

计算机

开源模型重磅出击，设备更新&消费品以旧换新政策加码

投资要点：

➤ Llama3.1 系列模型表现出色

Llama3.1 是由 Meta 公司发布的开源模型系列，最大参数量达到 4050 亿。它在多个 benchmark 测试中表现出色，与 gpt 4o 和 claude 3.5 sonnet 等闭源 SOTA 模型性能相当，甚至在某些方面实现了超越。Llama3.1 系列模型的上下文长度扩展到 128K，支持八种语言，并且在训练过程中，Meta 公司使用了超过 15 万亿个 token，优化了训练堆栈，并采用了多种技术手段提升模型性能和训练效率。

➤ Mistral Large 2: 千亿参数级的六边形战士

Mistral Large 2 是一个拥有 1230 亿参数的模型，它在代码生成、数学和推理等方面能力显著增强。虽然参数量不及 Llama3.1 的 1/3，但其综合性能与 Llama3.1 相当，甚至在某些基准测试中接近 GPT-4o 和 Claude 3.5 Sonnet。Mistral Large 2 模型在优化模型幻觉方面也取得了进展，能够在不确定时承认无法回答问题，而不是编造答案，其回答风格简洁明了，生成回答的平均长度在可比模型中较低。

➤ 设备更新及消费品以旧换新力度加码

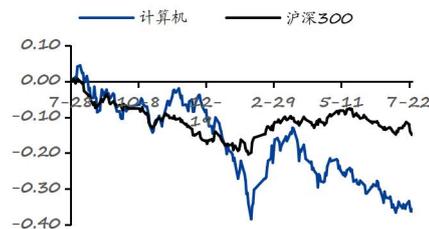
2024 年 7 月 24 日，中国国家发改委和财政部联合发布了《关于加力支持大规模设备更新和消费品以旧换新的若干措施》的通知，计划安排大约 3000 亿元的超长期特别国债资金，以支持大规模的设备更新和消费品以旧换新。政策覆盖了能源电力、老旧电梯等领域的设备更新，以及重点行业的节能降碳和安全改造，并为中小企业提供了支持。同时，对家电产品和汽车报废更新提供了财政补贴，以促进消费升级和环境保护。政策的实施有望加速设备更新和消费品以旧换新的进程。

➤ 风险提示

政策实施效果不及预期的风险；开源模型持续迭代不及预期的风险等。

强于大市（维持评级）

一年内行业相对大盘走势



团队成员

分析师：钱劲宇(S0210524040006)
QJY3773@hfzq.com.cn

相关报告

- 1、微软再回应蓝屏事件，影响全球近 850 万台设备——2024.07.21
- 2、百度萝卜快跑爆火出圈，特斯拉或推迟发布 Robotaxi——2024.07.14
- 3、“车路云一体化”试点城市名单公布——2024.07.07



正文目录

1 Llama3.1&Mistral 新模型发布，开源综合性能直逼闭源 SOTA.....	3
1.1 Llama3.1 系列模型表现出色.....	3
1.2 Mistral Large 2: 千亿参数级的六边形战士.....	4
2 设备更新及消费品以旧换新力度加码.....	5
3 AI 周度数据更新（20240719-20240725）.....	6
4 风险提示.....	7

图表目录

图表 1: Llama3.1 405B 基准测试结果.....	3
图表 2: Llama3.1 8B&70B 基准测试结果.....	3
图表 3: Llama3.1 训练流程.....	4
图表 4: Mistral Large 2 &Llama3.1 MMLU 基准测试对比.....	4
图表 5: Mistral Large 2 代码基准能力测试.....	5
图表 6: Mistral Large 2 生成回答的平均长度对比.....	5
图表 7: 海外 AI 应用周度下载量（单位：次）.....	6
图表 8: 国内 AI 应用周度下载量（单位：次）.....	7

1 Llama3.1&Mistral 新模型发布，开源综合性能直逼闭源 SOTA

7月24日及25日，开源模型界迎来重磅更新，Meta正式发布Llama3.1系列模型，最大参数量达4050亿；Mistral紧随其后发布Mistral Large 2，代码生成、数学和推理等能力显著增强，模型参数量仅为1230亿，不及Llama3.1的1/3。两者的综合性能均能和gpt 4o以及claude 3.5 sonnet对齐，开源模型和闭源SOTA之间的差距正加速缩小。

1.1 Llama3.1 系列模型表现出色

综合能力上，llama3.1 405B模型与gpt 4o和claude 3.5 sonnet相当，部分benchmark已实现超越。8B和70B小型模型在与相似参数量的模型中优势明显。上下文长度上，此次系列模型将上下文长度扩展到128K，达到主流模型水平，并增加对八种语言的支持。

图表 1: Llama3.1 405B 基准测试结果

Category Benchmark	Llama 3.1 405B	Nemotron 4 340B Instruct	GPT-4 (0-shot)	GPT-4 Omni	Claude 3.5 Sonnet
General	88.6	78.7 (non-CoT)	85.4	88.7	88.3
MMLU (0-shot, CoT)	73.3	62.7	64.8	74.0	77.0
MMLU PRO (0-shot, CoT)	88.6	85.1	84.3	85.6	88.0
IFEval	88.6	85.1	84.3	85.6	88.0
Code					
HumanEval (0-shot)	89.0	73.2	86.6	90.2	92.0
MBPP EvalPlus (Base) (0-shot)	88.6	72.8	83.6	87.8	90.5
Math					
GSM8K (0-shot, CoT)	96.8	92.3 (0-shot)	94.2	96.1	96.4 (0-shot)
MATH (0-shot, CoT)	73.8	41.1	64.5	76.6	71.1
Reasoning					
ARC Challenge (0-shot)	96.9	94.6	96.4	96.7	96.7
GPQA (0-shot, CoT)	51.1	-	41.4	53.6	59.4
Tool use					
BFCL	88.5	86.5	88.3	80.5	90.2
Nexus	58.7	-	50.3	56.1	45.7
Long context					
ZeroSCROLLS/QUALITY	95.2	-	95.2	90.5	90.5
InfiniteBench/EnMC	83.4	-	72.1	82.5	-
NH/Multi-needle	98.1	-	100.0	100.0	90.8
Multilingual					
Multilingual MGSM (0-shot)	91.6	-	85.9	90.5	91.6

来源：机器之心，华福证券研究所

图表 2: Llama3.1 8B&70B 基准测试结果

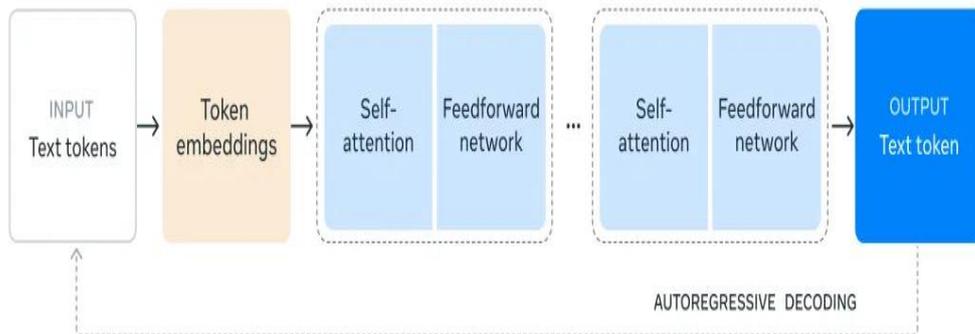
Category Benchmark	Llama 3.1 8B	Gemma 2 9B IT (5-shot, non-CoT)	Mistral 7B Instruct	Llama 3.1 70B	Mistral 8x22B Instruct	GPT 3.5 Turbo
General	73.0	72.3 (5-shot, non-CoT)	60.5	86.0	79.9	69.8
MMLU (0-shot, CoT)	48.3	-	36.9	66.4	56.3	49.2
MMLU PRO (0-shot, CoT)	80.4	-	57.6	87.5	72.7	69.9
IFEval	80.4	-	57.6	87.5	72.7	69.9
Code						
HumanEval (0-shot)	72.6	54.3	40.2	80.5	75.6	68.0
MBPP EvalPlus (Base) (0-shot)	72.8	71.7	49.5	86.0	78.6	82.0
Math						
GSM8K (0-shot, CoT)	84.5	76.7	53.2	95.1	88.2	81.6
MATH (0-shot, CoT)	51.9	44.3	13.0	68.0	54.1	43.1
Reasoning						
ARC Challenge (0-shot)	83.4	87.6	74.2	94.8	88.7	83.7
GPQA (0-shot, CoT)	32.8	-	28.8	46.7	33.3	30.8
Tool use						
BFCL	76.1	-	60.4	84.8	-	85.9
Nexus	38.5	30.0	24.7	56.7	48.5	37.2
Long context						
ZeroSCROLLS/QUALITY	81.0	-	-	90.5	-	-
InfiniteBench/EnMC	65.1	-	-	78.2	-	-
NH/Multi-needle	98.8	-	-	97.5	-	-
Multilingual						
Multilingual MGSM (0-shot)	68.9	53.2	29.9	86.9	71.1	51.4

来源：机器之心，华福证券研究所

模型训练上，Meta使用超15万亿个token进行训练。为了实现高效的训练，Meta优化了整个训练堆栈，并使用超过16000个H100的GPU。为解决训练中预见的难点，Meta在训练设计上做出以下优化：一方面仅进行少量调整的标准解码器Transformer模型架构，而不是MOE模型，以最大限度地提高训练稳定性；采用迭代后训练程序，每轮都使用监督微调和直接偏好优化。这使Meta能够为每轮创建最高质量的合成数据，并提高每项功能的性能。此外，Meta将模型从BF16量化为FP8，有效降低所需的计算需求，并允许模型在单个服务器节点上运行。为提升小模型的能力，Meta还是用405B的模型来提升小模型的后训练质量。



图表 3: Llama3.1 训练流程

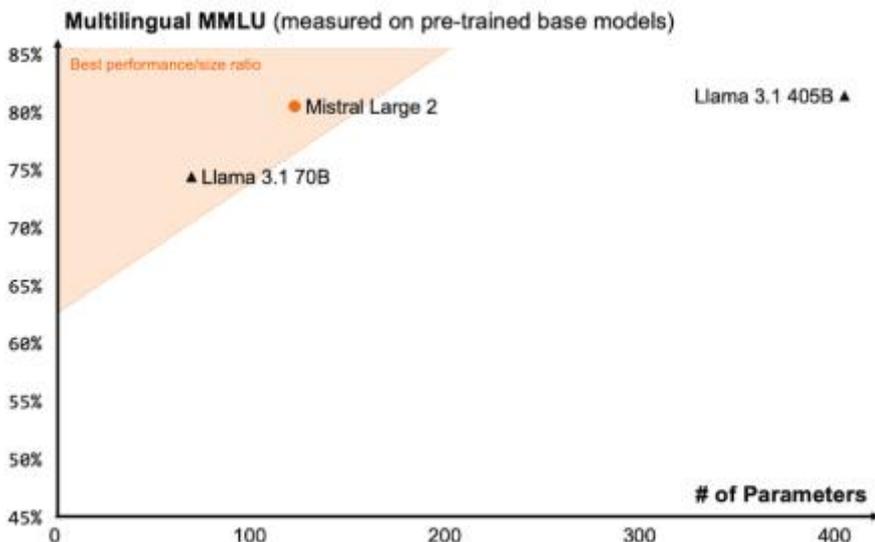


来源: 机器之心, 华福证券研究所

1.2 Mistral Large 2: 千亿参数级的六边形战士

Mistral Large 2 拥有 1230 亿参数, 128k 上下文窗口, 可以在单个节点上以高吞吐量运行, 支持包括中文在内的数 10 种语言、80 多种编程语言, 在多项基准测试中接近 GPT-4o 和 Claude 3.5 Sonnet, 预训练版本的多任务语言理解准确率高达 84%, 击败 Llama 3.1 70B。

图表 4: Mistral Large 2 & Llama3.1 MMLU 基准测试对比

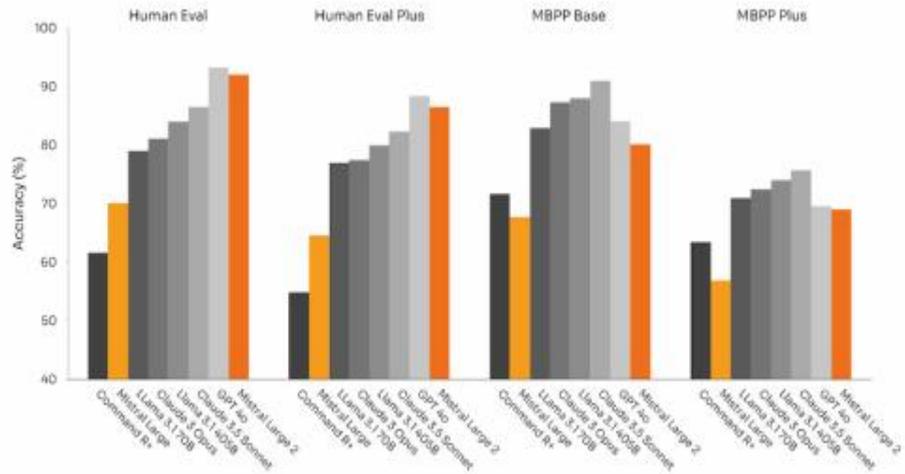


来源: 智东西, 华福证券研究所

较前代版本, Mistral Large 2 的代码生成能力显著提升, 综合能力接近 GPT-4o、Claude3 Opus 和 Llama 3 405B。



图表 5: Mistral Large 2 代码基准能力测试



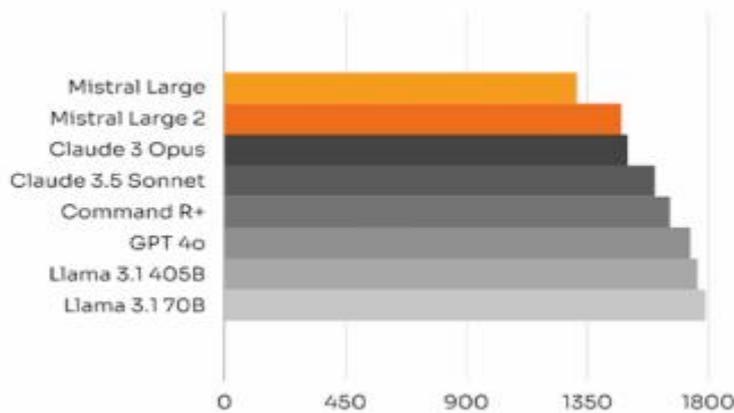
Performance accuracy on code generation benchmarks (all models were benchmarked through the same evaluation pipeline)

来源: 智东西, 华福证券研究所

Mistral Large 2 显著优化了模型幻觉。通过训练, Mistral Large 2 能在自己无法找到解决方案, 或没有足够自信能提供准确答案时, 会承认无法回答该问题, 而不是继续编造答案。为加快交互速度, Mistral Large 2 的回答风格简洁明了, 生成回答的平均长度在可比模型中较低。

图表 6: Mistral Large 2 生成回答的平均长度对比

MT Bench (dev) - Average Generation Length



来源: 智东西, 华福证券研究所

2 设备更新及消费品以旧换新力度加码

2024 年 7 月 24 日, 国家发改委、财政部印发《关于加力支持大规模设备更新和消费品以旧换新的若干措施》的通知, 统筹安排 3000 亿元左右超长期特别国债资金, 加大力度支持大规模设备更新和消费品以旧换新。

设备更新上, 安排超长期特别国债大规模设备更新专项资金加大对设备更新的支持力度。在传统设备更新的基础上, 支持范围扩大到能源电力、老旧电梯等领域



设备更新以及重点行业节能降碳和安全改造，并结合实际动态调整。并去除 1 亿元的申报门槛，支持中小企业设备更新。报废更新上，支持老旧营运船舶、老旧营运货车、农业机械的报废更新。设备更新贷款财政贴息比例上，提升中央财政贴息到 1.5%，贴息期限 2 年，总规模 200 亿元。

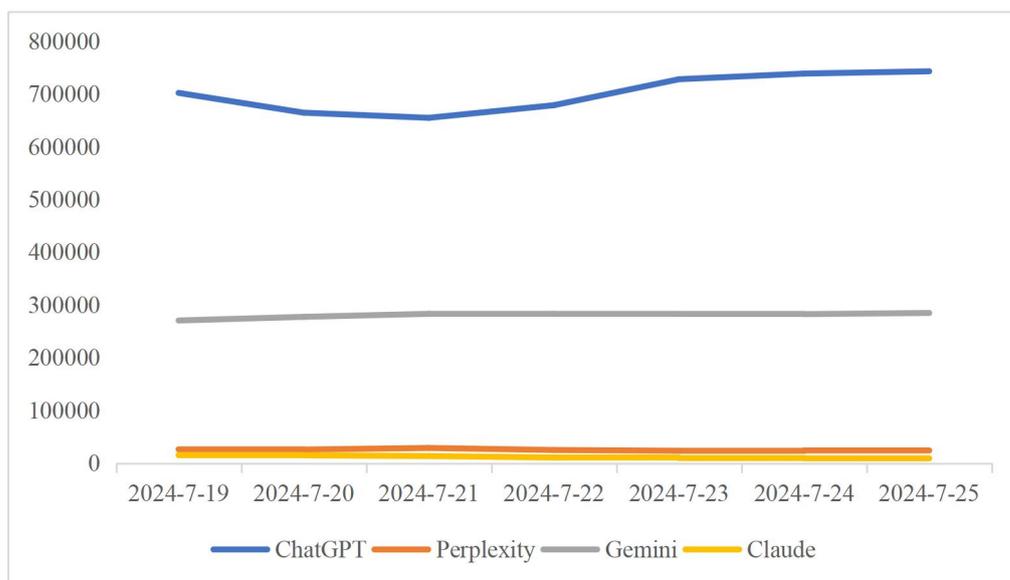
消费品以旧换新上，直接安排超长期特别国债资金，对冰箱、洗衣机、电视等八类家电产品给予以旧换新补贴，其中二级能效补贴 15%，一级能效补贴 20%，每个消费者补贴一件，补贴金额不超过 2000 元。汽车报废更新上，报废国三及以下排放标准燃油乘用车或 2018 年 4 月 30 日（含当日）前注册登记的新能源乘用车，并购买车型目录的新能源车或 2.0 升及以下排量的燃油车，补贴标准分别提高至 2 万元、1.5 万元。

此次设备更新资金用途分配明确，央地共担比例清晰。超出额度部分由地方资金支持。截至 24 年 12 月 31 日未用完的资金将会收回中央。综合结合资金支持力度，时间节点，设备及消费品更换周期来看，此次政策发布有望加速设备更新和消费品以旧换新的落地。

3 AI 周度数据更新（20240719-20240725）

海外 AI 应用下载量上，Chatgpt 仍遥遥领先，峰值日度下载量突破 74 万次；Gemini 下载量稳中有升，周内突破 28 万大关；Perplexity 和 Claude 下载量整体不高，Claude 后半周有所下滑，当前日下载量已不足 1 万次。

图表 7: 海外 AI 应用周度下载量（单位：次）

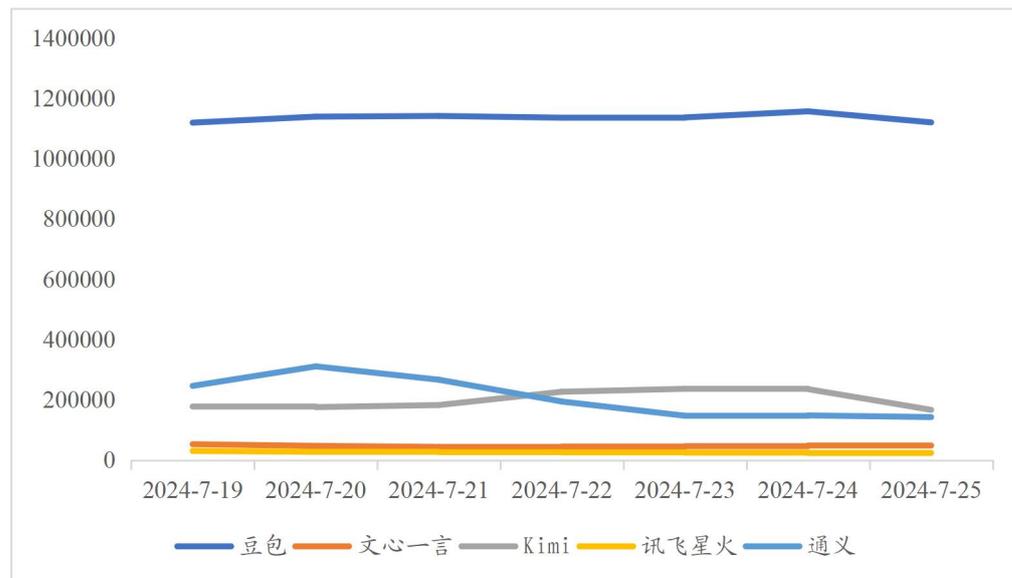


来源：点点数据，华福证券研究所

国内 AI 应用下载量上，豆包依旧遥遥领先，全平台合计下载量已突破 110 万次；Kimi 和通义自周中分别突破 23 万次、30 万次后开始下滑，当前下载量已恢复稳态水平；文心一言和讯飞星火下载量整体保持稳定。



图表 8: 国内 AI 应用周度下载量 (单位: 次)



来源: 点点数据, 华福证券研究所

4 风险提示

政策实施效果不及预期的风险; 开源模型持续迭代不及预期的风险等。

分析师声明

本人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，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将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。

一般声明

华福证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。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。本公司不保证本报告所含信息及资料保持在最新状态，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。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。

本报告版权归“华福证券有限责任公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

类别	评级	评级说明
公司评级	买入	未来 6 个月内，个股相对市场基准指数涨幅在 20%以上
	持有	未来 6 个月内，个股相对市场基准指数涨幅介于 10%与 20%之间
	中性	未来 6 个月内，个股相对市场基准指数涨幅介于-10%与 10%之间
	回避	未来 6 个月内，个股相对市场基准指数涨幅介于-20%与-10%之间
	卖出	未来 6 个月内，个股相对市场基准指数涨幅在-20%以下
行业评级	强于大市	未来 6 个月内，行业整体回报高于市场基准指数 5%以上
	跟随大市	未来 6 个月内，行业整体回报介于市场基准指数-5%与 5%之间
	弱于大市	未来 6 个月内，行业整体回报低于市场基准指数-5%以下

备注：评级标准为报告发布日后的 6~12 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）

联系方式

华福证券研究所 上海

公司地址：上海市浦东新区浦明路 1436 号陆家嘴滨江中心 MT 座 20 层

邮编：200120

邮箱：hfjys@hfzq.com.cn