

Lin Yang
lin.yang@htisec.com

从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海

投资要点：

- **紧抓痛点+局部技术领先，带来 Kimi 的高速增长。**2023 年 10 月 10 日，月之暗面发布了首个支持输入 20 万汉字的智能助手产品 Kimi Chat。根据月之暗面官微的口径，20 万汉字是当时（2023 年 10 月 10 日）全球市场上能够产品化使用的大模型服务中所能支持的最长上下文输入长度，标志着月之暗面在“长文本”这一重要技术上取得了世界领先水平。2024 年 3 月 18 日，月之暗面宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度提升了一个数量级到 200 万字。月之暗面相信，大模型无损上下文长度的数量级提升，也会进一步帮助大家打开对 AI 应用场景的想象力，包括完整代码库的分析理解、可以自主帮人类完成多步骤复杂任务的智能体 Agent、不会遗忘关键信息的终身助理、真正统一架构的多模态模型等等。根据 AI 产品榜统计，Kimi 的 2 月上榜访问量达到 305 万，仅次于阿里通义千问的 365 万和百度文心一言的 1006 万，排名国内总榜第三，月之暗面工程副总裁许欣然更是表示，目前 Kimi 平均每个月可能都有 100% 以上的增速。我们认为，Kimi 在长文本的技术突破，标志着国内大模型在理解和处理长文内容迈出了重要的一步，为实现大模型更深层次的智能化应用和创新提供了技术支持，加速了 AI 应用的商业落地。
- **国产大模型持续发展，性能逐渐接近世界第一梯队。**我们认为，好的 AI 应用与好的 AI 大模型是分不开的，除了 Kimi 背后的月之暗面自研大模型，国内其他大模型目前也正在紧追海外的步伐，部分领军者已经基本达到了世界一流的水平。2024 年 1 月 30 日，讯飞星火认知大模型 V3.5 正式发布，其各项能力越来越逼近现在全球最好的 GPT-4 Turbo 水平，并且已在语言理解、数学能力上超过了 GPT-4 Turbo；2023 年 10 月 17 日，百度发布文心大模型 4.0，李彦宏在宣布文心大模型 4.0 发布时表示，其综合能力“与 GPT-4 相比毫不逊色”；2024 年 01 月 16 日，智谱 AI 也推出新一代基座大模型 GLM-4，其整体性能相比上一代大幅提升，十余项指标逼近或达到 GPT-4；阿里也在 2023 年 10 月 31 日发布了千亿级参数大模型通义千问 2.0。在 10 个权威测评中，通义千问 2.0 综合性能超过 GPT-3.5，正在加速追赶 GPT-4；此外，2024 年 2 月 2 日，商汤发布了“日日新 SenseNova 4.0”，多维度全面升级大模型体系，日日新·商量大语言模型-通用版本（SenseChat V4），支持 128K 语境窗口长度，综合整体评测成绩水平比肩 GPT-4，相较 GPT-3.5 已经实现全面超越。
- **从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海。**我们判断，伴随着大模型长文本处理能力提升，金融领域、法律领域、教育领域、医疗领域等多个行业有望成为 AI 技术升级的核心受益者，我们认为，除了上述几个行业意外，其实对于基本所有存在事务性工作的行业，都有可能在这波 AI 升级的浪潮中受益，大模型能够在短时间内快速分析和总结大量工作文件，如合同、报告、会议记录等，帮助员工提取关键信息，减少人工阅读时间，提高决策效率，从而持续提升各行业运转效率，我们认为，伴随国产大模型的持续升级，2024 年有望成为国产大模型全面商业落地的元年，先进的大模型能力也将逐渐转化为落地实际场景的产品应用，从而带动大模型在金融、教育、法律、医疗等众多垂直行业的持续落地，AI2B 与 AI2C 的需求天花板也有望全面打开，在这个过程中，AI 的投资也从映射和跟随，变为引领。我们相信，从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海。
- **建议关注。**金山办公、福昕软件、华宇软件、通达海、科大讯飞、卫宁健康、创业慧康、久远银海、嘉和美康、恒生电子。
- **风险提示。**AI 技术的发展不及预期，AI 商业落地不及预期，市场竞争加剧的风险。

本研究报告由海通国际分销，海通国际是由海通国际研究有限公司，海通证券印度私人有限公司，海通国际株式会社和海通国际证券集团其他各成员单位的证券研究团队所组成的全球品牌，海通国际证券集团各成员分别在其许可的司法管辖区内从事证券活动。关于海通国际的分析师证明，重要披露声明和免责声明，请参阅附录。（Please see appendix for English translation of the disclaimer）

目 录

1. 紧抓痛点+局部技术领先，带来 Kimi 的高速增长.....	4
2. 国产大模型持续发展，性能逐渐接近世界第一梯队.....	14
3. 从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海.....	23
4. 建议关注与风险提示.....	25

图目录

图 1	月之暗面定位为“寻求将能源转化为智能的最优解”.....	4
图 2	Kimi 官网示例.....	5
图 3	利用 Kimi 分析英伟达 Q2 2024 的财报和电话会议.....	7
图 4	Kimi 在 AI 产品榜统计的 2 月国内总榜中排名第三.....	9
图 5	Kimi 在 AI 产品榜统计的 2 月全球增速榜中排名第一.....	10
图 6	Kimi 智能助手启动 200 万字无损上下文内测.....	11
图 7	星火认知大模型 V3.5 七大能力全面提升.....	14
图 8	人民数据《AI 大模型综合能力测评报告》测评整体情况一览表.....	15
图 9	文心 4.0 具备更强的理解、生成、逻辑和记忆能力.....	16
图 10	GLM-4 大模型各能力测评结果.....	17
图 11	GLM-4 网页浏览功能示例.....	18
图 12	通义千问 2.0 综合性能超过 GPT-3.5，正在加速追赶 GPT-4.....	19
图 13	InternLM-123B 在主要评测集上的表现及与国际其他领先模型的比较.....	20
图 14	日日新·商量大语言模型 SenseChat V4 综合评测成绩比肩 GPT-4.....	21

1. 紧抓痛点+局部技术领先，带来 Kimi 的高速增长

月之暗面 (Moonshot AI) 创立于 2023 年 3 月，致力于寻求将能源转化为智能的最优解，通过产品与用户共创智能，2023 年 10 月推出全球首个支持输入 20 万汉字的智能助手产品 Kimi。

创始团队核心成员参与了 Google Gemini、Google Bard、盘古 NLP、悟道等多个大模型的研发，多项核心技术被 Google PaLM、Meta LLaMa、Stable Diffusion 等主流产品采用。

图1 月之暗面定位为“寻求将能源转化为智能的最优解”



资料来源：月之暗面官网，HTI

根据 21 世纪经济报道官方百家号，月之暗面团队成立不到 2 个月就获得了红杉中国、今日资本、真格基金、砺思资本等知名 VC 投资，目前团队人数已超 80 人。

创始人杨植麟被多家头部 VC 寄予厚望饱受期待。杨植麟是 Transformer-XL 与 XLNet 等热门论文的第一作者，两者均为大语言模型领域的核心技术。同时也是中国 35 岁以下 NLP 领域引用最高的研究者；另外两位联合创始人周昕宇和吴育昕，则有超过 10000+ 的 Google Scholar 引用。

根据月之暗面官微，Monolith 砺思资本创始合伙人曹曦表示：“杨植麟是全球大模型领域里最被认可的华人技术专家，其团队在人工智能技术，特别是大语言模型 LLM 领域拥有深厚的技术积累，并已在国际上获得了广泛认可”。

真格基金合伙人戴雨森也表示：“杨植麟作为 XLNet 等多项知名科研工作的第一作者，具备相当丰富的科研和实践经验，多年来他一直坚信通过大模型实现对高维数据的压缩是人工智能发展的必经之路，也团结了一支人才密度超高，配合默契，又充满挑战巨头摇滚精神的创业团队。”

月之暗面在多模态相关技术上多有储备，团队成员曾发明了包括基于对比学习的视

觉预训练范式 MoCo、最高效的视觉网络结构之一 ShuffleNet 等，也有团队成员主导开发了 detectron2，后者是一个被广泛使用的视觉开源项目，曾被集成到 Meta 全线 VR/AR 产品中。

创始人杨植麟毕业于清华大学，师从唐杰教授；之后获得卡内基梅隆大学计算机博士学位，师从苹果公司人工智能负责人 Ruslan Salakhutdinov。“月之暗面”这个名字，来源于摇滚乐队 PinkFloyd 的著名专辑《The Dark Side of the Moon》（月之暗面）。据创始人杨植麟介绍，这个名字象征着神秘与未知，与公司在人工智能领域的探索精神相契合，此外公司成立的日期正好是这张专辑发行的 50 周年。

根据钛媒体 AGI 官微，融资方面，成立不到一年内，月之暗面就已完成两笔融资，总额超过 13 亿美元，投资方包括红杉中国、真格基金、小红书、美团、阿里等。2024 年 2 月的一笔融资成为国内 AI 大模型公司迄今获得的单轮最大金额融资。两轮投资后，目前，月之暗面公司估值或达 25 亿美元。

技术产品层面，成立至今，月之暗面已完成从通用大模型到上层应用的布局。其中，大模型基础层，月之暗面已训练千亿级别的自研通用大模型，并获得了国内大模型备案审批。

而在应用层，2023 年 10 月 10 日，月之暗面发布了首个支持输入 20 万汉字的智能助手产品 Kimi Chat。（注：Kimi Chat 后改名为 Kimi 智能助手，下文 Kimi Chat、Kimi、Kimi 智能助手均指同一产品）

图2 Kimi 官网示例



资料来源：Kimi 官网，HTI

根据月之暗面官微的口径，20 万汉字是当时（2023 年 10 月 10 日）全球市场上能够产品化使用的大模型服务中所能支持的最长上下文输入长度，标志着月之暗面在“长文本”这一重要技术上取得了世界领先水平。

从技术视角看，参数量决定了大模型支持多复杂的“计算”，而能够接收多少文本输入（即长文本技术）则决定了大模型有多大的“内存”，两者共同决定模型的应用效果。支持更长的上下文意味着大模型拥有更大的“内存”，从而使得大模型的应用更加深入和广泛：比如通过多篇财报进行市场分析、处理超长的法务合同、快速梳理多篇文章或多

个网页的关键信息、基于长篇小说设定进行角色扮演等等，都可以在超长文本技术的支持下，成为人们工作和生活的一部分。

相比当前市面上以英文为基础训练的大模型服务，Kimi Chat 具备较强的多语言能力。例如，Kimi Chat 在中文上具备显著优势，实际使用效果能够支持约 20 万汉字的上下文，2.5 倍于 Anthropic 公司的 Claude-100k（实测约 8 万字），8 倍于 OpenAI 公司的 GPT-4-32k（实测约 2.5 万字）。

同时，Kimi Chat 通过创新的网络结构和工程优化，在千亿参数下实现了无损的长程注意力机制，不依赖于滑动窗口、降采样、小模型等对性能损害较大的“捷径”方案。

在月之暗面公司看来，当前大模型输入长度普遍较低的现状对其技术落地产生了很大制约。例如：

目前大火的虚拟角色场景中，由于长文本能力不足，虚拟角色会轻易忘记重要信息，例如在 Character AI 的社区中用户经常抱怨“因为角色在多轮对话后忘记了自己的身份，所以不得不重新开启新的对话”。

对于大模型开发者来说，输入 prompt 长度的限制约束了大模型应用的场景和能力的发挥，比如基于大模型开发剧本杀类游戏时，往往需要将数万字甚至超过十万字的剧情设定以及游戏规则作为 prompt 加入应用，如果模型输入长度不够，则只能削减规则和设定，从而无法达到预期游戏效果。

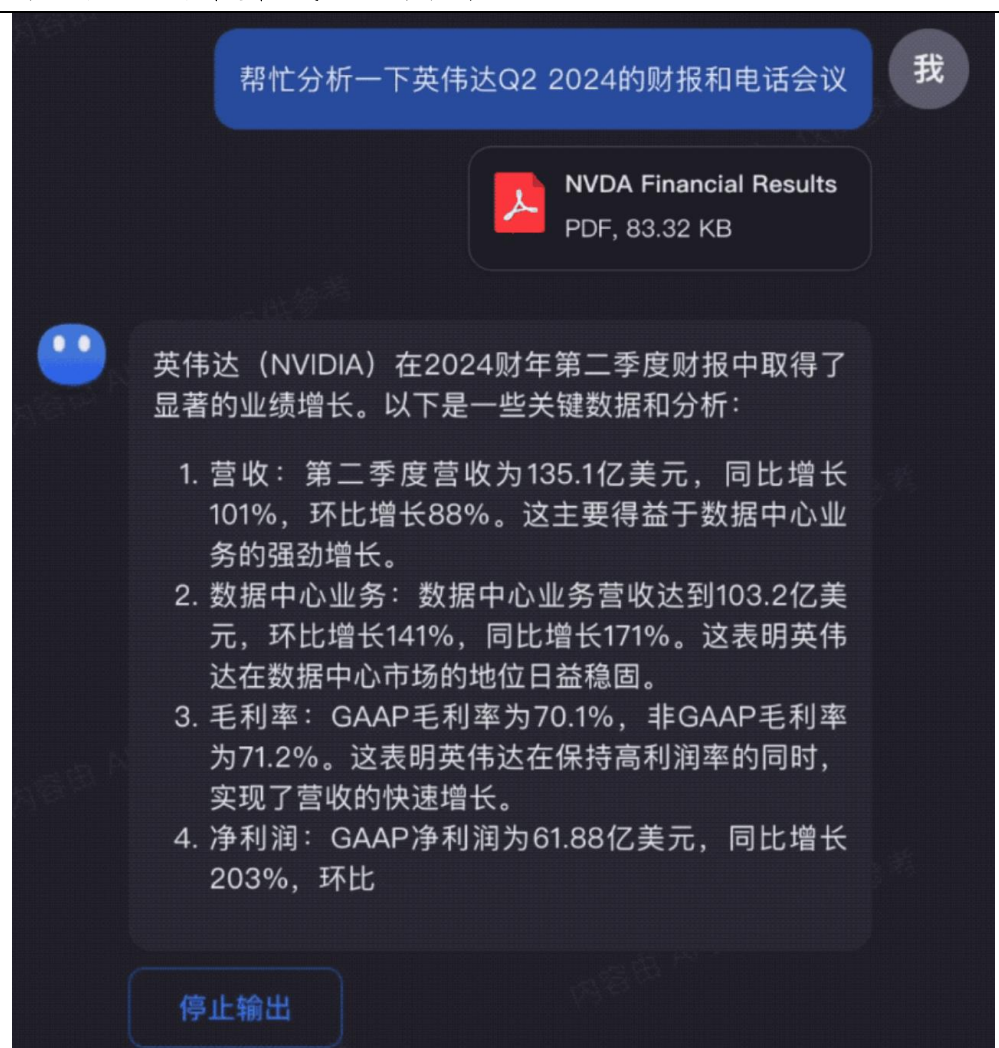
在另一个大模型应用的主要方向——Agent 中，由于 Agent 运行需要自动进行多轮规划和决策，且每次行动都需要参考历史记忆信息才能完成，这会带来了模型输入的快速增加，同时也意味着不能处理更长上下文的模型将因为无法全面准确的基于历史信息进行新的规划和决策从而降低 Agent 运行成功的概率。

在使用大模型作为工作助理完成任务的过程中，几乎每个深度用户都遇到过输入长度超出限制的情况。尤其是律师、分析师、咨询师等职业的用户，由于常常需要分析处理较长的文本内容，使用大模型时受挫的情况发生频率很高。

而上述所有的问题在大模型拥有足够长的上下文输入后都将会迎刃而解。

Kimi 可以快速总结分析公众号长文；Kimi 可以快速完成上市公司财报关键信息分析；Kimi 可以快速整理多份发票中所需要的信息；发现了新的算法论文时，Kimi 能够直接帮你根据论文复现代码；Kimi 可以理解网页内容；Kimi 还可以快速阅读理解整本图书。

图3 利用 Kimi 分析英伟达 Q2 2024 的财报和电话会议



资料来源：月之暗面官微，HTI

当模型可以处理的上下文变得更长后，大模型的能力能够覆盖到更多使用场景，真正在人们的工作、生活、学习中发挥作用，而且由于可以直接基于全文理解进行问答和信息处理，大模型生成的“幻觉”问题也可以得到很大程度的解决。

其实长文本技术的开发，存在一些对效果损害很大的“捷径”，主要包含以下几个方面：

“金鱼”模型，特点是容易“健忘”。通过滑动窗口等方式主动抛弃上文，只保留对最新输入的注意力机制。模型无法对全文进行完整理解，无法处理跨文档的比较和长文本的综合理解（例如，无法从一篇 10 万字的用户访谈录音转写中提取最有价值的 10 个观点）。

“蜜蜂”模型，特点是只关注局部，忽略整体。通过对上下文的降采样或者 RAG（检索增强的生成），只保留对部分输入的注意力机制。模型同样无法对全文进行完整理解（例如，无法从 50 个简历中对候选人的画像进行归纳和总结）。

“蝌蚪”模型，特点是模型能力尚未发育完整。通过减少参数量（例如减少到百亿参数）来提升上下文长度，这种方法会降低模型本身的能力，虽然能支持更长上下文，但是大量任务无法胜任。

月之暗面公司认为，走这些捷径无法达到理想的产品化效果。为了真正做出可用、好用的产品，就应该直面挑战。

具体来看。训练层面，想训练得到一个支持足够长上下文能力的模型，不可避免地要面对如下困难：

如何让模型能在几十万的上下文窗口中，准确的 Attend 到所需要的内容，不降低其原有的基础能力？已有的类似滑动窗口和长度外推等技术对模型性能的损害比较大，在很多场景下无法实现真正的上下文。

在千亿参数级别训练长上下文模型，带来了更高的算力需求和严重的显存压力，传统的 3D 并行方案已经难以无法满足训练需求。

缺乏充足的高质量长序列数据，如何提供更多的有效数据给模型训练？

推理层面，在获得了支持超长上下文的模型后，如何让模型能服务众多用户，同样要面临艰巨挑战：

Transformer 模型中自注意力机制（Self Attention）的计算量会随着上下文长度的增加呈平方级增长，比如上下文增加 32 倍时，计算量实际会增长 1000 倍，这意味着如果只是用朴素的方式实现，用户需要等待相当长的时间才能获得反馈。

超长上下文导致显存需求进一步增长：以 1750 亿参数的 GPT-3 为例，目前最高单机配置（80 GiB * 8）最多只能支持 64k 上下文长度的推理，超长文本对显存的要求可见一斑。

显著的显存带宽压力：英伟达 A800 或 H800 的显存带宽高达 2-3 TiB/s，但面对如此长的上下文，朴素方法的生成速度只能达到 2~5 tokens/s，使用的体验相当卡顿。

在过去 Kimi 发布前半年多的时间里，月之暗面的技术团队进行了极致的算法和工程优化，克服上述重重困难，终于完成了大内存模型的产品化，带来了首个支持 20 万字输入的千亿参数 LLM 产品。

月之暗面创始人杨植麟此前在接受采访时曾表示，无论是文字、语音还是视频，对海量数据的无损压缩可以实现高程度的智能。

无损压缩的进展曾深度依赖「参数为王」模式，该模式下压缩比直接与参数量相关，这大大增加了模型的训练成本和应用门槛，而月之暗面认为：**大模型的能力上限（即无损压缩比）是由单步能力和执行的步骤数共同决定的。单步能力与参数量正相关，而执行步骤数即上下文长度。**

月之暗面相信，更长的上下文长度可以为大模型应用带来全新的篇章，促使大模型从 LLM 时代进入 Long LLM (LLLM)时代：

每个人都可以拥有一个具备终身记忆的虚拟伴侣，它可以在生命的长河中记住与你交互的所有细节，建立长期的情感连接。

每个人都可以拥有一个在工作环境与你共生（co-inhabit）的助手，它知晓公域（互联网）和私域（企业内部文档）的所有知识，并基于此帮助你完成 OKR。


每个人都可以拥有一个无所不知的学习向导，不仅能够准确的给你提供知识，更能引导你跨越学科间的壁垒，更加自由的探索与创新。

当然，更长的上下文长度只是月之暗面在下一代大模型技术上迈出的第一步。月之暗面计划凭借该领域的领先技术，加速大模型技术的创新和应用落地，不断取得更多突破。

Kimi 智能助手初次亮相后，凭借约 20 万汉字的无损上下文能力，帮助用户解锁了很多新的使用场景，包括专业学术论文的翻译和理解、辅助分析法律问题、一次性整理几十张发票、快速理解 API 开发文档等，获得了良好的用户口碑和用户量的快速增长。

根据 AI 产品榜统计，Kimi 的 2 月上榜访问量已经达到 305 万，仅次于阿里通义千问的 365 万和百度文心一言的 1006 万，排名国内总榜第三。

图4 Kimi 在 AI 产品榜统计的 2 月国内总榜中排名第三




国内排名	产品名 AI产品榜	分类 aicpb.com	2月上榜 访问量	2月上榜 变化
1	百度文心一言	AI ChatBots	10.06M	-33.43%
2	阿里通义千问	AI ChatBots	3.65M	-45.05%
3	Kimi (Moonshot)	AI ChatBots	3.05M	107.60%
4	稿定AI	AI Design Tool	2.47M	-11.88%
5	魔音工坊	AI Audio Editing Tools	2.09M	-4.26%
6	讯飞星火	AI ChatBots	1.92M	-29.16%
7	火山方舟	Model Training & Deplo	1.83M	-23.22%
8	抖音豆包	AI ChatBots	1.73M	-2.18%
9	清华智谱清言	AI ChatBots	1.71M	27.47%
10	ProcessOn	AI Mind Map Generato	1.42M	-29.31%

资料来源：AI 产品榜官网，HTI

与此同时，Kimi 访问量增速也在国内超百万月上榜访问量的 AI 产品中位居第一，2 月达到 107.60% 的环比增速，远超同级别产品。而在全球增速榜中，Kimi 的上榜访问量增速依然排名第一。

根据钛媒体 AGI 官微，月之暗面工程副总裁许欣然更是表示，目前 Kimi 平均每个月可能都有 100% 以上的增速。

图5 Kimi 在 AI 产品榜统计的 2 月全球增速榜中排名第一



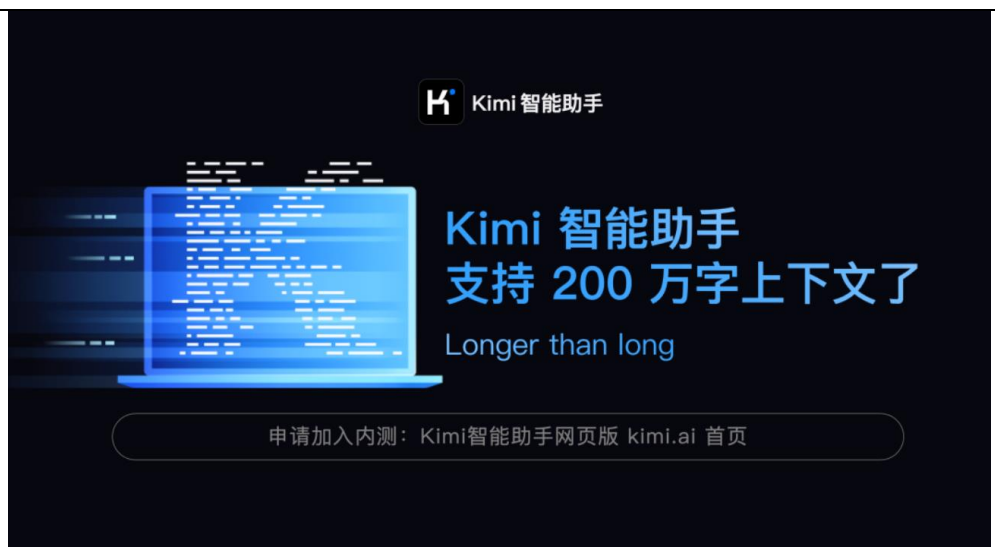
全球排名	产品名 AI产品榜	分类 aicpb.com	2月上榜 访问量	2月上榜 变化
1	Kimi (Moonshot)	AI ChatBots	3.05M	107.60%
2	Chaport	AI ChatBots	5.03M	80.83%
3	ai-novel	AI Story Generator	7.49M	71.25%
4	CheggMate	AI Tools for Education	52.55M	59.40%
5	Toolify	AI navigation site	3.94M	43.94%
6	Usemotion	Life assistant	2.17M	43.78%
7	Coze	AI ChatBots	2.71M	42.52%
8	Dreamgf.ai	AI Character Generator	6.82M	38.01%
9	Video Enhancer	AI Video Editing	1.53M	37.97%
10	Krea	AI Image Generator	6.08M	35.60%

资料来源：AI 产品榜官网，HTI

2024 年 3 月 18 日，月之暗面宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度提升了一个数量级到 200 万字。与此同时，支持 200 万字上下文的 Kimi 已启动“内测”。对大模型超长无损上下文能力有需求的用户，可到 Kimi 智能助手网页版 kimi.ai 首页申请抢先体验。

从 20 万字到 200 万字，由于没有采用常规的渐进式提升路线，月之暗面团队遇到的技术难度也是指数级增加的。为了达到更好的长窗口无损压缩性能，月之暗面的研发和技术团队从模型预训练到对齐、推理环节均进行了原生的重新设计和开发，不走“滑动窗口”、“降采样”等技术捷径，攻克了很多底层技术难点。

图6 Kimi 智能助手启动 200 万字无损上下文内测



资料来源：月之暗面官微，HTI

月之暗面相信，大模型无损上下文长度的数量级提升，也会进一步帮助大家打开对AI应用场景的想象力，包括完整代码库的分析理解、可以自主帮人类完成多步骤复杂任务的智能体Agent、不会遗忘关键信息的终身助理、真正统一架构的多模态模型等等。

月之暗面也放出了几个超长无损上下文的使用场景示例：

用户上传几十万字的经典德州扑克长篇教程后，让Kimi扮演德州专家为自己提供出牌策略的指导；上传一份完整的近百万字中医诊疗手册，让Kimi针对用户的问题给出诊疗建议；上传英伟达过去几年的完整财报，让Kimi成为英伟达财务研究专家，帮用户分析总结英伟达历史上的重要发展节点；上传一个代码仓库里的源代码，可以询问Kimi关于代码库的所有细节，即便是毫无注释的陈年老代码也能帮助你快速梳理出代码的结构。

根据月之暗面的口径，过去要10000小时才能成为专家的领域，现在只需要10分钟，Kimi就能接近任何一个新领域的初级专家水平。用户可以跟Kimi探讨这个领域的问题，让Kimi帮助自己练习专业技能，或者启发新的想法。有了支持200万字无损上下文的Kimi，快速学习任何一个新领域都会变得更加轻松。

快速整理大量的资料是很多用户在工作中经常遇到的挑战。现在Kimi能够一口气精读500个，甚至更多数量的文件，帮助用户快速分析所有文件的内容，并且支持通过自然语言进行信息查询和筛选，大大提高了信息处理效率。例如，公司HR可以基于业务需求，快速从最近的500份简历中，让Kimi快速找出有某个行业从业经历，同时从计算机类专业毕业的求职者，更加高效地筛选和识别合适的候选人。

从长篇小说、故事或剧本中，重新发现值得玩味的蛛丝马迹，进行深层细节的挖掘，是很多影视娱乐IP的爱好者热衷的事情。如果将甄嬛传全集几十万字的剧本传给Kimi，Kimi能在不同时间段、各个场景的故事情节中，深入挖掘甄嬛、果郡王的情感线以及和他们孩子的真相。

“通往通用人工智能（AGI）的话，无损的长上下文将会是一个很关键的基础技术。从word2vec到RNN、LSTM，再到Transformer，历史上所有的模型架构演进，本质上都是在提升有效的、无损的上下文长度。”月之暗面创始人杨植麟博士此前在接受采访时表示，“上下文长度可能存在摩尔定律，但需要同时优化长度和无损压缩水平两个指

标，才是有意义的规模化。”

从很多 Kimi 智能助手用户的反馈来看，20 万字的无损长上下文帮助他们打开了新的 AI 应用世界，带来了更大的价值，但随着他们尝试更复杂的任务、解读更长的文档，依然会遇到对话长度超出限制的情况。这也是大模型产品的无损上下文长度需要继续提升的一个直接原因。

此外，Kimi 智能助手的智能搜索，同样离不开大模型的无损长上下文能力。Kimi 主动搜索得到的多篇资料，会作为上下文的一部分交给模型去推理。正是因为 Kimi 大模型支持的上下文窗口足够长，窗口内的信息损失足够低，Kimi 智能助手才能输出高质量的结果，为用户带来截然不同的搜索体验。

Kimi 可以根据用户的问题，主动去互联网上搜索、分析和总结最相关的多个页面，生成更直接、更准确的答案。例如，用户可以让 Kimi 主动去搜索和对比两家同领域上市公司的最新财报数据，直接生成对比表格，节省大量的资料查找时间。而传统的搜索引擎通常只能根据用户的问题，返回一些参杂着广告信息网页链接。

另一项与大模型的无损上下文能力息息相关的指标是指令遵循（Instruction Following）能力。指令遵循能力主要体现在两个方面：第一，模型在对话中是否能够始终遵循用户的指令，理解用户的需求；第二，模型是否能够遵循复杂指令，有时候复杂指令可能长达几千、上万字。从产品推出以来的用户反馈来看，Kimi 智能助手的多轮交互和超长指令遵循能力，也是产品的一项核心优势。

根据钛媒体 AGI 官微，月之暗面方面表示，从很多 Kimi 智能助手用户的反馈来看，20 万字的无损长上下文帮助他们打开了新的 AI 应用世界，带来了更大的价值，但随着他们尝试更复杂的任务、解读更长的文档，依然会遇到对话长度超出限制的情况。这也是大模型产品的无损上下文长度需要继续提升的一个直接原因。此外，Kimi 智能助手的智能搜索，更是离不开大模型的无损长上下文能力。

月之暗面指出，正是因为用户对 Kimi 发出的指令越来越复杂，因此团队也一直致力于提升 Kimi 能遵循的指令的复杂度、信息检索能力。同时由于用户的使用场景从工作逐渐扩展到生活的方方面面，团队在网页端之外补齐了微信小程序、iOS 端以及安卓端等。

月之暗面工程副总裁许欣然透露，Kimi 充分发挥其作为“硅基生命”的优势，夜间也在持续不断地自我进化中。月之暗面联合创始人周昕宇强调，出于用户共创的考虑，Kimi 的定位比起“聊天机器人”更像一个“智能助手”，因为普通对话对大模型自身迭代的帮助有限。周昕宇表示，对于呼声很高的提示词教程，目前正在准备当中，预计在四月左右发布；同时，多模态模型也在不断研发，而音频处理能力、海外版同样在需求池内。

月之暗面工程副总裁许欣然向钛媒体 App 表示，月之暗面的 AI Infra（基础设施）团队也在持续提升能效比，均是采用自研技术。比起刚发布时，Kimi 在完全相同的硬件条件下响应速度提升了 3 倍。

目前，Kimi 智能助手仍完全免费。但随着用户群体的扩大和使用量的增长，难免会出现算力不足的情况。月之暗面工程副总裁许欣然透露，2024 年上半年预计开启商业化模式

除了持续发展 Kimi，根据界面新闻官方百家号，月之暗面联合创始人周昕宇表示，在 Sora 发布之前月之暗面就已经有多模态方面的研发，目前正在按照固有节奏推进，预计在 2024 年将会有相关产品发布。

我们认为，月之暗面代表着中国 AI 另一种可能性，即初创公司在某个大模型的细分领域带来的世界级领先，在 2023 年 10 月，月之暗面就在“长文本”这一重要技术上取得了世界领先水平，而伴随“200 万字”级别 Kimi 的发布，月之暗面的领先也有望持续扩大。而 Kimi 则是抓住了某个细分领域的“痛点”的产品，即长文本的分析，这种实际的需求，叠加月之暗面自研的大模型在这个领域的领先技术，共同创造出了 Kimi。

月之暗面创始人兼 CEO 杨植麟在接受腾讯新闻《潜望》专访中提到：“我肯定要做 AGI 嘛，这是接下来十年唯一有意义的事。但不是说我们不做应用。或者，不应该把它定义成一个‘应用’。‘应用’听起来好像你有一个技术，你想把它用在什么地方，有商业化闭环。但‘应用’不是准确的词。它跟 AGI 是相辅相成的。它本身是实现 AGI 的手段，也是实现 AGI 的目的。反过来，你如果只关注应用，不关注模型能力迭代，不关注 AGI，贡献也有限。”

我们认为，伴随月之暗面在未来持续深耕 AGI，其大模型的技术迭代和应用的落地有望同步推进，共同促进国产 AI 的发展。

2. 国产大模型持续发展，性能逐渐接近世界第一梯队

我们认为，好的 AI 应用与好的 AI 大模型是分不开的，除了 Kimi 背后的月之暗面自研大模型，国内其他大模型目前也正在紧追海外的步伐，部分领军者已经基本达到了世界一流的水平。

2024 年 1 月 30 日，讯飞星火认知大模型 V3.5 正式发布，讯飞星火 V3.5 实现了文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力、多模态能力等七大能力的全面提升。

根据与长三角人工智能产业链联盟、中国科学院人工智能产学研创新联盟等科研机构以及企业共同形成的通用认知智能大模型测评体系的 481 个维度，科大讯飞采用科学测评方法和随机动态数据对比检测，讯飞星火 V3.5 的各项能力越来越逼近现在全球最好的 GPT-4 Turbo 水平，并且讯飞星火 V3.5 已在语言理解、数学能力上超过了 GPT-4 Turbo。

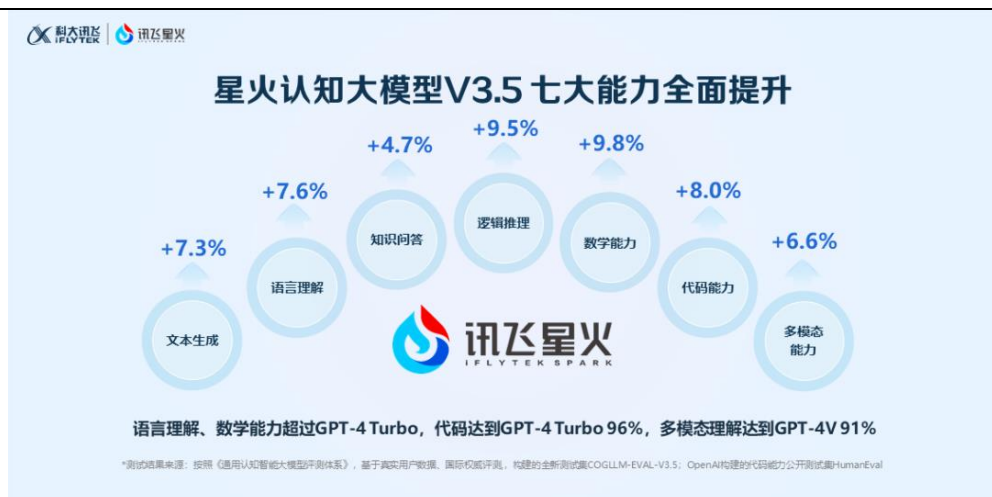
在代码功能上，讯飞星火 V3.5 已经达到了 GPT-4 Turbo 的 96%。在科研最常用的 Python 语言领域，不仅在 HumanEval 的测试集上，而且在真实的测试集上，讯飞星火 V3.5 技术测试效果已经实现超过 (GPT-4 Turbo)。另外在多模态的理解上，讯飞星火 V3.5 达到了 GPT-4V 的 91%，其中语音的多模态能力已经超过 GPT-4 了。

在要素抽取和问题生成方面，讯飞星火 V3.5 已经实现超越 GPT-4 Turbo，在概念理解、知识推理和图文生成方面还有差距，但也在迎头赶上，像概念理解这些关键的能力与 GPT-4 Turbo 相比已经差距很小。

时空推理和逻辑推理方面，讯飞星火 V3.5 相比 V3.0 已经有了显著提升，尽管与 GPT-4 Turbo 相比还存在一些差距，但差距并不是很大。如果将其限定在特定领域，并提供专业领域的知识进行学习和训练，当讯飞星火 V3.5 了解相对明确的应用场景后，其评分可能从 40-50 分提高到 90 多分，甚至在部分应用场景中可以保证万无一失。

总的来讲，讯飞星火 V3.5 在各个维度的能力都有提升，对标国际最先进水平。数学能力已经超过了 GPT-4 Turbo，代码能力在 Python 上超过了 GPT-4 Turbo，多模态能力相当于 GPT-4V 的 91% 水平。

图7 星火认知大模型 V3.5 七大能力全面提升



资料来源：科大讯飞官微，HTI

百度则早在 2023 年 3 月 16 日发布知识增强大语言模型文心一言。文心一言从数万亿数据和数万亿知识中融合学习，得到预训练大模型，在此基础上采用有监督精调、人

类反馈强化学习、提示等技术，具备知识增强、检索增强和对话增强的技术优势。

文心一言的基础模型 2023 年 5 月升级至文心大模型 3.5，在基础模型升级、精调技术创新、知识点增强、逻辑推理增强、插件机制等方面创新突破，取得效果和效率的提升。根据百度官微援引人民数据发布的《AI 大模型综合能力测评报告》显示，文心一言不仅综合评分超越 ChatGPT，位居全球第一，更在内容生态、数据认知、知识问答三大维度评分超越 ChatGPT，且六大维度评分均位列国内大模型榜首。同时，文心一言近 20 项细分测评指标排名全球第一，遥遥领先其他国产大模型。多个公开测评显示，文心大模型 3.5 版支持下的文心一言中文能力突出，甚至有超出 GPT-4 的表现；综合能力在评测中超过 ChatGPT，遥遥领先于其他大模型。例如，全球领先的 IT 市场研究和咨询公司 IDC 最新发布的《AI 大模型技术能力评估报告，2023》显示，根据百度官微援引百度文心大模型 3.5 拿下 12 项指标的 7 个满分，得到“综合评分第一，算法模型第一，行业覆盖第一”三个绝对第一；根据百度官微援引新华网的《国内 LLM 产品测试报告》中，百度文心一言整体领先，在内容安全、阅读理解、常识问答，数学运算等维度得分远超 ChatGPT3.5、讯飞星火和 ChatGLM。

图8 人民数据《AI 大模型综合能力测评报告》测评整体情况一览表

维度	文心一言	讯飞星火	通义千问	ChatGPT	平均分
内容生态	4.00★	3.70★	3.20★	3.40★	3.58★
数据认知	4.00★	3.50★	3.50★	3.75★	3.69★
言语理解	3.90★	3.70★	3.60★	4.00★	3.80★
知识问答	4.30★	4.30★	4.30★	4.10★	4.25★
逻辑推理	4.25★	3.50★	3.00★	4.50★	3.81★
助力科研	3.69★	3.63★	3.56★	4.31★	3.80★
平均分	4.02★	3.72★	3.53★	4.01★	3.82★

资料来源：百度官微援引人民数据《AI 大模型综合能力测评报告》，HTI

2023 年 10 月 17 日，百度发布文心大模型 4.0，相比 3.5 版本，理解、生成、逻辑、记忆四大能力都有显著提升。其中理解和生成能力的提升幅度相近，而逻辑和记忆能力的提升则更大，逻辑的提升幅度达到理解的近 3 倍，记忆的提升幅度也达到了理解的 2 倍多。百度基于文心大模型研制了智能代码助手 Comate，从内部应用效果来看，整体的代码采纳率达到 40%，高频用户的代码采纳率达到 60%。

文心大模型 4.0 在 2023 年 9 月已开始小流量上线，2023 年 9 月-2023 年 10 月一个多月间，效果又提升了近 30%。训练算法效率自 2023 年 3 月以来已累计提升 3.6 倍，周均的训练有效率超过 98%。文心大模型 4.0 基本技术架构与 3.0 和 3.5 版本一脉相承，并在多个关键技术方向上进一步创新突破。

此外，文心大模型 4.0 在输入和输出阶段都进行知识点增强。一方面，对用户输入的问题进行理解，并拆解出回答问题所需的知识点，然后在搜索引擎、知识图谱、数据库中查找准确知识，最后把这些找到的知识组装进 Prompt 送入大模型，准确率好，效率也高；另一方面，对大模型的输出进行反思，从生成结果中拆解出知识点，然后再利用搜索引擎、知识图谱、数据库，以及大模型本身进行确认，进而对有差错的点进行修正。李彦宏在宣布文心大模型 4.0 发布时表示，这是迄今为止最强大的文心大模型，实现了基础模型的全面升级，在理解、生成、逻辑和记忆能力上都有着显著提升，综合能力“与 GPT-4 相比毫不逊色”。

图9 文心 4.0 具备更强的理解、生成、逻辑和记忆能力



资料来源：百度官微，HTI

2024 年 01 月 16 日，智谱 AI 也推出新一代基座大模型 GLM-4。GLM-4 支持更长上下文；更强的多模态；支持更快推理速度，更多并发，大大降低推理成本；同时 GLM-4 增强了智能体能力。

基础能力 (英文): GLM-4 在 MMLU、GSM8K、MATH、BBH、HellaSwag、HumanEval 等数据集上，分别达到 GPT-4 94%、95%、91%、99%、90%、100%的水平。

指令跟随能力: GLM-4 在 IFEval 的 prompt 级别上中、英分别达到 GPT-4 的 88%、85%的水平，在 Instruction 级别上中、英分别达到 GPT-4 的 90%、89%的水平。

对齐能力: GLM-4 在中文对齐能力上整体超过 GPT-4。

长文本能力: 在 LongBench (128K) 测试集上对多个模型进行评测，GLM-4 性能超过 Claude 2.1; 在「大海捞针」(128K) 实验中，GLM-4 的测试结果为 128K 以内全绿，做到 100%精准召回。

图10 GLM-4 大模型各能力测评结果


资料来源：智谱 AI 官网，HTI

GLM-4 实现自主根据用户意图，自动理解、规划复杂指令，自由调用网页浏览器、Code Interpreter 代码解释器和多模态文生图大模型，以完成复杂任务。

简单来讲，即只需一个指令，GLM-4 会自动分析指令，结合上下文选择决定调用合适的工具。

All Tools - 文生图：GLM-4 能够结合上下文进行 AI 绘画创作（CogView3，其在文生图多个评测指标上，相比 DALLE3 约在 91.4%~99.3%的水平之间。），如下图所示，大模型能够遵循人的指令来不断修改生成图片的结果。

All Tools - 代码解释器：GLM-4 能够通过自动调用 python 解释器，进行复杂计算（例如复杂方程、微积分等），在 GSM8K、MATH、Math23K 等多个评测集上都取得了接近或同等 GPT-4 All Tools 的水平。

同样 GLM-4 也可以完成文件处理、数据分析、图表绘制等复杂任务，支持处理 Excel、PDF、PPT 等格式文件。

All Tools - 网页浏览：GLM-4 能够自行规划检索任务、自行选择信息源、自行与信息源交互，在准确率上能够达到 78.08，是 GPT-4 All Tools 的 116%。

图11 GLM-4 网页浏览功能示例

	准确率
GPT-4 All Tools	67.12
GLM-4 All Tools	78.08
GLM-4 / GPT-4	116%



资料来源：智谱 AI 官网，HTI

All Tools - Function Call: GLM-4 能够根据用户提供的 Function 描述，自动选择所需 Function 并生成参数，以及根据 Function 的返回值生成回复；同时也支持一次输入进行多次 Function 调用，支持包含中文及特殊符号的 Function 名字。这一方面 GLM-4 All Tools 与 GPT-4 Turbo 相当。

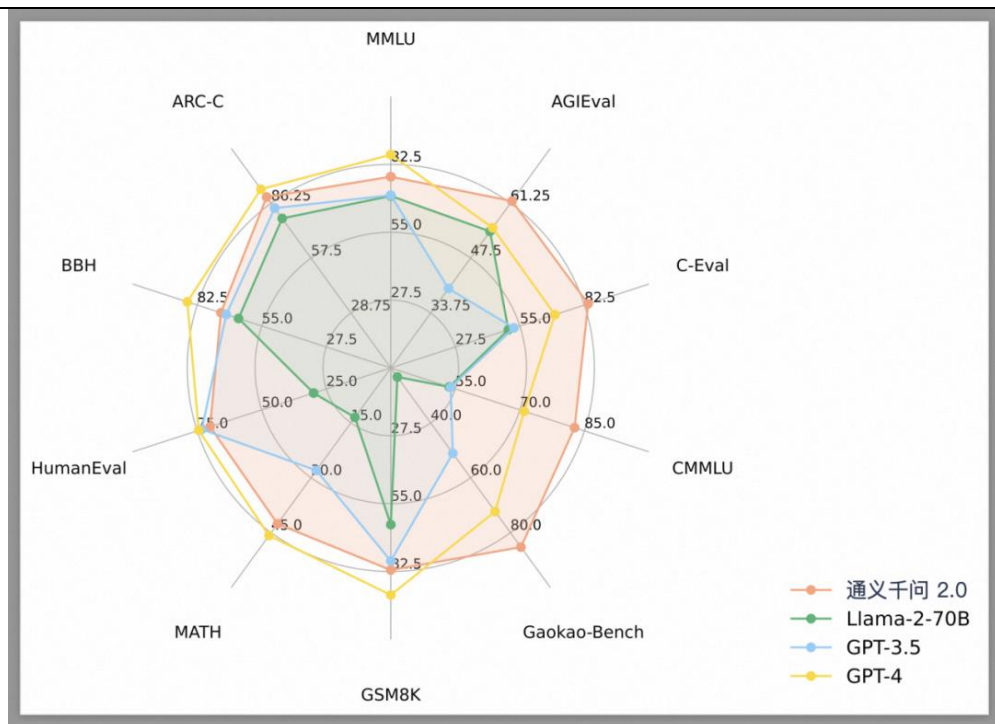
All Tools - 多工具自动调用: 除了以上单项工具自动调用外，GLM-4 同样能够实现多工具自动调用，例如结合 网页浏览、CogView3、代码解释器等调用方式。

总体来讲，GLM-4 的整体性能相比上一代大幅提升，十余项指标逼近或达到 GPT-4。

而阿里也在 2023 年 10 月 31 日发布了千亿级参数大模型通义千问 2.0。在 10 个权威测评中，通义千问 2.0 综合性能超过 GPT-3.5，正在加速追赶 GPT-4。

通义千问 2.0 在性能上取得巨大飞跃，相比 2023 年 4 月发布的 1.0 版本，通义千问 2.0 在复杂指令理解、文学创作、通用数学、知识记忆、幻觉抵御等能力上均有显著提升。通义千问的综合性能已经超过 GPT-3.5，加速追赶 GPT-4。

图12 通义千问 2.0 综合性能超过 GPT-3.5，正在加速追赶 GPT-4



资料来源：阿里云官微，HTI

在 MMLU、C-Eval、GSM8K、HumanEval、MATH 等 10 个主流 Benchmark 测评集上，通义千问 2.0 的得分整体超越 Meta 的 Llama-2-70B，相比 OpenAI 的 Chat-3.5 是九胜一负，相比 GPT-4 则是四胜六负，与 GPT-4 的差距进一步缩小。

中英文理解能力是大语言模型的基本功。英语任务方面，通义千问 2.0 在 MMLU 基准的得分是 82.5，仅次于 GPT-4，通过大幅增加参数量，通义千问 2.0 能更好地理解 and 处理复杂的语言结构和概念；中文任务方面，通义千问 2.0 以明显优势在 C-Eval 基准获得最高得分，这是由于模型在训练中学习了更多中文语料，进一步强化了中文理解和表达能力。

在数学推理、代码理解等领域，通义千问 2.0 进步明显。在推理基准测试 GSM8K 中，通义千问排名第二，展示了强大的计算和逻辑推理能力；在 HumanEval 测试中，通义千问得分紧跟 GPT-4 和 GPT-3.5，该测试主要衡量大模型理解和执行代码片段的能力，这一能力是大模型应用于编程辅助、自动代码修复等场景的基础。

2024 年 3 月 14 日，阿里通义千问还推出了免费的文档解析功能，可解析网页、文档、论文、图书，突破当前大模型长文档处理的天花板。针对单个文档，通义千问能够处理超万页的超长资料，换算成中文篇幅约 1000 万字；针对多个文档，可一键速读 100 份不同格式的资料；还可解析在线网页。

此外，商汤更是早在 2023 年 6 月便联合多家国内顶尖科研机构发布了预训练大语言模型 InternLM，成为国内首个超越 GPT-3.5-turbo 性能的基模型。

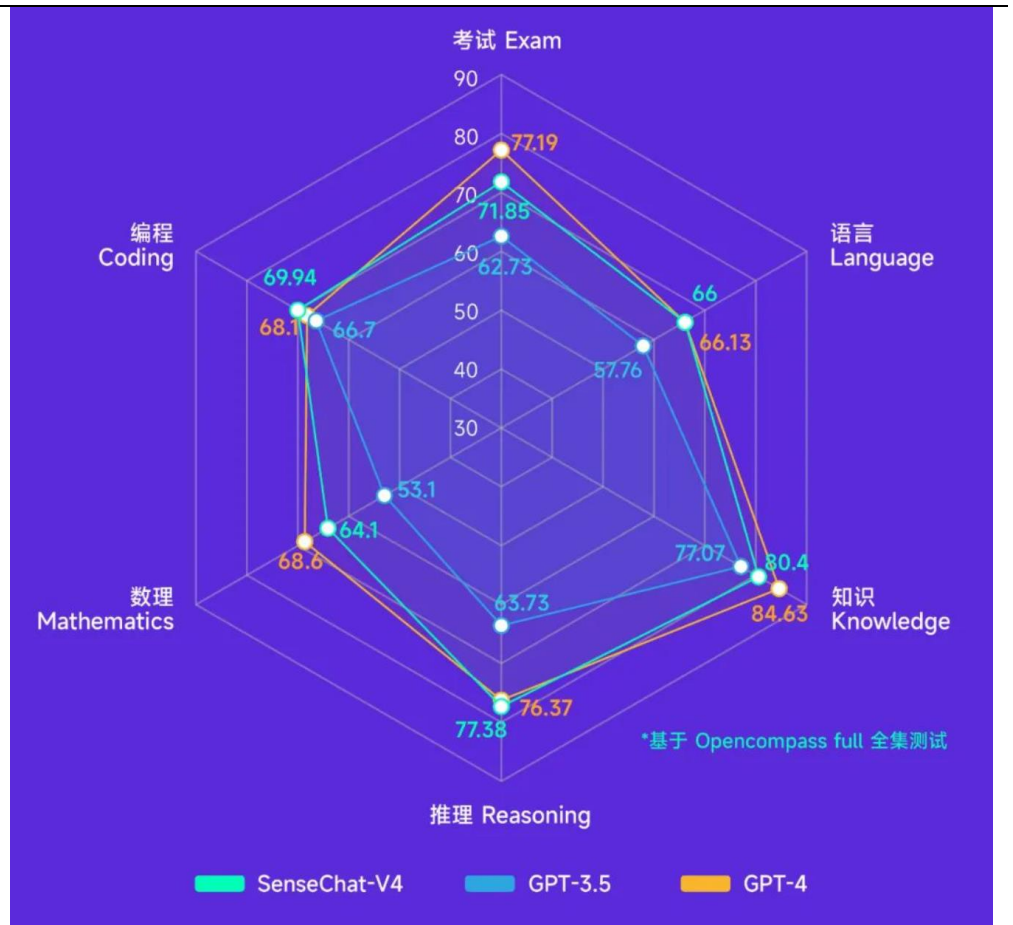
图13 InternLM-123B 在主要评测集上的表现及与国际其他领先模型的比较

评测集		InternLM-104B	InternLM-123B	ChatGPT	GPT-4	LLaMA-2-70B
綜合考試	MMLU	64.3	72.9	69.1	83.0	69.8
	C-Eval	54.4	67.5	52.5	69.9	50.1
	AGIEval	37.1	57.8	39.9	55.1	40.0
	ARC-c	76.6	90.2	83.7	93.6	78.3
	ARC-e	85.5	91.9	88.9	95.4	85.9
知識問答	CommonSenseQA	70.5	88.5	80.2	88.3	78.3
	NaturalQuestions	29.6	36.8	27.0	40.4	34.2
閱讀理解	C3	62.8	95.5	85.6	95.1	79.0
	CMRC	40.0	62.9	40.1	42.6	43.7
	RACE (Middle)	83.4	94.9	85.6	93.7	81.6
	RACE (High)	79.2	91.7	80.3	91.0	79.9
	LAMBADA	77.7	87.0	57.5	65.5	78.9
推理	WinoGrande	68.7	86.4	58.8	78.9	69.8
	StoryCloze	78.6	99.8	97.2	99.5	89.7
	HellaSwag	81.3	91.8	79.5	91.4	82.3
	StrategyQA	73.0	81.7	66.1	79.7	74.7
	PIQA	82.2	86.3	81.7	89.2	82.5
	SIQA	62.6	82.4	72.4	70.1	64.8
	BigBench-Hard	59.5	71.1	70.1	86.7	64.9
數學	GSM8K	52.7	76.4	78.2	91.4	63.3

资料来源：商汤集团 2023 半年报，HTI

2024 年 2 月 2 日，商汤发布了“日日新 SenseNova 4.0”，多维度全面升级大模型体系。“日日新 SenseNova 4.0”拥有更全面的知识覆盖、更可靠的推理能力，更优越的长文本理解力及更稳定的数字推理能力和更强的代码生成能力，并支持跨模态交互。日日新·商量大语言模型-通用版本（SenseChat V4），支持 128K 语境窗口长度，综合整体评测成绩水平比肩 GPT-4，相较 GPT-3.5 已经实现全面超越。

图14 日日新·商量大语言模型 SenseChat V4 综合评测成绩比肩 GPT-4



资料来源：商汤集团官微，HTI

在大模型发展的背景下，各大科技公司也高度重视 AI 应用的发展与落地。“大模型发展，应用才是硬道理。”科大讯飞董事长刘庆峰强调。讯飞星火自 2023 年 5 月诞生以来，不断迭代升级其大模型能力，深耕千行百业的应用刚需。星火赋能个人应用打造，目前基于讯飞听见、讯飞星火 APP、讯飞输入法等应用，已累计赋能亿万用户。为加速企业大模型应用价值落地，讯飞星火 V3.5 将提供全栈自主可控的优化套件。基于全国产化算力打造的讯飞星火 V3.5 支持异构算力调度，可实现行业大模型训练提效 90%，支持 23 个企业应用场景的敏捷优化。截止 2024 年 1 月，星火开发者超 35 万，生态增长迅猛，打造个人应用赋能亿万用户。讯飞星火赋能千行百业，携手保险、银行、能源、汽车、通信等领域龙头企业，打造大模型赋能的应用标杆。此外，深度适配国产算力的讯飞星火开源大模型“星火开源-13B”首次发布，场景应用效果领先，昇思开源社区联合首发上线。

百度创始人、董事长兼首席执行官李彦宏也表示：“我们一定要去卷 AI 原生应用，要把这个东西做出来，模型才有价值。”2023 年 10 月 17 日，百度发布了国内首家一站式交易的 AI 原生应用商店——百度智能云千帆 AI 原生应用商店。千帆 AI 原生应用商店上线百天时，累计上线 AI 原生应用超 100 款，涵盖文案智能创作、AI 作画、代码生成、数字人等应用场景，应用数量指数级增长中。一个个新生的 AI 原生应用在这里实现商业化，月活跃用户超 40 万，应用周订单量超 300%增长，部分优质应用购买转化率达 15% 以上。

商汤也将先进的大模型能力转化为落地实际场景的产品应用，基于最新发布的日日新·商量大语言模型 Function call & Assistants API 版本，完成开发并发布数据分析工具“办

公小浣熊”。通过自然语言输入，办公小浣熊结合商汤大模型体系的意图识别、逻辑理解、代码生成能力，自动将数据转化为有意义的分析结果和可视化图表。未来，商汤“日日新 SenseNova”大模型体系及相关产品和工具将为实现通用人工智能(AGI)提供精准着力点，助力全场景、多产业实现“大模型+”，拓宽大模型应用边界。

我们认为，2023年是国产大模型快速发展的元年，在过去一年多的发展时间中，部分领先的国产大模型目前已经实现了对 GPT-3.5 的超越，正在接近世界第一梯队 GPT-4 的水平。国产大模型目前仍在持续发展的过程中，例如科大讯飞董事长刘庆峰就表示：“星火认知大模型 2024 年将继续保持快速升级，预计上半年达到 GPT-4 Turbo 当前最好水平”，而国产大模型在发展中对于算力、数据的需求仍将不断扩大，而伴随大模型的升级迭代，应用的落地也有望全面铺开。

3. 从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海

我们认为，通过针对长文本的快速学习，大模型在某些细分领域能够起到较为突出的辅助作用，正如月之暗面官微中的介绍：“过去要 10000 小时才能成为专家的领域，现在只需要 10 分钟，Kimi 就能接近任何一个新领域的初级专家水平。用户可以跟 Kimi 探讨这个领域的问题，让 Kimi 帮助自己练习专业技能，或者启发新的想法。有了支持 200 万字无损上下文的 Kimi，快速学习任何一个新领域都会变得更加轻松。”

我们判断，伴随着大模型长文本处理能力提升，以下多个行业有望成为 AI 技术升级的核心受益者：

1、金融领域：

根据新华网转载的经济日报的《大模型技术对金融业意味着什么》一文，从匹配度上看，金融业是典型的数据密集型产业，而大模型技术的一大特征就是具有强大的数据洞察理解能力，可以缩短数据间发生连接与被计算的时间，提高数据创造价值的效率。如果把大模型的能力放在金融业中去处理原有的一些任务，会对很多工作产生提质增效的效果。

我们认为，在金融行业日常的工作中，信息处理是一项无可避免的流程，以投研领域为例，分析师往往需要处理大量的信息，包括公司公告、行业新闻、市场交易数据等等，而许多信息都相当繁杂，例如金山办公 2023 年年报就长达 305 页，完整细读将会消耗大量的时间，即使是选取年报中重点领域进行有针对性的研究总结，在一份年报所花费的时间也是以小时计的，这些基础工作将会给分析师带来繁重的负担。

而伴随着大模型长文本处理能力提升，金融领域信息整理归纳的效率也有望迅速提升，举例来说，Kimi 智能助手目前处理能力上限高达 200w 字，我们简单做一个假设，一篇年报为 10 万字，则我们可以一次性让 Kimi 分析 20 份年报，而通义千问目前能够处理超万页的超长资料，换算成中文篇幅约 1000 万字，即在我们的假设下其可以同时分析超过 200 份年报，这将大大提升分析师信息整理归纳的效率，从而最终提升整体工作效率。

而且，根据月之暗面官微，过去要 10000 小时才能成为专家的领域，现在只需要 10 分钟，Kimi 就能接近任何一个新领域的初级专家水平。我们认为，这对于需要提供大量 C 端用户服务的金融领域将带来全面的革新，以保险领域为例，AI 应用能够读取和分析冗长的保险条款、索赔文件等，快速帮助保险人员快速确定保险责任和赔偿范围，简化管理流程，提升客户满意度，可以说，伴随长文本处理能力的升级，未来 AI 能够在短时间内学习更大量的知识，从而成为部分细分业务领域的“初级专家”，从而减少许多过去许多人类进行的工作量。

我们认为，金融业作为数据密集型行业，在长期业务发展中，金融机构积累了海量的数据基础，而且金融业良好的数据资源禀赋与下沉到 C 端用户服务的特征决定了其很可能是大模型技术率先渗透的领域之一。此外，由于金融相关文档通常包含大量的专业术语和复杂的数据，对文本处理的准确性和理解能力提出了较高要求，伴随着大模型长文本处理能力提升，金融行业有望持续受益。

2、法律领域

我们认为，由于法律的规范化程序和强制性规定，不可避免地产生很多事务性的工作，例如涉案人员信息的记录、合同的审查、案件的处理顺序、证据的概括、案件卷宗的整理归档等等。但是这背后带来的是海量的卷宗，举例来说，司法部官网提及的山西李增虎案，仅仅这个案件中，专案组就完成了 1500 余册卷宗的审查，形成 13000 余页、650 万字的办案材料。这就使得法律行业各参与者在“阅卷”这一必要程序中需要花费巨

大的时间和精力。

根据中共瑞安市委办公室法治瑞安官微转载的李杰（江苏省泰州市中级人民法院）所撰写的《论刑事案件法官心证的形成——基于庭审实质化视角下说服责任的考察》一文，阅卷在审理过程中是相当重要的一步。一些法官在发表的访谈和撰写的审判经验文章里都会明确强调庭前阅卷的重要性，详细的阅卷是庭前做好充分庭审准备的必备功课，更是法官形成、验证心证的基本依赖和主要来源。但是阅卷往往需要花费大量的时间，李杰调取了部分样本案件的开庭时间，如果将法定审限划分为“前—中—后”三段来看，开庭时间多集中在审限的中后段，占比高达 92.86%。审限的大量时间用于阅卷和撰写文书，真正的庭审时间并不太长，庭审的定案功能被虚化，这与庭审实质化的要求相背离。

我们认为，小数量的文本信息可以通过人工来处理，但是面对成千上万的海量文件，如何进行归纳整理，快速提取文件中的信息，这是现今法律行业面临的巨大困扰。而伴随着大模型长文本处理能力提升，对海量数据的快速整理归纳却成为了可能，举例来说，通义千问目前能够处理超万页的长资料，换算成中文篇幅约 1000 万字，上文提到的办案材料高达 13000 余页、650 万字的山西李增虎案能够一次性被通义千问处理完毕，并在短时间内给出归纳总结。

而且，我国司法部门也认识到了人工智能在法律界的重要意义，根据最高人民法院在 2022 年 12 月发布的《最高人民法院关于规范和加强人工智能司法应用的意见》（以下简称《意见》），总体目标就是到 2025 年，基本建成较为完备的司法人工智能技术应用体系，为司法为民、公正司法提供全方位智能辅助支持，显著减轻法官事务性工作负担，有效保障廉洁司法，提高司法管理水平，创新服务社会治理。到 2030 年，建成具有规则引领和应用示范效应的司法人工智能技术应用和理论体系，为司法为民、公正司法提供全流程高水平智能辅助支持，应用规范原则得到社会普遍认可，大幅减轻法官事务性工作负担，高效保障廉洁司法，精准服务社会治理，应用效能充分彰显。并且《意见》在应用范围中专门谈到“加强人工智能辅助事务性工作。支持电子卷宗自动分类归目、案件信息自动回填、案件繁简分流、送达地址及方式自动推荐、司法活动笔录自动生成、执行财产查控辅助、电子卷宗自动归档等智能化应用，降低各类人员工作负担，提高司法效率。”

我们认为，对于天然需要在海量文件中快速进行归纳整理、信息提取的法律行业，大模型长文本处理能力的提升恰恰直击了行业痛点，有望持续提升行业各类从业者的工作效率，从而带动整个行业良性发展。

3、教育领域:

我们认为，伴随着大模型长文本处理能力提升，大模型有望在教育领域承担多个细分任务。

从学生的角度来看，我们认为，伴随教育层次的不断提升，学生往往也需要接触大量的学习资料，以医学生为例，根据大连医科大学官微，《生理学》有 122 万字，《外科学》有 227.5 万字，而《内科学》更是有 248.3 万字，过去大模型很难一次性整理、理解这些资料和内容，但是伴随着大模型长文本处理能力提升，例如 Kimi 就可以通过一次性读完《生理学》这本课本，通义千问更是可以一次性读完上述三本课本，并且在短时间内完成重点的整理归纳，从而提升学生的学习效率。根据阿里云官微，考试复习人群还可以把大模型当做学习助手，做摘要、划重点，进行有的放矢的问答学习。

从教师的角度来看，我们认为，大模型现在可以同时分析多份资料，能够帮助教师短时间内通过日常作业分析来识别出不同学生的学习状况，进行学情分析，而且大模型在学习了海量资料后，还能够成为“初级专家”，辅助老师合学生学习现状和学习表现对教学内容及活动进行个性化定制，此外，在考完试后，大模型也能够同时分析多份试卷，汇总各群体考试结果，生成考试的相关情况说明并归结错因用于辅助老师的精准教学。

我们认为，伴随着大模型长文本处理能力提升，AI 与教育的结合有望更加紧密，各

类 AI 辅助类的教育应用效果也有望进一步提升，从而持续提升 AI 在教育行业各细分领域的渗透率。

4、医疗领域:

根据田伟院士等发表在《中国工程科学》2023 年第 25 卷第 5 期的《人工智能与机器人辅助医学发展研究》，医学 AI 的独特之处在于，通过 AI 算法处理并分析大量的医疗数据（如医学图像、病历记录、基因组学数据），可以辅助医生快速且准确地作出诊断和治疗决策；也可依托自然语言处理、机器学习等技术，从海量的医学文献中快速提取有价值的信息，为医生提供前沿进展与治疗方案建议。随着 AI 技术的加速发展，医学 AI 应用更为丰富，甚至变革了疾病的检测、诊断、治疗模式，成为居民健康的增量保障。

我们认为，伴随着大模型长文本处理能力提升，各类 AI 应用能够通过同时处理多份医疗检测报告、病历记录等长文本资料，达到辅助医生快速理解多个不同患者的健康状况的效果，从而进一步为诊断和治疗提供参考，提高医疗服务的效率和质量。

上文我们提到的只是部分有望在大模型长文本处理能力提升背景下受益的行业，我们认为，其实对于基本所有存在事务性工作的行业，都有可能在这波 AI 升级的浪潮中受益，大模型能够在短时间内快速分析和总结大量工作文件，如合同、报告、会议记录等，帮助员工提取关键信息，减少人工阅读时间，提高决策效率，从而持续提升各行业运转效率。

我们认为，伴随国产大模型的持续升级，2024 年有望成为国产大模型全面商业落地的元年，先进的大模型能力也将逐渐转化为落地实际场景的产品应用，从而带动大模型在金融、教育、法律、医疗等众多垂直行业的持续落地，AI2B 与 AI2C 的需求天花板也有望全面打开，在这个过程中，AI 的投资也从映射和跟随，变为引领。

我们相信，从 Kimi 展望国产大模型和应用，“月之暗面”的背后是星辰大海。

4. 建议关注与风险提示

建议关注：金山办公、福昕软件、华宇软件、通达海、科大讯飞、卫宁健康、创业慧康、久远银海、嘉和美康、恒生电子。

风险提示：AI 技术的发展不及预期，AI 商业落地不及预期，市场竞争加剧的风险。

APPENDIX 1

Summary

Investment Highlights:

Kimi's rapid growth is driven by addressing pain points and local technological leadership. On October 10, 2023, Dark Side of the Moon released Kimi Chat, the first intelligent assistant supporting input of 200,000 Chinese characters. This marked a world-leading achievement in long-text technology. By March 18, 2024, they further broke through, extending context length to 2 million characters without loss. This advancement is expected to unlock AI's potential in various applications. In February, Kimi's visits reached 3.05 million, ranking third in China. Dark Side of the Moon's VP, Xu Xinran, stated Kimi's monthly growth could exceed 100%. Kimi's breakthroughs in long-text processing signify a step forward for domestic foundation models, supporting deeper AI applications and accelerating commercialization.

Domestic foundation models are catching up with global leaders. Good AI applications rely on strong foundation models. China's models are closing in on the global pace, with some matching world-class standards. Iflytek's Starfire Cognition V3.5, released on January 30, 2024, approaches GPT-4 Turbo and surpasses it in language understanding and math. Baidu's Ernie Bot 4.0, SenseTime's SenseNova 4.0, and others are also making significant strides, challenging GPT-4's capabilities.

Looking ahead, industries like finance, law, education, and healthcare are poised to benefit from AI advancements. Foundation models can analyze documents quickly, aiding decision-making and efficiency. 2024 could mark the year of widespread commercial deployment of domestic models, driving AI applications across various sectors.

Watchlist: Beijing Kingsoft Office Software, Fujian Foxit Software Development, Winning Health Technology Group, Nanjing Tdh Technology Co.,Ltd., Iflytek, Goodwill E-Health Info, Sichuan Jiuyuan Yin Hai Software, B-Soft, Hundsun Technologies.

Risk Warning: AI development or commercialization may fall short of expectations, and increased market competition poses risks.

附录 APPENDIX

重要信息披露

本研究报告由海通国际分销，海通国际是由海通国际研究有限公司(HTIRL)，Haitong Securities India Private Limited (HSIPL)，Haitong International Japan K.K. (HTIJKK)和海通国际证券有限公司(HTISCL)的证券研究团队所组成的全球品牌，海通国际证券集团(HTISG)各成员分别在其许可的司法管辖区内从事证券活动。

IMPORTANT DISCLOSURES

This research report is distributed by Haitong International, a global brand name for the equity research teams of Haitong International Research Limited (“HTIRL”), Haitong Securities India Private Limited (“HSIPL”), Haitong International Japan K.K. (“HTIJKK”), Haitong International Securities Company Limited (“HTISCL”), and any other members within the Haitong International Securities Group of Companies (“HTISG”), each authorized to engage in securities activities in its respective jurisdiction.

HTIRL 分析师认证 Analyst Certification:

我，杨林，在此保证 (i) 本研究报告中的意见准确反映了我们对本研究中提及的任何或所有目标公司或上市公司的个人观点，并且 (ii) 我的报酬中没有任何部分与本研究报告中表达的具体建议或观点直接或间接相关；及就此报告中所讨论目标公司的证券，我们（包括我们的家属）在其中均不持有任何财务利益。我和我的家属（我已经告知他们）将不会在本研究报告发布后的 3 个工作日内交易此研究报告所讨论目标公司的证券。I, Lin Yang, certify that (i) the views expressed in this research report accurately reflect my personal views about any or all of the subject companies or issuers referred to in this research and (ii) no part of my compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this research report; and that I (including members of my household) have no financial interest in the security or securities of the subject companies discussed. I and my household, whom I have already notified of this, will not deal in or trade any securities in respect of the issuer that I review within 3 business days after the research report is published.

利益冲突披露 Conflict of Interest Disclosures

海通国际及其某些关联公司可从事投资银行业务和/或对本研究中的特定股票或公司进行做市或持有自营头寸。就本研究报告而言，以下是有关该等关系的披露事项（以下披露不能保证及时无遗漏，如需了解及时全面信息，请发邮件至 ERD-Disclosure@htisec.com）

HTI and some of its affiliates may engage in investment banking and / or serve as a market maker or hold proprietary trading positions of certain stocks or companies in this research report. As far as this research report is concerned, the following are the disclosure matters related to such relationship (As the following disclosure does not ensure timeliness and completeness, please send an email to ERD-Disclosure@htisec.com if timely and comprehensive information is needed).

评级定义（从 2020 年 7 月 1 日开始执行）：

海通国际（以下简称“HTI”）采用相对评级系统来为投资者推荐我们覆盖的公司：优于大市、中性或弱于大市。投资者应仔细阅读 HTI 的评级定义。并且 HTI 发布分析师观点的完整信息，投资者应仔细阅读全文而非仅看评级。在任何情况下，分析师的评级和研究都不能作为投资建议。投资者的买卖股票的决策应基于各自情况（比如投资者的现有持仓）以及其他因素。

分析师股票评级

优于大市，未来 12-18 个月内预期相对基准指数涨幅在 10%以上，基准定义如下

中性，未来 12-18 个月内预期相对基准指数变化不大，基准定义如下。根据 FINRA/NYSE 的评级分布规则，我们会将中性评级划入持有这一类别。

弱于大市，未来 12-18 个月内预期相对基准指数跌幅在 10%以上，基准定义如下

各地股票基准指数：日本 - TOPIX, 韩国 - KOSPI, 台湾 - TAIEX, 印度 - Nifty100, 美国 - SP500; 其他所有中国概念股 - MSCI China.

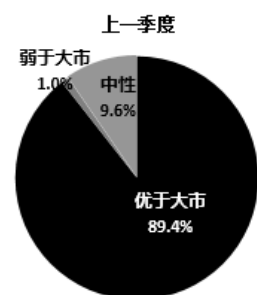
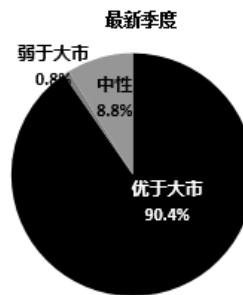
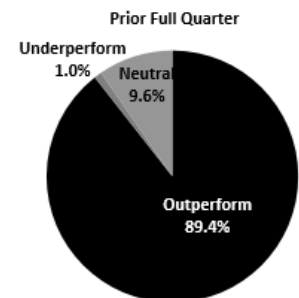
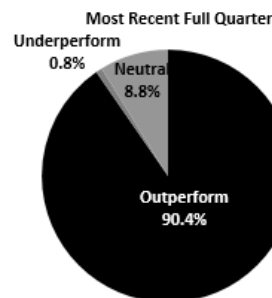
Ratings Definitions (from 1 Jul 2020):

Haitong International uses a relative rating system using Outperform, Neutral, or Underperform for recommending the stocks we cover to investors. Investors should carefully read the definitions of all ratings used in Haitong International Research. In addition, since Haitong International Research contains more complete information concerning the analyst's views, investors should carefully read Haitong International Research, in its entirety, and not infer the contents from the rating alone. In any case, ratings (or research) should not be used or relied upon as investment advice. An investor's decision to buy or sell a stock should depend on individual circumstances (such as the investor's existing holdings) and other considerations.

Analyst Stock Ratings

Outperform: The stock's total return over the next 12-18 months is expected to exceed the return of its relevant broad market benchmark, as indicated below.

评级分布 Rating Distribution



Neutral: The stock's total return over the next 12-18 months is expected to be in line with the return of its relevant broad market benchmark, as indicated below. For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category.

Underperform: The stock's total return over the next 12-18 months is expected to be below the return of its relevant broad market benchmark, as indicated below.

Benchmarks for each stock's listed region are as follows: Japan – TOPIX, Korea – KOSPI, Taiwan – TAIEX, India – Nifty100, US – SP500; for all other China-concept stocks – MSCI China.

截至 2024 年 3 月 31 日海通国际股票研究评级分布

	优于大市	中性 (持有)	弱于大市
海通国际股票研究覆盖率	90.4%	8.8%	0.8%
投资银行客户*	3.3%	4.9%	0.0%

*在每个评级类别里投资银行客户所占的百分比。

上述分布中的买入，中性和卖出分别对应我们当前优于大市，中性和落后大市评级。

只有根据 FINRA/NYSE 的评级分布规则，我们才将中性评级划入持有这一类别。请注意在上表中不包含非评级的股票。

此前的评级系统定义（直至 2020 年 6 月 30 日）：

买入，未来 12-18 个月内预期相对基准指数涨幅在 10% 以上，基准定义如下

中性，未来 12-18 个月内预期相对基准指数变化不大，基准定义如下。根据 FINRA/NYSE 的评级分布规则，我们会将中性评级划入持有这一类别。

卖出，未来 12-18 个月内预期相对基准指数跌幅在 10% 以上，基准定义如下

各地股票基准指数：日本 – TOPIX, 韩国 – KOSPI, 台湾 – TAIEX, 印度 – Nifty100; 其他所有中国概念股 – MSCI China.

Haitong International Equity Research Ratings Distribution, as of March 31, 2024

	Outperform	Neutral (hold)	Underperform
HTI Equity Research Coverage	90.4%	8.8%	0.8%
IB clients*	3.3%	4.9%	0.0%

*Percentage of investment banking clients in each rating category.

BUY, Neutral, and SELL in the above distribution correspond to our current ratings of Outperform, Neutral, and Underperform.

For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category. Please note that stocks with an NR designation are not included in the table above.

Previous rating system definitions (until 30 Jun 2020):

BUY: The stock's total return over the next 12-18 months is expected to exceed the return of its relevant broad market benchmark, as indicated below.

NEUTRAL: The stock's total return over the next 12-18 months is expected to be in line with the return of its relevant broad market benchmark, as indicated below. For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category.

SELL: The stock's total return over the next 12-18 months is expected to be below the return of its relevant broad market benchmark, as indicated below.

Benchmarks for each stock's listed region are as follows: Japan – TOPIX, Korea – KOSPI, Taiwan – TAIEX, India – Nifty100; for all other China-concept stocks – MSCI China.

海通国际非评级研究：海通国际发布计量、筛选或短篇报告，并在报告中根据估值和其他指标对股票进行排名，或者基于可能的估值倍数提出建议价格。这种排名或建议价格并非为了进行股票评级、提出目标价格或进行基本面估值，而仅供参考使用。

Haitong International Non-Rated Research: Haitong International publishes quantitative, screening or short reports which may rank stocks according to valuation and other metrics or may suggest prices based on possible valuation multiples. Such rankings or suggested prices do not purport to be stock ratings or target prices or fundamental values and are for information only.

海通国际 A 股覆盖：海通国际可能会就沪港通及深港通的中国 A 股进行覆盖及评级。海通证券（600837.CH），海通国际于上海的母公司，也会于中国发布中国 A 股的研究报告。但是，海通国际使用与海通证券不同的评级系统，所以海通国际与海通证券的中国 A 股评级可能有所不同。

Haitong International Coverage of A-Shares: Haitong International may cover and rate A-Shares that are subject to the Hong Kong Stock Connect scheme with Shanghai and Shenzhen. Haitong Securities (HS; 600837 CH), the ultimate parent company of HTISG based in Shanghai, covers and publishes research on these same A-Shares for distribution in mainland China. However, the rating system employed by HS differs from that used by HTI and as a result there may be a difference in the HTI and HS ratings for the same A-share stocks.

海通国际优质 100 A 股 (Q100) 指数：海通国际 Q100 指数是一个包括 100 支由海通证券覆盖的优质中国 A 股的计量产品。这些股票是通过基于质量的筛选过程，并结合对海通证券 A 股团队自下而上的研究。海通国际每季对 Q100 指数成分作出复审。

Haitong International Quality 100 A-share (Q100) Index: HTI's Q100 Index is a quant product that consists of 100 of the highest-quality A-shares under coverage at HS in Shanghai. These stocks are carefully selected through a quality-based screening process in combination with a review of the HS A-share team's bottom-up research. The Q100 constituent companies are reviewed quarterly.

盟浪义利 (FIN-ESG) 数据通免责声明条款: 在使用盟浪义利 (FIN-ESG) 数据之前, 请务必仔细阅读本条款并同意本声明:

第一条 义利 (FIN-ESG) 数据系由盟浪可持续数字科技有限责任公司 (以下简称“本公司”) 基于合法取得的公开信息评估而成, 本公司对信息的准确性及完整性不作任何保证。对公司述的评估结果造成的任何直接或间接损失负责。

第二条 盟浪并不因收到此评估数据而将收件人视为客户, 收件人使用此数据时应根据自身实际情况作出自我独立判断。本数据所载内容反映的是盟浪在最初发布本数据日期当日的判断, 盟浪有权在不发出通知的情况下更新、修订与发出其他与本数据所载内容不一致或有不同结论的数据。除非另行说明, 本数据 (如财务业绩数据等) 仅代表过往表现, 过往的业绩表现不作为日后回报的预测。

第三条 本数据版权归本公司所有, 本公司依法保留各项权利。未经本公司事先书面许可授权, 任何个人或机构不得将本数据中的评估结果用于任何营利性目的, 不得对本数据进行修改、复制、编译、汇编、再次编辑、改编、删减、缩写、节选、发行、出租、展览、表演、放映、广播、信息网络传播、摄制、增加图标及说明等, 否则因此给盟浪或其他第三方造成损失的, 由用户承担相应的赔偿责任, 盟浪不承担责任。

第四条 如本免责声明未约定, 而盟浪网站平台载明的其他协议内容 (如《盟浪网站用户注册协议》《盟浪网用户服务 (含认证) 协议》《盟浪网隐私政策》等) 有约定的, 则按其他协议的约定执行; 若本免责声明与其他协议约定存在冲突或不一致的, 则以本免责声明约定为准。

SusallWave FIN-ESG Data Service Disclaimer: Please read these terms and conditions below carefully and confirm your agreement and acceptance with these terms before using SusallWave FIN-ESG Data Service.

1. FIN-ESG Data is produced by SusallWave Digital Technology Co., Ltd. (In short, SusallWave)'s assessment based on legal publicly accessible information. SusallWave shall not be responsible for any accuracy and completeness of the information. The assessment result is for reference only. It is not for any investment advice for any individual or institution and not for basis of purchasing, selling or holding any relative financial products. We will not be liable for any direct or indirect loss of any individual or institution as a result of using SusallWave FIN-ESG Data.
2. SusallWave do not consider recipients as customers for receiving these data. When using the data, recipients shall make your own independent judgment according to your practical individual status. The contents of the data reflect the judgment of us only on the release day. We have right to update and amend the data and release other data that contains inconsistent contents or different conclusions without notification. Unless expressly stated, the data (e.g., financial performance data) represents past performance only and the past performance cannot be viewed as the prediction of future return.
3. The copyright of this data belongs to SusallWave, and we reserve all rights in accordance with the law. Without the prior written permission of our company, none of individual or institution can use these data for any profitable purpose. Besides, none of individual or institution can take actions such as amendment, replication, translation, compilation, re-editing, adaption, deletion, abbreviation, excerpts, issuance, rent, exhibition, performance, projection, broadcast, information network transmission, shooting, adding icons and instructions. If any loss of SusallWave or any third-party is caused by those actions, users shall bear the corresponding compensation liability. SusallWave shall not be responsible for any loss.
4. If any term is not contained in this disclaimer but written in other agreements on our website (e.g. *User Registration Protocol of SusallWave Website, User Service (including authentication) Agreement of SusallWave Website, Privacy Policy of Susallwave Website*), it should be executed according to other agreements. If there is any difference between this disclaimer and other agreements, this disclaimer shall be applied.

重要免责声明:

非印度证券的研究报告: 本报告由海通国际证券集团有限公司 (“HTISGL”) 的全资附属公司海通国际研究有限公司 (“HTIRL”) 发行, 该公司是根据香港证券及期货条例 (第 571 章) 持有第 4 类受规管活动 (就证券提供意见) 的持牌法团。该研究报告在 HTISGL 的全资附属公司 Haitong International (Japan) K.K. (“HTIJK”) 的协助下发行, HTIJK 是由日本关东财务局监管为投资顾问。

印度证券的研究报告: 本报告由从事证券交易、投资银行及证券分析及受 Securities and Exchange Board of India (“SEBI”) 监管的 Haitong Securities India Private Limited (“HTSIPL”) 所发行, 包括制作及发布涵盖 BSE Limited (“BSE”) 和 National Stock Exchange of India Limited (“NSE”) 上市公司 (统称为「印度交易所」) 的研究报告。HTSIPL 于 2016 年 12 月 22 日被收购并成为海通国际证券集团有限公司 (“HTISG”) 的一部分。

所有研究报告均以海通国际为名作为全球品牌, 经许可由海通国际证券股份有限公司及/或海通国际证券集团的其他成员在其司法管辖区发布。

本文件所载信息和观点已被编译或源自可靠来源, 但 HTIRL、HTISGL 或任何其他属于海通国际证券集团有限公司 (“HTISG”) 的成员对其准确性、完整性和正确性不做任何明示或暗示的声明或保证。本文件中所有观点均截至本报告日期, 如有更改, 恕不另行通知。本文件仅供参考使用。文件中提及的任何公司或其股票的说明并非意图展示完整的内容, 本文件并非/不应被解释为对证券买卖的明示或暗示地出价或征价。在某些司法管辖区, 本文件中提及的证券可能无法进行买卖。如果投资产品以投资者本国货币以外的币种进行计价, 则汇率变化可能会对投资产生不利影响。过去的表现并不一定代表将来的结果。某些特定交易, 包括设计金融衍生工具的, 有产生重大风险的可能性, 因此并不适合所有的投资者。您还应认识到本文件中的建议并非为您量身定制。分析师并未考虑到您自身的财务情况, 如您的财务状况和风险偏好。因此您必须自行分析并在适用的情况下咨询自己的法律、税收、会计、金融和其他方面的专业顾问, 以期在投资之前评估该项建议是否适合于您。若由于使用本文件所载的材料而产生任何直接或间接的损失, HTISG 及其董事、雇员或代理人对此均不承担任何责任。

除对本文内容承担责任的分析师外, HTISG 及其关联公司、高级管理人员、董事和雇员, 均可不时作为主事人就本文件所述的任何证券或衍生品持有长仓或短仓以及进行买卖。HTISG 的销售员、交易员和其他专业人士均可向 HTISG 的相关客户和公司提供与本文件所述意见相反的口头或书面市场评论意见或交易策略。HTISG 可做出与本文件所述建议或意见不一致的投资决策。但 HTIRL 没有义务来确保本文件的收件人了解到该等交易决定、思路或建议。

请访问海通国际网站 www.equities.htisec.com, 查阅更多有关海通国际为预防和避免利益冲突设立的组织和行政安排的内容信息。

非美国分析师披露信息: 本项研究首页上列明的海通国际分析师并未在 FINRA 进行注册或者取得相应的资格, 并且不受美国 FINRA 有关与本项研究目标公司进行沟通、公开露面和自营

证券交易的第 2241 条规则之限制。

IMPORTANT DISCLAIMER

For research reports on non-Indian securities: The research report is issued by Haitong International Research Limited (“HTIRL”), a wholly owned subsidiary of Haitong International Securities Group Limited (“HTISGL”) and a licensed corporation to carry on Type 4 regulated activity (advising on securities) for the purpose of the Securities and Futures Ordinance (Cap. 571) of Hong Kong, with the assistance of Haitong International (Japan) K.K. (“HTIJKK”), a wholly owned subsidiary of HTISGL and which is regulated as an Investment Adviser by the Kanto Finance Bureau of Japan.

For research reports on Indian securities: The research report is issued by Haitong Securities India Private Limited (“HSIPL”), an Indian company and a Securities and Exchange Board of India (“SEBI”) registered Stock Broker, Merchant Banker and Research Analyst that, inter alia, produces and distributes research reports covering listed entities on the BSE Limited (“BSE”) and the National Stock Exchange of India Limited (“NSE”) (collectively referred to as “Indian Exchanges”). HSIPL was acquired and became part of the Haitong International Securities Group of Companies (“HTISG”) on 22 December 2016.

All the research reports are globally branded under the name Haitong International and approved for distribution by Haitong International Securities Company Limited (“HTISCL”) and/or any other members within HTISG in their respective jurisdictions.

The information and opinions contained in this research report have been compiled or arrived at from sources believed to be reliable and in good faith but no representation or warranty, express or implied, is made by HTIRL, HTISCL, HSIPL, HTIJKK or any other members within HTISG from which this research report may be received, as to their accuracy, completeness or correctness. All opinions expressed herein are as of the date of this research report and are subject to change without notice. This research report is for information purpose only. Descriptions of any companies or their securities mentioned herein are not intended to be complete and this research report is not, and should not be construed expressly or impliedly as, an offer to buy or sell securities. The securities referred to in this research report may not be eligible for purchase or sale in some jurisdictions. If an investment product is denominated in a currency other than an investor's home currency, a change in exchange rates may adversely affect the investment. Past performance is not necessarily indicative of future results. Certain transactions, including those involving derivatives, give rise to substantial risk and are not suitable for all investors. You should also bear in mind that recommendations in this research report are not tailor-made for you. The analyst has not taken into account your unique financial circumstances, such as your financial situation and risk appetite. You must, therefore, analyze and should, where applicable, consult your own legal, tax, accounting, financial and other professional advisers to evaluate whether the recommendations suits you before investment. Neither HTISG nor any of its directors, employees or agents accepts any liability whatsoever for any direct or consequential loss arising from any use of the materials contained in this research report.

HTISG and our affiliates, officers, directors, and employees, excluding the analysts responsible for the content of this document, will from time to time have long or short positions in, act as principal in, and buy or sell, the securities or derivatives, if any, referred to in this research report. Sales, traders, and other professionals of HTISG may provide oral or written market commentary or trading strategies to the relevant clients and the companies within HTISG that reflect opinions that are contrary to the opinions expressed in this research report. HTISG may make investment decisions that are inconsistent with the recommendations or views expressed in this research report. HTI is under no obligation to ensure that such other trading decisions, ideas or recommendations are brought to the attention of any recipient of this research report.

Please refer to HTI's website www.equities.htisec.com for further information on HTI's organizational and administrative arrangements set up for the prevention and avoidance of conflicts of interest with respect to Research.

Non U.S. Analyst Disclosure: The HTI analyst(s) listed on the cover of this Research is (are) not registered or qualified as a research analyst with FINRA and are not subject to U.S. FINRA Rule 2241 restrictions on communications with companies that are the subject of the Research; public appearances; and trading securities by a research analyst.

分发和地区通知:

除非下文另有规定, 否则任何希望讨论本报告或者就本项研究中讨论的任何证券进行任何交易的收件人均应联系其所在国家或地区的海通国际销售人员。

香港投资者的通知事项: 海通国际证券股份有限公司 (“HTISCL”) 负责分发该研究报告, HTISCL 是在香港有权实施第 1 类受规管活动 (从事证券交易) 的持牌公司。该研究报告并不构成《证券及期货条例》(香港法例第 571 章) (以下简称“SFO”) 所界定的要约邀请, 证券要约或公众要约。本研究报告仅提供给 SFO 所界定的“专业投资者”。本研究报告未经过证券及期货事务监察委员会的审查。您不应仅根据本研究报告中所载的信息做出投资决定。本研究报告的收件人就研究报告中产生或与之相关的任何事宜请联系 HTISCL 销售人员。

美国投资者的通知事项: 本研究报告由 HTIRL, HSIPL 或 HTIJKK 编写。HTIRL, HSIPL, HTIJKK 以及任何非 HTISG 美国联营公司, 均未在美国注册, 因此不受美国关于研究报告编制和研究分析人员独立性规定的约束。本研究报告提供给依照 1934 年“美国证券交易法”第 15a-6 条规定的豁免注册的「美国主要机构投资者」 (“Major U.S. Institutional Investor”) 和「机构投资者」 (“U.S. Institutional Investors”)。在向美国机构投资者分发研究报告时, Haitong International Securities (USA) Inc. (“HTI USA”) 将对报告的内容负责。任何收到本研究报告的美国投资者, 希望根据本研究报告提供的信息进行任何证券或相关金融工具买卖的交易, 只能通过 HTI USA。HTI USA 位于 340 Madison Avenue, 12th Floor, New York, NY 10173, 电话 (212) 351-6050。HTI USA 是在美国于 U.S. Securities and Exchange Commission (“SEC”) 注册的经纪商, 也是 Financial Industry Regulatory Authority, Inc. (“FINRA”) 的成员。HTIUSA 不负责编写本研究报告, 也不负责其中包含的分析。在任何情况下, 收到本研究报告的任何美国投资者, 不得直接与分析师直接联系, 也不得通过 HSIPL, HTIRL 或 HTIJKK 直接进行买卖证券或相关金融工具的交易。本研究报告中出现的 HSIPL, HTIRL 或 HTIJKK 分析师没有注册或具备 FINRA 的研究分析师资格, 因此可能不受 FINRA 第 2241 条规定的与目标公司的交流, 公开露面和分析师账户持有的交易证券等限制。投资本研究报告中讨论的任何非美国证券或相关金融工具 (包括 ADR) 可能存在一定风险。非美国发行的证券可能没有注册, 或不受美国法规的约束。有关非美国证券或相关金融工具的信息可能有限制。外国公司可能不受审计和汇报的标准以及与美国境内生效相符的监管要求。本研究报告中以美元以外的其他货币计价的任何证券或相关金融工具的投资或收益的价值受汇率波动的影响, 可能对该等证券或相关金融工具的价值或收入产生正面或负面影响。美国收件人的所有问询请联系:

Haitong International Securities (USA) Inc.
340 Madison Avenue, 12th Floor
New York, NY 10173
联系人电话: (212) 351 6050

DISTRIBUTION AND REGIONAL NOTICES

Except as otherwise indicated below, any Recipient wishing to discuss this research report or effect any transaction in any security discussed in HTI's research should contact the Haitong International salesperson in their own country or region.

Notice to Hong Kong investors: The research report is distributed by Haitong International Securities Company Limited ("HTISCL"), which is a licensed corporation to carry on Type 1 regulated activity (dealing in securities) in Hong Kong. This research report does not constitute a solicitation or an offer of securities or an invitation to the public within the meaning of the SFO. This research report is only to be circulated to "Professional Investors" as defined in the SFO. This research report has not been reviewed by the Securities and Futures Commission. You should not make investment decisions solely on the basis of the information contained in this research report. Recipients of this research report are to contact HTISCL salespersons in respect of any matters arising from, or in connection with, the research report.

Notice to U.S. investors: As described above, this research report was prepared by HTIRL, HSIPL or HTIJKK. Neither HTIRL, HSIPL, HTIJKK, nor any of the non U.S. HTISG affiliates is registered in the United States and, therefore, is not subject to U.S. rules regarding the preparation of research reports and the independence of research analysts. This research report is provided for distribution to "major U.S. institutional investors" and "U.S. institutional investors" in reliance on the exemption from registration provided by Rule 15a-6 of the U.S. Securities Exchange Act of 1934, as amended. When distributing research reports to "U.S. institutional investors," HTI USA will accept the responsibilities for the content of the reports. Any U.S. recipient of this research report wishing to effect any transaction to buy or sell securities or related financial instruments based on the information provided in this research report should do so only through Haitong International Securities (USA) Inc. ("HTI USA"), located at 340 Madison Avenue, 12th Floor, New York, NY 10173, USA; telephone (212) 351 6050. HTI USA is a broker-dealer registered in the U.S. with the U.S. Securities and Exchange Commission (the "SEC") and a member of the Financial Industry Regulatory Authority, Inc. ("FINRA"). HTI USA is not responsible for the preparation of this research report nor for the analysis contained therein. Under no circumstances should any U.S. recipient of this research report contact the analyst directly or effect any transaction to buy or sell securities or related financial instruments directly through HSIPL, HTIRL or HTIJKK. The HSIPL, HTIRL or HTIJKK analyst(s) whose name appears in this research report is not registered or qualified as a research analyst with FINRA and, therefore, may not be subject to FINRA Rule 2241 restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account. Investing in any non-U.S. securities or related financial instruments (including ADRs) discussed in this research report may present certain risks. The securities of non-U.S. issuers may not be registered with, or be subject to U.S. regulations. Information on such non-U.S. securities or related financial instruments may be limited. Foreign companies may not be subject to audit and reporting standards and regulatory requirements comparable to those in effect within the U.S. The value of any investment or income from any securities or related financial instruments discussed in this research report denominated in a currency other than U.S. dollars is subject to exchange rate fluctuations that may have a positive or adverse effect on the value of or income from such securities or related financial instruments. All inquiries by U.S. recipients should be directed to:

Haitong International Securities (USA) Inc.
340 Madison Avenue, 12th Floor
New York, NY 10173
Attn: Sales Desk at (212) 351 6050

中华人民共和国的通知事项: 在中华人民共和国(下称“中国”,就本报告目的而言,不包括香港特别行政区、澳门特别行政区和台湾)只有根据适用的中国法律法规而收到该材料的人员方可使用该材料。并且根据相关法律法规,该材料中的信息并不构成“在中国从事生产、经营活动”。本文件在中国并不构成相关证券的公共发售或认购。无论根据法律规定或其他任何规定,在取得中国政府所有的批准或许可之前,任何法人或自然人均不得直接或间接地购买本材料中的任何证券或任何实益权益。接收本文件的人员须遵守上述限制性规定。

加拿大投资者的通知事项: 在任何情况下该等材料均不得被解释为在任何加拿大的司法管辖区内出售证券的要约或认购证券的要约邀请。本材料中所述证券在加拿大的任何要约或出售行为均只能在豁免向有关加拿大证券监管机构提交招股说明书的前提下由 Haitong International Securities (USA) Inc. ("HTI USA") 予以实施,该公司是一家根据 National Instrument 31-103 Registration Requirements, Exemptions and Ongoing Registrant Obligations ("NI 31-103") 的规定得到「国际交易商豁免」("International Dealer Exemption") 的交易商,位于艾伯塔省、不列颠哥伦比亚省、安大略省和魁北克省。在加拿大,该等材料在任何情况下均不得被解释为任何证券的招股说明书、发行备忘录、广告或公开发行。加拿大的任何证券委员会或类似的监管机构均未审查或以任何方式批准该等材料、其中所载的信息或所述证券的优点,任何与此相反的声明即属违法。在收到该等材料时,每个加拿大的收件人均将被视为属于 National Instrument 45-106 Prospectus Exemptions 第 1.1 节或者 Securities Act (Ontario) 第 73.3(1) 节所规定的「认可投资者」("Accredited Investor"), 或者在适用情况下 National Instrument 31-103 第 1.1 节所规定的「许可投资者」("Permitted Investor")。

新加坡投资者的通知事项: 本研究报告由 Haitong International Securities (Singapore) Pte Ltd ("HTISSPL") [公司注册编号 201311400G] 于新加坡提供。HTISSPL 是符合《财务顾问法》(第 110 章) ("FAA") 定义的豁免财务顾问,可 (a) 提供关于证券,集体投资计划的部分,交易所衍生品合约和场外衍生品合约的建议 (b) 发行或公布有关证券、交易所衍生品合约和场外衍生品合约的研究分析或研究报告。本研究报告仅提供给符合《证券及期货法》(第 289 章) 第 4A 条项下规定的机构投资者。对于因本研究报告而产生的或与之相关的任何问题,本研究报告的收件人应通过以下信息与 HTISSPL 联系:

Haitong International Securities (Singapore) Pte. Ltd
50 Raffles Place, #33-03 Singapore Land Tower, Singapore 048623
电话:(65) 6536 1920

日本投资者的通知事项: 本研究报告由海通国际证券有限公司所发布,旨在分发给从事投资管理的金融服务提供商或注册金融机构(根据日本金融机构和交易法("FIEL")) 第 61 (1) 条,第 17-11 (1) 条的执行及相关条款)。

英国及欧盟投资者的通知事项: 本报告由从事投资顾问的 Haitong International Securities Company Limited 所发布,本报告只面向有投资相关经验的专业客户发布。任何投资或与本报告相关的投资行为只面对此类专业客户。没有投资经验或相关投资经验的客户不得依赖本报告。Haitong International Securities Company Limited 的分支机构的净长期或短期金融权益可能超过本研究报告中提及的实体已发行股本总额的 0.5%。特别提醒有些英文报告有可能此前已经通过中文或其它语言完成发布。

澳大利亚投资者的通知事项: Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited 和 Haitong International Securities (UK) Limited 分别根据澳大利亚证券和投资委员会(以下简称"ASIC")公司(废除及过度性)文书第 2016/396 号规章在澳大利亚分发本项研究,该等规章免除了根据 2001 年《公司法》在澳大利亚为批发客户提供金融服务时海通国际需持有澳大利亚金融服务许可的要求。ASIC 的规章副本可在以下网站获取: www.legislation.gov.au。海通国际提供的金融服务受外国法律法规规定的管制,该等法律与在澳大利亚所适用的法律存在差异。

印度投资者的通知事项: 本报告由从事证券交易、投资银行及证券分析及受 Securities and Exchange Board of India ("SEBI") 监管的 Haitong Securities India Private Limited ("HTSIPL") 所

发布, 包括制作及发布涵盖 BSE Limited (“BSE”) 和 National Stock Exchange of India Limited (“NSE”) (统称为「印度交易所」) 研究报告。

研究机构名称: Haitong Securities India Private Limited

SEBI 研究分析师注册号: INH000002590

地址: 1203A, Floor 12A, Tower 2A, One World Center

841 Senapati Bapat Marg, Elphinstone Road, Mumbai 400 013, India

CIN U74140MH2011FTC224070

电话: +91 22 43156800 传真: +91 22 24216327

合规和申诉办公室联系人: Prasanna Chandwaskar; 电话: +91 22 43156803; 电子邮箱: prasanna.chandwaskar@htisec.com

“请注意, SEBI 授予的注册和 NISM 的认证并不保证中介的表现或为投资者提供任何回报保证”。

本项研究仅供收件人使用, 未经海通国际的书面同意不得予以复制和再次分发。

版权所有: 海通国际证券集团有限公司 2019 年。保留所有权利。

People’s Republic of China (PRC): In the PRC, the research report is directed for the sole use of those who receive the research report in accordance with the applicable PRC laws and regulations. Further, the information on the research report does not constitute “production and business activities in the PRC” under relevant PRC laws. This research report does not constitute a public offer of the security, whether by sale or subscription, in the PRC. Further, no legal or natural persons of the PRC may directly or indirectly purchase any of the security or any beneficial interest therein without obtaining all prior PRC government approvals or licenses that are required, whether statutorily or otherwise. Persons who come into possession of this research are required to observe these restrictions.

Notice to Canadian Investors: Under no circumstances is this research report to be construed as an offer to sell securities or as a solicitation of an offer to buy securities in any jurisdiction of Canada. Any offer or sale of the securities described herein in Canada will be made only under an exemption from the requirements to file a prospectus with the relevant Canadian securities regulators and only by Haitong International Securities (USA) Inc., a dealer relying on the “international dealer exemption” under National Instrument 31-103 Registration Requirements, Exemptions and Ongoing Registrant Obligations (“NI 31-103”) in Alberta, British Columbia, Ontario and Quebec. This research report is not, and under no circumstances should be construed as, a prospectus, an offering memorandum, an advertisement or a public offering of any securities in Canada. No securities commission or similar regulatory authority in Canada has reviewed or in any way passed upon this research report, the information contained herein or the merits of the securities described herein and any representation to the contrary is an offence. Upon receipt of this research report, each Canadian recipient will be deemed to have represented that the investor is an “accredited investor” as such term is defined in section 1.1 of National Instrument 45-106 Prospectus Exemptions or, in Ontario, in section 73.3(1) of the Securities Act (Ontario), as applicable, and a “permitted client” as such term is defined in section 1.1 of NI 31-103, respectively.

Notice to Singapore investors: This research report is provided in Singapore by or through Haitong International Securities (Singapore) Pte Ltd (“HTISSPL”) [Co Reg No 201311400G. HTISSPL is an Exempt Financial Adviser under the Financial Advisers Act (Cap. 110) (“FAA”) to (a) advise on securities, units in a collective investment scheme, exchange-traded derivatives contracts and over-the-counter derivatives contracts and (b) issue or promulgate research analyses or research reports on securities, exchange-traded derivatives contracts and over-the-counter derivatives contracts. This research report is only provided to institutional investors, within the meaning of Section 4A of the Securities and Futures Act (Cap. 289). Recipients of this research report are to contact HTISSPL via the details below in respect of any matters arising from, or in connection with, the research report:

Haitong International Securities (Singapore) Pte. Ltd.

10 Collyer Quay, #19-01 - #19-05 Ocean Financial Centre, Singapore 049315

Telephone: (65) 6536 1920

Notice to Japanese investors: This research report is distributed by Haitong International Securities Company Limited and intended to be distributed to Financial Services Providers or Registered Financial Institutions engaged in investment management (as defined in the Japan Financial Instruments and Exchange Act (“FIEL”) Art. 61(1), Order for Enforcement of FIEL Art. 17-11(1), and related articles).

Notice to UK and European Union investors: This research report is distributed by Haitong International Securities Company Limited. This research is directed at persons having professional experience in matters relating to investments. Any investment or investment activity to which this research relates is available only to such persons or will be engaged in only with such persons. Persons who do not have professional experience in matters relating to investments should not rely on this research. Haitong International Securities Company Limited’s affiliates may have a net long or short financial interest in excess of 0.5% of the total issued share capital of the entities mentioned in this research report. Please be aware that any report in English may have been published previously in Chinese or another language.

Notice to Australian investors: The research report is distributed in Australia by Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited, and Haitong International Securities (UK) Limited in reliance on ASIC Corporations (Repeal and Transitional) Instrument 2016/396, which exempts those HTISG entities from the requirement to hold an Australian financial services license under the Corporations Act 2001 in respect of the financial services it provides to wholesale clients in Australia. A copy of the ASIC Class Orders may be obtained at the following website, www.legislation.gov.au. Financial services provided by Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited, and Haitong International Securities (UK) Limited are regulated under foreign laws and regulatory requirements, which are different from the laws applying in Australia.

Notice to Indian investors: The research report is distributed by Haitong Securities India Private Limited (“HSIPL”), an Indian company and a Securities and Exchange Board of India (“SEBI”) registered Stock Broker, Merchant Banker and Research Analyst that, inter alia, produces and distributes research reports covering listed entities on the BSE Limited (“BSE”) and the National Stock Exchange of India Limited (“NSE”) (collectively referred to as “Indian Exchanges”).

Name of the entity: Haitong Securities India Private Limited

SEBI Research Analyst Registration Number: INH000002590

Address : 1203A, Floor 12A, Tower 2A, One World Center

841 Senapati Bapat Marg, Elphinstone Road, Mumbai 400 013, India

CIN U74140MH2011FTC224070

Ph: +91 22 43156800 Fax:+91 22 24216327

Details of the Compliance Officer and Grievance Officer : Prasanna Chandwaskar : Ph: +91 22 43156803; Email id: prasanna.chandwaskar@htisec.com

“Please note that Registration granted by SEBI and Certification from NISM in no way guarantee performance of the intermediary or provide any assurance of returns to investors”.

This research report is intended for the recipients only and may not be reproduced or redistributed without the written consent of an authorized signatory of HTISG.

Copyright: Haitong International Securities Group Limited 2019. All rights reserved.

<http://equities.htisec.com/x/legal.html>
