


# 汇洲智能：卡位大模型的数据预训练新秀

 证券研究报告

## 投资评级:增持(维持)

基本数据	2023-12-25
收盘价(元)	3.80
流通股本(亿股)	19.88
每股净资产(元)	1.13
总股本(亿股)	20.00

### 最近 12 月市场表现



分析师 程兵

SAC 证书编号: S0160523060001  
chengbing@ctsec.com

分析师 余炜超

SAC 证书编号: S0160522080002  
shewc@ctsec.com

### 相关报告

1. 《Q3 营收增速趋缓，布局数据预训练，重塑发展动能》 2023-11-30

## 核心观点

❖ **数据预训练是 AI 应用降本增效的重要因素，随着场景落地，数据将成为大模型竞争力提升的高效燃料，预训练数据服务商有望伴随着 AI 场景落地而快速成长。**

前期 AI 模型搭建以过往数据为主，预训练数据服务需求并未伴随模型发展而快速增长，由此带来错觉：预训练数据服务在 AI 中不需要。**事实恰恰相反，预训练数据服务是 AI 模型降本增效，加速迭代重要因素。**

**预训练数据服务需求将伴随 AI 场景落地而迅速增长。**预训练数据服务有助于优化算力配置；有助于更精准实现客户需求；有助于 AI 模型快速迭代。因此，随着 AI 场景落地，使用更多全新数据，预训练数据服务需求由此快速增长。

**数据反向赋能应用场景，数据和场景落地有望共振。**终端场景应用落地加速数据要素多模态扩容，将进一步提升预训练数据环节的重要性，高质量预训练数据集是大模型自反馈强化学习机制的前提，数据与场景将实现共振。

❖ **承齐重之积淀，机床成为公司发展压舱石。**

“齐一”品牌优势显著，多年积累成就“4+3+N”创新体系巩固研发优势。近年公司双向发力，纵向承接风电行业需求释放研发系列新品。横向针对航空航天、船舶、能源等行业进行高端品升级。

❖ **公司切入数据预训练业务，发力智能标注和自动驾驶，重塑发展动能。**

平台化统筹、智能化、技术先进性是自动驾驶场景数据预训练的准入壁垒，公司当前已具备舱内语音、舱外图像、视频等多类型数据的标注能力，Enable AI 智能平台支持 3D 点云数据的连续帧标注、点云融合。

自动驾驶数据需求随算法迭代和落地车型量产的时间呈现周期性收敛，根据德勤测算，2027E 年自动驾驶带来的 AI 预训练数据服务需求有望达到 83 亿元，2022E-2027E 五年复合增速为 37%。

❖ **投资建议：**我们预计公司 2023-2025 年实现营业收入 8.86/11.06/13.72 亿元，归母净利润 1.82/2.66/3.54 亿元。对应 PE 分别为 41.69/28.53/21.45 倍，维持“增持”评级。

❖ **风险提示：**大模型产业结合不及预期；智能数据标注市场竞争加速；智能标注平台落地不及预期。

盈利预测：

	2021A	2022A	2023E	2024E	2025E
营业收入(百万元)	837.65	702.66	886.10	1106.12	1371.75
收入增长率(%)	16.06	-16.12	26.11	24.83	24.01
归母净利润(百万元)	-727.49	85.10	182.29	266.34	354.33
净利润增长率(%)	-554.78	111.70	114.19	46.11	33.04
EPS(元/股)	-0.37	0.04	0.09	0.13	0.18
PE	—	72.22	41.69	28.53	21.45
ROE(%)	-98.96	4.04	7.92	10.36	12.12
PB	4.71	2.94	3.30	2.96	2.60

数据来源：wind 数据，财通证券研究所

## 内容目录

1	机床龙头，发力数据预训练再塑成长动能.....	5
1.1	夯实重工基础，机床为发展压舱石.....	5
1.2	发力 AI 数据预训练，重塑成长动能.....	8
2	场景驱动，AI 预训练数据有望增长.....	10
2.1	预训练数据是 AI 产业链的基石.....	10
2.2	多模态趋势显著，提升数据预训练的重要性.....	12
2.3	场景落地，驱动数据预训练需求释放.....	14
3	布局自动驾驶，快速铸就行业壁垒.....	15
3.1	发力智能标注，助力降本增效.....	15
3.2	切入自动驾驶，享受更多行业增量.....	17
4	盈利预测与估值.....	20
4.1	盈利预测.....	20
4.2	估值与投资建议.....	21
5	风险提示.....	22

## 图表目录

图 1.	汇洲智能股权结构（数据截止到 2023H1）.....	5
图 2.	公司搭建“4+3+N”创新体系巩固研发壁垒.....	6
图 3.	公司研发支出持续增长.....	7
图 4.	研发人员规模不断扩张（单位：人）.....	7
图 5.	技术研发团队从业时间较长（截止 2022 年底）.....	7
图 6.	AI 数据服务子公司的股权结构（数据截止到 2023H1）.....	8
图 7.	公司内容审核业务矩阵完善.....	9
图 8.	公司已形成多层次高粘性的客户矩阵.....	10
图 9.	数据是 AI 产业链的基石.....	11
图 10.	数据处理过程占据 AI 项目 80% 的时间.....	11
图 11.	数据预训练服务商是 AI 数据产业链中承上启下的关键一环.....	12
图 12.	海内外大模型均趋向多模态.....	13
图 13.	全球 AI 训练数据市场规模 2027 年有望达到 1574 亿元.....	14

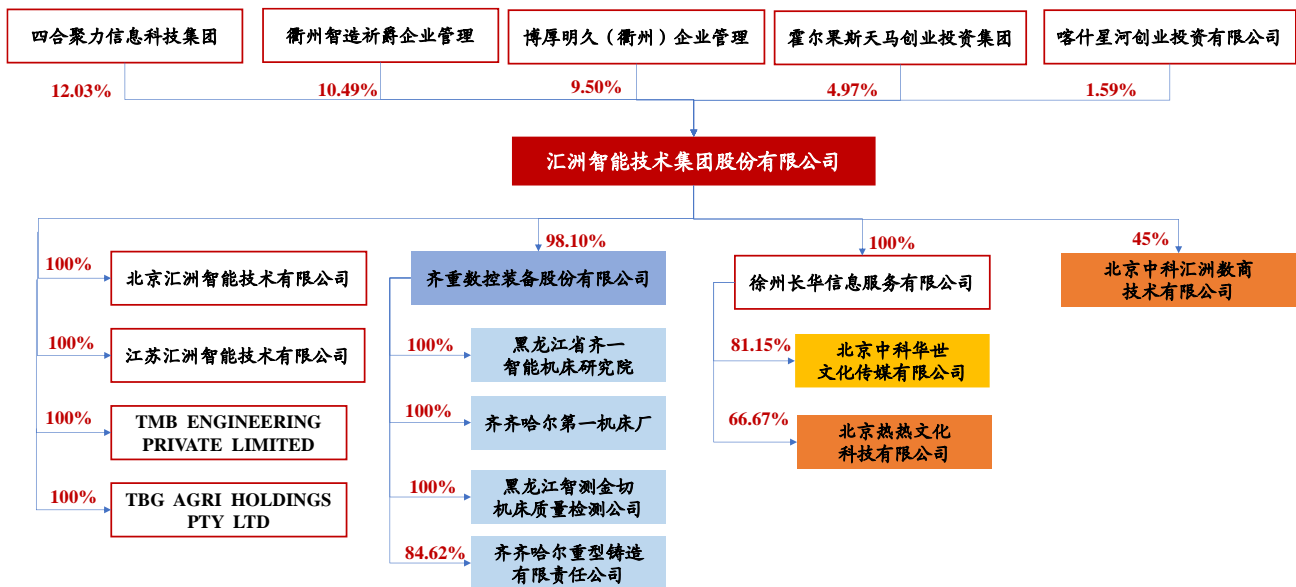
图 14. 公司的智能化体系.....	16
图 15. AI 智能标注助力专业预训练数据服务商和模型开发商互相赋能，形成闭环.....	17
图 16. Enable AI 智能标注平台 3D 点云标注操作视图.....	19
图 17. 数据需求释放进度随算法迭代和落地车型量产的时间呈现周期性收敛.....	20
表 1. 公司机床产品矩阵完善.....	6
表 2. 公司技术产品矩阵行业前列水平.....	9
表 3. Llama-13B 与 GPT-3 数据对比.....	13
表 4. Llama-13B 与 GPT-3 性能测试对比.....	14
表 5. 智能驾驶是数据预训练未来五年弹性最大的应用场景.....	15
表 6. 数据预训练头部企业纷纷切入智能化平台.....	17
表 7. Enable AI 智能标注平台支持多类型标注工具、标注任务.....	18
表 8. 盈利预测（单位：百万元、%）.....	21
表 9. 可比公司估值对比分析.....	22

## 1 机床龙头，发力数据预训练再塑成长动能

汇洲智能是我国传统机床龙头。公司机床业务起家，旗下控股子公司齐重系机床等高端装备制造业务的经营主体，历经 73 年发展，成为我国传统机床龙头企业。

为迎接 AI 浪潮，发力 AI 大模型数据预训练业务，重塑新的成长动能。公司为迎接 AI 发展浪潮，通过收购积极布局 AI 大模型数据预训练业务。2019 年收购长华文化，取得热热文化控制权，作为 AI 数据与训练业务的经营主体，同时将控股子公司中科汇洲数商作为 AI 数据预训练业务的技术研发主体。

图1.汇洲智能股权结构（数据截止到 2023H1）



数据来源：wind，汇洲智能公司公告，财通证券研究所

### 1.1 夯实重工基础，机床为发展压舱石

公司机床业务产品矩阵完善。公司始建于 1950 年，是国家一五时期重点建设项目之一，高端制造底蕴丰厚。历经 73 年积累，已成为我国重要的数控机床生产基地。截至目前已形成 38 万平方米的机床生产基地，包含 10 大类，26 系列，600 多品种的机床产品矩阵。

公司多项技术弥补海内外空白。截至目前，公司 400 多项具备自主知识产权的产品填补国内空白；立式车床加工直径最小可达 0.5 米，最大可到填补国际空白的 25 米；卧式车床加工直径最小可达 1 米，最高到填补国际空白的 6.3 米。广泛用于船舶、汽车、风电等下游行业。

“齐一”品牌优势显著。“齐一”数控机床是中国知名机床品牌，部分高端机床产品已成功出口欧美、日韩等 30 多个国家和地区，截至 2023H1 重型车床市场占有率 40% 到 50%，重型深孔钻镗床市场占有率 100%。

表1.公司机床产品矩阵完善

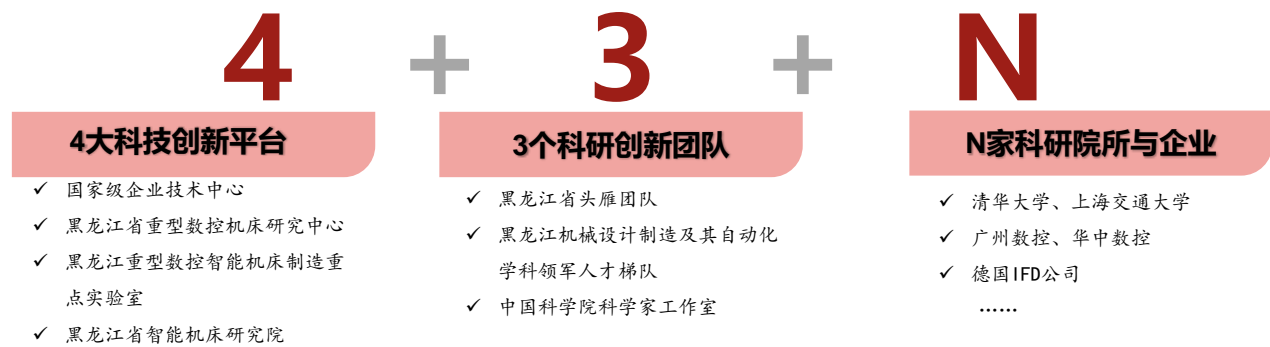
产品	应用范围	系列
立式车床	适用 <b>船舶、能源、汽车</b> 等行业； 用于 <b>回转类零件</b> 切削加工	SVT系列 单柱立式（铣）车床
		DVT系列 双柱立式（铣）车床
		CKE系列 双柱立式车床
		CK51E系列 双柱立式车床
卧式车床	适用 <b>船舶、能源、轧钢</b> 等行业； 用于 <b>轴类、盘类</b> 等回转类零件加工； 是 <b>大型电机、发动机主轴</b> 及各类转子必不可少的加工设备	HT III系列 重型卧式车床（四轨）
		HT III、II P系列 重型卧式车床（三轨）
		HT I系列 重型卧式车床（两轨）
滚齿机	使用范围广泛，适用于 <b>齿形</b> 加工	YK系列滚齿机
龙门铣床	适用 <b>风电、重机、机车、造船、发电、机床、印刷、轻纺、模具</b> 等制造行业； 用于 <b>重型、超重型基础零件</b> 的加工	XKA28系列 数控龙门移动式动梁镗铣车复合机床
		XK28系列 数控龙门移动式动梁镗铣车
		XK27系列 数控龙门移动式定梁镗铣车
高精智能产品	接轨国际高端标准； 其中数控重型卧式车床用于 <b>核电电机转子、水轮机转子、燃料容器</b> 等零件加工	高精度数控龙门镗铣车加工中心
		高精度立式铣车复合加工中心
		数控重型卧式车床

数据来源：齐重数控公司官网，财通证券研究所

持续扩大研发支出，搭建“4+3+N”的创新体系巩固研发优势。历经 73 年发展，公司机床产品研发经验丰富。近年来，公司发力构建“4+3+N”的创新体系：搭建 4 个科技创新平台，培育 3 个科技创新团队，引进 N 家科研院所及企业，高平台化研发能力，巩固研发优势。截至 2023H1，公司是我国重型卧式车床、重型深孔钻镗床、重型立式车床的国家检验标准制定单位，主持和参与制修订标准累计 77 项，专利授权 302 项。

图2.公司搭建“4+3+N”创新体系巩固研发壁垒

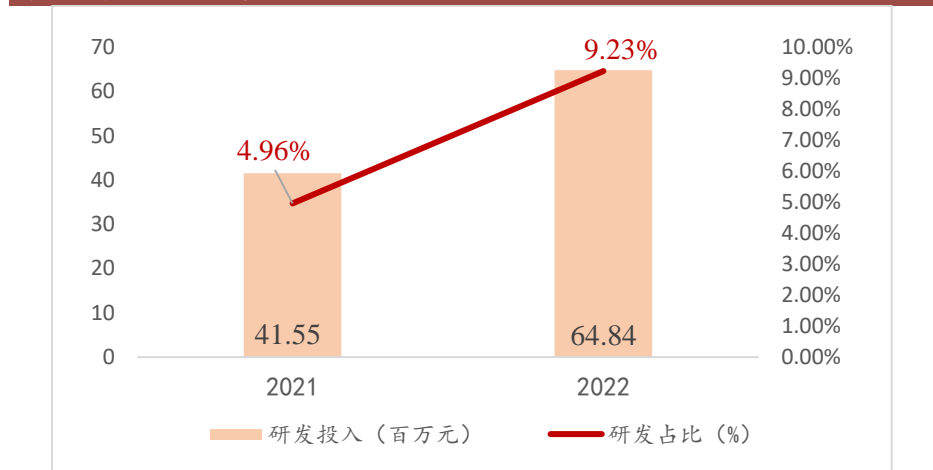
完善的科研创新体系：参与制修订国家/行业/团体标准累计77项，累计专利授权302项



数据来源：汇洲智能公司公告，财通证券研究所

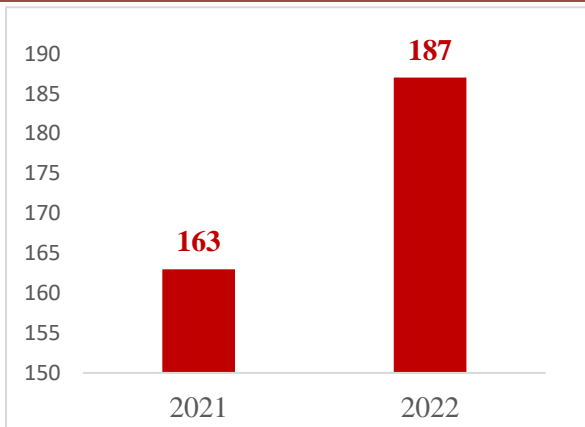
公司技术团队经验丰富，持续扩大研发支出。截至 2022 年底，公司研发人员数量 187 人，同比 2021 年提升 15%，其中 40 岁以上占比 57%，多数研发人员拥有多年研发经验。公司持续扩大研发支出，2022 年研发支出 64.84 百万元，同比增长 56.05%，研发支出占比营收 9.23%，同比提升 4.27pct。

图3.公司研发支出持续增长



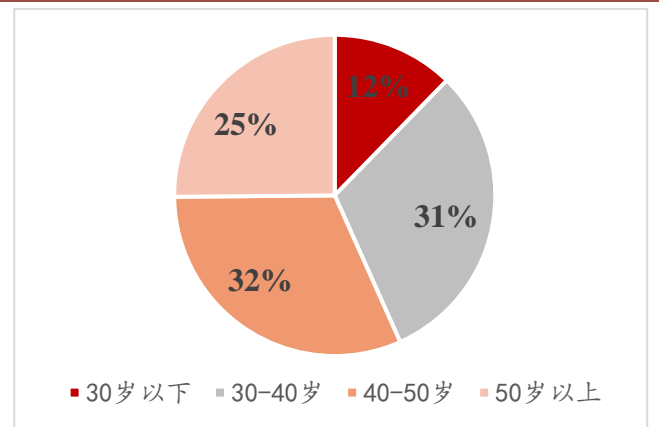
数据来源：汇洲智能公司公告，财通证券研究所

图4.研发人员规模不断扩张（单位：人）



数据来源：汇洲智能公司公告，财通证券研究所

图5.技术研发团队从业时间较长（截止 2022 年底）



数据来源：汇洲智能公司公告，财通证券研究所

**2020 年起，公司紧跟下游需求变动双向发力，实现机床产品的全面升级。**

公司纵向紧跟需求变动，向下游热门行业拓展。为承接下游风电行业的需求释放，公司快速针对风电主轴、法兰、回转支承研发了数控专用卧车、定梁立车、硬车和滚齿机等一系列产品，2022 年风电行业产品新增合同额占比高达 60% 以上。



公司横向紧跟行业技术趋势，向高端品升级。2022 年公司发力研发适用于下游航空航天、船舶、能源行业的高精度机床，完成全新高端产品设计 26 台，技术准备 147 项，电气设计 190 项，工艺设计 221 项。

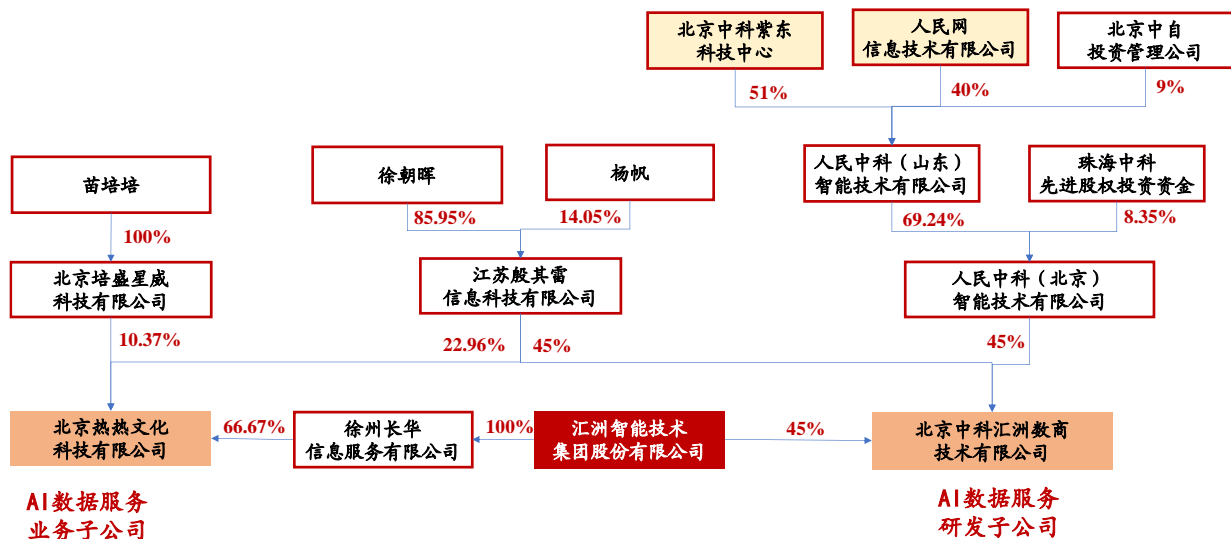
截至 2022 年底，公司仍存在 8 项高端产品在研项目。其中超声微锻造辅助激光增材制造项目，有望弥补国内技术空白；基于 5G 通讯的智能机床在研项目，有望助力公司产品的智能化转型升级；数控重型卧式镗床针对风电主轴内孔，国内领先，下游应用前景广泛。

### 1.2 发力 AI 数据预训练，重塑成长动能

公司受益于大模型赛道的长坡厚雪，依靠三大核心壁垒：丰富的技术和产品矩阵、高粘性的优质客户、经验丰富的核心技术团队，快速成为 AI 数据服务商新秀。

公司布局数据预训练业务，重塑新的发展动能。公司于 2019 年收购徐州长华 100% 股权，取得了对热热文化的控制权，构建 AI 数据预训练业务经营主体，同步设立控股子公司人民中科数商为 AI 数据预训练研发主体，提前研发布局，承接下游数据标注需求的释放。

图6.AI 数据服务子公司的股权结构（数据截止到 2023H1）



数据来源：wind，汇洲智能公司公告，财通证券研究所

公司数据业务矩阵分为互联网内容审核和 AI 数据标注两大板块，数据标注产品和技术已进入行业前列水平。公司已形成行业领先的数据标注技术矩阵和数据集



产品矩阵，截至当前，已实现针对文本、图片、语音、视频等数据类型的标注，其中计算机视觉相关的视频追踪、打点、连续帧技术；2D 图像语义分割；3D 点云融合、连续帧等数据标注技术，适用于下游自动驾驶、人形机器人等热门终端场景；沉淀的可复卖的标品数据集涵盖文本、图片、音频、视频；并可针对自动驾驶、城市规划、医学影像诊断等场景制定解决方案。

**表2.公司技术产品矩阵行业前列水平**

企业	标注技术矩阵	数据产品矩阵
汇洲智能	1. <b>计算机视觉</b> ：视频（追踪、打点、连续帧）；图像（2D矩形框、语义分割、关键点）；3D点云标注。 2. <b>语音工程</b> ：ASR转写、方言翻译、小语种标注。 3. <b>自然语言处理</b> ：文本分类OCR转写、情感标注、意图判断。	<b>标品</b> ：数据集全覆盖文本、图片、音视频。 <b>定制</b> ：自动驾驶（轮外）、城市规划、医学影像诊断、视频会议监督、图像分割等场景。
海天瑞声	1. <b>计算机视觉</b> ：图片、2D3D联合标注、语义分割、点云标注分割追踪、点云连续帧、全景分割、多模态标注。 2. <b>语音工程</b> ：ASR(正字转写、语音转写、发音词典、环境性别领域低地域标注)、TTS(正字转写、发音校对，词性韵律因素歌曲标注、发音词典制作)。 4. <b>自然语言处理</b> ：OCR标注、词性标注、情感标注、文本分类聚类正则化泛化、知识图谱等。	<b>标品</b> ：超1,498个成品数据集。NLP数据集302个；计算机视觉数据集155个；智能语音数据集1041个，具有独家词典数据集。 <b>定制</b> ：智能驾驶、家居、教育、安防、物联网。
澳鹏	1. <b>计算机视觉</b> ：多语言2D图像标注（包含普通目标检测、语义分割等），3D图像（立体框）标注，Lidar标注（如点云目标检测、联合标注、点云语义分割、点云跟踪标注等）；视频分类、视频审核，视频按规则抽帧再标注；3D点云标注（点云目标检测、联合标注、点云语义分割、点云跟踪标注）。 2. <b>语音工程</b> ：语音切割、转写。 3. <b>自然语言处理</b> ：文本语义识别、智能客服机器人数据标注。	<b>标品</b> ：超400个成品数据集。250+已标注标品数据集，音频、图像、测试、情绪和兴趣点数据的定制数据集； <b>定制</b> ：智能驾驶、医疗健康、人体2D建模、智能家居、智慧金融、新零售

数据来源：海天瑞声官网，澳鹏官网，热热文化官网，海天瑞声公司公告，财通证券研究所

内容审核业务在人员规模、内容覆盖和准确率多维度领先同业。热热文化 2016 年成立，发展至今已在北京、成都、绥化、枣庄、金华设立五个规模型审核、标注基地，审核团队规模超 5 千人，累计培养审核人才超 5 万人，累计审核数据 20 亿条；在网络内容的安全审核业务事故率和遗漏率均较低。

**图7.公司内容审核业务矩阵完善**


数据来源：热热文化官网，财通证券研究所

公司核心团队自带 AI 数据预训练产业经验，背景丰富奠定研发优势。数据预训练研发主体中科汇洲数商人民网+中科院北京紫东科技中心等多方牵头创办研发平台，自带 AI 数据技术背景和产业经验。业务主体热热文化总经理李刚是前阿里云互联网事业部技术负责人，公司监事、技术团队负责人李兵是中科院自动化所博导，人民中科首席科学家，曾成功开发跨模态预训练智能搜索引擎“白泽”模型。

公司已形成多层次、高粘性的客户矩阵。公司内容审核业务起家，与人民网渊源深厚，起点高叠加深耕基础数据服务数年，公司依靠高质量服务已成功建立多层次高粘性的客户矩阵，遍布下游互联网、人工智能科技企业、航天、重工、金融、政务、高校等行业。

图8.公司已形成多层次高粘性的客户矩阵



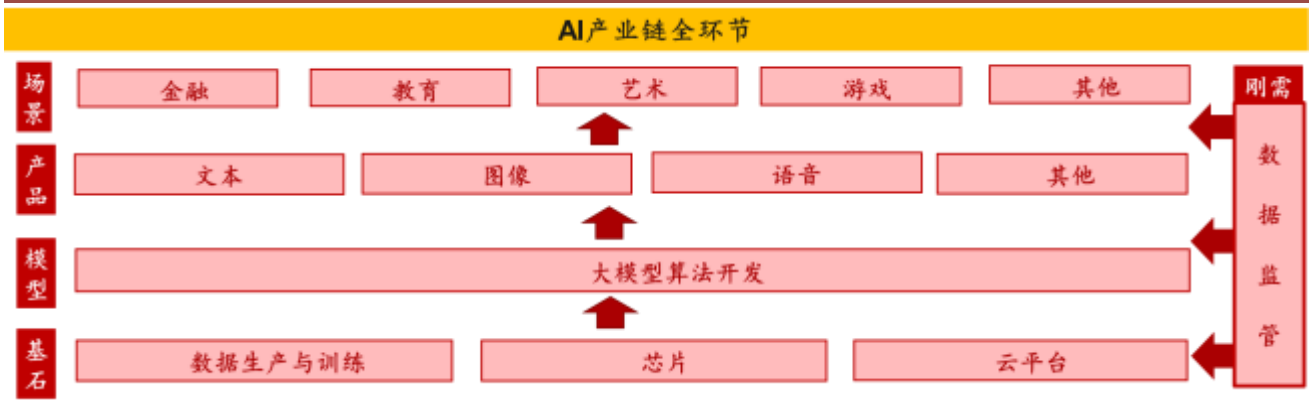
数据来源：热热文化官网，财通证券研究所

## 2 场景驱动，AI 预训练数据有望增长

### 2.1 预训练数据是 AI 产业链的基石

数据是 AI 产业链上游的基石。ChatGPT 大模型引领本轮生成式 AI 浪潮，纵观当前的 AI 产业链，上游数据来源于终端场景的采集，是算法感知世界的起点；中游模型开发是数据应用的工具；下游场景端拉动模型算法迭代；监管是贯穿上下游全环节的刚需。

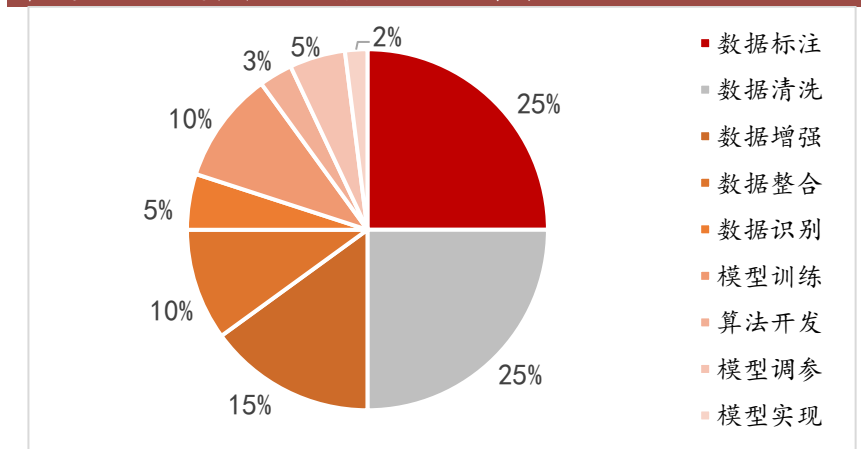
图9.数据是 AI 产业链的基石



数据来源：《2023 大模型和 AIGC 产业图谱》（信通院），财通证券研究所

根据 AI 分析公司 Congnilytica 研究显示，AI 项目中数据处理过程占据 80% 的时间，其中数据标注占比 25%，针对复杂场景数据的有效预处理，可缩短数据识别、整合、增强、清洗、标注全环节的时间周期，为模型开发节约成本。

图10.数据处理过程占据 AI 项目 80% 的时间



备注：数据处理过程为数据标注、清洗、增强、整合、识别过程，预测时间 2022 年  
数据来源：Congnilytica，观研天下，财通证券研究所

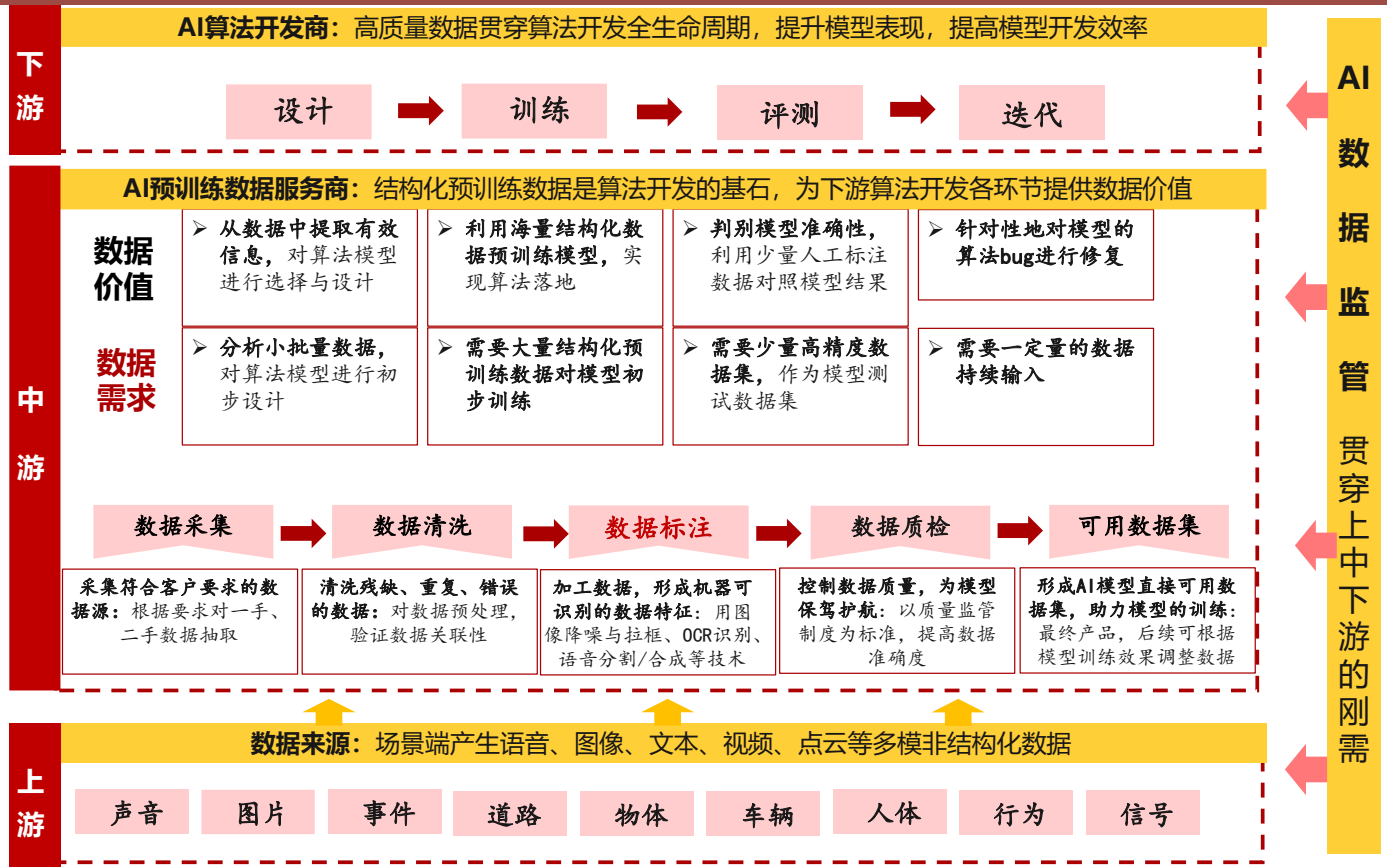
纵观 AI 数据链，AI 数据预训练环节是承上启下的关键抓手。

**预训练数据服务承接上游数据来源：**非结构数据只有经过预训练处理才能激活其价值。AI 预训练数据服务将场景端语音、图像、文本、视频、点云等非结构化数据进行采集、清洗、标注、质检，形成 AI 模型可直接利用的有效预训练数据集。

预训练数据助力下游算法设计、训练、评测、迭代全生命周期降本增效。

(1) 算法设计环节，利用小批量预训练数据对算法初步的设计进行验证，减少模型设计的方向性偏差。(2) 算法训练环节，有效简化模型参数规模，节约算法开发时间。(3) 算法测评环节，少量人工标注的预训练数据可作为模型输出结果的对照组，有效评定模型的准确性。(4) 算法迭代环节，针对 bug 对预训练数据进行精确标注处理，进而对模型精准修复，有效提升模型性能。

图11.数据预训练服务商是 AI 数据产业链中承上启下的关键一环



数据来源：德勤，艾瑞咨询，财通证券研究所

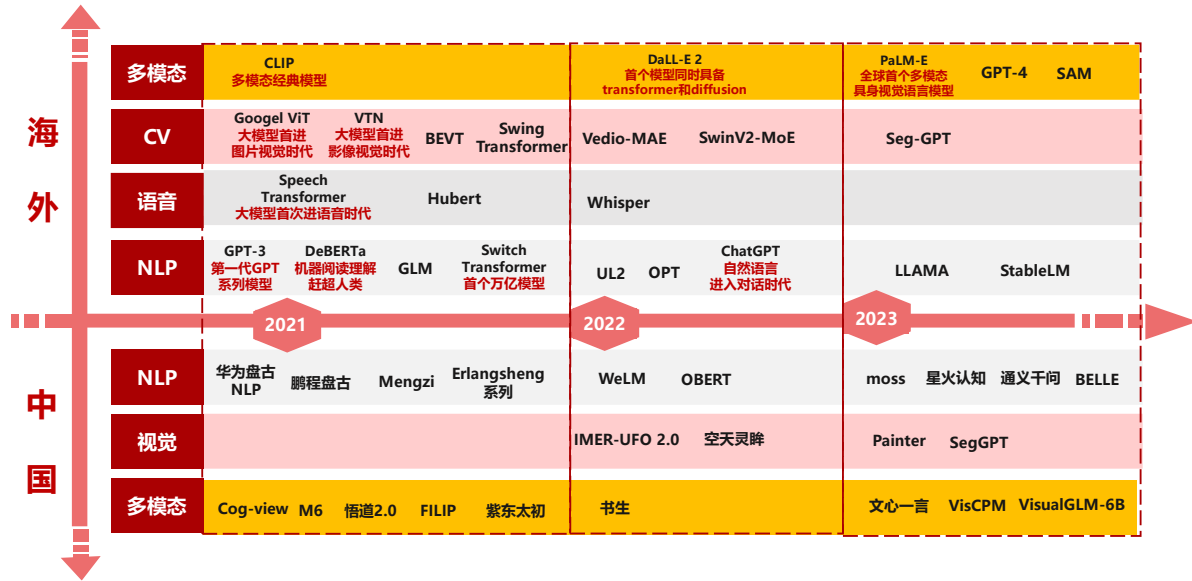
## 2.2 多模态趋势显著，提升数据预训练的重要性

多模态预训练数据是解决 AI 应用长尾问题的关键，大模型与垂直领域的产业结合趋势带来的多模态技术迁移，将进一步提升预训练数据环节的重要性。

海内外模型多模态趋势显著，输入数据从海量语言信息、文本信息，发展为多类垂直领域的多模态数据。叠加模型底层是模型通过对指令的理解，建立起不同模态数据，如：文本、语音、视频、图像等数据的关键特征，并建立多维映射。因

此，模型训练和优化过程需要海量的多模态数据。数据预训练通过对非结构化多模态数据进行跨模态特征的提取、对齐和融合，解决产业结合下多模态数据难以有效识别和语义信息深度利用的痛点。

图12.海内外大模型均趋向多模态



数据来源：中国科学技术部，财通证券研究所

模型趋向特定方向上的优化迭代，竞争焦点从参数规模转移到与数据质量。

伴随大模型与垂直领域的产业结合更多模型或将采用类强化学习模式来进行特定领域或特定方向上的优化迭代，因此，在模型预训练环节、微调环节，高质量的标注过的指令数据是模型精确度、泛化能力的基础。

竞争焦点从参数规模的竞争到数据质量的竞争。如大模型诞生初期，主流观点是参数规模是模型效果增强的核心要素，模型参数越大，性能表现越好，当前这一观点正逐渐被打破。如 Llama-13B 参数规模为 GPT-3 的 1/13，依靠模型训练数据规模，最终常识推理、闭卷问答、阅读理解等方面表现略优于 GPT-3。

表3.Llama-13B 与 GPT-3 数据对比

模型	参数规模	训练数据规模
GPT-3	175B	570GB
Llama-13B	13B	4.7TB+BPE分词(1T tokens)

数据来源：《LLaMA: Open and Efficient Foundation Language Models》(Meta AI)，财通证券研究所

表4.Llama-13B 与 GPT-3 性能测试对比

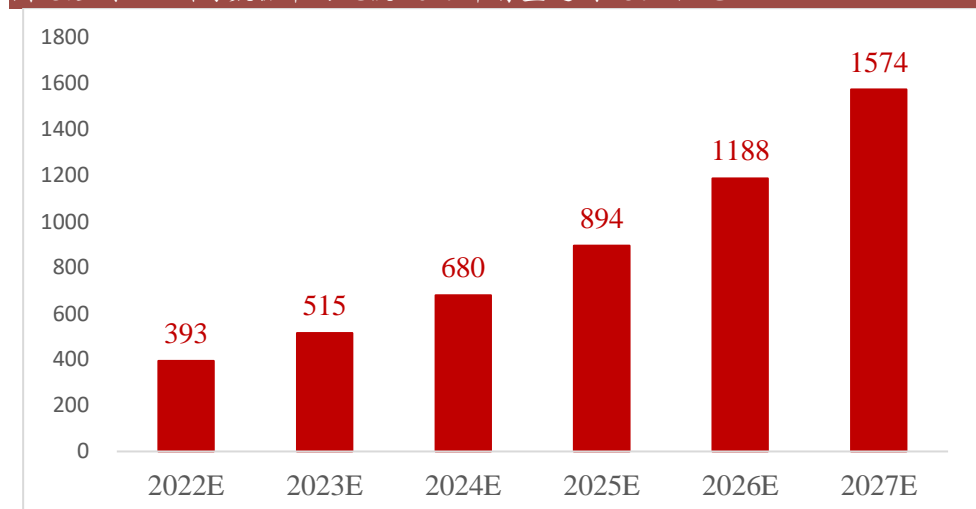
常识推理			
	BoolQ	HellaSwag	ARC-e
GPT-3	60.5	78.9	68.8
Llama-13B	78.1	79.2	74.8
闭卷问答			
	0-shot	1-shot	64-shot
GPT-3	14.6	23	29.9
Llama-13B	20.1	23.4	31.9
阅读理解			
	RACE-middle	RACE-high	
GPT-3	58.4	45.5	
Llama-13B	61.6	47.2	

数据来源：《LLaMA: Open and Efficient Foundation Language Models》(Meta AI)，财通证券研究所

### 2.3 场景落地，驱动数据预训练需求释放

模型演变趋向算法功能至上，数据成为场景落地重要的推动力量，全球范围内 AI 预训练数据需求快速增长。ChatGPT 诞生初期，模型预训练数据为截止到 2019 年 5 月的历史存量数据，本来伴随模型在垂直领域的结合，海量终端全新数据的标注需求有望释放。根据 Cognilytica 预测，2022E 全球 AI 训练数据市场规模为 393 亿元，2027E 年有望达到 1574 亿元，2022E-2027E 全球市场五年复合增速 31.98%。

图 13.全球 AI 训练数据市场规模 2027 年有望达到 1574 亿元



备注：美元兑人民币汇率采用 2023.11 月平均汇率 7.15  
数据来源：wind，中国人民银行，Cognilytica，财通证券研究所



智能驾驶是数据预训练未来五年弹性最大的应用场景，2022E-2027E 五年复合增速为 37%。根据德勤测算，2022E 年中国基础数据服务行业市场规模为 45 亿元，预计 2027E 年最高有望达到 160 亿元，2022E-2027E 五年复合增速为 29%。从终端场景出发，当前自动驾驶、智慧工业、互联网内容等终端场景占比数据服务行业较大市场份额。

车型迭代进展、量产进度、渗透率三个因素将有望催化数据处理需求呈现指数级增长。(1) 车型迭代过程中，不同传感器配置需要基础数据服务商定制不同的数据解决方案。(2) 量产进度带来终端场景数据加工规模的指数级增加。(3) 渗透率的提高加深了场景数据的复杂性，针对复杂场景多模态数据的加工精度对数据标注提出更高的要求。

表5.智能驾驶是数据预训练未来五年弹性最大的应用场景

	自动驾驶	智慧工业	互联网内容
<b>采集数据</b>	不同天气下的道路图像和视频等	生产环境,设备运行画面,产品状态图像等	用户端生成文章,搜索,直播,视频,图像等
<b>数据标注</b>	标注行驶车辆,路人等	标注生产状况,产品状态等级	标注敏感文字,视频图片人行为手势等动作
<b>潜在增长量</b>	功能迭代和场景拓展需海量数据	工业视觉为主要动能,数据需求存在一定增量空间	技术走向为无监督训练,长期看数据需求先增后跌
<b>当前规模(亿元)</b>	17	7	5
<b>2027年规模(亿元)</b>	83	26	11
<b>未来增速(%)</b>	37%	29%	20%
<b>市场份额</b>	增长	持平	下降

备注：2027 年规模为乐观预测下最高有望达到数值  
 数据来源：德勤，财通证券研究所

数据推动应用端和模型迭代的共振，未来有望反向赋能场景落地。数据预训练是整个大模型训练的知识灌输阶段，数据标注服务商为大模型提供大量标签数据，保证模型真正学习产业核心数据知识，进一步加深产业适配。高质量预训练数据集是大模型自反馈强化学习机制的前提，加速大模型迭代齿轮运转，长期实现场景落地和模型迭代的共振。

### 3 布局自动驾驶，快速铸就行业壁垒

#### 3.1 发力智能标注，助力降本增效

智能工具辅助标注，助力数据预训练业务降本增效。

公司通过智能预标注、人机交互辅助标注、智能质检等，实现降低人员成本，同时也能够快速提升 AI 模型能力。当前人工智能标注众包平台、人工智能巡检



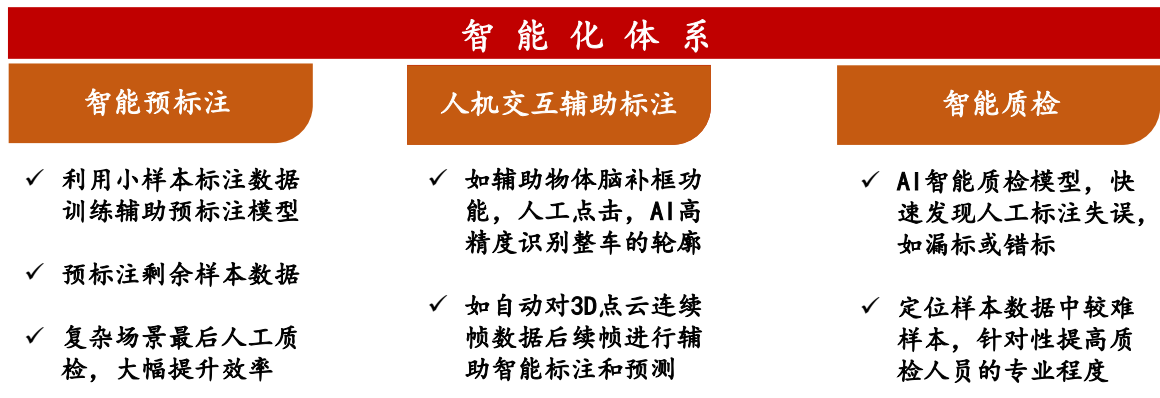
审核平台均进入运营阶段。

(1)人机交互辅助标注：公司上线 Enable AI 智能化标注平台，人机交互模式，提升复杂终端场景数据标注的效率和准确度。如人工点击，AI 高精度识别整车的轮廓；针对 3D 点云数据，自动对连续帧数据的后续帧进行智能标注和预测。

(2)智能预标注：针对较简单场景的数据，首先利用人工标注的小样本数据训练辅助预标注模型，其次模型预标注剩余样本数据，最后人工质检。

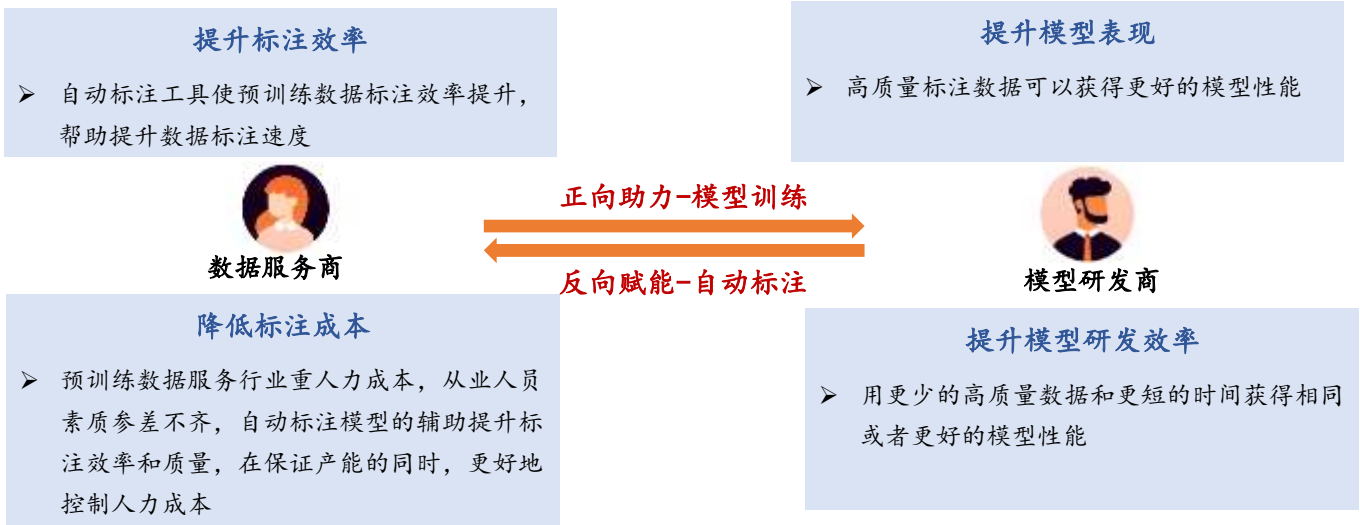
(3)智能质检：通过 AI 智能质检模型的巡检，一方面发现人工标注失误，如说漏标或错标，提高数据标注质量。另一方面定位样本数据中较难样本，针对性提高质检人员的专业程度，进而提升质检效率。

图14.公司的智能化体系



数据来源：热热文化官网，财通证券研究所

智能标注助力数据预训练和模型开发互相赋能，海内外龙头纷纷切入，降本增效效果显著。智能辅助标注一方面助力数据预训练服务商提升标注效率，降低标注成本；另一方面赋能模型开发商提升模型性能，提高模型研发效率，实现闭环。当前行业自动标注趋势显著，海内外龙头纷纷切入，降本增效效果显著。如海天瑞声上线一体化智能数据处理平台，并接入开源大模型；澳鹏自研智能辅助标注平台，通过数据预标注可提升效率 91.5%。

**图15.AI智能标注助力专业预训练数据服务商和模型开发商互相赋能，形成闭环**

**AI助力预训练数据服务商和模型开发商互相赋能，形成闭环**

AI辅助标注工具	AI自动预标注
<ul style="list-style-type: none"> <li>自动化程度较低</li> <li>人工标注为主，标注过程使用AI辅助工具，实现自动贴边、自动分割，减少人工操作环节，提升标注效率</li> </ul>	<ul style="list-style-type: none"> <li>自动化程度较高</li> <li>AI算法直接生成初步标注结果，后续进行人工审核和交叉验证</li> </ul>

数据来源：德勤，汇洲智能公司公告，汇洲智能投资者问答平台，财通证券研究所

**表6.数据预训练头部企业纷纷切入智能化平台**

企业	自动化平台
汇洲智能	自研Enable AI是高效、智能的数据标注系统，提供多种标注工具和流程管理功能，支持各种数据类型和标注任务，帮助企业实现降本增效。
海天瑞声	<b>具备自研和API接入的一体化智能数据处理平台</b> ：数据处理平台接入开源大模型，如智能驾驶平台DOTS-AD中接入开源的语义分割模型SAM(Segment Anything Model)，并对模型优化升级，有效提升了2D语义分割项目中的降本增效能力。
澳鹏	<b>自主研发的人工智能辅助数据标注平台，预标注提升标注员效率91.5%</b> ：如2D图像标注，3D点云标注，语音标注和文本标注等，对点、线、框标注，人脸关键点标注，语义分割，拉框标注，语音切分转写，NER等均有工具套组支持。

数据来源：海天瑞声公告，IDC，海天瑞声CEO公开演讲，澳鹏官方公号，澳鹏官网，热热文化官网，财通证券研究所

### 3.2 切入自动驾驶，享受更多行业增量

平台化统筹、智能化、技术先进性是自动驾驶场景数据预训练的的准入壁垒，公司始终紧跟 AI 大模型应用场景的转变，依靠技术、产品、研发优势，抢先布

局自动驾驶赛道，将会更好掌握智能驾驶数据市场的主动权，切分更多行业增量空间。

**(1) 平台化统筹能力：**公司针对自动驾驶场景，已成功积累项目统筹管理经验和数据预训练处理经验。当前完善的人员管理架构助力大型项目人员的高效分配；针对数据多样性和复杂性，设置合适的置信区间、算法引擎投票机制、置信区间等，提高数据预训练的质量和效率。

**(2) 智能化水平：**将激光雷达和深度摄像头等终端传感器获取的点云数据进行标注，可助力自动驾驶舱外算法和服务机器人实现预判，如精准的环境感知、高效的路径规划、可靠的障碍物检测，最终助力算法的行为决策。截止当前，公司智能辅助标注工具已实现点云连续帧、点云融合标注，为自动驾驶场景构建更加全面的三维环境模型。

表7.Enable AI 智能标注平台支持多类型标注工具、标注任务

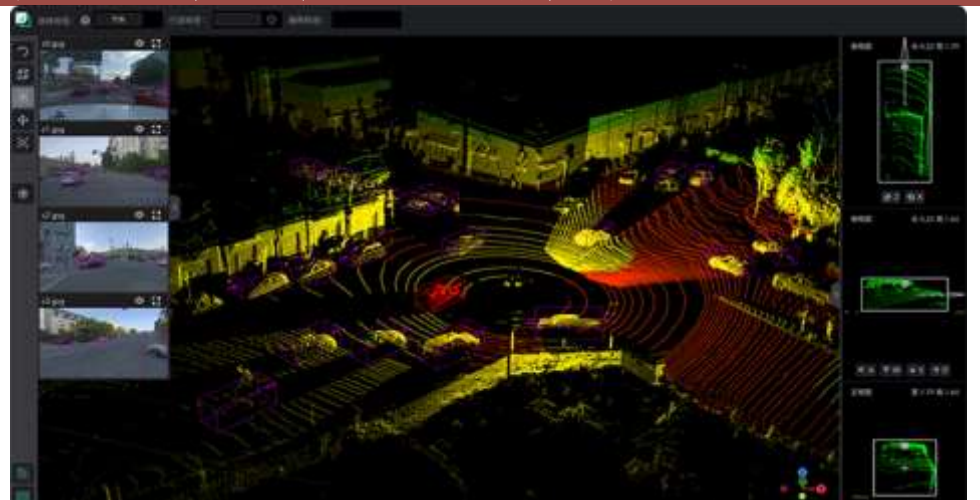
标注工具	自动驾驶场景	Enable AI智能平台
3D点云	标注点云数据中图像的形状、位置、方向、大小	
2D语义分割	标记城市规划部门、医疗机构、车企等行业数据	
车道线标注	标记交通管理系统、地图绘制、导航机构、车企、驾校等行业数据	
物体识别标注	标记车企、安防系统、医疗机构、机器人制造等行业数据	

数据来源：热热文化官网，财通证券研究所

### (3) 技术先进性

终端场景的变动，驱动数据预训练企业对终端软件层算法趋势和硬件层数据采集趋势形成自身洞察，前瞻性研发布局，不断迭代标注技术，持续保持技术先进性。截至目前，公司已具备自动驾驶方案解决能力，具备舱内语音、舱外图像、视频等多类型数据的标注能力。如表 5 所示，Enable AI 智能标注平台支持传感器 3D 点云数据的连续帧标注、不同数据的点云融合。

图 16.Enable AI 智能标注平台 3D 点云标注操作视图



数据来源：热热文化官网，财通证券研究所

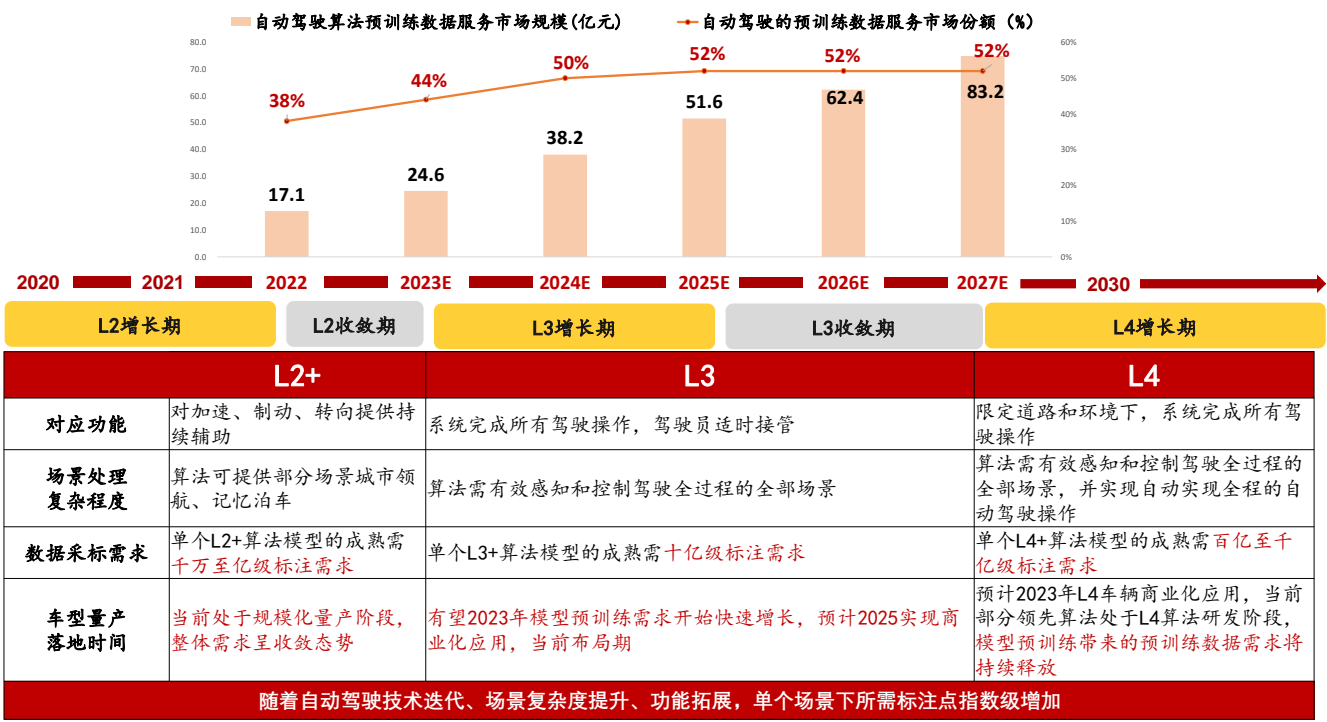
**智能驾驶场景有望领先释放预训练数据需求，需求释放进度随算法迭代和落地车型量产的时间呈现周期性收敛。**

考虑数据处理位于算法开发产业链上游，数据需求前置于终端场景，智能驾驶场景有望领先释放预训练数据服务行业。根据德勤测算，2027E 年智能驾驶带来的 AI 预训练数据服务需求有望达到 83 亿元，2022E-2027E 五年复合增速为 37%，占据市场份额 52%。

2022 年自动驾驶处于研发并推进 L2+级别的自动驾驶落地，2025 年有望实现 L3 级别自动驾驶的商业应用，2030 年有望实现 L4 级别自动驾驶的逐步落地。因此，目前将继续受益于 L2+向 L3 技术迭代带来的数据需求放量，2025 年之后基础数据需求将开始相对收敛。2027 年开始 L3+向 L4 升级迭代，算法迭代提升应用场景的复杂性，AI 预训练数据处理需求指数级上升，2027 年需求或将开始新一轮逐步释放。



图17.数据需求释放进度随算法迭代和落地车型量产的时间呈现周期性收敛



数据来源：德勤，财通证券研究所

## 4 盈利预测与估值

### 4.1 盈利预测

#### 主营业务收入假设：

(1) **机床业务**：2022 年公司该部分业务收入为 6.09 亿元，增长 28.61%。公司位于机床设备制造第一梯队，技术壁垒稳固。根据公司历史两年的销量数据，假设 2023-2025 年销量为 500、512、515 台，对应该部分业务营收分别为 7.61、9.29、11.33 亿元，对应增长率分别为 24.89%、22.21%、21.87%

(2) **数据预训练业务**：该部分业务主要为网络内容审核和大模型预训练数据标注，当前体量占比营收较小但弹性可观，有望成为未来业务增长的最大动能。当前自动标注模型研发进展顺利，叠加下游客户拓展已初见成效，我们假设 2023-2025 年该部分业务增速回调至 65.13%、99.15%、60.21%，对应 2023-2025 年营收为 0.50、1.00、1.60 亿元，对应增长率分别为 65.13%、99.15%、60.21%。

(3) **其他业务**：2022 年公司该部分业务收入和上年相比呈微增趋势，变化相对稳定。假设未来继续保持平稳增速，我们预测公司 2023-2025 年该部分业务营收

分别为 0.75、0.77、0.79 亿元。

综上预计 2023-2025 年营收为 8.86、11.06、13.72 亿元，同比增长 26.11%、24.83%、24.01%。

毛利率假设：

(1) **机床制造业务**：公司该部分业务工艺成熟，存在规模效应，且当前公司横向发力向中高端市场切换，毛利率有望持续改善，预计 2023-2025 年分别为 22.99%、24.23%、25.12%。

(2) **数据预训练业务**：公司该部分业务有望受智能驾驶场景应用的催化，业务从互联网内容标注切换到高附加值的预训练数据标注，未来随着自研模型落地，有望实现半自动化标注，进一步提升毛利率。预估毛利率相对稳定，预计 2023-2025 年分别为 21.12%、22.85%、25.16%。

(3) **其他业务**：假设毛利率水平平稳变动，预计 2023-2025 年为 22.86%、21.73%、20.92%。

综上，考虑业务占比推算公司 2023-2025 综合毛利率为 22.87%、23.93%、24.88%。

表8.盈利预测（单位：百万元、%）

	2021	2022	2023E	2024E	2025E
收入合计	837.65	702.66	886.10	1106.12	1371.75
yoy	16.06%	-16.12%	26.11%	24.83%	24.01%
毛利率	30.70%	22.64%	22.87%	23.93%	24.88%
机床制造业务	473.49	608.97	760.54	929.46	1132.74
yoy	50.08%	28.61%	24.89%	22.21%	21.87%
毛利率	23.66%	22.99%	22.99%	24.23%	25.12%
模型预训练业务	159.54	30.39	50.18	99.93	160.10
yoy	-20.98%	-80.95%	65.13%	99.15%	60.21%
毛利率	59.65%	20.45%	21.12%	22.85%	25.16%
其他业务	204.62	63.31	75.38	76.72	78.91
yoy	0.12%	-69.06%	19.06%	1.79%	2.85%
毛利率	24.42%	20.38%	22.86%	21.73%	20.92%

数据来源：wind，财通证券研究所

## 4.2 估值与投资建议

公司当前主营业务为机床业务，故选高端设备制造龙头秦川机床、海天精工和亚威股份。海天精工为最新预测数据，秦川机床和亚威股份为 wind 一致预测数据，

2023-2025 年可比公司的 PE 平均数值为 36.40X、27.32X、21.77X。

公司在传统核心机床制造业务技术壁垒稳固，优势明显。新切入基础数据服务赛道受益大模型的长坡厚雪，叠加公司前瞻性卡位智能数据标注和汽车自动驾驶，有望切分更多行业份额。综上所述，我们预计公司 2023-2025 年实现营业收入 8.86/11.06/13.72 亿元，归母净利润 1.82/2.66/3.54 亿元，对应 PE 分别为 41.69/28.53/21.45 倍，维持“增持”评级。

表9.可比公司估值对比分析

证券名称	证券代码	股价(元)	EPS(元)			PE(倍)		
			2023E	2024E	2025E	2023E	2024E	2025E
秦川机床	000837.SZ	10.64	0.24	0.34	0.45	44.57	31.16	23.39
海天精工	601882.SH	23.92	1.17	1.43	1.74	20.44	16.73	13.75
亚威股份	002559.SZ	11.11	0.25	0.33	0.39	44.17	34.08	28.18
平均	-	15.22	0.55	0.70	0.86	36.40	27.32	21.77
汇洲智能	002122.SZ	3.80	0.09	0.13	0.18	41.69	28.53	21.45

备注：收盘价日期为 2023.12.25，海天精工为最新预测数据，秦川机床和亚威股份为 wind 一致预期数据  
 数据来源：wind，财通证券研究所

## 5 风险提示

### 1、大模型产业结合不及预期

大模型产业结合的落地进度受到模型迭代水平，产业数据确权，产业数据采集标注难度增加等多方面的压力，模型产业结合任重道远。

### 2、智能数据标注市场竞争加剧

国内智能数据标注市场参与者为头部数据服务公司，大厂自研数据众包平台等，随着市场放量，多家先后自研数据标注模型，抢占先发红利，竞争逐渐加剧。

### 3、智能标注平台落地不及预期

公司自研的智能标注大模型仍在研发过程中，存在技术难关攻克难以匹及预期的风险。



**公司财务报表及指标预测**

利润表 (百万元)						财务指标					
	2021A	2022A	2023E	2024E	2025E		2021A	2022A	2023E	2024E	2025E
<b>营业收入</b>	<b>837.65</b>	<b>702.66</b>	<b>886.10</b>	<b>1106.12</b>	<b>1371.75</b>	<b>成长性</b>					
减:营业成本	580.52	543.55	683.42	841.40	1030.42	营业收入增长率	16.06%	-16.12%	26.11%	24.83%	24.01%
营业税费	11.04	12.48	13.56	16.48	19.89	营业利润增长率	-39.74%	327.69%	-71.68%	37.56%	30.05%
销售费用	21.55	29.92	42.62	52.21	64.33	净利润增长率	554.78%	111.70%	114.19%	46.11%	33.04%
管理费用	119.96	150.67	149.66	181.51	224.01	EBITDA 增长率	-6.34%	-144.46%	452.67%	36.91%	28.07%
研发费用	25.77	54.22	43.51	53.30	65.71	EBIT 增长率	-11.88%	-208.47%	283.13%	39.18%	30.65%
财务费用	21.56	18.76	20.16	21.22	22.45	NOPLAT 增长率	-6.70%	-256.93%	-241.37%	40.88%	30.97%
资产减值损失	-16.48	-932.29	-29.26	-18.24	-4.33	投资资本增长率	-41.03%	112.00%	8.19%	9.94%	11.91%
加:公允价值变动收益	-0.91	-20.26	—	—	—	净资产增长率	-46.78%	145.07%	8.02%	10.08%	12.21%
投资和汇兑收益	104.52	1711.88	251.18	300.77	356.41	<b>利润率</b>					
<b>营业利润</b>	<b>150.33</b>	<b>642.93</b>	<b>182.11</b>	<b>250.51</b>	<b>325.78</b>	毛利率	30.70%	22.64%	22.87%	23.93%	24.88%
加:营业外净收支	-841.71	-599.55	-30.25	-32.31	-35.41	营业利润率	17.95%	91.50%	20.55%	22.65%	23.75%
<b>利润总额</b>	<b>-691.39</b>	<b>43.38</b>	<b>151.86</b>	<b>218.21</b>	<b>290.37</b>	净利率	-83.67%	9.05%	19.40%	22.61%	24.32%
减:所得税	9.51	-20.24	-20.08	-31.85	-43.18	EBITDA/营业收入	14.63%	-7.75%	21.68%	23.78%	24.56%
<b>净利润</b>	<b>-727.49</b>	<b>85.10</b>	<b>182.29</b>	<b>266.34</b>	<b>354.33</b>	EBIT/营业收入	10.34%	-13.37%	19.41%	21.65%	22.80%
资产负债表 (百万元)						运营效率					
	2021A	2022A	2023E	2024E	2025E		2021A	2022A	2023E	2024E	2025E
货币资金	241.99	342.58	811.01	1266.42	1818.48	固定资产周转天数	134.55	158.73	121.81	94.03	73.34
交易性金融资产	72.19	417.62	923.03	718.03	513.03	流动营业资本周转天数	-235.74	-333.15	154.08	60.76	-36.16
应收账款	148.03	111.11	157.53	189.20	234.76	流动资产周转天数	774.19	830.04	999.38	1101.42	1044.29
应收票据	17.26	5.08	12.10	14.02	16.59	应收帐款周转天数	68.86	66.38	54.57	56.42	55.63
预付帐款	37.65	34.68	44.45	54.35	66.58	存货周转天数	297.71	392.17	362.28	347.17	354.85
存货	523.99	660.25	715.25	907.55	1123.80	总资产周转天数	1842.22	1959.73	1758.75	1768.47	1606.23
其他流动资产	119.93	13.81	113.81	93.81	73.81	投资资本周转天数	661.06	912.07	1023.16	894.23	800.17
可供出售金融资产	—	—	—	—	—	<b>投资回报率</b>					
持有至到期投资	—	—	—	—	—	ROE	-98.96%	4.04%	7.92%	10.36%	12.12%
长期股权投资	974.59	752.46	901.76	1027.94	1103.77	ROA	-17.61%	2.42%	3.55%	4.65%	5.44%
投资性房地产	17.06	16.47	21.87	27.21	31.58	ROIC	7.69%	-5.69%	7.44%	9.53%	11.16%
固定资产	313.31	306.32	293.35	284.48	274.40	<b>费用率</b>					
在建工程	10.88	5.60	5.65	5.29	4.24	销售费用率	2.57%	4.26%	4.81%	4.72%	4.69%
无形资产	229.76	152.90	122.64	103.27	96.75	管理费用率	14.32%	21.44%	16.89%	16.41%	16.33%
其他非流动资产	1.39	0.14	0.14	0.14	0.14	财务费用率	2.57%	2.67%	2.28%	1.92%	1.64%
<b>资产总额</b>	<b>4130.23</b>	<b>3519.93</b>	<b>5138.01</b>	<b>5729.38</b>	<b>6511.42</b>	三费/营业收入	19.47%	28.37%	23.98%	23.05%	22.66%
短期债务	112.58	109.14	122.27	131.59	140.05	<b>偿债能力</b>					
应付帐款	209.37	271.80	300.63	375.88	467.13	资产负债率	77.26%	34.60%	40.86%	44.49%	45.28%
应付票据	1.55	11.76	8.00	11.37	14.70	负债权益比	339.69%	52.90%	84.42%	93.11%	95.99%
其他流动负债	133.21	23.93	18.93	13.93	8.93	流动比率	0.49	1.51	1.53	1.44	1.49
长期借款	—	—	—	—	—	速动比率	0.24	0.91	1.10	1.02	1.05
其他非流动负债	—	—	—	—	—	利息保障倍数	2.53	-3.08	5.16	7.50	9.58
<b>负债总额</b>	<b>3190.89</b>	<b>1217.86</b>	<b>2099.35</b>	<b>2548.76</b>	<b>2948.63</b>	<b>分红指标</b>					
少数股东权益	204.19	194.40	184.05	167.77	146.99	DPS(元)	—	—	—	—	—
股本	1210.47	1987.39	1994.01	1994.01	1994.01	分红比率	—	—	—	—	—
留存收益	-515.32	-430.21	-247.22	19.82	374.85	股息收益率	—	—	—	—	—
<b>股东权益</b>	<b>939.35</b>	<b>2302.07</b>	<b>2486.68</b>	<b>2737.44</b>	<b>3071.69</b>	<b>业绩和估值指标</b>					
<b>现金流量表 (百万元)</b>						2021A	2022A	2023E	2024E	2025E	
净利润	-727.49	85.10	182.29	266.34	354.33	EPS(元)	-0.37	0.04	0.09	0.13	0.18
加:折旧和摊销	35.93	39.46	20.11	23.63	24.08	BVPS(元)	0.61	1.06	1.15	1.29	1.46
资产减值准备	21.63	980.77	29.26	18.24	4.33	PE(X)	—	72.22	41.69	28.53	21.45
公允价值变动损失	0.91	20.26	—	—	—	PB(X)	4.71	2.94	3.30	2.96	2.60
财务费用	20.74	23.19	33.36	31.94	32.65	P/FCF	—	—	—	—	—
投资收益	-104.52	-1711.88	41.12	52.89	63.08	P/S	4.13	8.82	8.55	6.85	5.52
少数股东损益	26.60	-21.48	-10.35	-16.28	-20.78	EV/EBITDA	27.93	-109.68	36.01	24.61	17.60
营运资金的变动	811.13	629.72	-50.32	-50.32	-50.32	CAGR(%)	—	—	—	—	—
<b>经营活动产生现金流量</b>	<b>74.50</b>	<b>25.68</b>	<b>273.93</b>	<b>357.19</b>	<b>440.89</b>	PEG	—	0.65	0.37	0.62	0.65
<b>投资活动产生现金流量</b>	<b>206.86</b>	<b>-133.11</b>	<b>202.92</b>	<b>120.14</b>	<b>134.65</b>	ROIC/WACC	—	—	—	—	—
<b>融资活动产生现金流量</b>	<b>-352.28</b>	<b>86.35</b>	<b>-9.11</b>	<b>-22.62</b>	<b>-24.19</b>	REP	—	—	—	—	—

资料来源: wind, 财通证券研究所

## 信息披露

### ● 分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格，并注册为证券分析师，具备专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解。本报告清晰地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，作者也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。

### ● 资质声明

财通证券股份有限公司具备中国证券监督管理委员会许可的证券投资咨询业务资格。

### ● 公司评级

以报告发布日后 6 个月内，证券相对于市场基准指数的涨跌幅为标准：

买入：相对同期相关证券市场代表性指数涨幅大于 10%；

增持：相对同期相关证券市场代表性指数涨幅在 5%~10% 之间；

中性：相对同期相关证券市场代表性指数涨幅在 -5%~5% 之间；

减持：相对同期相关证券市场代表性指数涨幅小于 -5%；

无评级：由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。

A 股市场代表性指数以沪深 300 指数为基准；香港市场代表性指数以恒生指数为基准；美国市场代表性指数以标普 500 指数为基准。

### ● 行业评级

以报告发布日后 6 个月内，行业相对于市场基准指数的涨跌幅为标准：

看好：相对表现优于同期相关证券市场代表性指数；

中性：相对表现与同期相关证券市场代表性指数持平；

看淡：相对表现弱于同期相关证券市场代表性指数。

A 股市场代表性指数以沪深 300 指数为基准；香港市场代表性指数以恒生指数为基准；美国市场代表性指数以标普 500 指数为基准。

### ● 免责声明

本报告仅供财通证券股份有限公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告的信息来源于已公开的资料，本公司不保证该等信息的准确性、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投资标的邀请或向他人作出邀请。

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。

本公司通过信息隔离墙对可能存在利益冲突的业务部门或关联机构之间的信息流动进行控制。因此，客户应注意，在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的情况下，本公司的员工可能担任本报告所提到的公司的董事。

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。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司不对任何人使用本报告中的任何内容所引致的任何损失负任何责任。

本报告仅作为客户作出投资决策和公司投资顾问为客户提供投资建议的参考。客户应当独立作出投资决策，而基于本报告作出任何投资决定或就本报告要求任何解释前应咨询所在证券机构投资顾问和服务人员的意见；

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用，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。