

蓝耘科技 (871169.NQ) 专注 GPU 算力服务, 立足 AI 模型算法驱动盈利增长

2024 年 05 月 21 日

——新三板公司研究报告

日期	2024/5/20
当前股价(元)	11.98
一年最高最低(元)	14.01/4.95
总市值(亿元)	11.07
流通市值(亿元)	8.79
总股本(亿股)	0.92
流通股本(亿股)	0.73
近 3 个月换手率(%)	7.95

诸海滨 (分析师)	赵昊 (分析师)
zhuhaibin@kysec.cn	zhaohao@kysec.cn
证书编号: S0790522080007	证书编号: S0790522080002

北交所研究团队

● 聚焦 GPU 算力服务, 盈利整体呈增长态势

蓝耘科技是一家专注于为企业级规模的 GPU 加速工作负载提供算力云服务的公司。公司的主营业务为 IT 系统集成业务、技术服务业务、GPU 算力云服务业务, 2023 年公司算力云解决方案、GPU 算力云服务业务稳步推进, 其中算力云解决方案收入占比为 66.14%, GPU 算力云服务收入占比从 2022 年的 21.16% 稳步提升至 33.86%, 公司业务结构持续优化, GPU 算力云服务业务营收规模保持稳健增长, 经营质效稳步提升, 且毛利率达到 39.63%。2023 年实现营收 4.09 亿元、归母净利润 5128.17 万元; 2020-2023 年 3 年营收 CAGR 为 5.54%, 归母净利润 CAGR 达到 15.75%, 毛利率/净利率分别为 25.79%/12.38%。

● AI 驱动算力云发展加速, IT 系统集成需求增长

根据 IDC 数据预计 2023 年, 中国人工智能服务器市场规模将达 91 亿美元, 同比增长 82.5%, 2027 年将达到 134 亿美元, 五年年复合增长率为 21.8%。2022 年中国通用算力规模达 54.5EFLOPS, 预计到 2027 年通用算力规模将达到 117.3EFLOPS。预计 2022-2027 年期间, 中国智能算力规模年复合增长率达 33.9%, 同期通用算力规模年复合增长率为 16.6%。智算需求增长带动 AI 云计算服务, 百度、腾讯、阿里等服务商通过 aaS (as a Service) 服务提供 AI 平台和 AI 服务, 2022 年国内 AI 公有云服务市场增长 80.6% 至 79.7 亿元逆势高增, 其中计算机视觉占据约一半市场份额, 同时私有云也正快速发展。

● 立足 AI 模型算法研发, 产业合作加速算力资源建设

蓝耘科技研发费用整体呈现增长态势, 自 2018 年的 288 万元上升至 2022 年的 1700 万元以上, 2022 年、2023 年研发费用分别为 1738.51 万元、1595.56 万元, 研发费用率分别为 4.41%、3.91%。截至 2023 年 12 月 31 日, 公司正在申请专利 14 项, 目前拥有计算机软件著作权 125 项。2023 年, 蓝耘科技共有研发项目 10 个, 主要围绕 AI 模型算法开发、AI 模型预训练、AI 推理等场景的平台和工具的研发。并且开始在平台之上开始构建 AIGC 等相关应用。为用户从开发、训练到应用部署提供全流程支撑。研发的平台和工具降低了 AI 技术人员的使用门槛, 也降低了应用落地的门槛, 提高了应用部署的效率。提升了公司在行业内的竞争力。

● 蓝耘科技同行可比公司 PE2023 均值为 61.34X

蓝耘科技同行可比公司 PE2023 均值为 61.34X。蓝耘科技专注于为企业级规模的 GPU 加速工作负载提供算力云服务, 服务于高校科研、生命科学、人工智能、自动驾驶、工业设计、智慧城市等领域, 目前与产业方广泛合作并通过融资租赁等方式加速 AI 时代布局, 有望拓展算力版图, 不断提升产业地位, 在 LLM 等需求驱动下前景广阔。目前公司可比公司 PE 2023 均值 61.34X, 蓝耘科技 PE 2023 为 21.56X, 估值相对较低, 同时毛利率高于可比公司均值, 具备一定可比优势。

● 风险提示: 算力需求波动、产业政策变化、GPU 渠道风险

相关研究报告

《蓝耘科技: AI 大模型时代算力综合服务商, GPU 采购加速产业布局——新三板公司研究报告》-2023.10.14

目 录

1、 业务：聚焦 GPU 算力服务，盈利整体呈增长态势.....	4
1.1、 产品：GPU 算力业务占比持续提升，2023 年毛利率达 39.63%	4
1.2、 商业模式：自建智算云平台，服务高性能计算需求行业龙头企业.....	5
1.3、 财务：2023 年实现营收 4.09 亿元，归母净利润 5128.17 万元.....	7
2、 行业：AI 驱动算力云发展加速，IT 系统集成需求增长.....	9
2.1、 GPU 算力云：我国 AI 服务器市场达 91 亿美元，智能算力需求高景气.....	9
2.2、 IT 系统集成：市场结构变迁，AI、边缘等服务器增量持续驱动.....	13
3、 看点：立足 AI 模型算法研发，产业合作加速算力资源建设	16
4、 估值对比：同行可比公司 PE2023 均值为 61.34X.....	21
5、 风险提示.....	21

图表目录

图 1： GPU 算力云服务业务营收规模保持稳健增长，经营质效稳步提升.....	4
图 2： 2023 年算力云服务收入占比稳步提升至 33.86%	5
图 3： 2023 年 GPU 算力云服务业务营收达到 1.38 亿元.....	5
图 4： 蓝耘科技集成软硬件以提供 GPU 算力云服务和系统集成服务.....	6
图 5： 2023 年实现营收 4.09 亿元	7
图 6： 2023 年实现归母净利润 5128.17 万元	7
图 7： 2023 年毛利率上升至 25.79%.....	7
图 8： 2021-2022 算力云服务毛利率远高于系统集成业务	8
图 9： 2023 年 GPU 算力云服务毛利为 39.63%.....	8
图 10： 2023 年公司业务规模扩大，期间费用率为 12.80%.....	8
图 11： 2023 年净利率为 12.38%.....	8
图 12： 2025-2030 年中国人工智能技术将在各个应用场景逐步发展成熟	9
图 13： 预计 2022-2027 年期间，中国智能算力规模年复合增长率达 33.9%.....	10
图 14： 智能算力是驱动计算机视觉产业链发展的基础.....	11
图 15： 2021 年 GPU 应用于中国 89% 的计算加速场景	11
图 16： 预计 AI 服务器用于推理计算比例将提升至 6 成	11
图 17： 2022 年全球智能算力规模增长 25.7%/EFLOPS.....	12
图 18： 2022 年中国智能算力规模增长 41.4%/EFLOPS.....	12
图 19： 2022 年 AI 云服务市场随下游需求逆势增长至 79.7 亿元，互联网巨头主导.....	12
图 20： 服务器系统集成属于数字化基建产业的基础层领域.....	13
图 21： 服务器集成商需要在架构等硬件层的设计经验上有深厚积累.....	14
图 22： 服务器集成市场主要面向 AI、通用、存储、边缘等各类服务器.....	14
图 23： 预计中国 AI 服务器市场 2026 年达 123 亿美元	15
图 24： 预计中国边缘服务器市场 2026 年达 92 亿美元.....	15
图 25： 2023 年研发费用率为 3.91%.....	16
图 26： 蓝耘科技提供具备优势的英伟达 GPU 算力资源解决方案.....	18
图 27： GPU 云计算集群服务于大模型计算，带动公司 GPU 云计算服务需求.....	18
图 28： GPU、CPU 云计算服务可提升生命科学等科研、工程领域效率.....	19
图 29： 蓝耘科技利用裸金属 K8S 部署实现推理服务的性能优势.....	20

图 30: 蓝耘科技携手战略合作伙伴中联数据集团共同打乌兰察布智算中心.....	21
表 1: 2023 年蓝耘科技向前五大客户销售金额为 1.78 亿元.....	5
表 2: 截至 2023 年 12 月 31 日, 公司正在申请专利 14 项.....	16
表 3: 2023 年公司研发支出前五名的研发项目共研发支出 1205.28 万元.....	17
表 4: 同行业可比公司 PE2023 均值为 61.34X.....	21

1、业务：聚焦 GPU 算力服务，盈利整体呈增长态势

蓝耘科技是一家专注于为企业级规模的 GPU 加速工作负载提供算力云服务的公司，基于灵活的基础设施与大规模的 GPU 算力资源，为客户提供开放、高性能、高性价比的算力云服务。

公司成立于 2004 年，总部位于北京，在上海、广州、成都、厦门等地设有分子公司，于 2017 年挂牌新三板。公司拥有专业技术团队，从基础架构建设到云计算，从实施到售后，均有丰富的经验；已成为 国家高新技术企业，北京市专精特新小巨人企业。

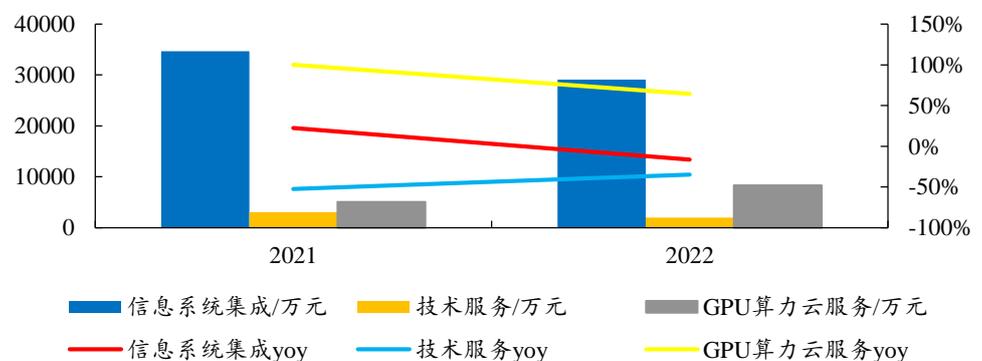
发展至今，蓝耘形成了以 GPU 算力解决方案为基础，GPU 算力云服务为核心的业务版图，为相关高校、科研机构、生命科学、人工智能、自动驾驶、工业设计、智慧城市等领域有高性能计算需求的客户，提供低成本高质量的 GPU 算力云服务，满足客户业务应用和 GPU 算力按需扩展的双重需求。

1.1、产品：GPU 算力业务占比持续提升，2023 年毛利率达 39.63%

公司的主营业务为 IT 系统集成业务、技术服务业务、GPU 算力云服务业务，其中 GPU 算力云服务业务自 2021 年以来高速增长，近年来占比不断上升，2023 年公司算力云解决方案、GPU 算力云服务业务稳步推进，其中算力云解决方案收入占比为 66.14%，GPU 算力云服务收入占比从 2022 年的 21.16% 稳步提升至 33.86%，公司业务结构持续优化，GPU 算力云服务业务营收规模保持稳健增长，经营质效稳步提升，且毛利率达到 39.63%，高于公司传统的系统集成业务。

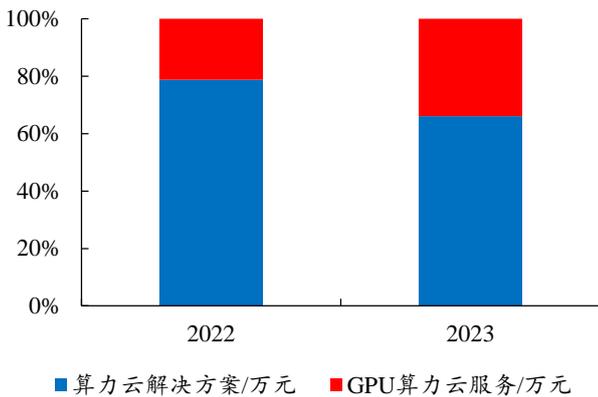
从业务内容来看，IT 系统集成业务主要产品与服务包括企业私有云解决方案、容灾备份解决方案、高性能计算 HPC 解决方案以及分布式存储系统解决方案；技术服务业务主要包括 IT 系统运维服务，软件系统开发服务；GPU 算力云服务主要是基于蓝耘 GPU 云 IAAS 基础设施，根据客户业务对 CPU 及 GPU 算力的需求，通过算力云平台为客户提供远程 CPU 或 GPU 多核并行算力，客户按需购买算力核时或卡时预付费给蓝耘科技。蓝耘算力云平台可以向用户提供虚拟化的计算资源、存储资源、网络资源等，这些资源能够根据用户的需求动态的分配。

图1：GPU 算力云服务业务营收规模保持稳健增长，经营质效稳步提升



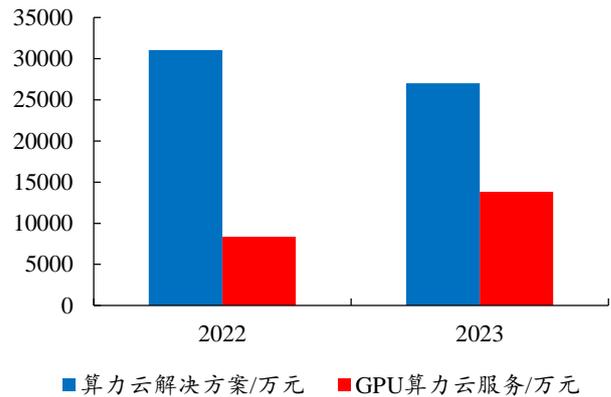
数据来源：Wind、公司财报、开源证券研究所

图2：2023年算力云服务收入占比稳步提升至33.86%



数据来源：Wind、公司财报、开源证券研究所

图3：2023年GPU算力云服务业务营收达到1.38亿元



数据来源：Wind、公司财报、开源证券研究所

1.2、商业模式：自建智算云平台，服务高性能计算需求行业龙头企业

蓝耘算力云业务以 **IaaS+Pass** 为基础架构，依托数据中心，交付给用户高效、灵活、可扩展的基础设施资源，用户无需购买、维护硬件设备和相关系统软件，就可以直接在蓝耘云平台上构建自己的平台和应用。蓝耘算力云平台可以向用户提供虚拟化的计算资源、存储资源、网络资源等。这些资源能够根据用户的需求动态的分配。支撑该服务的技术体系包括虚拟化技术和相关资源动态管理与调度技术。

蓝耘 GPU 算力云业务目标用户包括个人用户与企业用户，个人用户的算力云服务应用场景包括机器学习、计算机视觉、自然语言处理等 AI 领域的培训与学习，以及高校相关专业学生的毕业论文设计、科研项目基础研究等，涉及目标院系包括计算机、人工智能、物理化学、生命科学、环境科学、地理勘测、航空航天等。企业用户的算力云服务应用场景主要集中在自动驾驶、AI 大模型、生命科学、工业设计以及教科研领域的大型科研项目等。

2022-2023 年，蓝耘科技向前五大客户销售金额分别为 1.46 亿元、1.78 亿元，前五大客户收入占比分别为 37.05%、43.57%。

核心客户方面，公司系统集成业务的核心客户以数字化服务、数据服务、网络通信等领域的服务商和集成商龙头企业为主。2022 年，第一、第四和第五大客户，主要交易内容为服务器（配件）及安装服务，不直接提供算力服务；而 2022 年 GPU 算力云服务的主要客户包括第二大客户（软件和信息技术服务商）、第三大客户（科技推广和应用服务商，致力于开发全球最大的中文搜索引擎之一的 ROI 媒体平台）这类具有较大 GPU 算力加速需求的厂商。

表1：2023年蓝耘科技向前五大客户销售金额为1.78亿元

客户排名	收入/万元	销售占比
2023		
客户 1	4778.76	11.70%
客户 2	4341.58	10.63%
客户 3	3146.11	7.70%
客户 4	2851.65	6.98%

客户排名	收入/万元	销售占比
客户 5	2680.93	6.56%
合计	17799.03	43.57%
2022		
客户 1	4675.75	11.87%
客户 2	2796.88	7.10%
客户 3	2680.93	6.80%
客户 4	2607.19	6.62%
客户 5	1838.46	4.67%
合计	14599.22	37.05%

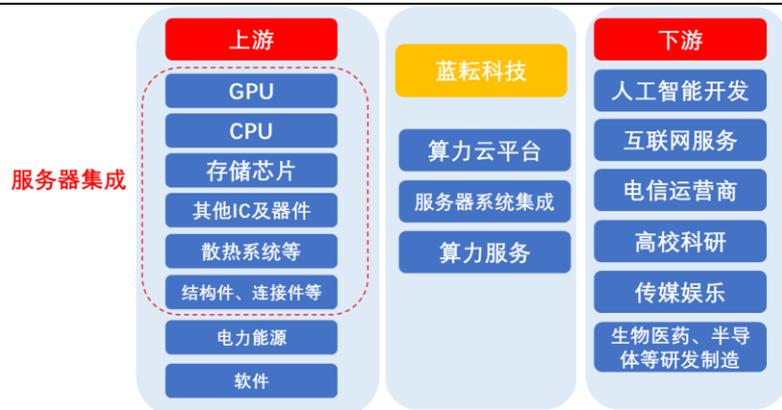
数据来源：公司年报、开源证券研究所

经营模式方面，蓝耘目前正围绕 GPU 算力打造“系统集成及服务”+“算力资源”+“云服务”为一体的综合业务版图。

业务流程来看，算力云服务平台是基于 Kubernetes 原生云设计，专为大规模 GPU 加速工作量身打造，搭载高可配置高可用的英伟达高端 GPU 系列，支持按需灵活计费、轻松调整存储容量及扩展容量，客户可直接线上付费购买算力资源、进行服务器租借并实现定制化服务等等；蓝耘科技以自建云平台为主，需要采购各类应用和基础软件。此外，公司目前虽然系统集成业务收入占比最高，但承接的系统集成业务主要也是 GPU 服务器相关的 IT 系统集成，获取订单后由公司售前及商务人员配合根据客户需求进行拆单和选型，之后根据公司库存情况将需进行采购的部件向供应商下单采购。最后 IT 硬件到达公司仓库或客户现场后，由公司实施人员进行集成，集成部分工作较为简单，主要为软硬件的连接与调试；公司把集成中的服务分块成标准服务，并通过视频或现场指导外包服务团队按标准提供服务的方式完成集成，最后完成客户交付。公司业务过程中向不同客户提供的都是高度相似的类标准化服务，供应链体系和商务体系较为完善，在人员配备(1-2 人)上无需较大投入。

采购项目方面，公司采购的主要为 IT 系统集成业务和 GPU 算力云服务业务项目所需的计算机信息系统相关的软件、硬件产品（如服务器、交换机、小型机管理控制台）、技术服务等；主要供应商包括戴尔、英迈电子、北京同方信息、联想、中铁信息工程等。

图4：蓝耘科技集成软硬件以提供 GPU 算力云服务和系统集成服务

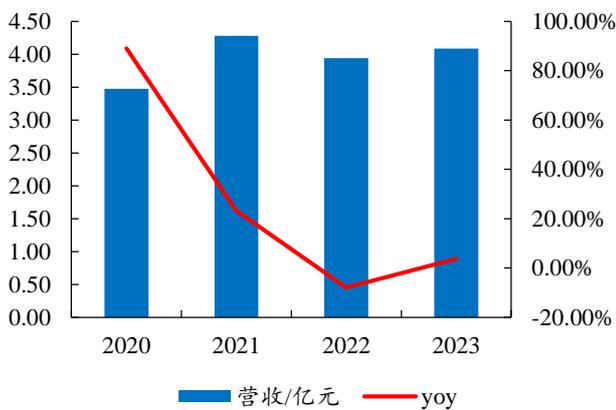


资料来源：公司官网、开源证券研究所

1.3、财务：2023 年实现营收 4.09 亿元，归母净利润 5128.17 万元

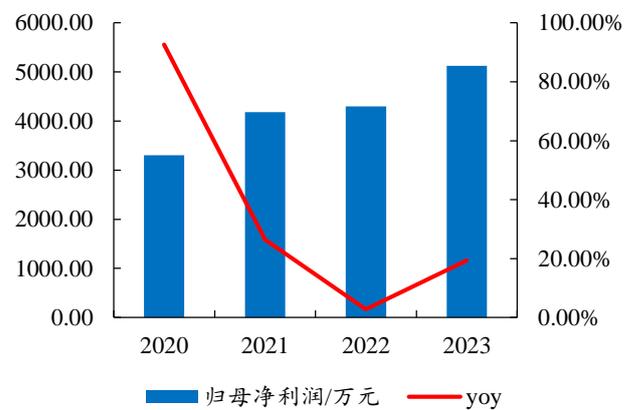
公司营收和盈利规模稳定增长，成长性良好。受益算力建设需求，公司业务快速增长，2023 年实现营收 4.09 亿元、归母净利润 5128.17 万元；2020-2023 年 3 年营收 CAGR 为 5.54%，归母净利润 CAGR 达到 15.75%。公司算力云解决方案、GPU 算力云服务业务稳步推进，蓝耘智算云平台于 2022 年 4 月正式上线，基于蓝耘云基础设施，面向高校、科研院所、企事业单位等高性能计算需求的客户，满足客户业务应用和 GPU 算力按需扩展的双重需求。另外 2024 年 4 月公司签订重大销售合同，合同金额为 1.04 亿元。

图5：2023 年实现营收 4.09 亿元



数据来源：Wind、开源证券研究所

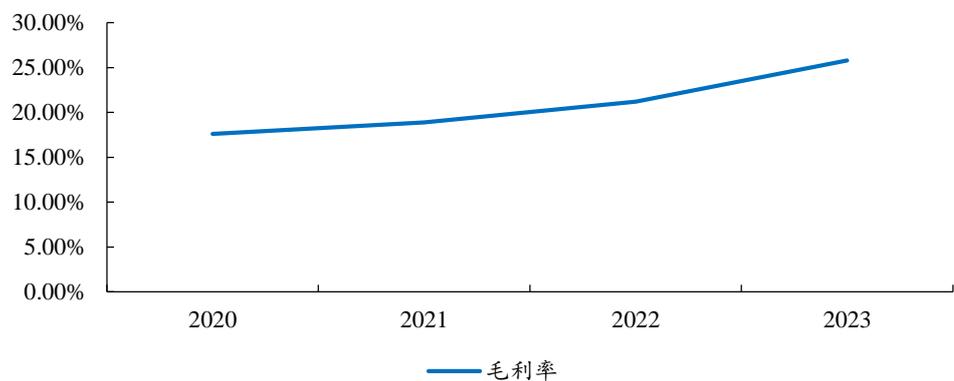
图6：2023 年实现归母净利润 5128.17 万元



数据来源：Wind、开源证券研究所

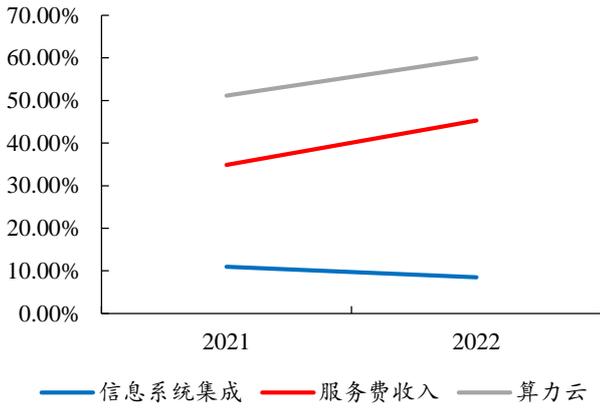
2021 年公司业务开始结构性优化，着力发展高毛利率的 GPU 算力云业务，推动公司盈利能力快速上升，2023 年毛利率上升至 25.79%。

图7：2023 年毛利率上升至 25.79%



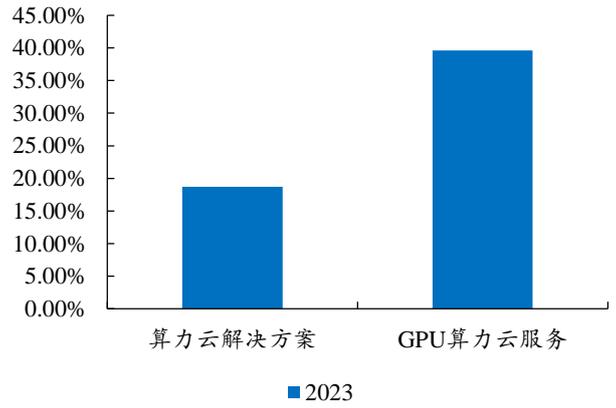
数据来源：Wind、开源证券研究所

图8: 2021-2022 算力云服务毛利率远高于系统集成业务



数据来源: 公司年报、开源证券研究所

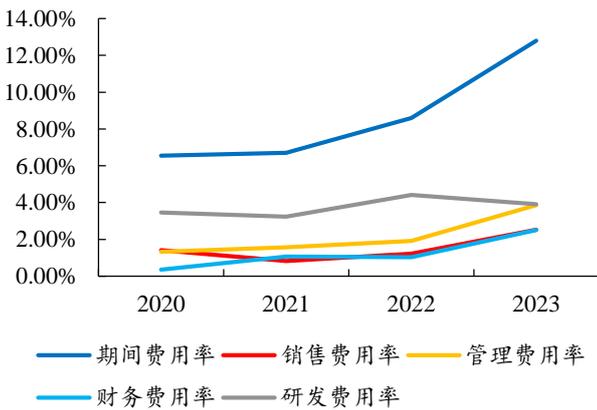
图9: 2023 年 GPU 算力云服务毛利为 39.63%



数据来源: Wind、开源证券研究所

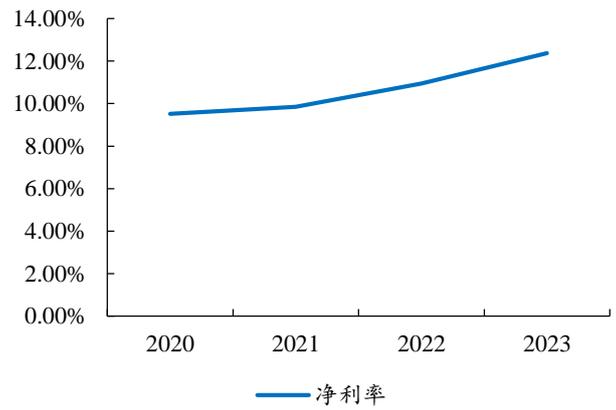
期间费用率呈上升态势,2020-2023 年,期间费用率分别为 6.55%、6.70%、8.59%、12.80%,主要原因为公司业务规模扩大,营销费用及人员薪酬费用上升。净利率呈现上升趋势,2023 年净利率达到 12.38%。

图10: 2023 年公司业务规模扩大,期间费用率为 12.80%



数据来源: Wind、开源证券研究所

图11: 2023 年净利率为 12.38%



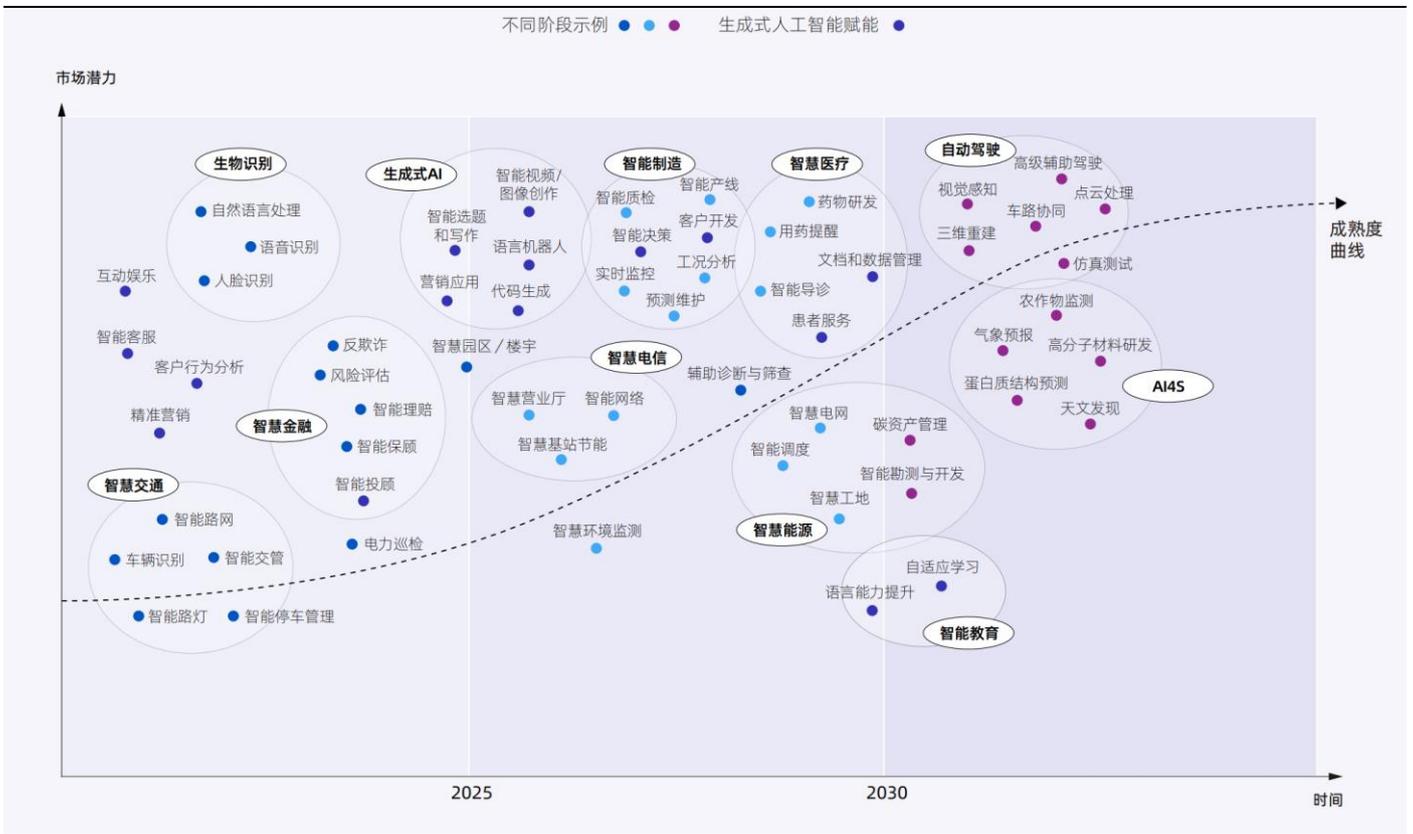
数据来源: Wind、开源证券研究所

2、行业：AI 驱动算力云发展加速，IT 系统集成需求增长

2.1、GPU 算力云：我国 AI 服务器市场达 91 亿美元，智能算力需求高景气

近年来国内数字经济快速发展，云计算行业作为最重要的基础设施产业之一对下游增长起到关键的促进作用，据中国信息通信院发布的《中国算力发展指数白皮书（2022 年）》测算，算力每投入 1 元，将带动 3-4 元的经济产出；而随着人工智能、数字孪生、元宇宙等新兴领域的崛起，各类算力需求规模正快速增长，其中 GPU 算力对应“智能算力”的主要部分，将首先受益发展。智能算力即人工智能算力，由 GPU（图形处理器）、ASIC（专用集成电路）、FPGA（现场可编程逻辑门阵列）、NPU（神经网络处理器）等各类专用芯片承担人工智能场景应用的计算，在专门应用下具有性能更优、能耗更低等优点。

图12：2025-2030 年中国人工智能技术将在各个应用场景逐步发展成熟



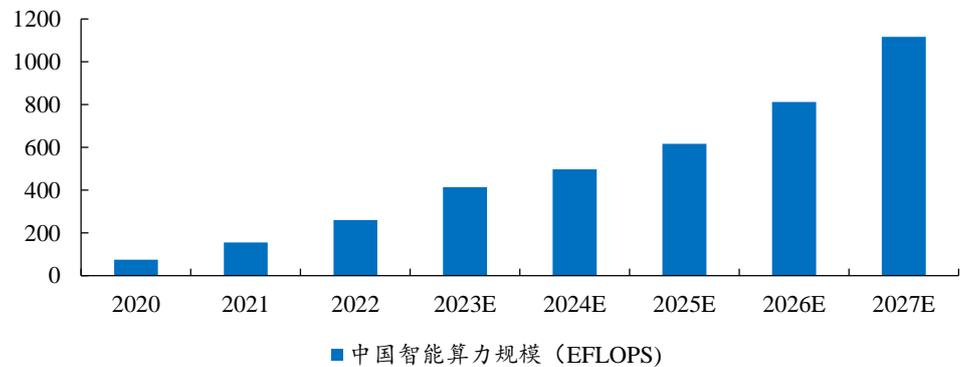
资料来源：《2023-2024 中国人工智能算力发展评估报告》

截至 2022 年年底，我国算力核心产业规模达到 1.80 万亿元，算力总规模达到 180EFLOPS，居全球第二，近 5 年保持 30.00%左右的年均增速，对 GDP 增长的拉动作用显著。随着 AI 人工智能、ChatGPT、OPENAI 等应用不断涌现，智能算力规模急速攀升，具备海量数据并行计算能力、擅长加速计算处理的 GPU 算力需求持续高涨。

据国际数据公司 IDC 与浪潮信息联合发布的《2023-2024 年中国人工智能算力发展评估报告》显示，未来市场需求量也将会实现大幅度上升，预计 2023 年，中国人工智能服务器市场规模将达 91 亿美元，同比增长 82.5%，2027 年将达到 134 亿

美元，五年年复合增长率为 21.8%。2022 年中国通用算力规模达 54.5EFLOPS，预计到 2027 年通用算力规模将达到 117.3EFLOPS。2022 年中国智能算力规模达 259.9 EFLOPS，2023 年将达到 414.1 EFLOPS，预计到 2027 年将达到 1117.4 EFLOPS。预计 2022-2027 年期间，中国智能算力规模年复合增长率达 33.9%，同期通用算力规模年复合增长率为 16.6%。

图13：预计 2022-2027 年期间，中国智能算力规模年复合增长率达 33.9%

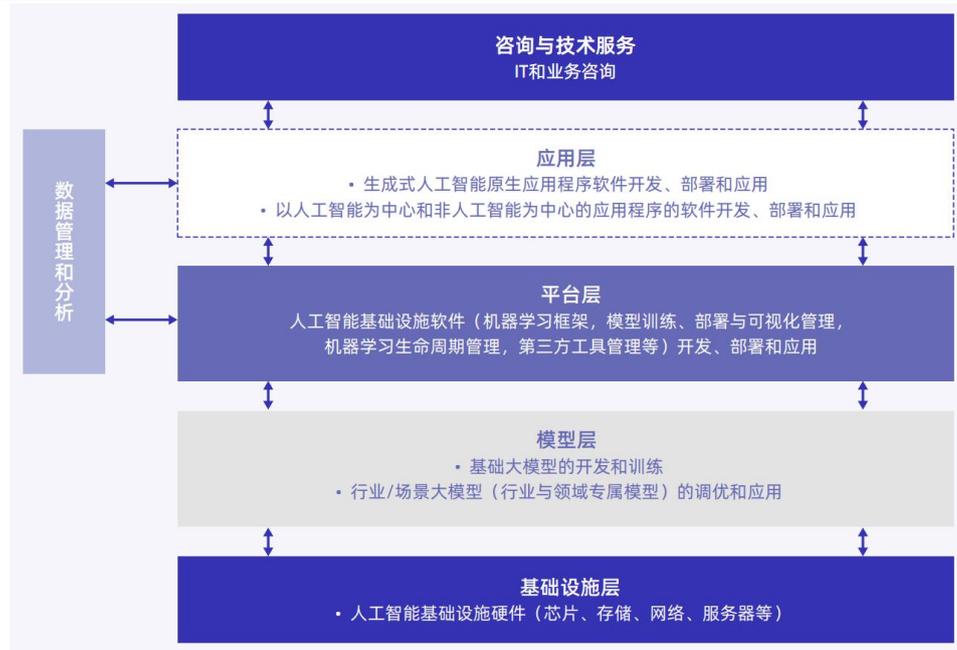


数据来源：IDC 与浪潮信息联合发布《2023-2024 年中国人工智能算力发展评估报告》、开源证券研究所

智能算力在 AI 深度学习中的应用主要体现在以下几个方面：**一是训练模型**，深度学习模型具有复杂的结构和大量的参数，需要大量的计算资源进行训练，智能算力通过 GPU、TPU 等高性能的计算设备，加速深度学习模型的训练过程。**二是推断推理**，智能算力通过高性能计算设备和专门的推理芯片加速深度学习模型的推断过程，提高了模型的实时性和稳定性。**三是模型优化**，通过智能算力可以对模型进行自动化的超参数调优、网络结构搜索、模型剪枝等操作，进一步提高模型的精度和效率。

以**计算机视觉**为例，需要 GPU 加速作为算力基础，实现大量数据训练和 AI 模型训练，比如通过卷积神经网络技术进行图像分类、目标检测和图像分割等任务，从而达成图像分类和目标检测、人脸识别、图像生成和风格迁移、三维重建和虚拟现实、视频分析和行为识别等具体应用。

图14：智能算力是驱动计算机视觉产业链发展的基础

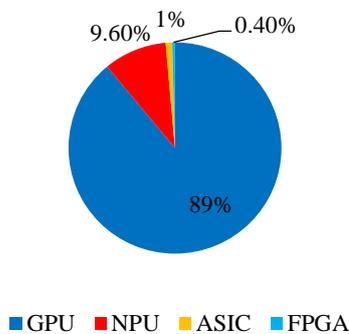


资料来源：IDC、《2023-2024 年中国人工智能算力发展评估报告》

GPU 芯片多用于图形图像处理、复杂的数学计算等场景，可较好支持高度并行的工作负载，常用于数据中心的模型训练，也可以用于边缘侧和端侧的推理工作负载。2021 年中国仍以 GPU 为主实现数据中心计算加速，据 IDC 测算 GPU 市场占有率近 90%，而 ASIC、FPGA、NPU 等非 GPU 芯片市场占有率超过 10%，其中 NPU 芯片设计逻辑更为简单，常用于边缘侧和端侧的模型推理，并生成结果，在处理推理工作负载时，具有显著的能耗节约优势，因此增长较快。

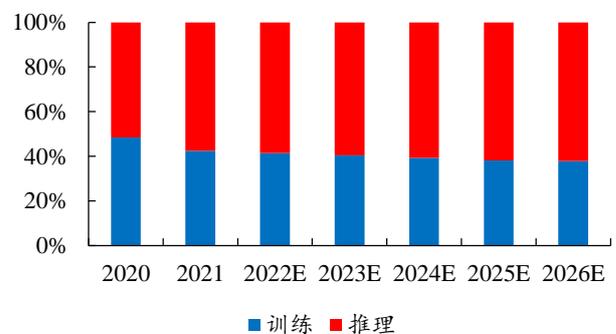
从工作负载角度而言，据 IDC 数据，2021 年中国数据中心用于推理的 AI 服务器占比已达 57.6%，预计这一比例将在 2026 年提升至 6 成。伴随企业人工智能应用成熟度逐步递增，企业将把精力更多从人工智能训练转移到人工智能推理工作负载上，这意味着人工智能模型将逐步进入广泛投产模式，因此智算中心的基础设施规划和运营支出规划更加需要追求服务器利用率提升和性能利用率的提升，以实现更好经济性。

图15：2021 年 GPU 应用于中国 89%的计算加速场景



数据来源：IDC、开源证券研究所

图16：预计 AI 服务器用于推理计算比例将提升至 6 成



数据来源：IDC、开源证券研究所

当下全球智能算力快速增长，据中国信通院数据，2022 年底全球算力总规模达到 650EFLOPS，其中通用算力规模为 498EFLOPS，智能算力规模为 142EFLOPS，

超算算力规模为 10EFLOPS；智能算力规模同比增长 25.7%，规模占比达 21.9%。国内市场方面，2022 年算力总规模为 180EFLOPS，排名全球第二，其中通用算力规模为 137EFLOPS，智能算力规模为 41EFLOPS，超算算力规模为 2EFLOPS。中国智能算力同比增长 41.4%，规模占比达 22.8%。而 2023 年 10 月工业和信息化部、中央网信办、教育部、国家卫生健康委、中国人民银行、国务院国资委等六部门联合印发《算力基础设施高质量发展行动计划》，提出到 2025 年实现国内算力规模超 300EFLOPS，其中智能算力占比达到 35%（105EFLOPS），预计 2022-2025 实现 36.8% CAGR。

图17: 2022 年全球智能算力规模增长 25.7%/EFLOPS

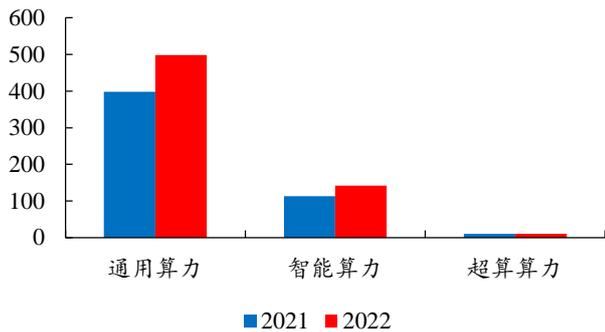
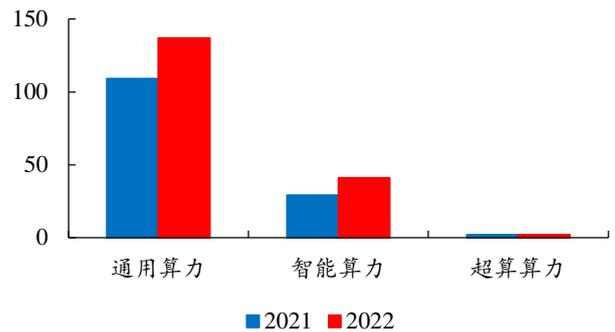


图18: 2022 年中国智能算力规模增长 41.4%/EFLOPS

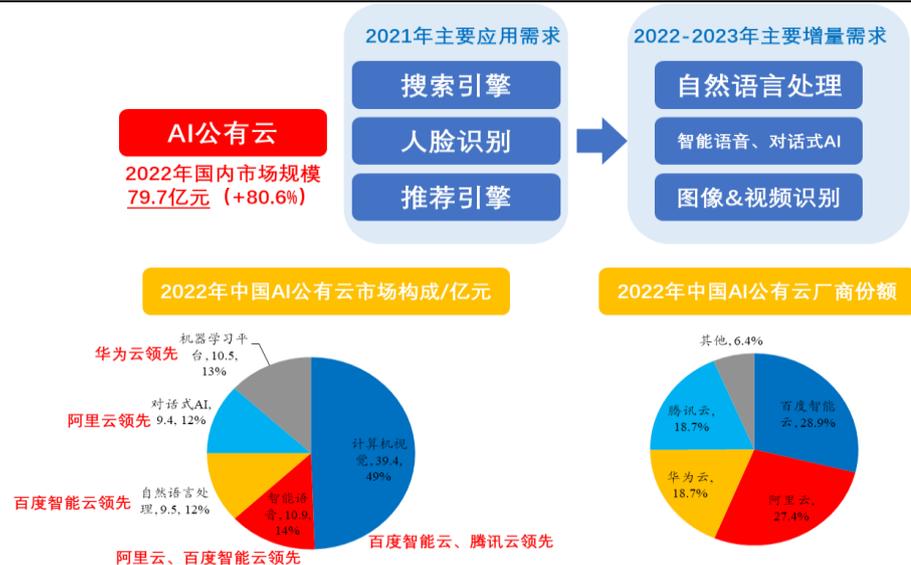


数据来源: Gartner、IDC、中国信通院、开源证券研究所

数据来源: 《中国算力白皮书》、中国信通院、开源证券研究所

智算需求增长背景下，云计算服务能提供更丰富的算力支持，如百度、腾讯、阿里等互联网巨头及蓝耘科技这类专门服务商通过 aaS (aaSService) 服务提供 AI 平台和 AI 服务，从而实现快速的产品迭代能力和丰富的场景化人工智能能力。据 IDC 数据，2021 年国内人工智能公有云服务市场规模达到 44.1 亿元人民币，占整体人工智能软件市场的 13.4%，增速远超人工智能软件整体市场增速；2022 年在 AI 市场整体遇冷背景下，国内人工智能公有云服务市场规模增长 80.6% 至 79.7 亿元，计算机视觉占据约一半市场份额。另一方面，IDC 也预计 AI 私有化部署仍是市场主流。

图19: 2022 年 AI 云服务市场随下游需求逆势增长至 79.7 亿元，互联网巨头主导

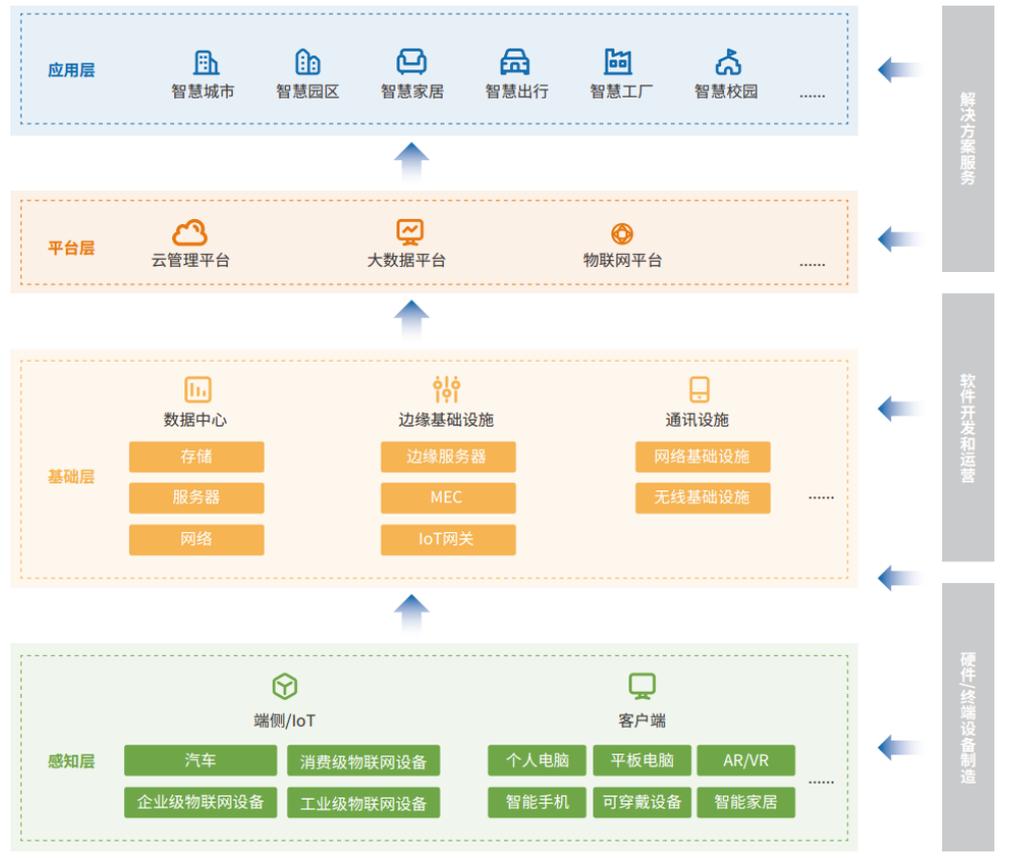


数据来源: IDC、开源证券研究所注: “领先”指该公司市场份额领先

2.2、IT 系统集成：市场结构变迁，AI、边缘等服务器增量持续驱动

蓝耘科技长期从事系统集成业务，在数字化产业链中处于基础层的方案解决商地位，主要受到服务器建设和数字化转型等下游需求驱动；近年来 AI 服务器建设需求增长背景下，蓝耘也更多将系统集成业务向 GPU 服务器切入，与自身算力业务打造资源和技术积累上的协同。

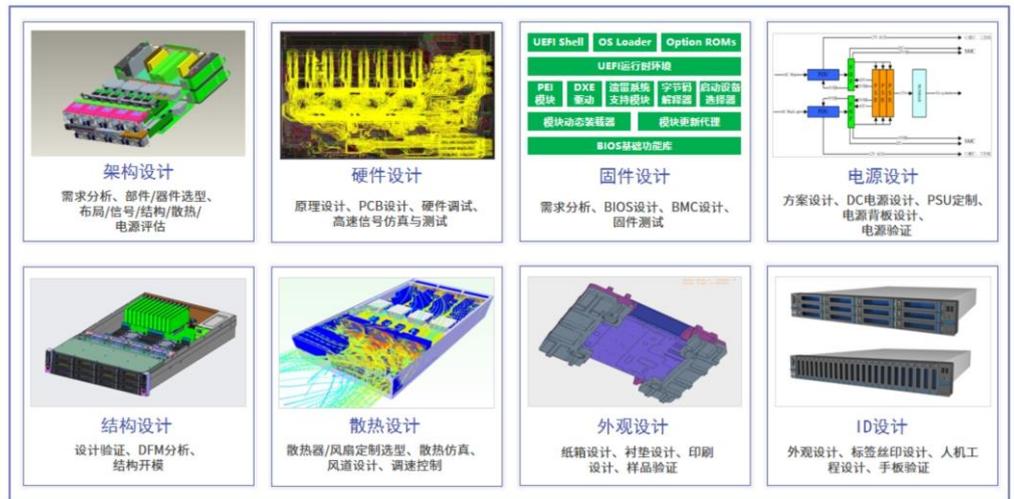
图20：服务器系统集成属于数字化基建产业的基础层领域



资料来源：IDC

系统集成行业具备较强的综合性，其中 IT 设备的部署与集成服务需要基于行业客户的数字化转型等实际需求，结合厂商自身服务经验，通过先进的架构设计、硬件整合和方案设计、电源和散热等辅助设备优化等综合解决方案能力来帮助客户打造高效、安全可靠、绿色的数字基础设施。

图21: 服务器集成商需要在架构等硬件层的设计经验上有深厚积累



资料来源: 坤前

从服务器的系统集成类型来看, 传统市场以通用服务器为主, 主要需求来自互联网、金融、工业、通信等领域对数据处理和运算有较高需要的大型公司对数据中心的建设; 而目前 AI 服务器需求涌现, 服务器集成需求增量从 CPU 向 GPU 切换, AI 服务器系统集成市场有望迎来整体增量及格局变迁。

图22: 服务器集成市场主要面向 AI、通用、存储、边缘等各类服务器



资料来源: 坤前

随着人工智能产业化加速, IDC 数据显示 2021 年全球 AI 服务器市场同比增长 39.1%, 其中中国市场增长至 59.22 亿美元。预计到 2026 年, 全球人工智能服务器市场将达到 347 亿美元, 将占整个服务器市场的 21.7%, 较 2022 年的 16.6% 占比持续提升, 且中国市场预计将达 123.38 亿美元。据 IDC 数据, 2021 年浪潮信息、戴尔、HPE 分别以 20.9%、13.0%、9.2% 的市占率位列全球前三, 三家厂商总市场份额占比达 43.1%。

另一方面, 以端 (智能终端/物联网) -边 (边缘计算) -云 (云计算) -网 (5G 和高速光纤网络) -智 (行业智能) 为核心的新型 IT 技术架构不断完善, 边缘计算已经成为当下重要的业务模式, 其中包括基础设施边缘和设备边缘等概念。作为边缘计算基础设施的服务器建设也将随之发展, 尤其受云网边端融合趋势持续驱动, IDC 预

计中国边缘计算服务器市场规模将从 2022 年的 40 亿美元增长至 2026 年的 92 亿美元，系统集成市场需求也将迎来增长。

图23: 预计中国 AI 服务器市场 2026 年达 123 亿美元

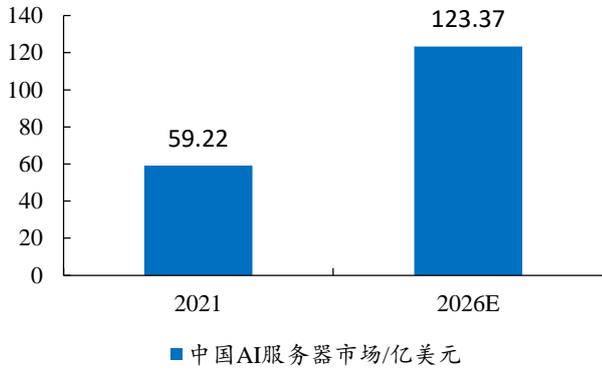
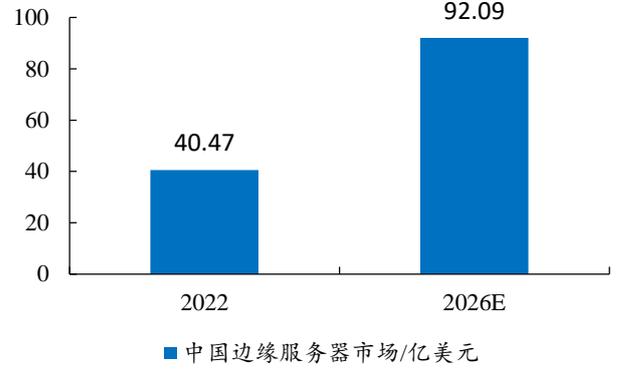


图24: 预计中国边缘服务器市场 2026 年达 92 亿美元



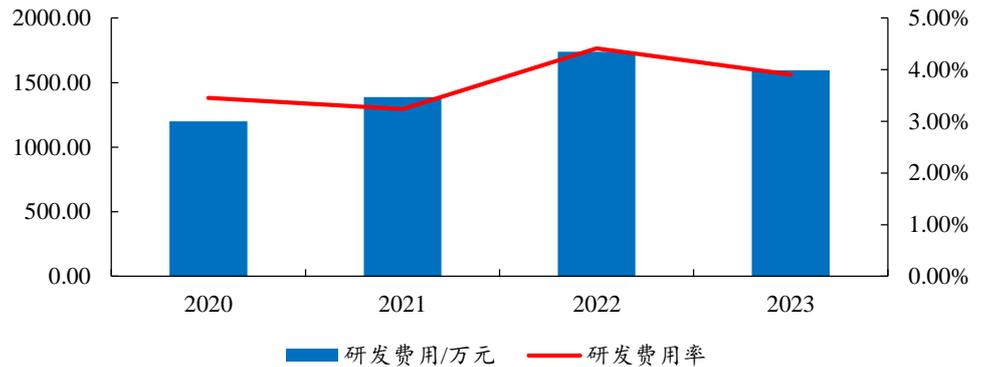
数据来源: IDC、开源证券研究所

数据来源: IDC、开源证券研究所

3、看点：立足 AI 模型算法研发，产业合作加速算力资源建设

技术创新方面，公司研发费用呈增长态势，自 2018 年的 288 万元上升至 2022 年的 1700 万元以上，2022 年、2023 年研发费用分别为 1738.51 万元、1595.56 万元，研发费用率分别为 4.41%、3.91%，维持稳定。公司获得北京市“专精特新”中小企业、北京市专精特新“小巨人”企业、国家高新技术企业等称号。

图25：2023 年研发费用率为 3.91%



数据来源：Wind、开源证券研究所

公司的研发模式是以自主研发和外包研发的模式进行。截至 2023 年 12 月 31 日，公司正在申请专利 14 项，目前拥有计算机软件著作权 125 项。

表2：截至 2023 年 12 月 31 日，公司正在申请专利 14 项

专利类型	申请号/专利号	发明名称	申请日
发明专利	2023114259538	一种动态调整 CPU 和 GPU 缓存的方法	2023/10/31
发明专利	2023111637099	一种云计算系统中的资源分配方法	2023/9/11
发明专利	2023111637084	一种高安全性的云计算方法	2023/9/11
发明专利	2023110276591	一种集群 GPU 资源调度方法	2023/8/16
发明专利	2023110276587	基于 CPU 与 GPU 协作的计算方法	2023/8/16
发明专利	2022116734457	企业终端智能安全管理技术平台	2022/12/26
发明专利	2022116744660	gpu 虚拟化实施方法	2022/12/26
发明专利	2022116740693	多维并行调度系统	2022/12/26
发明专利	2022103107787	一种基于云计算移动网络资源管理的 HPC 云计算系统	2022/3/28
发明专利	2021103233234	一种云计算系统中的资源分配方法	2021/3/26
发明专利	2021103233249	一种高安全性的云计算方法	2021/3/26

专利类型	申请号/专利号	发明名称	申请日
发明专利	2021102330415	一种基于 CPU 与 GPU 协作的计算方法	2021/2/23
发明专利	202110233042X	一种云计算环境中 GPU 资源调度系统	2021/2/23
发明专利	2021102330434	一种集群 GPU 资源调度系统	2021/2/23

数据来源：公司年报、开源证券研究所

2023 年，蓝耘科技共有研发项目 10 个，主要围绕 AI 模型算法开发、AI 模型预训练、AI 推理等场景的平台和工具的研发。并且开始在平台之上开始构建 AIGC 等相关应用。为用户从开发、训练到应用部署提供全流程支撑。研发的平台和工具降低了 AI 技术人员的使用门槛，也降低了应用落地的门槛，提高了应用部署的效率。提升了公司在行业内的竞争力。

表3：2023 年公司研发支出前五名的研发项目共研发支出 1205.28 万元

序号	研发项目名称	总研发支出金额（万元）
1	GPU 算力云平台项目	357.89
2	冷冻电镜数据处理服务平台	281.40
3	面向在线服务的 AIGC 系统开发	210.64
4	人工智能博弈算法开发	201.06
5	云原生大模型训练平台	154.28
	合计	1205.28

数据来源：公司研报、开源证券研究所

蓝耘的 GPU 智算云服务具体包括三方面：1) GPU 虚拟云主机，面向 AI 训练、图形处理、科学计算等领域，为客户提供专享高性能 GPU 虚拟云主机，可大幅提升图形处理和计算能力；**2) HPC 平台**，为生命科学、CAE 仿真、物理化学、气象环境、石油勘探等领域的科研界提供 CPU、GPU 算力服务；**3) 裸金属 GPU 服务器**，提供高性能、资源独享、安全隔离的专属弹性裸金属服务器，同时具备云资源的敏捷性、灵活性和高可用性，蓝耘提供完备的设备管理权限及运维服务。

GPU 算力资源整合及优质云平台服务是蓝耘科技在 AI 时代的核心壁垒。蓝耘 GPU 算力云服务基于 Kubernetes 原生云设计，搭载高可配置高可用的英伟达高端 GPU 系列，支持按需灵活计费、轻松调整存储容量及扩展容量，协同网络架构水平扩展、系统调优及应用排错，可实现现有工作负载轻松迁移、快速部署和弹性拓展，平滑应对突发流量高峰，满足 AI 训练开发、机器学习、大规模数据处理、分析等计算密集型任务场景中对 GPU 算力灵活性及多样化的需求，也为集中涌现的 AI 推理场景应用构筑强大的计算力支撑底座。

图26：蓝耘科技提供具备优势的英伟达 GPU 算力资源解决方案

提供无与伦比的GPU加速工作负载性能

我们提供行业内广泛的NVIDIA GPU访问，让您可以根据工作负载的复杂性进行计算资源的扩展。我们基于Kubernetes的原生基础设施提供极快的启动时间、灵敏的自动扩展和现代化的网络架构，确保性能与您的需求相匹配。



根据工作负载实际需求进行适当调整

每个模型都是独特的，它们的计算需求也各不相同。蓝耘拥有行业内广泛的GPU选择，您可以更快、更高效地进行模型的训练、微调和服务。



通过Kubernetes实现裸金属性能

通过部署容器化的工作负载，可以摆脱堆栈中的虚拟化层。使您能够享受裸金属的好处，而无需让管理基础设施成为负担。



全栈机器学习专业知识

机器学习是我们的专长，而我们的基础设施也反映了这一点。无论您是在训练还是部署模型，我们构建了蓝耘Cloud来简化设置过程并提高性能。

资料来源：公司官网

大模型便是基于智算算力驱动之下诞生，在云场景下需要 GPU 算力服务结合高速网络、高性能文件存储应对大规模分布式训练场景，蓝耘科技正广泛开拓该领域客户。得益于模型泛化能力强、长尾数据的低依赖性以及下游模型使用效率的提升，大模型被认为具备了“通用智能”的雏形。

大模型的技术基础是 transformer 架构、迁移学习和自监督学习，transformer 架构应用于 NLP、计算机视觉等领域。按“算力当量”（PetaFlops/s-day，PD，即每秒千万亿次的计算机完整运行一天消耗的算力总量）计算，AI+Science 领域的 AlphaFold2、自动驾驶系统、GPT-3 等模型训练需要几百到几千 PD 算力支持，如 GPT-3 训练需要 3640PD 的算力，且 GPT-4、GPT-5 需求还将指数级上升。

图27：GPU 云计算集群服务于大模型计算，带动公司 GPU 云计算服务需求



资料来源：火山引擎官网

在科研和工程仿真领域，GPU 云计算可以在生命科学、宇宙天文、量子物理、计算化学和天体物理材料设计、气候气象，海洋探测等学科中匹配上层应用的高性能计算集群。**蓝耘科技服务于各类科研客户，在教育科研领域已有多多个大型项目成功落地**，如 2022 年中标北京生命科学研究所以**冷冻电镜数据处理及存储配套计算集群**项目等，为高校和研究机构提供软硬云一体化 HPC 交付方案和一站式高性能计算整体解决方案等，以协助高校和科研机构数字化转型升级。

以冷冻电镜为例，实验分析中的数据处理和计算需要大量的计算能力，未来算力需求广阔。冷冻电镜主要是用于扫描电镜的超低温冷冻制样及传输技术(Cryo-SEM)，可实现直接观察液体、半液体及对电子束敏感的样品，如生物、高分子材料等，观测精度达到 1/10 纳米，使科学家能够以亚纳米分辨率观察生物分子。冷冻电镜工作流程中需要采集大量的数据，然后通过对这些数据经过处理和计算来重构出生物大分子的三维结构，图像典型大小 6-8GB、可以缩小到 1-3GB，一套数据通常有**5-10TB 数据量**，一台主流 300kV 电镜一年产生**1-2PB 数据**，算力需求潜力广阔。针对生命科学研究以及冷冻电镜的高性能计算需求，蓝耘可为客户提供包括混合式企业 HPC 平台、软硬云一体化 HPC 交付方案、算力云订阅等服务，涵盖基因测序分析、样本筛选、蛋白质结构预测、分子动力学模拟、分子设计、药物靶点发现、药物晶型预测等应用场景。

图28：GPU、CPU 云计算服务可提升生命科学等科研、工程领域效率



资料来源：火山引擎官网

具体技术方案上，蓝耘科技积累了多年经验，通过硬件层、平台架构层、软件及数据层三大方面实现 AI 算力服务的优化升级。

1) **硬件层**：基础设施硬件架构是 AI 算力优化要跨越的第一道“天堑”，蓝耘通过服务器配置调优、国际领先的 AI 加速卡、高速网卡、IB 卡搭载和管理网络、计算网络、存储网络的“三网分离”模式，以及针对 AI 大模型训练场景计算网络量身打造的专用 1:1 收敛的 Infiniband 网络，保障集群内任意 GPU 节点之间都能实现无阻塞的 RDMA 快速通信，为用户带来高带宽、低延迟的优异性能体验;在存储方面，蓝

耘基于多年 HPC 系统集成运维、优化经验及 NVIDIA 专业团队技术支持和大模型客户的实际使用诉求，在 GPU 算力集群中优化部署高性能并行文件系统，可为 AI 大模型训练提供高达 100GB/s 的吞吐能力，充分满足 AI 训练、推理、微调及高性能计算场景中对计算、网络、存储等环节灵活性和多样化的需求。

2) 架构层：平台架构层也是突破 AI 算力升级壁垒的重要一环。蓝耘自研的裸金属调度平台，针对 AI 大模型训练场景优化了架构设计和系统镜像，配合全生命周期的管理，让大规模裸金属集群部署变得更加简单易用。蓝耘基于 Kubernetes 平台二次开发的容器调度平台，针对大模型训练过程中驱动、GPU 硬件、物理服务器硬件等故障导致的训练停止痛点，拥有无人工干预即可自动拉起失败任务的强大“自愈”功能，大大降低了大模型训练用户的运维难度。

3) 软件层：就软件栈和预训练数据层面而言，蓝耘 AI 大模型训练平台面向深度学习等开发场景，可实现计算资源统一分配调度、训练数据集中管理并加速，通过预置集成开箱即用的国内外主流大模型训练框架、SDK、加速器、预训练数据及容器镜像等，轻松实现集群环境快速部署，高效支撑 AI 创新研发。

图29：蓝耘科技利用裸金属 K8S 部署实现推理服务的性能优势



资料来源：公司官网

联手中联数据集团打造新型智算中心，产业合作加速升级。蓝耘乌兰察布智算中心面对高频算力需求，携手战略合作伙伴中联数据集团共同打造，通过提供面向人工智能应用的算力服务，不断提升智能计算效能和速度，满足 AI 训练、推理等计算密集型应用场景下大规模、多线并行的多元计算需求，坐落于内蒙古亚信数据港 1 号。乌兰察布智算中心采用高效浸没式液冷技术+智能动态高密风墙技术为 GPU 服务器提供全量制冷，单机柜功率密度可达 30kw 以上，支持 350W 以上高功率芯片冷却，同时依托乌兰察布超长的自然冷却时长及丰富的可再生能源优势，实现较低成本、高效率和高性能。这一产业合作案例的建成将为蓝耘未来携手更多数据中心、地方政府打造新型智算中心，进一步扩充蓝耘底层算力资源池，以推动智算服务更好地赋能 AI 创新发展提供有益的参考和借鉴。

图30：蓝耘科技携手战略合作伙伴中联数据集团共同打造乌兰察布智算中心


资料来源：投资界

4、估值对比：同行可比公司 PE2023 均值为 61.34X

蓝耘科技专注于为企业级规模的 GPU 加速工作负载提供算力云服务，服务于高校科研、生命科学、人工智能、自动驾驶、工业设计、智慧城市等领域，主要业务包括 IT 系统集成业务和 GPU 算力云服务业务两大模块，GPU 算力云服务业务自 2021 年以来高速增长，目前与产业方广泛合作并通过融资租赁等方式加速 AI 时代布局，有望拓展算力版图，不断提升产业地位，在 LLM 等需求驱动下前景广阔。目前公司可比公司 PE 2023 均值 61.34X，蓝耘科技 PE 2023 为 21.56X，估值相对较低，同时毛利率高于行业，具备一定可比优势。

表4：同行业可比公司 PE2023 均值为 61.34X

公司名称	股票代码	市值/亿元	PE 2023	PE TTM	2023 年营收/亿元	2023 年归母净利润/万元	2023 年毛利率
鸿博股份	002229.SZ	87.36	-321.77	-142.61	6.2	-5,444.29	24.95%
并行科技	839493.NQ	27.48	-40.74	-50.57	4.96	-8,049.95	25.63%
云赛智联	600602.SH	173.56	76.07	82.15	52.64	19,291.16	17.99%
浪潮信息	000977.SZ	579.58	46.6	30.46	658.67	178,278.43	10.04%
均值		173.605	61.34	56.31	180.62	98784.80	19.65%
蓝耘科技	871169.NQ	11.38	21.56	22.19	4.09	5,128.17	25.79%

数据来源：Wind、开源证券研究所 注：数据截至 20240509，PE2023、PE TTM、2023 年归母净利润/万元均值计算剔除了负值

5、风险提示

算力需求波动、产业政策变化、GPU 渠道风险

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

分析师承诺

负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及开源证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（outperform）	预计相对强于市场表现 5%~20%；
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；
	减持	预计相对弱于市场表现 5%以下。
行业评级	看好（overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡	预计行业弱于整体市场表现。
备注：评级标准为以报告日后的6~12个月内，证券相对于市场基准指数的涨跌幅表现，其中A股基准指数为沪深300指数、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普500或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。		

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用，如接收人并非开源证券客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，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、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。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所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的开源证券网站以外的地址或超级链接，开源证券不对其内容负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所**上海**

地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼10层
邮编：200120
邮箱：research@kysec.cn

深圳

地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

北京

地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

西安

地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn