仅供内部参考，请勿外传！

GPT-6 Astra：Computer Use突破，AI从Copilot迈向端到端Agent

AI系列专题研究

行业研究 · 专题报告

传媒·计算机

投资评级：优于大市（维持）

证券分析师：张衡
021-60875160
zhangheng2@guosen.com.cn
S0980517060002

证券分析师：熊莉
021-61761067
xiongli@guosen.com.cn
S0980519030002

联系人：侯睿
hourui3@guosen.com.cn

仅供内部参考，请勿外传！

核心要点：从Copilot迈向端到端Agent，AI应用与Inference需求有望进入新阶段

- **GPT-6 Astra能力继续跃升，模型开始从“会回答”走向“能完成工作”。** Astra在数学、科学及复杂推理能力上进一步提升，FrontierMath Tier 4达到约98%，ARC-AGI-3达到99.9%；模型在部分复杂Agent及软件工程任务中能够以更少输出Token、更高成功率完成任务。大模型竞争指标有望从单纯关注Benchmark及单Token价格，逐渐转向Task Completion Rate、单任务成本、执行时间及人工干预次数，模型的经济价值开始更多体现为“完成一项真实工作需要多少成本”。

- **Computer Use与端到端专业工作能力迎来明显升级，AI正在从Copilot进一步演进为Agent。** Astra可以直接操作浏览器及计算机界面，完成网页操作、CRM更新、数据分析、软件安装测试、Coding及QA等任务；OSWorld 2.0得分由GPT-5.6 Sol的65.7%提升至72.6%，同时模拟单任务耗时由约75分钟下降至40分钟，结合新一代Codex Agent Harness后，Mind2Web任务完成速度提升至上一代体验的约1.9倍。Astra开始能够将检索、分析、代码执行、软件操作、结果验证以及Word/Excel/PPT等最终交付物制作整合到同一个Workflow中，意味着AI开始由“生成中间内容”向“直接交付结果”演进。

- **Agent有望率先重构软件开发及企业SaaS，软件竞争壁垒从UI和Seat数量进一步转向Data、Workflow与Permission。** 在软件开发领域，AI能力正由代码补全向需求理解、开发、测试、Debug、浏览器验证及部署等完整Software Engineering流程扩展，有望持续提升软件工程人均产出；在企业软件领域，人机交互方式可能逐渐由“用户→SaaS UI”演化为“用户→Agent→多个软件/API”。拥有核心企业数据、深度Workflow、系统权限及Agent Orchestration能力的平台型软件厂商有望强化竞争优势，而缺乏核心数据及流程壁垒、主要依赖Per-seat收费的软件可能面临Seat压缩压力，商业模式或进一步向Consumption、Agent Seat及Outcome-based Pricing迁移。

- **Agent将AI计算需求从“人类调用”进一步升级为“自主循环调用”，Inference需求天花板有望继续打开。** 传统Chatbot一次用户请求通常只对应有限模型调用，而复杂Agent为完成一个任务，需要持续进行推理、搜索、代码执行、Computer Use、结果验证及错误修正，单任务背后的模型调用次数可能由个位数提升至几十甚至数百次，任务持续时间也由秒级/分钟级向数十分钟甚至更长周期延伸。因此即使单位Token成本持续下降，只要Agent可执行任务数量、任务复杂度及经济价值持续提升，整体Inference Compute需求仍有望保持较快增长，持续利好GPU/ASIC、HBM、网络与光模块、数据中心及电力等基础设施环节。Agent有望成为下一阶段AI应用商业化和推理算力需求增长的重要驱动力。

- **风险提示：**大模型能力提升及Agent技术进展低于预期风险，Agent商业化及企业渗透低于预期风险，AI推理算力需求增长低于预期风险，AI应用商业模式及付费意愿低于预期风险，软件行业竞争格局变化风险，AI安全、权限及监管要求趋严风险等。

仅供内部参考，请勿外传！

GPT-6 Astra发布：模型从“会回答”进一步走向“能解决问题”

- **OpenAI发布GPT-6 Astra，模型在数学、科学、软件工程等高难度任务上继续明显提升，FrontierMath Tier 4达到约98%，ARC-AGI-3达到99.9%。**但相比单项Benchmark刷新，我们认为本轮更值得关注的是模型能力结构的变化：Astra不仅推理能力进一步增强，在长周期任务、工具调用、Computer Use以及端到端专业工作上的能力同步提升，大模型开始由单轮问答和内容生成进一步走向复杂问题解决。

- **模型效率评价体系有望从单Token价格向单任务的成本转变。**虽然Astra API单Token价格高于GPT-5.6 Sol，但在部分复杂Agent和软件工程任务中，由于输出Token减少、任务成功率提升以及执行效率改善，完成一个有效任务的综合成本反而可以下降。因此模型竞争的核心指标有望由“每百万Token价格”，转向“完成一项真实任务所需要的总Token、时间、成功率以及人工干预次数”。

图：能力跃迁对比

能力维度	GPT-5.6 Sol	GPT-6 Astra	意义
FrontierMath Tier 4	83.0%	97.6%	前沿数学推理继续提升
ARC-AGI-3	7.8%	99.9%	陌生环境探索与抽象推理跃迁
Terminal-Bench 4.0	37.3%	57.9%	复杂软件工程任务能力提升
OSWorld 2.0	65.7%	72.6%	Computer Use能力增强
AutomationBench	18.1%	41.4%	真实工作流自动化能力提升
MRCR 512K–1M	73.8%	96.3%	长上下文、长任务保持能力提升

资料来源：OpenAI，国信证券经济研究所整理

仅供内部参考，请勿外传！

Computer Use迎来代际升级，AI从Copilot进一步迈向Agent

- **Astra本轮最重要的能力突破之一来自Computer Use**。模型已经可以直接操作浏览器和计算机界面，完成网页表单填写、CRM信息更新、日历管理、数据分析、软件安装测试、网站开发及QA等任务。在OSWorld 2.0测试中，Astra得分由GPT-5.6 Sol的65.7%提升至72.6%，同时模拟单任务耗时由约75分钟下降至40分钟；结合新一代Codex Agent Harness后，Mind2Web任务完成速度达到上一代体验的约1.9倍。

- **Computer Use的意义并不仅是“模型会操作鼠标和键盘”，而是AI开始能够将推理能力直接转化为现实工作流执行**。过去Copilot更多提供建议、代码或内容，最终仍需要人完成软件操作；未来Agent则可以直接调用浏览器、Office、代码环境及专业软件，自主完成多步骤任务。GUI在一定程度上成为AI连接不同软件系统的“通用接口”，大幅降低Agent进入真实企业工作流所需要的系统集成门槛

图：工作方式：从辅助人执行，到AI直接完成工作流



资料来源：OpenAI，国信证券经济研究所整理

表：性能提升明显

指标	GPT-5.6 Sol	GPT-6 Astra	变化
OSWorld 2.0	65.7%	72.6%	+6.9pct
ScreenSpot-Pro	76.9%	92.7%	+15.8pct
单任务耗时	~75分钟	~40分钟	-47%
Mind2Web速度	1.0x	~1.9x	近翻倍

资料来源：OpenAI，国信证券经济研究所整理

仅供内部参考，请勿外传！

专业工作能力增强，Agent评价标准有望从“回答准确率”转向“任务完成率”

- Astra针对复杂知识工作进行了进一步训练，可以在一个任务中同时完成信息检索、分析、软件操作，以及Word、Excel、PPT等最终交付物制作，并能够遵循企业已有模板、视觉风格及业务规则。与此同时，其长任务上下文保持能力进一步增强，Codex开始支持跨Context Window保存工作笔记并重新检索此前对话和工具调用结果，从而降低长周期任务中由于上下文压缩造成的信息丢失。

- Agent有望从“生成中间内容”向“交付最终结果”演进。过去评价大模型更多关注知识问答、Coding Accuracy或单次推理正确率，而在Agent时代，更关键的指标将逐渐变成Task Completion Rate：模型能否在较少人工干预下，稳定完成一个包含几十甚至数百个步骤的完整Workflow。模型的经济价值也将更多取决于能够独立承担多少比例的真实业务流程。

表：Astra推动Agent从“内容生成工具”向“端到端数字劳动力”演进

能力环节	传统 Copilot / Chatbot	GPT-6 Astra / Professional Agent	产业意义
任务输入	接收单轮问题或明确指令	接收高层目标与复杂任务	从Prompt执行转向目标驱动
信息获取	依赖用户提供材料	自主搜索、浏览、调用工具	减少人工资料准备
分析推理	输出答案、代码或建议	连续推理并动态调整执行路径	支持长周期复杂任务
软件操作	人工复制、粘贴、执行	直接操作浏览器、Office、IDE等	AI从建议者变成执行者
上下文管理	Context压缩后容易丢失历史信息	跨Context Window保留Notes，并可检索历史工具结果	长任务连续性明显改善
结果交付	生成文字、代码等中间产物	直接生成Word、Excel、PPT等最终交付物	从“生成内容”走向“完成工作”
评价指标	Accuracy / Benchmark Score	Task Completion Rate	评价标准发生变化
人工角色	人负责组织流程、检查与收尾	人更多负责目标设定、审核和例外处理	Human-in-the-loop显著减少

资料来源：OpenAI，国信证券经济研究所整理

仅供内部参考，请勿外传！

Cyber首次进入Critical等级，自主解决未知问题能力开始显现

- **GPT-6 Astra成为OpenAI首个在Preparedness Framework中达到Cybersecurity “Critical”能力等级的模型**。Astra在ExploitBench上达到100%，明显高于GPT-5.6 Sol的78.5%；更值得关注的是，在采用近期漏洞构建的内部测试中，模型还发现并利用了两个此前未知的Zero-day漏洞。其能力已经由代码生成进一步延伸至漏洞发现、利用开发、测试验证以及连续迭代。

- Cyber Critical意义在于证明模型在长周期执行、工具使用、自主探索和错误修正等能力上已经达到新的水平。模型正从“回答一个已知问题”逐渐走向“面对未知环境自主寻找解决路径”。对应到产业端，AI在软件开发中的价值也有望由Coding继续扩展至代码审计、漏洞扫描、自动修复及安全运维等更完整的软件生命周期。**Cyber Critical的重要性在于它首次验证了AI在陌生复杂环境中持续探索、调用工具、修正错误并找到未知解决路径的能力**。

表：Astra首次进入Cybersecurity Critical能力等级

维度	GPT-6 Astra表现	意义
Preparedness等级	Cybersecurity首次达到 Critical	模型网络安全能力进入新的风险等级
ExploitBench	100%，GPT-5.6 Sol为78.5%	已知漏洞利用能力显著提升
Zero-day发现	内部测试中发现并利用 2个此前未知漏洞	开始具备未知漏洞探索能力
任务模式	可连续完成代码分析、漏洞定位、Exploit开发、测试与修正	从单点辅助走向端到端Cyber Agent
自主性	在较少人工干预下持续探索攻击路径	体现长周期Agent、自主探索和纠错能力
AI Self-Improvement	仍低于High门槛	Critical不等于AGI或RSI
产业机会	代码审计、漏洞扫描、自动修复、安全运维	AI Security应用范围扩大
新增约束	权限隔离、身份认证、行为审计、轨迹监控	Agent安全基础设施重要性上升

资料来源：OpenAI，国信证券经济研究所整理

仅供内部参考，请勿外传！

产业影响：软件开发与SaaS率先进入Agentic阶段

- 软件开发可能成为最早被Agent深度重构的专业行业之一。过去Copilot主要提升程序员单点编码效率，而Astra已经能够理解代码库、修改代码、运行测试、操作浏览器验证结果并持续Debug，AI可覆盖的价值链开始由Coding向需求理解、开发、测试、Debug乃至部署扩展。由此软件工程人均产出有望继续提升，简单开发和外包业务可能承受更明显的价格压力。

- 企业软件的交互方式也可能随之变化。过去员工直接操作CRM、ERP、Office等应用，未来可能逐步转向“用户→Agent→多个软件/API”。短期来看，拥有企业数据、Workflow和权限体系的平台型SaaS厂商更具优势；长期来看，单纯依赖Seat数量收费、缺乏核心数据和Workflow壁垒的软件可能面临Seat压缩压力，商业模式有望进一步向Consumption、Agent Seat乃至Outcome-based Pricing迁移。

图：AI从“辅助开发”走向“重构软件生产与交互方式”



资料来源：OpenAI，国信证券经济研究所整理

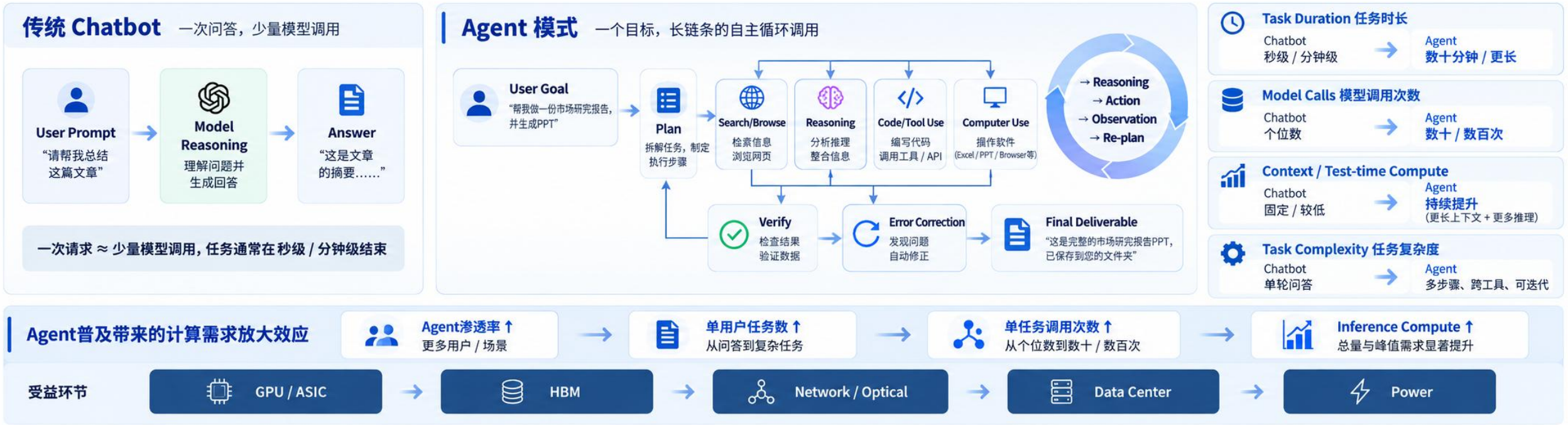
仅供内部参考，请勿外传！

产业影响：Agent拉长任务链条，Inference需求天花板进一步打开

- 传统Chatbot一次用户请求通常只对应有限模型调用，而复杂Agent为完成一个完整任务，需要持续进行推理、搜索、代码执行、Computer Use以及多轮验证，单个业务任务背后可能对应几十甚至数百次模型及工具调用。随着Astra能够处理耗时几十分钟甚至更长的工作流，模型推理时间、上下文长度及Test-time Compute都有望持续增加，AI Token需求正在从“人类主动调用”进一步转向“Agent自主调用”。

- 因此，即使单位Token成本继续下降，总体Inference Compute需求仍可能保持较快增长。更关键的是，未来算力需求可能不再简单取决于用户数量，而取决于“每个用户委托给Agent多少工作”。当AI能够完成更高价值、更长周期的任务后，用户也更有动力为更高推理成本买单。由此GPU/ASIC、HBM、网络、光模块、电力和数据中心等基础设施仍有望持续受益于Agent渗透率提升。

图：Agent将AI计算需求从“人类调用”升级为“自主循环调用”



资料来源：OpenAI，国信证券经济研究所整理

仅供内部参考，请勿外传！

投资观点：Agent有望成为下一阶段AI应用与Inference需求增长的核心驱动力

- **GPT-6 Astra进一步强化了我们对于AI产业从“模型能力竞争”走向“Agent应用落地”的判断**。本轮能力升级的重点并不只是数学、科学、Coding等Benchmark继续提升，而是Computer Use、长周期任务执行、复杂工具调用以及端到端专业工作能力同步增强。过去大模型更多承担信息检索、内容生成和局部辅助，最终仍需要人完成软件操作、流程组织和结果交付；Astra开始具备“理解目标—拆解任务—调用工具—操作软件—验证结果—持续修正—最终交付”的完整闭环，意味着AI正在由Copilot式生产力工具进一步向可独立承担Workflow的数字劳动力演进。我们认为，下一阶段AI产业竞争的核心指标将逐渐由单项Benchmark和单Token价格，转向Task Completion Rate、单任务综合成本、任务执行时间以及人工干预比例，AI商业价值的计量单位也有望从“每百万Token”逐步转向“完成一项真实工作需要多少钱”。

- **从产业链受益顺序来看，我们认为Inference基础设施仍是当前确定性最高的方向，而Agent应用具备更大的中长期弹性**。传统Chatbot一次用户请求通常只对应有限次数的模型调用，而复杂Agent为完成一个真实任务，可能持续进行推理、搜索、代码执行、Computer Use、结果验证和错误修正，单个任务对应的模型调用次数有望由个位数扩展至几十次甚至数百次，同时任务持续时间由秒级、分钟级向数十分钟乃至更长周期延伸。即便单位Token成本持续下降，只要Agent可完成任务的数量、复杂度和经济价值持续提升，整体Inference Compute需求仍有望保持较快增长。因此，GPU/ASIC、HBM、网络与光模块、数据中心及电力等基础设施仍是Agent渗透最直接、确定性最高的受益环节。应用侧则重点关注Coding、企业办公、SaaS、安全以及专业知识工作等高价值场景，其中软件开发由于任务数字化程度高、验证标准明确、工具链成熟，有望成为Agent最早实现规模化渗透的行业之一。

- **中期来看，Agent更可能推动企业软件竞争格局和商业模式进一步分化**。 当用户与软件之间的交互方式由“用户直接操作SaaS UI”逐渐演变为“用户→Agent→多个软件/API”后，软件入口的重要性可能部分下降，而企业数据、Workflow、权限体系、系统连接能力以及Agent Orchestration的重要性进一步上升。拥有核心业务数据、深度嵌入企业流程并掌握系统权限的平台型软件厂商，更有能力成为Agent执行任务的基础设施和控制层；相反，缺乏核心数据及Workflow壁垒、主要依赖Seat数量收费的标准化工具软件，可能面临用户席位压缩和价格压力。商业模式也可能逐步由传统Per-seat Subscription向Consumption、Agent Seat以及Outcome-based Pricing迁移。总体而言，我们认为未来2-3年AI产业的核心主线将从“谁拥有最好的Chatbot”逐步转向“谁能够让AI安全、稳定、低成本地完成最多真实世界任务”，Agent有望同时成为AI应用商业化和推理算力需求增长的重要驱动力。。

仅供内部参考，请勿外传！

风险提示

大模型能力提升及Agent技术进展低于预期风险。若后续前沿模型能力迭代速度放缓，复杂任务成功率、执行稳定性或长任务持续能力提升低于预期，可能导致Agent由Demo向真实生产环境落地的节奏慢于预期，并影响相关AI应用及基础设施需求增长。

Agent商业化及企业渗透低于预期风险。 当前Agent仍处于从辅助工具向端到端工作流执行演进的早期阶段，企业客户对于模型准确率、可控性、数据安全、系统集成及投资回报率仍存在较高要求。若Agent在真实业务流程中的Task Completion Rate、成本效率或可靠性改善不及预期，企业可能继续以Copilot等低自治形态部署AI，从而导致Agent用户渗透率、付费率及相关软件收入增长低于预期。

AI推理算力需求增长低于预期风险。若模型推理效率及算法优化速度显著快于Agent任务量增长，或企业通过小模型、缓存、蒸馏、路由及本地推理等方式显著降低单位任务算力消耗，则Agent渗透对整体Inference Compute需求的拉动可能低于预期，进而影响GPU/ASIC、HBM、网络、光模块、数据中心及电力等相关基础设施需求。

AI应用商业模式及付费意愿低于预期风险。若Agent的单位任务综合成本下降速度有限，或客户对于Consumption、Agent Seat及Outcome-based Pricing等新收费模式接受程度低于预期，软件厂商可能难以将模型能力提升有效转化为收入及利润增长，AI应用商业化进程可能慢于市场预期。

软件行业竞争格局变化风险。 Agent可能改变传统企业软件的用户入口、Seat数量及价值分配方式，但具体影响仍存在较大不确定性。若基础模型厂商或通用Agent平台直接进入企业应用层，可能压缩部分传统SaaS厂商的产品入口及议价能力；另一方面，若企业数据、Workflow、权限体系及行业Know-how仍高度依赖现有软件平台，传统龙头也可能进一步强化竞争优势。不同软件公司的受益程度可能明显分化。

AI安全、权限及监管要求趋严风险。 随着模型Computer Use、代码执行及自主工具调用能力增强，Agent误操作、越权访问、数据泄露、模型欺骗及网络安全等潜在风险同步上升。若监管机构进一步加强对高自治AI系统、企业数据调用、模型责任认定及网络安全能力的限制，或企业显著提高Human-in-the-loop、权限审批及审计要求，可能降低Agent自主程度并增加部署成本，从而影响高自治Agent商业化速度。

技术路线及行业发展判断存在不确定性风险。 当前AI技术迭代速度较快，模型架构、Agent Framework、Computer Use、Memory、Context Engineering及软硬件协同方式仍处于快速演进阶段，未来可能出现不同于当前判断的新技术路线或商业模式。

仅供内部参考，请勿外传！

免责声明

国信证券 GUOSEN SECURITIES

国信证券投资评级

投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数(HSI.HI)作为基准；美国市场以标普500指数(SPX.GI)或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者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，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

仅供内部参考，请勿外传！

国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046　　总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032