



上海证券
SHANGHAI SECURITIES

证券研究报告

2023年5月9日

行业：传媒

增持（维持）

图像篇专题（一）：开源模型高速迭代，Stable Diffusion促进AI技术民主化

分析师：陈旻 SAC编号：S0870522020001

主要观点

我们将“AI+传媒”的研究框架体系定义为“通用大模型”+“行业小样本”的技术架构，“AI+传媒”在应用层表现效力优劣的关键取决于通用大模型对垂直应用的适配程度及迭代速度，

1、适配程度是指：多模态的输入及输出是否匹配应用层的输入及输出。比如GPT-4属于“图+文”多模态输入+“文”单模态输出，因此输入模态为“图或文”且输出模态为“文”的垂直应用更适配GPT-4。

2、迭代速度是指：应用层产生的“行业小样本”的数据量是否匹配大模型的迭代要求。根据我们对GPT模型的理解，比如Bing AI产生的“行业小样本”源自Bing的搜索结果，ChatGPT产生的“行业小样本”源自用户的反馈和互动。因此我们认为，对于超出GPT所使用的预训练数据库范围（2021年9月前）的事实性表述，Bing AI反馈的是搜索的结果，ChatGPT反馈的是用户主动的观点，Bing AI反馈的效果比ChatGPT更好。

我们认为“行业小样本”的价值取决于数据数量及数据质量，数量大且质量高（多模态）的应用场景复用及迭代AI能力的效力更强，因此更进一步理解我们的研究框架，我们将“行业小样本”的结构分层（中层小模型+下层应用及内容），并将“行业小样本”的结合方式分类（调用+训练）：

1、“行业小样本”的数据集来自小模型或应用及内容：AI产业链包括上层大模型、中层小模型、下层应用及内容，包括应用及内容直接接入大模型或通过小模型接入大模型两种方式，即“大模型+应用及内容”或“大模型+小模型+应用或内容”，其中具备特定功能的AIGC软件产品及MaaS我们理解为“小模型”+“应用”的技术范式，本身具备较高质量的AI能力，若接入匹配的多模态大模型，有望实现能力上的质变突破。



2、“行业小样本”的结合方式包括“能力调用”及“能力训练”两类：

（1）“能力调用”是指下游垂类场景直接调用通用大模型的通用能力，并基于垂类场景内产生的特性化数据不断提升调用能力在垂类场景内的适配程度。我们认为现阶段下游应用及内容主要采取此类方式接入大模型能力，此类方式可高效快速调用大模型先进能力，在时间上及成本上具备优势。我们认为“能力调用”匹配“AI+传媒”的第一层利好，即通过AI降本增效，大幅提高数据及内容的供给量。内容产业本质由供给决定需求，因此内容供给量的明显提升将有效带动传媒基本面拐点及增量空间出现。

（2）“能力训练”是指下游垂类场景将通用大模型针对特性化数据集进行再训练，从而形成垂类场景专属大模型。例如彭博社利用自身丰富的金融数据源，基于开源的GPT-3框架再训练，开发出了金融专属大模型BloombergGPT。我们认为“能力训练”匹配“AI+传媒”的第二层利好，即下游垂类场景本身的数据或内容反过来“再训练”通用大模型（或开源大模型），形成传媒内容场景专属大模型，形成更稳定且高质的内容输出。我们认为训练难度文本<图片<视频<影视<游戏，且内容数量逐步递减但内容质量逐步递增，即偏后端的影视、游戏在内容数量上训练量级不足，因此高质量的内容形态首先通过“能力调用”输出AIGC内容，再将AIGC内容“再训练”大模型以解决高质量内容数量不足的问题（合成数据“再训练”范畴）。



主要观点

从投资的角度，按照我们的研究框架，传媒对应垂类场景的“行业小样本”，其核心价值取决于数据与内容，第一层对应数据与内容的输入模态是否匹配大模型的输出模态；第二层对应数据与内容的数量及质量是否匹配大模型的能力再训练：

1、按照“模态匹配”的逻辑，AI+文本/虚拟人预计率先兑现案例及业绩，其次AI+图片可通过“大模型”+“小模型”组合方式实现（如GPT+Stable Diffusion、GPT+Midjourney）。随着未来GPT-5提供更多模态的输入及输出，下游垂类场景的适配范围有望扩大，通过“能力调用”适配的应用及内容场景更为丰富，因此后续“AI+视频/影视/游戏”的案例兑现度存在新的催化空间。

OpenAI 最新发布的GPT-4核心特征包括：（1）多模态输入（图+文），单模态输出（文），可以阅读并总结论文内容、解答较高难度的物理题目、具备较强的OCR能力（如识别网页草稿并按要求反馈网页代码）、理解人类社会常识；（2）具备长文字处理及推理判断能力，GPT-4上下文上限约2.5万字，允许使用长格式内容创建、扩展对话以及文档搜索和分析等，能够阅读并记忆更多信息，且具备更高的推理判断能力；（3）可靠性大幅提升，分辨能力提高，有效减少“虚构”或“有害”信息输出。

2、按照“能力再训练”的逻辑，AI+内容/IP预计空间及价值更大，其价值核心取决于数据与内容/IP的数量及质量的高低。本周，我们认为最核心意义为大幅降低垂类场景专属大模型的训练门槛，小模型层及应用层有望明显受益。掌握数据及优质内容（多模态数据）的下游场景具备核心竞争力，因此内容及IP（版权）的价值有望重估。

DeepSpeed-Chat集成预训练语言大模型完整三个步骤，其中针对第三步RLHF训练集成了高效且经济的DeepSpeed-RLHF系统，使复杂的RLHF训练变得快速、经济并且易于大规模推广（相比现有系统提速15倍以上，且大幅降低算力要求及成本）。



主要内容

本文将选取国外AI图像生成领域的龙头之一进行解析，Stable Diffusion是一个文本到图像的潜在扩散模型(Latent Diffusion Models)，于2022年8月推出。截至2022年底已被全球超20万开发者下载和授权，成为当前可用性高的开源模型。To C端，Stability AI 面向消费者的产品 DreamStudio自8月推出后迅速成长，截至2023年4月11日，已为来自50多个国家/地区的超过100万注册用户创造了超过1.7亿张图像。

- ◆ **代码技术迭代迅速，ControlNet插件促进AI绘画革命性进步。** Stable Diffusion的开源优势在于能够吸引大量的开发者，最大程度的把模型用起来，开源社区会共同推进 prompt engineering，解决技术难题，这使得代码的迭代速度非常快，优化效率远远高于闭源系统，使得文生图行业快速成长和普及。同时社区成员会创建新的UI，通过扩展现有的功能创造新的用例，在Stable Diffusion模型上创建的应用逐渐繁荣。
- ◆ **历经多次版本迭代，消费级GPU即刻快速部署使图像生成民主化。** Stable Diffusion V1在消费类GPU上运行的VRAM低于10GB，几秒钟内即可生成512*512像素的图像，显著降低部署门槛，使得图像生成民主化。2022年11月V2版本发布，从四个方面进行版本升级：1) 全新文本编码器（OpenCLIP）训练的文生图模型显著提升了生成图像的质量；2) Upscaler Diffusion模型将生成图像的分辨率提高了4倍；3) 使用Depth2img模型用来推理输入图像的深度，保持图像生成的连贯性和深度；4) 引入新的text-guided（文本引导修复模型）使得用户可以快速智能地切换图像内容。
- ◆ **商业模式仍在积极探索，版权问题未来有待解决。** 目前公司实现的营收暂不足以覆盖巨额的服务器和人才招聘的费用，而训练模型所需的计算资源和成本非常高昂，且2023年3月，Stability AI斥资收购了成像工具Init ML。版权问题是图像生成应用发展遇到的普遍问题，由于牵涉到艺术版权的问题较多，且对外开源的特性使得对图像生成的监管程度更低，未来发展中版权问题有待解决。

风险提示：

- 宏观经济风险；地缘政治风险，技术发展不及预期风险；AIGC行业发展不及预期风险



目录

Content

- 一、Stable Diffusion
- 二、投资建议
- 三、风险提示

一、Stable Diffusion

1.1 当前可用性高的开源模型，Stable Diffusion风靡开源社区

- ◆ **Stable Diffusion**是一个文本到图像的潜在扩散模型(Latent Diffusion Models)，由Runway、LMU Munich、Eleuther AI、LAION、Stability AI的研究人员和工程师创建。Stable Diffusion的项目基础是由慕尼黑大学机器视觉与学习研究小组和Runway的研究人员，基于CVPR2022的一篇论文《High-Resolution Image Synthesis with Latent Diffusion Models》，并与其他社区团队合作开发的一款开源模型。
- ◆ **Stable Diffusion 于 2022 年 8 月推出，所有人均可在本地训练和部署 AI 模型**。2022年底，Stability AI 宣布融资 1.01 亿美元，并宣称 **Stable Diffusion 已被全球超过 200,000 名开发者下载和授权，成为当前可用性高的开源模型**。而据《福布斯》报道，每天有1000万人使用 Stable Diffusion——比使用 OpenAI 的 DALL-E 2 的人数还要多。同时，Stability AI 面向消费者的产品 DreamStudio 自8月推出后迅速成长，截至2022年11月9日，已为来自 50 多个国家/地区的超过 100 万注册用户创造了超过 1.7 亿张图像。

图1: Stable Diffusion官网示例图



资料来源: Stable Diffusion官网, 上海证券研究所

图2: “A high tech solarpunk utopia in the Amazon rainforest” 命令对应图



资料来源: Stable Diffusion官网, 上海证券研究所



1.2 代码技术迭代迅速，ControlNet插件促进AI绘画革命性进步

- ◆ 开源模式吸引大量开发者，代码及技术迭代迅速，推动行业整体发展。Stable Diffusion的开源优势在于能够吸引大量的开发者，最大程度的把模型用起来。开源社区会齐心协力地完善模型文档，共同推进 prompt engineering，解决技术难题。这使得代码的迭代速度非常快，优化效率远远高于闭源系统，使得文生图行业快速成长和普及。同时社区成员会创建新的UI，通过扩展现有的功能创造新的用例，在Stable Diffusion模型上创建的应用逐渐繁荣。
- ◆ **ControlNet插件显著提升扩散模型生成实用性和稳定性，AI绘画革命性进步。**2023年2月，斯坦福大学研究人员发布论文《Adding Conditional Control to Text-to-Image Diffusion Models》并提出一种神经网络结构ControlNet来控制预训练的大型扩散模型以支持其他输入条件。ControlNet通过端到端的方式满足特定任务的条件，如线图成图、分割图成图、Pose生图等，极大提高AI生成图片的可控性和实用性。

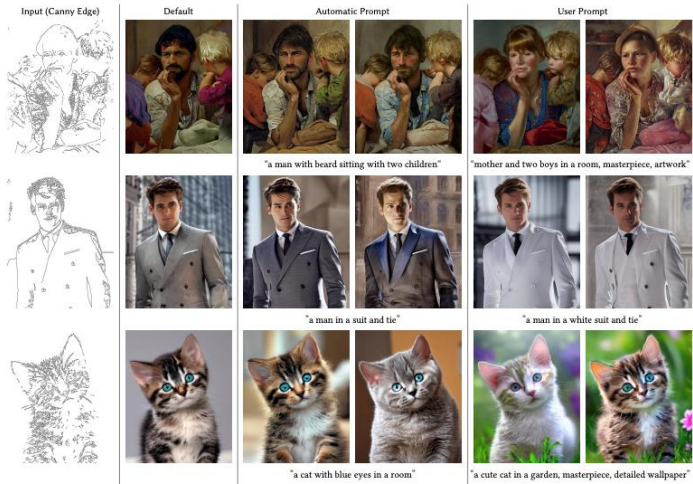
图3: 论文《Adding Conditional Control to Text-to-Image Diffusion Models》

图4: Stable Diffusion的ControlNet插件输入线图即可输出图片

Adding Conditional Control to Text-to-Image Diffusion Models

Lvmin Zhang and Maneesh Agrawala
Stanford University

资料来源:《Adding Conditional Control to Text-to-Image Diffusion Models》, Lvmin Zhang, Maneesh Agrawala, 上海证券研究所

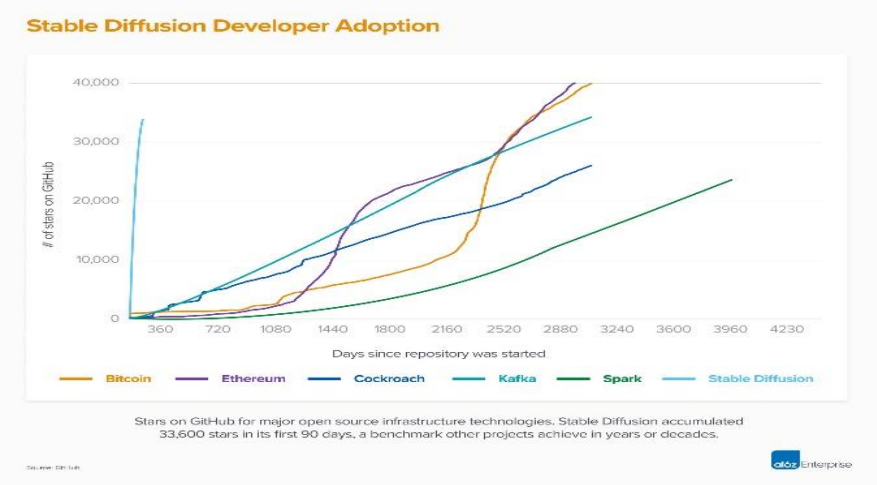


资料来源:《Adding Conditional Control to Text-to-Image Diffusion Models》, Lvmin Zhang, Maneesh Agrawala, 上海证券研究所

1.3 历经多次版本迭代，使图像生成民主化

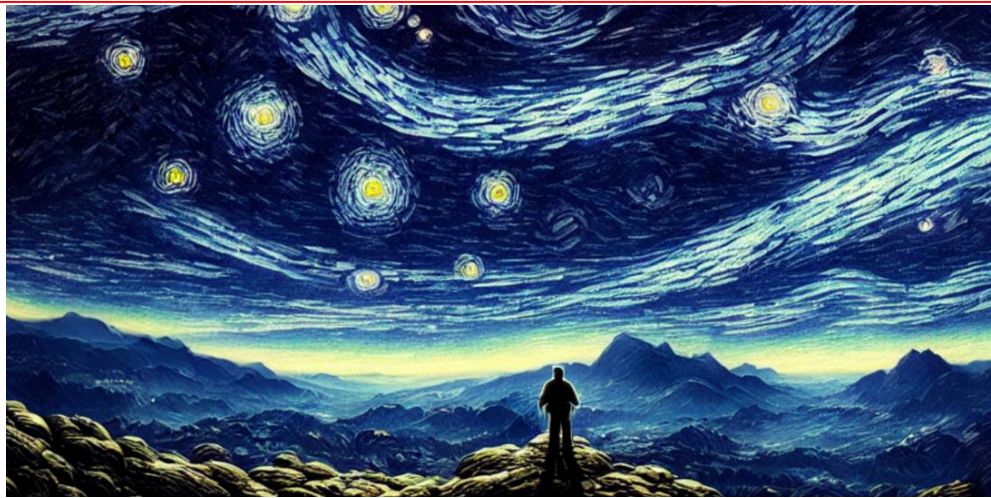
◆ 2022年8月，Stable Diffusion V1版本发布，在Github 排行榜不到两个月的时间内飙升至 33K star，使图像生成更加民主化。Stable Diffusion在消费类GPU上运行的VRAM低于10GB，几秒钟内生成512x512像素的图像。这将使研究人员和公众都可以进行文生图操作，使图像生成民主化。V1亮相之初，在Github 排行榜，Stable Diffusion为所有软件中攀升至10K star最快的其中之一，在不到两个月的时间内飙升至 33K star。

图5: Stable DifusionV1版本在Github上Stars数的变化趋势



资料来源: A16z和Github, 上海证券研究所

图6: Stable Diffusion V1版本生成图像



资料来源: Stability AI官网, 上海证券研究所



1.3 历经多次版本迭代，使图像生成民主化

- ◆ 2022年11月24日，Stable Diffusion V2.0版本发布，从四个方面对模型进行升级。
- 1) 使用全新文本编码器（OpenCLIP）训练的text-to-image模型。该编码器由LAION开发且得到Stability AI的支持，该编码器显著提高了生成图像的质量。此版本中的文本到图像模型可生成默认分辨率为512x512像素和768x768像素的图像。
- 2) 使用Upscaler Diffusion（超分辨率扩散）模型。该模型将生成图像的分辨率提高了4倍。
- 3) 使用Depth2img（depth-guided stable diffusion，即深度引导稳定扩散）模型。扩展了V1版本中image-to-image的特性，为创意应用提供了全新的可能性。该模型用来推理输入图像的深度，然后使用文本和深度信息生成新图像。Depth2img提供了不同创新性应用，虽然生成的图像与原始图像有很大的不同，但仍然保持了图像的连贯性和深度。
- 4) 引入新的text-guided（文本引导修复模型）。用户可以较为智能、快速地切换图像部分内容。

图7: Stable Diffusion 2.0版本官网示例图



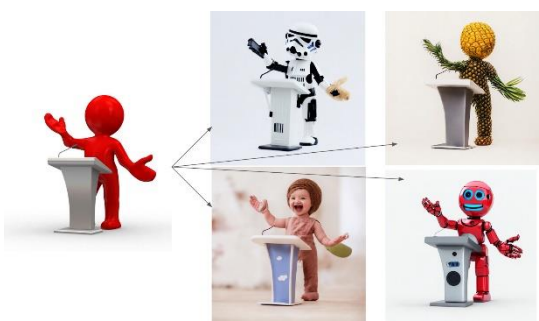
资料来源: Stability AI官网, 上海证券研究所

图8: 使用Upscaler Diffusion 模型效果对比图



资料来源: Stability AI官网, 上海证券研究所

图9: 使用Depth2img效果示例图



资料来源: Stability AI官网, 上海证券研究所



1.3 历经多次版本迭代，使图像生成民主化

- ◆ 2022年12月7日，Stable Diffusion V2.1版本发布，相对于之前发布的V2.0版本有以下三大亮点：
 - 1) **减少人像过滤**。Stability AI听取了用户的反馈，调整过滤器以减少限制，在与LAION-5B的开发者合作分析了NSFW过滤器及其对训练数据的影响后，将设置调整得更加平衡。
 - 2) **图像质量飞跃式提升**。V2.1版本相比于V2.0版本，在建筑、室内设计、野生动物和景观场景方面的图像质量上，进行了又一次飞跃。
 - 3) **加强“否定提示”使图像更精致**。该版本加强了反向提示词的应用，用于消除不需要的细节，进行图像微调，用户可以使用提示加权优化整体图像，以增加或减少合成元素，使用户能够更好地控制图像合成。

图10: V2.1版本人像示例图



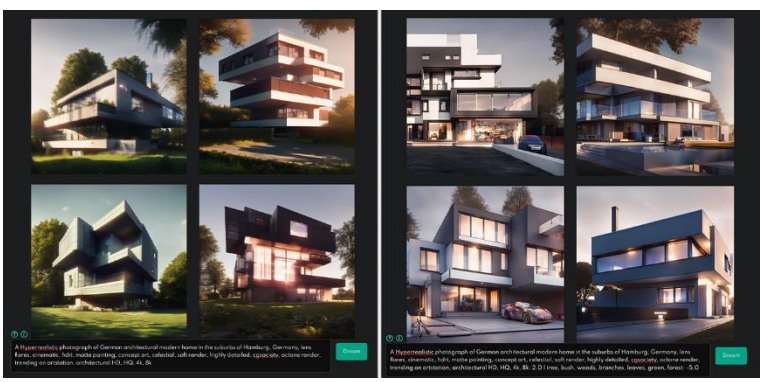
资料来源：Stability AI官网，上海证券研究所

图11: V2.1版本非标准分辨率图像渲染效果图



资料来源：Stability AI官网，上海证券研究所

图12: V2.1版本非否定提示（左）VS否定提示（右）



资料来源：Stability AI官网，上海证券研究所



1.4 商业模式仍在积极探索，版权问题未来有待解决

- ◆ **Stable Diffusion并不是Stability AI完全独立开发出来的商品，结合了两种形式的人工智能。** Stable Diffusion并不是公司 Stability AI 完全独立开发出来的商品，项目基础是由慕尼黑大学机器视觉与学习研究小组和Runway的研究人员，基于CVPR2022的一篇论文《High-Resolution Image Synthesis with Latent Diffusion Models》，并与其他社区团队合作开发的一款开源模型。在此之前，这些研究人员已经开始研究如何将自然语言描述绘制成图片，Stability AI找到并资助他们进行AI开源大模型等其他研究。
- ◆ **版权问题不仅仅是Stable Diffusion的发展的瓶颈，基本上是所有AI绘画工具的制约因素。** 由于Stable Diffusion的使用过于泛滥，牵涉到艺术版权的问题较多，对外开源的特性导致对生成图像的监管程度更低。目前AI画作版权归属不明的问题，企业如果想利用这类技术进行商业服务，就要在法务合规等方面做出额外的准备。同时，受制于To C提供服务时潜在的版权风险，企业可以更倾向于在To B场景中间接使用生成式AI的结果。比如通过其生成的图像结果辅助设计师构图，或通过其产出的代码为软件工程师提供代码设计建议，从而做到规避其结果的风险。
- ◆ **营收难以覆盖模型训练及人力成本，Stable Diffusion背后的公司Stability AI仍在探索商业模式。** 目前公司实现的营收不足以覆盖巨额的服务器和人才招募的费用。训练模型所需的计算资源和成本非常高昂。在2023年3月，Stability AI 还斥资收购了成像工具 Init ML。据彭博社报道，Stability AI 正在谋求新一轮融资，希望能将公司的估值翻4番至40亿美元。



二、投资建议

基于我们“大模型+小模型+应用及内容”的研究框架，我们认为从基本面受益的角度上看，需沿着目前成熟大模型模态输出的范围选择标的，重点关注AI+文本/虚拟人板块，其次关注AI+图片（大模型+小模型）；从价值重估的角度上看，需沿着掌握优质数据或内容（多模态数据）的范围选择标的，重点关注AI+内容/IP/版权板块。

- 1、建议关注A股稀缺的优质内容型平台公司【芒果超媒】，有望成为传媒估值中枢锚。
- 2、建议关注拥有海外用户/业务，有望接入GPT的优质标的，如【汤姆猫】（全球用户）、【昆仑万维】（Opera）、【神州泰岳】（游戏出海+NLP）、【华凯易佰】（跨境电商）、【焦点科技】（跨境电商）、【蓝色光标】（出海）。
- 3、建议关注中国版Discord【创梦天地】。
- 4、建议关注可与生成式图像AI Midjourney对标的【浙文互联】（米画）、【视觉中国】。
- 5、建议关注国内电商类的【新华都】（电商代运营）、【值得买】（内容测评）、【遥望科技】（虚拟人）、【壹网壹创】（电商代运营）、【青木股份】（电商代运营）、【若羽臣】（电商代运营）、【丽人丽妆】（电商代运营）、【返利科技】。
- 6、建议关注AI+游戏，如【姚记科技】、【盛天网络】、【三七互娱】、【完美世界】、【吉比特】、【世纪华通】、【巨人网络】；建议关注AI+影视，如【百纳千成】、【欢瑞世纪】、【光线传媒】、【华策影视】、【博纳影业】、【上海电影】、【慈文传媒】等；建议关注AI+出版，如【中国科传】、【中国出版】等；建议关注AI+IP，如【中文在线】、【奥飞娱乐】等。
- 7、建议关注【福昕软件】、【光云科技】、【力盛体育】、【汇纳科技】。



- ◆ 宏观经济风险：AI相关技术研发及应用成本高企，且后续资本支出需求较大，若宏观环境变化导致相关公司经营情况波动、现金流出现问题，可能会影响研发进展；
- ◆ 地缘政治风险：若相关公司涉及到数据安全等领域的问题可能会引起政治风险；
- ◆ 技术发展不及预期风险：GPT与下游应用结合表现存在不及预期的可能；
- ◆ AIGC行业发展不及预期风险：若底层通用大模型发展不及预期，可能影响下游小模型发展。



行业评级与免责声明

分析师声明

作者具有中国证券业协会授予的证券投资咨询资格或相当的专业胜任能力，以勤勉尽责的职业态度，独立、客观地出具本报告，并保证报告采用的信息均来自合规渠道，力求清晰、准确地反映作者的研究观点，结论不受任何第三方的授意或影响。此外，作者薪酬的任何部分不与本报告中的具体推荐意见或观点直接或间接相关。

公司业务资格说明

本公司具备证券投资咨询业务资格。

投资评级体系与评级定义

股票投资评级：	分析师给出下列评级中的其中一项代表其根据公司基本面及（或）估值预期以报告日起6个月内公司股价相对于同期市场基准指数表现的看法。	
	买入	股价表现将强于基准指数20%以上
	增持	股价表现将强于基准指数5-20%
	中性	股价表现将介于基准指数±5%之间
	减持	股价表现将弱于基准指数5%以上
	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
行业投资评级：	分析师给出下列评级中的其中一项代表其根据行业历史基本面及（或）估值对所研究行业以报告日起12个月内的基本面和行业指数相对于同期市场基准指数表现的看法。	
	增持	行业基本面看好，相对表现优于同期基准指数
	中性	行业基本面稳定，相对表现与同期基准指数持平
	减持	行业基本面看淡，相对表现弱于同期基准指数
相关证券市场基准指数说明：A股市场以沪深300指数为基准；港股市场以恒生指数为基准；美股市场以标普500或纳斯达克综合指数为基准。		

投资评级说明：

不同证券研究机构采用不同的评级术语及评级标准，投资者应区分不同机构在相同评级名称下的定义差异。本评级体系采用的是相对评级体系。投资者买卖证券的决定取决于个人的实际情况。投资者应阅读整篇报告，以获取比较完整的观点与信息，投资者不应以分析师的投资评级取代个人的分析与判断。



行业评级与免责声明

免责声明

。本公司不会因接收人收到本报告而视其为客户。

本报告版权归本公司所有，本公司对本报告保留一切权利。未经书面授权，任何机构和个人均不得对本报告进行任何形式的发布、复制、引用或转载。如经过本公司同意引用、刊发的，须注明出处为上海证券有限责任公司研究所，且不得对本报告进行有悖原意的引用、删节和修改。

在法律许可的情况下，本公司或其关联机构可能会持有报告中涉及的公司所发行的证券或期权并进行交易，也可能为这些公司提供或争取提供多种金融服务。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见和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、价值或投资收入可升可跌。过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见或推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告中的内容和意见仅供参考，并不构成客户私人咨询建议。在任何情况下，本公司、本公司员工或关联机构不承诺投资者一定获利，不与投资者分享投资收益，也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责，投资者据此做出的任何投资决策与本公司、本公司员工或关联机构无关。

市场有风险，投资需谨慎。投资者不应将本报告作为投资决策的唯一参考因素，也不应当认为本报告可以取代自己的判断。

