）	0.47	-0.50	0.33	0.43
PE（现价）	28	/	38	30
PB	1.8	1.9	1.8	1.7

资料来源：，民生证券研究院预测；（注：股价为 2023 年 04 月 10 日收盘价）

4.3 数据港

手握核心资源，聚焦算力高需求地区。分业务看，2022 年 IDC 服务业、IDC 解决方案、云销售分别实现收入 14.5 亿元/281 万元/27 万元。分地区看，长三角、大湾区、京津冀收入占比 35.5%/21.3%/40.0%，毛利率分别为 11.1%/37.7%/38.1%。目前公司已建成 35 座数据中心，运营 IT 兆瓦数达 371MW，按 5KW 标准机柜折算合计约 7.42 万个标准柜。根据中国信通院联合发布的《中国第三方数据中心运营商分析报告（2022 年）》，数据港总体规模已跃升至行业第二。公司紧跟国家“东数西算”政策，在东西部国家枢纽节点均有布局，我们认为随着公司未来廊坊、闵行等项目建设完成，整体规模将得到进一步提升，客户群体更趋于多元化。

泛在算力互联有望增多，算力网络创新加速。金融、交通、教育、工业等行业对泛在算力场景的需求不断增长，用户需要随时随地可接入的更优惠、更高质的算力服务，基于算力泛在使用和交易需求衍生的算力网络成为下一代算力基础设施发展的重点。**中国联通**在全国多地开展算力网络建设试点，通过算力网络业务链、网络切片、资源感知实现算力资源调度和感知，形成云网安一体服务。**中国电信**在算力网络建设过程中重点强调“云”为核心，侧重网络、算力和存储三大资源相互融合，推进天翼云持续升级，实现天翼云节点和天翼边缘节点统一

管理调度，云网融合、云边协同是电信云的优势。**中国移动**重点改造底层算力基础设施，基于 X86 和 ARM 架构打造通用算力网络，同时基于 GPU、ASIC 不断丰富智能算力。我们认为，我国正处于各行业数字化转型的加速期，以数据中心为代表的数字基础设施应用场景仍将进一步扩大，数据中心产业将迎来更大机遇，发展前景将更为广阔。

投资建议：我们看好“东数西算”+AI 大参数模型背景下，下游企业需求复苏。我们认为随着公司多地数据中心扩建平稳落地，多元的客户结构也有望增强公司收入规模化效应，并带来更平稳的经营性现金流。预计 23 年~25 年公司收入 16.6/18.5/19.9 亿元，EBITDA 11.0/13.3/15.2 亿元，对应 4 月 10 日收盘价 EV/EBITDA 15/12/11x，归母净利润 1.6/2.4/2.6 亿元，对应 4 月 10 日收盘价 P/E 74/51/46x。维持“推荐”评级。

风险提示：国内云厂商收入增速放缓，新客户拓展受阻风险。

表6：数据港盈利预测与财务指标

项目/年度	2022A	2023E	2024E	2025E
营业收入（百万元）	1,455	1,655	1,848	1,993
增长率（%）	16.9	13.7	11.7	7.9
归属母公司股东净利润（百万元）	115	162	235	260
增长率（%）	-5.4	41.3	45.0	10.6
每股收益（元）	0.35	0.49	0.72	0.79
PE（现价）	104	74	51	46
PB	4.0	3.8	3.5	3.3

资料来源：，民生证券研究院预测；（注：股价为 2023 年 04 月 10 日收盘价）

4.4 英维克

公司发布 2022 年度业绩。2023 年 4 月 10 日，公司发布 2022 年度业绩报告，全年实现收入 29.2 亿元，同比增长 31.2%，归母净利润 2.8 亿元，同比增长 36.7%。单季度看，4Q22 公司实现收入 14.4 亿元，同比增长 98.5%，归母净利润 1.6 亿元，同比增长 204.7%。

液冷技术进展顺利。2022 年公司机房温控节能、机柜温控产品分别实现收入 14.4/11.0 亿元，同比分别增长。公司的机房温控节能产品主要针对数据中心、算力设备、通信机房、高精度实验室等领域的房间级专用温控节能。在算力设备以及数据中心机房的高热密度趋势和高能效散热要求的双重推动下，液冷技术加速导入。公司对冷板、浸没等液冷技术平台长期投入。公司已推出针对算力设备和数据中心的 Coolinside 液冷机柜及全链条液冷解决方案。我们认为 AI 等新的应用场景出现，对算力设备和承载算力的数据中心均提出更高的要求，未来算力设备、数据中心机柜的热密度都将显著提高。公司 2022 年已部署冷板式液冷，22 年来自数据中心机房及算力设备的液冷技术相关营业收入在报告期内约为上一年的 5 倍左右。

投资建议：我们看好“东数西算”+AI 大参数模型背景下，下游企业需求复苏。我们认为随着公司多地数据中心扩建平稳落地，多元的客户结构也有望增强公司收入规模化效应，并带来更平稳的经营性现金流。预计 23 年~25 年公司收入 38.6/49.4/61.6 亿元，归母净利润 3.7/4.7/5.9 亿元，对应 4 月 10 日收盘价 P/E 41/33/26x。维持“推荐”评级。

风险提示：下游 5G 或数据中心建设不及预期；行业竞争加剧；原材料大幅涨价导致毛利率承压。

表7：英维克盈利预测与财务指标

项目/年度	2022A	2023E	2024E	2025E
营业收入（百万元）	2,923	3,860	4,937	6,164
增长率（%）	31.2	32.0	27.9	24.9
归属母公司股东净利润（百万元）	280	366	464	588
增长率（%）	36.7	30.8	26.5	26.8
每股收益（元）	0.64	0.84	1.07	1.35
PE（现价）	54	41	33	26
PB	7.2	6.3	5.5	4.7

资料来源：，民生证券研究院预测；（注：股价为 2023 年 04 月 10 日收盘价）

4.5 申菱环境

服务行业主流客户，数据中心业务快速增长。公司产品已广泛服务于中国移动、电信、联通、华为、腾讯、阿里巴巴、百度、美团、字节跳动、快手、京东、秦淮数据、世纪互联、万国数据、曙光、国防科大、浪潮等众多知名客户的数据中心。截至 1H22 公司数据服务空调营收同比增长 42.44%，同时在该业务领域的工程服务及材料等其他产品增长约 386.22%，整体数据服务领域业务增长约 72.61%，主要是由于 H 公司和其他数据服务客户的业务都实现了较好的增长。间接蒸发冷却产品营收同比增长 63.62%，液冷温控系统营收同比增长 110.56%，并实现了海外大型液冷项目规模化的交付。

持续研发创新，构筑产品壁垒。公司的整体技术水平先进，尤其在水电领域（高效节能水电空调）、轨道交通领域（隧道嵌装式全工况高效能空调系统）、信息通信领域（数据中心液/气双通道精准高效制冷系统关键技术及应用）、核电领域（三代核电站宽环境温度风冷冷水机组）的技术达到了国际领先或国际先进水平。公司中国移动南方基地液/气双通道制冷项目有效解决了高热高湿的南方地区数据中心制冷效能瓶颈问题，全年平均 PUE < 1.2，数据中心性能良好，运营平稳可靠。

4.6 依米康

数字基础设施全生命周期解决方案服务商。公司专注精密环境可靠、高效、节能热管理技术和应用，为数据中心和精密环境提供关键制冷设备。作为边缘计

算市场的主流增量产品，公司推进模块化数据中心产品方案的快速演进，上半年智能一体柜及微模块产品均顺利完成产品升级换代，进一步提升了产品的客户体验及市场竞争力。公司积极开拓非数据中心市场应用领域，开发第二曲线：战略性切入轨道交通温控行业，完成了针对地铁冷源产品的设计交付及样机验收；针对快速上行的国内外储能市场，进一步扩充丰富了储能空调产品家族，积极推进完成多系列产品的 CE/UL 认证；依托液冷技术平台衍生的创新型储能液冷产品，也处于快速开发推进状态。

入选国家绿色数据中心名单，发展水冷技术。在低碳绿色制冷方面，依米康制定了面向大型 IDC 的空调节能演进方案，从工程应用成熟的传统房间冷冻水方案，到延续主流暖通方案的风墙方案，再到工程预制化、充分利用自然冷源的间接蒸发冷方案、整体式氟泵方案，以及创造极致能效的液冷方案，依米康持续攻坚高效制冷方案升级，推动制冷系统能耗逐步降低，成就数据中心理想 PUE 值。

5 风险提示

1) 市场竞争加剧。数据中心属于互联网的基础设施，随着互联网的发展特别是 5G、大数据、云计算的到来，市场对数据中心的需求将进一步增加，与此同时，也会吸引更多的资本进入该行业，市场竞争将更加激烈。市场竞争的加剧，一方面使得公司面临市场份额被竞争对手抢夺的风险；另一方面，竞争加剧可能导致 IDC 行业的整体毛利率下降，从而给公司经营业绩带来不利影响。

2) 技术更新风险。数据中心行业的技术演进特征主要表现为优先关注安全可靠、稳定的运营服务能力，新技术的应用成为主流通常需要较长的成熟期。在合同有效期内，新技术的采用导致部分数据中心服务能力的过时或滞销的可能性较低。

3) 电价波动风险。电力成本占数据中心成本的比例较高，其价格变动对业务毛利率的影响较大。若电力价格大幅上涨，将直接增加行业内公司的成本压力。如公司不能同步通过服务价格的调整覆盖成本上涨的风险，将给公司的盈利能力造成不利影响。

插图目录

图 1: 引入数学知识引擎后, ChatGPT 具备较为精确的计算能力	3
图 2: ChatGPT 可线上链接多个第三方库资源	4
图 3: 我国 AI 软件及应用市场规模	5
图 4: 我国数据量规模全球领先	5
图 5: 百度文心大模型全景图	6
图 6: 华为 ModelArts 平台	7
图 7: 太极-HCF ToolKit 加速大模型落地	7
图 8: 太极-HCF ToolKit 加速大模型落地	8
图 9: 阿里巴巴通义大模型落地	8
图 10: OpenAI 官网访问量	10
图 11: ChatGPT 日活	10
图 12: 主流 AI 大模型训练运算量	12
图 13: 英伟达 DGX A100 服务器架构	13
图 14: AI 服务器出货量稳步增长	14
图 15: 全球 A 服务器采购量占比	14
图 16: 百度文心一言访问频次	14
图 17: Midjourney 访问频次	14
图 18: 未来国内 AI 模型有望呈现“大模型带小模型”模式	15
图 19: 海外 IDC 公司收入及增速	15
图 20: 万国数据、秦淮数据、世纪互联上架率回升	15
图 21: 我国数据中心市场规模	16
图 22: 我国数据中心机架规模	16
图 23: 上海传统 IDC 业务市场空间	17
图 24: 北京及周边地区 IDC 机柜供给快速增长	17
图 25: 中国科大瀚海 20 超级计算系统配置	18
图 26: 智算中心投资成本	19
图 27: 未来高功率机柜需求约 7.5~10.4 万架	19
图 28: 我国数据中心电量占全社会电量比重持续上升	21
图 29: 我国数据中心 PUE 情况	21
图 30: 数据中心冷却技术分类	22

表格目录

重点公司盈利预测、估值与评级	1
表 1: NLP 模型参数规模及训练所需算力	11
表 2: CV 模型参数规模及训练所需算力	11
表 3: GPU、ASIC、FPGA 对比	12
表 4: 单相浸没式液冷技术具有较好的应用前景	22
表 5: 光环新网盈利预测与财务指标	24
表 6: 数据港盈利预测与财务指标	25
表 7: 英维克盈利预测与财务指标	26

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并登记为注册分析师，基于认真审慎的工作态度、专业严谨的研究方法与分析逻辑得出研究结论，独立、客观地出具本报告，并对本报告的内容和观点负责。本报告清晰地反映了研究人员的研究观点，结论不受任何第三方的授意、影响，研究人员不曾因、不因、也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

评级说明

投资建议评级标准		评级	说明
以报告发布日后的 12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。	公司评级	推荐	相对基准指数涨幅 15%以上
		谨慎推荐	相对基准指数涨幅 5% ~ 15%之间
		中性	相对基准指数涨幅-5% ~ 5%之间
		回避	相对基准指数跌幅 5%以上
	行业评级	推荐	相对基准指数涨幅 5%以上
		中性	相对基准指数涨幅-5% ~ 5%之间
		回避	相对基准指数跌幅 5%以上

免责声明

民生证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。

本公司境内客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅为参考之用，并不构成对客户的投资建议，不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，客户应当充分考虑自身特定状况，不应单纯依靠本报告所载的内容而取代个人的独立判断。在任何情况下，本公司不对任何人因使用本报告中的任何内容而导致的任何可能的损失负任何责任。

本报告是基于已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及预测仅反映本公司于发布本报告当日的判断，且预测方法及结果存在一定程度局限性。在不同时期，本公司可发出与本报告所刊载的意见、预测不一致的报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问、咨询服务等相关服务，本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

若本公司以外的金融机构发送本报告，则由该金融机构独自为此发送行为负责。该机构的客户应联系该机构以交易本报告提及的证券或要求获悉更详细的信息。本报告不构成本公司向发送本报告金融机构之客户提供的投资建议。本公司不会因任何机构或个人从其他机构获得本报告而将其视为本公司客户。

本报告的版权仅归本公司所有，未经书面许可，任何机构或个人不得以任何形式、任何目的进行翻版、转载、发表、篡改或引用。所有在本报告中使用的商标、服务标识及标记，除非另有说明，均为本公司的商标、服务标识及标记。本公司版权所有并保留一切权利。

民生证券研究院：

上海：上海市浦东新区浦明路 8 号财富金融广场 1 幢 5F；200120

北京：北京市东城区建国门内大街 28 号民生金融中心 A 座 18 层；100005

深圳：广东省深圳市福田区益田路 6001 号太平金融大厦 32 层 05 单元；518026