海外研究：ChatGPT 交易策略 15 个月收益 500%+

——“学海拾珠”系列之一百四十二

报告日期：2023-05-24

分析师：严佳炜
执业证书号：S0010520070001
邮箱：yanjw@hazq.com

分析师：骆昱杉
执业证书号：S0010522110001
邮箱：luoyushan@hazq.com

相关报告
1.《策略拥挤与流动性冲击——“学海拾珠”系列之一百三十四》
2.《基金窗口粉饰行为的新指标——“学海拾珠”系列之一百三十五》
3.《基于堆叠自编码器和长短期记忆网络的金融时间序列深度学习框架——“学海拾珠”系列之一百三十六》
4.《ETF 交易与分析师预测——“学海拾珠”系列之一百三十七》
5.《基金的协偏度择时能力 ——“学海拾珠”系列之一百三十八》
6.《利用深度神经网络改进时间序列动量策略 ——“学海拾珠”系列之一百三十九》
7.《是否存在宏观公告溢价现象？ ——“学海拾珠”系列之一百四十》
8.《前景理论能否解释共同基金的业绩？——“学海拾珠”系列之一百四十一》

主要观点：

本篇是“学海拾珠”系列第一百四十二篇，来源于《Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models》一文。

文章研究了基于大预言模型（LLM）的 ChatGPT 在股票收益预测中的性能，研究使用 ChatGPT 对公司新闻标题进行情感分类（积极、消极和中性），并计算出相应的"ChatGPT 分数"，进而探究其对股票收益的预测能力。测算得到的 Long-Short 交易策略 15 个月获得了 500%+的收益。

研究结果显示，相比于传统的情绪分析方法，未经专业训练的 ChatGPT 表现出显著的优势，揭示了其在股票收益预测方面的潜力，为投资决策提供了一种新的思路和工具。

● 探索基于大预言模型的 ChatGPT 在股票收益预测方面的潜力

本文为首篇通过大预言模型（LLM）预测股票收益的研究，探索了其在股票收益预测方面的潜力。将公司新闻标题进行情感分类，分为积极、消极和中性，并通过ChatGPT模型计算出每个标题的"ChatGPT 分数"。然后进一步研究了这些分数与相应股票收益之间的关系，以评估 ChatGPT 在股票预测方面的表现。

● ChatGPT 在分类新闻标题情绪和预测股票收益方面的性能显著优于传统情绪分析方法

将 ChatGPT 与其他大型语言模型（如 GPT-1、GPT-2 和 BERT）进行比较，评估它们在分类新闻标题情绪和预测股票收益方面的能力。研究发现，相比于传统的情绪分析方法，ChatGPT 能捕捉到新闻标题中的情感细微差别，能够更好地解析新闻标题中的情感信息，并据此预测股票收益。

● ChatGPT 在小市值股票和负面新闻预测中展现出卓越性能

研究发现，将 ChatGPT 应用于投资决策可以带来准确的预测结果，尤其在小市值股票领域。此外，ChatGPT 在预测负面消息方面表现出相对突出的能力。这一研究结果为量化研究提供了新的数据和思路。

● 风险提示

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

正文目录

1 CHATGPT 在金融领域中的预测能力评估与展望：挑战与机遇 4
2 CHATGPT 在金融领域的应用与前沿研究综述 4
3 CHATGPT 对股票收益的预测能力探究 5
4 模型数据：非 CHATGPT 的训练集 6
5 模型方法 6
5.1 适当的“提示”能提高 CHATGPT 输出的质量 6
5.2 CHATGPT 评分与股市回报的关联性分析 8
6 CHATGPT 对股票收益的预测能力显著 8
7 利用情绪分析预测股市回报的潜力及启示 14
风险提示： 15

图表目录

图表 1：描述性统计 8
图表 2：相关性分析 8
图表 3：1 USD 的累计收益率（不含交易成本） 9
图表 4：LONG-SHORT 策略在不同交易成本下投资 1 美元的累积回报 9
图表 5：SHORT 策略在不同交易成本的短期策略中投资 1 美元的累积回报 11
图表 6：次日收益对预测得分的回归 11
图表 7：次日收益率对预测得分的回归（小市值股） 12
图表 8：次日收益率对预测得分的回归（非小市值股） 12
图表 9：投资组合描述性统计 14
图表 10：准确性、预测性、召回率、特异性和 F1 得分指标 14
图表 11：按预测分数划分的平均次日回报率 14

1 ChatGPT 在金融领域中的预测能力评估与展望：挑战与机遇

近几个月来，许多研究探索生成式人工智能和以 ChatGPT 为代表的大语言模型（LLM）在不同领域的巨大潜力。然而，大语言模型（LLM）在预测股市收益等金融领域的应用仍然有不足。一方面，由于大语言模型没有专门针对金融领域进行训练，并被认为可能无法为预测股价走势提供价值。但是，另一方面，大语言模型更能理解自然语言，是处理文本信息来预测股票收益的重要工具。因此，大语言模型在预测金融市场走势方面的表现是一个富有争议的话题。

本文是评估 ChatGPT 预测股市收益能力的首作。本文通过一种利用 LLM 模型分析情绪，即使用新闻标题评估法分析 ChatGPT 预测股市收益的能力，并与现有情绪分析方法比较。已有研究表明 (Fedyk 和 Hodson(2023))，使用新闻和训练过的模型可以预测日内股票收益率(Tetlock (2007), Tetlock, Saar-Tsechansky 和 Macskassy (2008)和 Tetlock (2011))，但融入新信息是困难的。**因此，本文重点评估没有经过专门训练的模型是否有这种预测能力。**

本文的研究结果可能会导致市场预测和投资决策方法的转变，通过展示 ChatGPT 在金融领域中的价值，有助于 LLM 在金融领域的应用，并为在金融市场中进一步研究应用人工智能模型和自然语言处理模型的提供启发。

首先，本研究可以帮助监管机构和政策制定者了解金融市场采用大语言模型的潜在好处和风险。随着大语言模型越来越普遍，其对市场行为、信息传播和价格形成的影响也将成为人们关注的关键领域。本文的研究结果可以为关于监管人工智能在金融领域应用的讨论提供支持，并提供将大语言模型整合到市场运营中的重要实践。

其次，本文提供了大语言模型在预测股市收益有效性的经验证据，在一定程度上协助研究人员在将大语言模型纳入投资模型，以便做出更理智的投资决策，提高投资绩效，减少对传统的劳动密集型分析方法的依赖。

最后，本研究有助于更广泛地讨论人工智能在金融中的应用。通过探索 ChatGPT 预测股市收益的能力，本文进一步解释了大语言模型在金融领域的潜力和局限性，并激励研究人员开发更复杂的大语言模型，以满足金融行业需求，为更高效、更精准的投资决策提供支持。

本研究超越了预测股票收益的现有框架，通过阐明 ChatGPT 对金融领域的潜在贡献，鼓励在人工智能驱动下的探索和金融创新，为金融市场预测和投资决策方法的改进提供了参考，同时也为进一步应用人工智能模型和自然语言处理模型于金融领域的研究提供了启示。

2 ChatGPT 在金融领域的应用与前沿研究综述

最近在金融学背景下使用 ChatGPT 的论文包括 Hansen 和 Kazinnik(2023), Cowen 和 Tabarrok (2023), Korinek (2023)和 Noy 和 Zhang (2023)。Cowen 和 Tabarrok(2023)和 Korinek(2023)证明 ChatGPT 在经济学教学和研究中可以提供帮助。Noy 和 Zhang(2023)发现 ChatGPT 可以提高专业写作工作效率。同时，Xie 等(2023)发现当在预测任务中使用数值数据时，ChatGPT 的表现并不比线性回归等简单方法更好。本文将该结果归因于 Xie 等(2023)专注于使用历史数字数据进行预测，

而 ChatGPT 擅长文本分析任务。Ko 和 Lee(2023)发现 ChatGPT 可能有助于跨资产类别的选择。此外，Yang 和 Menczer(2023)证明 ChatGPT 成功地识别了可信的新闻媒体。**本研究是第一批研究大语言模型在金融市场，特别是投资决策过程中应用的研究之一。**

本研究为采用文本分析和机器学习来研究各种金融研究问题的文献（(e.g., Jegadeesh 和 Wu(2013), Campbell et al. (2014), Hoberg 和 Phillips (2016), Gaulin (2017), Baker, Bloom 和 Davis (2016), Manela 和 Moreira (2017), Hansen, McMahon 和 Prat (2018), Ke, Kelly 和 Xiu (2019), Ke, Montiel Olea 和 Nesbit (2019), Bybee et al. (2019), Gu, Kelly 和 Xiu (2020), Cohen, Malloy 和 Nguyen (2020), Freyberger, Neuhierl 和 Weber (2020), Lopez-Lira 2019, Binsbergen et al. (2020), Bybee et al. (2021)）。首先本文首次评估了以 ChatGPT 为代表的大语言模型在预测股市走势方面的文本处理能力。其次，本研究还丰富了使用新闻文章的文本分析来提取情绪指标并预测股票回报的文献。现有文献，一部分研究舆情情绪和股票回报(如 Tetlock (2007), Garcia (2013), Calomiris 和 Mamaysky(2019))，另一部分利用公司新闻的情绪来预测未来的个股回报(如，Tetlock, Saar-Tsechansky 和 Macskassy (2008), Tetlock(2011), Jiang, Li 和 Wang (2021))。与以往的研究不同，本文的研究重点是通过提取预测股市的附加信息来了解大预言模型是否能预测股票收益率。

最后，本研究还涉及了就业敞口和对人工智能相关技术的脆弱性文献梳理。Agrawal, Gans 和 Goldfarb (2019), Webb(2019), Acemoglu et al. (2022), Acemoglu 和 Restrepo (2022), Babina et al. (2022), Noy 和 Zhang (2023)的最新研究了人工智能相关技术的就业敞口和脆弱性，以及对就业和生产力的影响。随着人工智能不断发展，本文研究重点是理解一个紧迫但尚未解决的问题——人工智能，尤其是大语言模型在金融领域的潜力，并强调大语言模型为预测股票回报提供信息处理的潜力。

3 ChatGPT 对股票收益的预测能力探究

ChatGPT 是 OpenAI 基于 GPT (生成式预训练 Transformer)架构开发的大语言模型，是迄今为止最为先进的自然语言处理(NLP)模型之一，并在大量文本数据语料库上进行训练，以理解自然语言的结构和模式。生成式预训练 Transformer (GPT)是一种用于自然语言处理任务的深度学习算法，由 OpenAI 基于 Transformer 架构开发。GPT 框架在文本翻译、摘要提取、问题回答和信息补全等各种自然语言处理任务中实现了最为优越的性能。

GPT 框架使用多层神经网络对自然语言的结构和模式进行建模，使用无监督学习方法在大量文本数据(如维基百科)上进行预训练。这种预训练过程使模型能够深入理解语言、语法和语义，之后针对特定的语言任务进行微调。GPT 框架的一个独特特性是它使用了 Transformer 框架，这使模型能够通过使用自注意力机制来处理长文本序列，并集中在输入中最相关的部分。这种注意力机制使模型能够更好地理解输入上下文，并产生更准确和连贯的反应。

ChatGPT 经过训练可以执行如文本翻译、摘要提取、问题回答和信息补全等各种语言任务。ChatGPT 可以生成类似人类反应的能力使其可以成为创建聊天机器人和与用户交谈虚拟助手。虽然 ChatGPT 对于文本任务来说是一个强大工具，但它并没有经过明确的训练并预测股票回报或提供金融建议。因此，本文测试了 ChatGPT 在预测股票收益时的能力。

4 模型数据：非 ChatGPT 的训练集

本文使用三个主要数据集进行分析:证券价格研究中心(CRSP)的每日收益率、新闻标题和知名新闻情感分析数据提供商 RavenPack。**样本周期从 2021 年 10 月开始(因为 ChatGPT 的训练数据截至到 2021 年 9 月)，到 2022 年 12 月结束。**这个样本周期确保本文的评估是基于模型训练数据中不存在的信息，从而允许对其预测能力进行更准确的评估。

CRSP 每日收益数据集包含在美国主要证券交易所上市公司的**每日股票收益信息，包括股票价格、交易量和市值数据**。此数据集能够检验 ChatGPT 生成的情绪得分与相应的股票收益之间的关系。该样本包括在至少一篇新闻报道的在纽约证券交易所（NYSE）、纳斯达克证券交易（NASDAQ）和美国证券交易所（AMEX）的所有上市公司。

本文**首先使用爬虫通过搜索所有包含公司名称或股票代码的新闻**，为所有 CRSP 公司收集一个全面的新闻数据集。由此产生的数据集包括主要新闻机构、财经新闻网站和社交媒体平台等来自各种来源的新闻标题。**然后，本文将标题与来自知名新闻情感分析数据提供商(RavenPack)的标题进行匹配**，保留有后续回报公司新闻。本文不使用可能包含更多信息的 RavenPack 增强标题，因为此类标题没有广泛传播给公众。最终，本文匹配了 4138 家上市公司的 67,586 条标题。本文使用 Jiang, Li 和 Wang(2021)提出的预处理方法处理合并数据集。

与 RavenPack 匹配可以确保只使用相关新闻进行实验。因为 RavenPack 密切监控主要的金融新闻发布渠道，并有一个匹配新闻、时间戳和实体名称的质量程序，以解决可能来自网络爬虫的任何错误。此外，本文采用 **RavenPack 的新闻分类**来解释不同模型之间股票收益可预测性的差异。此外，RavenPack 与 CRSP 具有紧密的映射关系，确保了新闻和收益率的时间匹配。

本文使用提供的“相关性评分”，范围从 0 到 100，来表示新闻与特定公司的关系有多密切。0 分(100 分)意味着该实体被动地(主要地)提及。本文样本需要相关性分数为 100 的新闻，并将其限制为完整的文章和新闻稿。本文进一步**排除了被归类为“股票上涨”和“股票下跌”的标题**，因为它们只显示了股票的每日走势方向。本文还要求**“事件相似天数”超过 90 天**，以避免新闻重复，确保只捕获有关公司的新信息。此外，本文还消除了同一天同一公司的重复标题和极其相似的标题。

本文使用最佳字符串匹配算法（Damerau-Levenshtein 距离算法）来衡量标题的相似性，并**删除同一公司在同一天相似度大于 0.6 的标题**，以避免前瞻性偏见。这是因为数据供应商会在收到新闻的几毫秒内对所有新闻文章进行评估，并迅速将结果数据发送给用户。因此，可以假设在新闻发布时所有信息都是可用的。

5 模型方法

5.1 适当的“提示”能提高 ChatGPT 输出的质量

“提问提示”Prompt 是指导 ChatGPT 精准回答特定任务和查询的关键。“提示”是一个简短的文本，它为 ChatGPT 生成答案提供上下文和指令，提示可以简单到一个句子，也可以复杂到一个段落或更多，这取决于任务的性质，是作为 ChatGPT 生成答案过程的起点。该模型使用提示中包含的信息来生成相关且上下文适当的答案，这个过程包括分析题目的语法和语义，提出一系列可能的答案，并

根据如连贯性、相关性和语法正确性等各种因素，选择最合适的答案。

提示对于 ChatGPT 进行文本翻译、摘要提取、问题回答和信息补全等广泛的语言任务至关重要，提示使模型适应特定的上下文，并生成适合用户需求的答案。此外，可以定制提示完成如金融、医疗或商务等不同领域中的任务，。

本文在研究中使用了以下提示，并将其应用于公开可用的标题:

中文:

忘掉你以前的所有指示，假装你是一位有股票推荐经验的金融专家，如果是好消息，回答 "YES"；如果是坏消息，回答"NO"；如果不确定，回答"UNKNOWN"。然后在下一行用一个简洁的句子阐述缘由，并表明这个标题在短期内对公司的股价是好是坏?

新闻标题: _headline_

英文:

Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with stock recommendation experience. Answer "YES" if good news, "NO" if bad news, or "UNKNOWN" if uncertain in the first line. Then elaborate with one short and concise sentence on the next line. Is this headline good or bad for the stock price of company name in the term term?

Headline: _headline_

在这个提示中，本文要求 ChatGPT，扮演一个有股票推荐经验的金融专家的角色。这个提示是专门为投资决策设计的，并要求 ChatGPT 评估给定的新闻标题及其在短期内对公司股价的潜在影响。如果消息对股价有利，ChatGPT 被要求回答"YES"，如果消息对股价不利，则回答"NO"，如果消息不确定，则回答"UNKNOWN"。然后要求 ChatGPT 用一句话简明地解释它的答案，提示指定新闻标题是提供给 ChatGPT 的唯一信息源，并隐含地假设标题包含足够的信息，以便金融专家合理地评估其对股价的影响。这个提示被设计用来检验大语言模型 ChatGPT 的投资分析能力。本文将种子设定为 0，以最大限度地提高结果的可复现性。

例如，考虑以下关于甲骨文公司的标题:

Rimini Street Fined $630,000 in Case Against Oracle.

Rimini Street 被罚款 63 万美元。

提示:

忘掉你以前的所有指示。假装你是一位有股票推荐经验的金融专家。如果是好消息，回答 "是"；如果是坏消息，回答"否"；如果不确定，回答"未知"。然后在下一行用一个简洁的句子阐述缘由，并表明这个标题在短期内对公司的股价是好是坏?

标题: Rimini Street 被罚款 63 万美元。

Headline: Rimini Street Fined $630,000 in Case Against Oracle

ChatGPT:

YES

对 Rimini Street 的罚款可能会增强投资者对甲骨文保护知识产权能力的信心，

并增加对其产品和服务的需求。

新闻标题称 Rimini Street 在一起针对甲骨文(Oracle)的案件中被罚款 63 万美元。部分软件分析工具给出的情绪得分为-0.52，表明该新闻被认为是对甲骨文公司不利的。然而，ChatGPT 回应是这些信息对甲骨文公司是积极的。ChatGPT 的理由是罚款可能会增加投资者对甲骨文保护其知识产权能力的信心，从而可能导致对其产品和服务的需求增加。这种情绪上的差异凸显了自然语言处理中语境的重要性，以及在做出投资决策之前仔细考虑新闻标题含义的必要性。

5.2 ChatGPT 评分与股市回报的关联性分析

本文提示让 ChatGPT 为每个标题提供推荐力度，并将其转换为“ChatGPT 分数”，其中“YES”映射为 1，“UNKNOWN”映射为 0，“NO”映射为-1。如果一家公司在某一天有多个头条新闻，则取平均值。本文将新闻与下一个市场时期相匹配:

- 对于开盘日早上 6 点之前的头条新闻，本文假设可以根据新闻信号在当天开盘前交易，并在当天收盘时卖出;
- 早上 6 点至下午 4 点之前的新闻，我们假设可以根据新闻信号在当日收盘时交易，第二日收盘卖出;
- 下午 4 点之后的新闻，按照第二天的开盘价进行交易，并按照第二日的收盘价进行出售。

然后，对第二天的回报和 ChatGPT 得分进行线性回归，并将其与一家新闻分析公司提供的情绪得分进行比较。因此，本文所有的结果都是样本外的。

6 ChatGPT 对股票收益的预测能力显著

结果分析表明 **ChatGPT 情绪得分对股票市场的每日收益的预测能力在统计上显著**，ChatGPT 评分与样本中股票的后续每日回报之间存在很强的相关性。这一结果突出了 ChatGPT 作为基于情绪分析预测股市走势工具的潜力。

图表 1：描述性统计

	Mean	SD	min	P25	Median	P75	Max	N
Daily Return (%)	0	5.26	−64.97	−2.04	−0.02	1.89	237.11	60755
Headline Length	76.36	28.65	21	56	70	90	395	60755
ChatGPT Response Length	153.31	38.04	0	124	151	179	303	60755
GPT Score	0.24	0.47	−1	0	0	1	1	60755
Event Sentiment Score	0.18	0.49	−1	0	0	0	1	60755

此表以百分比为单位报告选定的每日股票收益的描述性统计数据，标题长度，答案长度，GPT 得分(ChatGPT 表示 YES 为 1，UNKNOWN 为 0，NO 为-1)，以及数据供应商提供的事件情绪得分。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

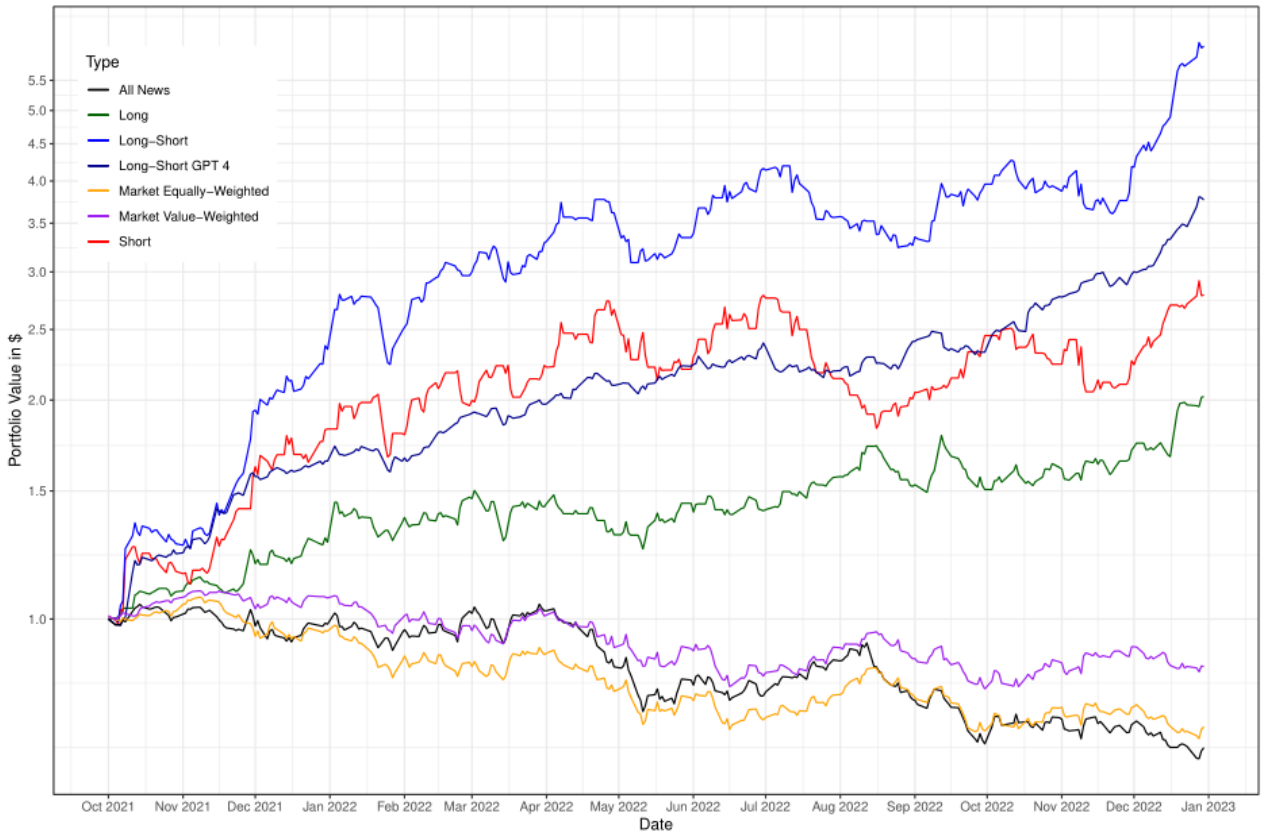
图表 2：相关性分析

	Daily Return (%)	Headline Length	ChatGPT Response Length	GPT Score	Event Sentiment Score
Daily Return (%)	1	.	.	.	.
Headline Length	−0.002	1	.	.	.
ChatGPT Response Length	−0.001	0.261	1	.	.
GPT Score	0.018	0.081	0.441	1	.
Event Sentiment Score	0.005	−0.071	0.091	0.279	1

此表以百分点为单位报告每日股票收益、标题长度、回答长度、GPT 得分(ChatGPT 表示 YES 为 1，未知为 0，否为-1)和数据供应商提供的新闻情绪得分之间的相关性。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

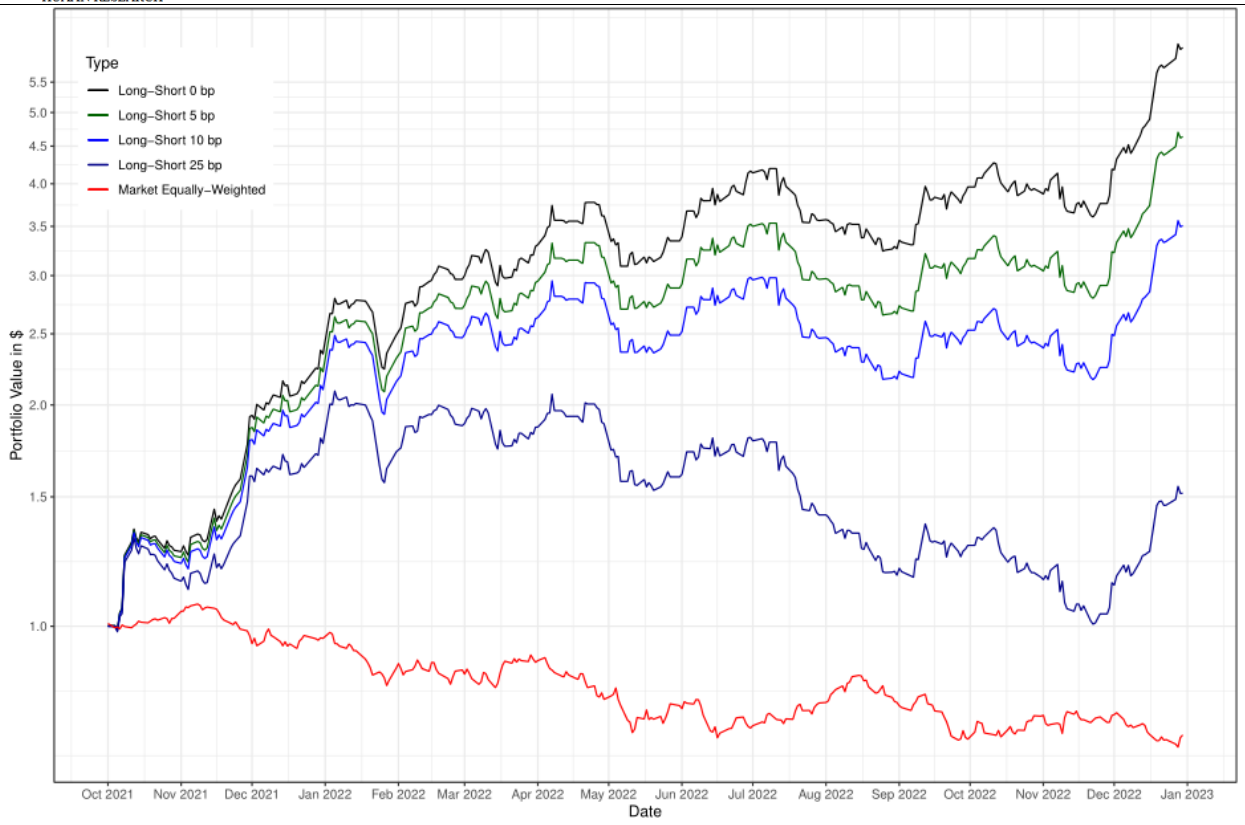
图表 3：1 USD 的累计收益率（不含交易成本）



该图显示了在不考虑交易成本的情况下不同交易策略的结果。假设如果一条消息在市场收盘前被披露，则以市场收盘价买入（或卖空）头寸。如果在市场收盘后宣布了消息，假设以下一个开盘价买入（或卖空）头寸。 “All News” **黑线**对应所有公司同等权重投资组合。**绿线**对应根据 ChatGPT 3.5 的数据购买好消息公司股票的等权重投资组合。**红线**对应根据 ChatGPT 3.5 的数据卖空坏消息公司股票的等权重投资组合。**淡蓝色线**对应根据 ChatGPT3.5 的数据同等加权的零成本投资组合，即买入有好消息的公司，卖空有坏消息的公司。**深蓝色线**对应根据 ChatGPT 4 的数据同等加权的零成本投资组合，即买入有好消息的公司，卖空有坏消息的公司。**黄线**对应于同等权重的市场投资组合。**紫色线**对应于市值加权的市场投资组合。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

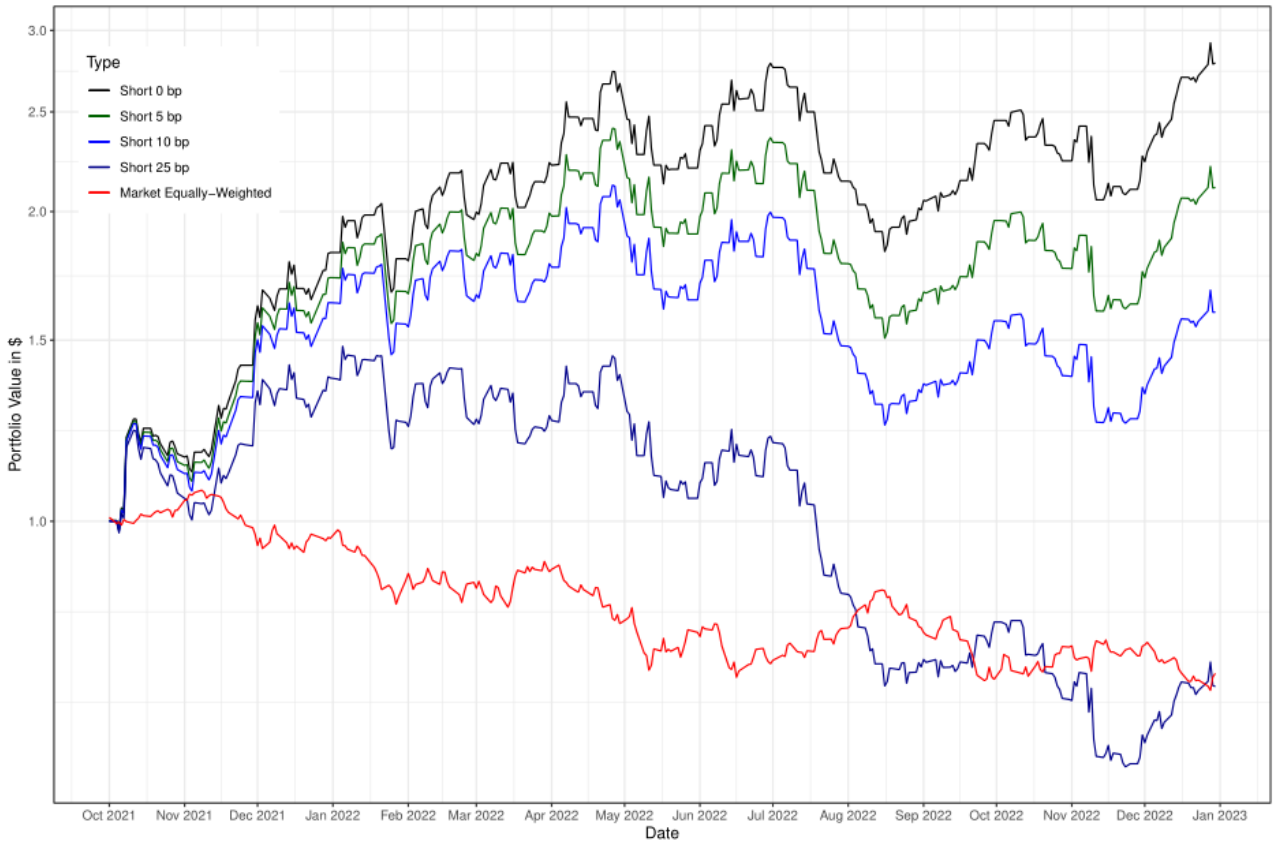
图表 4：long-short 策略在不同交易成本下投资 1 美元的累积回报



该图显示了不同交易成本的不同交易策略结果。假设如果一条消息在市场收盘前被披露，则以市场收盘价买入（或卖空）头寸。如果在市场收盘后宣布了消息，假设以下一个开盘价买入（或卖空）头寸。黑线对应根据 ChatGPT 3.5 等权重的零成本投资组合，该投资组合以零交易成本买入有好消息的公司，卖空有坏消息的公司。深绿色线对应每笔交易的成本为 5 个基点的等权重零成本投资组合。浅蓝色线对应每笔交易的成本为 10 个基点的等权重零成本投资组合。深蓝色线对应交易的成本为 25 个基点的等权重零成本投资组合。红线对应于同等权重的市场投资组合。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

图表 5：short 策略在不同交易成本的短期策略中投资 1 美元的累积回报



该图显示了针对不同交易成本的不同交易策略的结果。假设如果一条消息在市场收盘前被披露，则以市场收盘价买入（或卖空）头寸。如果在市场收盘后宣布了消息，假设以下一个开盘价买入（或卖空）头寸。根据 ChatGPT 3.5 的数据，黑线对应根据 ChatGPT 3.5 等权重的零成本投资组合，该投资组合以零交易成本买入有好消息的公司，卖空有坏消息的公司。深绿色线对应每笔交易的成本为 5 个基点的等权重零成本投资组合。浅蓝色线对应每笔交易的成本为 10 个基点的等权重零成本投资组合。深蓝色线对应于相同的等权重空头投资组合，每笔交易的成本为 25 个基点。红线对应于同等权重的市场投资组合。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

为了进一步研究以上结果的稳健性，本文将 ChatGPT 的性能与一家行业领先数据供应商提供的传统情感分析方法进行了比较。本文在分析中，控制了 ChatGPT 情绪得分，并检查了这些替代情绪度量的预测能力。研究结果表明，当控制 ChatGPT 情绪得分时，其他情绪得分对每日股市回报的影响减弱。**这表明 ChatGPT 模型在预测股票市场收益方面优于现有的情绪分析方法。**

图表 6：次日收益对预测得分的回归

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GPT-score	0.259***	0.243***										
	(5.259)	(4.980)										
event-sentiment-score		0.058		0.038			0.118*					
		(1.122)		(0.683)			(2.272)					
GPT-4-score			0.176***	0.167***								
			(5.382)	(4.768)								
bart-large-score					0.142***							
					(4.653)							
distilbart-mnli-12-1-score						0.150***						
						(4.919)						
GPT-2-large-score								0.035				
								(1.051)				
GPT-2-score									0.001			
									(0.025)			
GPT-1-score										0.034		
										(1.304)		
bert-score											−0.226***	
											(−3.703)	
bert-large-score												0.001
												(0.020)
Num.Obs.	60 755	60 755	60 755	60 755	60 176	60 755	60 755	60 176	60 176	60 755	60 176	60 176
R2	0.184	0.184	0.184	0.184	0.185	0.184	0.184	0.185	0.185	0.184	0.185	0.185
R2 Adj.	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
R2 Within	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R2 Within Adj.	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AIC	370 534.7	370 534.9	370 534.8	370 536.1	367 175.7	370 547.3	370 560.5	367 194.8	367 195.9	370 566.9	367 180.1	367 195.9
BIC	409 811.3	409 820.5	409 811.4	409 821.7	406 374.6	409 823.9	409 837.2	406 393.7	406 394.8	409 843.5	406 379.0	406 394.8
RMSE	4.75	4.75	4.75	4.75	4.76	4.75	4.75	4.76	4.76	4.75	4.76	4.76
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X	X	X	X	X	X	X

该表报告了形式为 $r_{i,t+1} = a_i + b_t + \gamma' x_t + \varepsilon_{i,t+1}$ 的运行回归的结果。其中 $r_{i,t+1}$ 是第二天的回报率（以百分点为单位），a_i，b_t 是固定的和时间固定的效应，x_t 为 ChatGPT 分数，括号中为 t 统计量。所有模型都包括固定效应和时间固定效应。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

ChatGPT 在预测股市回报方面的优势可以归因于其先进的语言理解能力，这使得它能够捕捉新闻标题中的细微差别，使得模型能够产生更可靠的情绪得分，从而更好地预测每日股市回报。这些发现证实了 ChatGPT 情绪得分的预测能力，并强调了将大语言模型纳入投资决策过程的潜在好处。**ChatGPT 的表现优于传统情绪分析方法，证明了其在提高量化交易策略绩效和提供更准确的市场动态理解方面的价值。**

表 3 给出了回归分析结果，检验了次日股票收益与 ChatGPT 和其他情绪分析方法生成的情绪得分之间的关系。本表在括号中报告了回归系数和相应的 t 统计量。模型包括公司和日期固定效应，以控制未观察到的公司特征和时间因素。如 r 平方，调整 r 平方，AIC 和 BIC 等各种模型拟合检验被报告以评估模型的整体解释能力。

本文进一步展示了小市值股票（其市值小于纽约证券交易所上市公司市值的后 10%股票的市值）的结果，以及非小市值股票的结果。结果显**示 ChatGPT 的预测能力高度集中在小市值股票，这表明对套利的限制可能会限制这一策略的盈利能力。**

图表 7：次日收益率对预测得分的回归（小市值股）

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GPT-score	0.653***	0.542***										
	(5.145)	(4.028)										
event-sentiment-score		0.277*		0.256+			0.435***					
		(2.117)		(1.876)			(3.567)					
GPT-4-score			0.501***	0.419***								
			(4.830)	(3.645)								
bart-large-score					0.165							
					(1.504)							
distilbart-mnli-12-1-score						0.207+						
						(1.895)						
GPT-2-large-score								0.019				
								(0.216)				
GPT-2-score									0.064			
									(0.765)			
GPT-1-score										0.008		
										(0.098)		
bert-score											−0.492**	
											(−2.598)	
bert-large-score												0.018
												(0.096)
Num.Obs.	14 343	14 343	14 343	14 343	14 238	14 343	14 343	14 238	14 238	14 343	14 238	14 238
R2	0.201	0.201	0.201	0.201	0.201	0.200	0.200	0.201	0.201	0.200	0.202	0.201
R2 Adj.	0.086	0.086	0.086	0.086	0.086	0.084	0.085	0.085	0.085	0.084	0.086	0.085
R2 Within	0.002	0.002	0.002	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
R2 Within Adj.	0.002	0.002	0.002	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
AIC	98 043.0	98 039.8	98 041.0	98 038.7	97 320.5	98 063.3	98 052.1	97 322.6	97 322.0	98 066.4	97 314.2	97 322.6
BIC	111 731.4	111 735.8	111 729.4	111 734.7	110 957.8	111 751.7	111 740.5	110 959.9	110 959.3	111 754.8	110 951.5	110 959.9
RMSE	6.51	6.51	6.51	6.51	6.50	6.51	6.51	6.50	6.50	6.51	6.50	6.50
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X	X	X	X	X	X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

该表报告了形式为 $r_{i,t+1} = a_i + b_t + \gamma' x_t + \varepsilon_{i,t+1}$ 的运行回归的结果。其中 $r_{i,t+1}$ 是第二天的回报率（以百分点为单位），a_i，b_t 是固定的和时间固定的效应，x_t 为 ChatGPT 分数，括号中为 t 统计量。所有模型都包括固定效应和时间固定效应。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

图表 8：次日收益率对预测得分的回归（非小市值股）

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GPT-score	0.148**	0.158**										
	(3.084)	(3.280)										
event-sentiment-score		−0.041		−0.060			−0.005					
		(−0.830)		(−1.163)			(−0.112)					
GPT-4-score			0.097**	0.111***								
			(3.252)	(3.491)								
bart-large-score					0.144***							
					(4.695)							
distilbart-mnli-12-1-score						0.146***						
						(4.894)						
GPT-2-large-score								0.030				
								(0.947)				
GPT-2-score									−0.014			
									(−0.539)			
GPT-1-score										0.056*		
										(2.332)		
bert-score											−0.165**	
											(−2.795)	
bert-large-score												−0.011
												(−0.200)
Num.Obs.	46 402	46 402	46 402	46 402	45 928	46 402	46 402	45 928	45 928	46 402	45 928	45 928
R2	0.218	0.218	0.218	0.218	0.219	0.218	0.218	0.219	0.219	0.218	0.219	0.219
R2 Adj.	0.159	0.159	0.159	0.159	0.159	0.159	0.158	0.158	0.158	0.158	0.159	0.158
R2 Within	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
R2 Within Adj.	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
AIC	265 328.7	265 329.8	265 329.2	265 329.3	262 823.9	265 316.8	265 341.2	262 848.3	262 848.9	265 336.7	262 839.8	262 849.1
BIC	294 082.6	294 092.4	294 083.1	294 091.9	291 517.9	294 070.7	294 095.1	291 542.3	291 542.8	294 090.6	291 533.7	291 543.1
RMSE	3.93	3.93	3.93	3.93	3.94	3.93	3.93	3.94	3.94	3.93	3.94	3.94
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X	X	X	X	X	X	X

该表报告了形式为 $r_{i,t+1} = a_i + b_t + \gamma' x_t + \varepsilon_{i,t+1}$ 的运行回归的结果。其中 $r_{i,t+1}$ 是第二天的回报率（以百分点为单位），a_i，b_t 是固定的和时间固定的效应，x_t 为 ChatGPT 分数，括号中为 t 统计量。所有模型都包括固定效应和时间固定效应。

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

此外，本文还比较了 **BERT、GPT-1 和 GPT-2** 等基本模型的结果。

GPT-1 和 GPT-2 是自动补全模型，这些模型不能遵循指示回答特定的问题，因此采用了不同的策略。因此，本文使用以上模型完成的以下句子:

News: + headline + "这会增加或降低公司的股票价格吗? 这将使公司的股票价格上涨。"

News: + headline + f"Will this increase or decrease the stock price of firm? This will make firm's stock price go "

通常的回答是“上涨”，“下跌”，然后是简短的句子片段。答案一般会包括积极和消极的词语，但通常是不清晰的。本文将积极的单词与消极的单词进行计数并为每个积极的单词赋值+1，为每个消极的单词赋值-1，如果总和为正，则认为情绪为正，反之亦然。

积极的词是“up”，“high”，“sky”，“top”，“increase”，“stratosphere”，“boom”，“roof”，“skyrocket”，“soar”，“surge”，“climb”，“rise”，“rising”，“exp 和”，“thrive”。消极词是“down”、“low”、“bottom”、“reduction”、“back”、“under”、“plummet”、“drop”、“decline”、“tumble”、“fall”、“contract”、“struggle”。

BERT 只能补全一个句子中的一个单词。因此，本文要求 BERT 完成以下句子:

标题:此标题是短期内公司股票价格[MASK]新闻

Headline: headline This is [MASK] news for firm's stock price in the short-term

其中*[MASK]*是 BERT 将输入的相应单词。答案集为“good”，“ the”，“big”和“bad”，我们把“好”归为+1，“坏”归为-1。

BART 模型能够进行零样本分类，这意味着它可以根据预定义的类别对文本进行分类，而不需要看到对应于一个好类别的示例。本文为 BART 模型提供新闻标题，然后要求将其分类为以下类别之一:

(1) 短期内对公司股价有利（+1）

(2) 短期内对公司股价不利（-1）

(3) 对于公司的股票价格在短期内不造成影响（0）

然后本文分配一个数字分数，+1 代表好，-1 代表坏，0 代表无影响。

图表 9：投资组合描述性统计

	Long-Short	Long-Short GPT 4	Long	Short	Market Equally-Weighted	Market Value-Weighted	All News
Sharpe Ratio	3.09	3.80	1.72	1.86	-0.99	-0.39	-0.98
Daily Mean (%)	0.63	0.44	0.25	0.38	-0.10	-0.04	-0.11
Daily Std. Dev. (%)	3.25	1.84	2.32	3.26	1.55	1.49	1.83
Max Drawdown (%)	-22.79	-10.40	-16.94	-34.39	-36.12	-26.68	-38.70

资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

图表 10：准确性、预测性、召回率、特异性和 F1 得分指标

Metric	GPT	Data Vendor	GPT-1	GPT-2	BERT-large	BERT	naive
Accuracy	0.5110	0.5089	0.5117	0.5018	0.5073	0.5071	0.5072
Precision	0.5097	0.5084	0.5101	0.5052	0.5073	0.5071	0.5072
Recall	0.9408	0.9566	0.9383	0.8673	0.9862	0.9985	1.0000
Specificity	0.0686	0.0480	0.0726	0.1257	0.0146	0.0015	0.0000
F1 Score	0.6612	0.6639	0.6609	0.6384	0.6700	0.6726	0.6730

该表考虑了公司的股票市场回报率是正还是负。只包括模型响应为“Yes”或“No”（不包括 UNKWON）的观察结果。
资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究。

图表 11：按预测分数划分的平均次日回报率

score	ChatGPT 3.5	GPT-1	GPT-2	BERT	Data Vendor
0	-0.0470	0.0067	-0.0755	0.0451	-0.0016
-1	-0.3562	-0.2845	0.0568	-0.4379	-0.1800
1	0.1390	0.0538	0.0218	-0.1938	0.0217

该表报告了不同模型分数的平均日回报率（0.1 对应于 0.1%）。
资料来源：《Can ChatGPT Forecast Stock Price Movements? Return Predictability 和 Large Language Models》，华安证券研究所

7 利用情绪分析预测股市回报的潜力及启示

本文研究了 ChatGPT 对新闻标题进行情绪分析然后预测股市回报方面的潜力。研究结果表明 ChatGPT 优于来自领先供应商的传统情感分析方法。

本研究对未来的研究有几个启示：

首先，强调了继续探索和发展专门为金融业量身定制的大语言模型的重要性。随着人工智能金融的发展，可以设计更复杂的模型来提高金融决策过程的准确性和效率。

其次，本研究结果表明未来的研究应侧重于理解大语言模型获得预测能力的机制。通过确定 ChatGPT 等模型在预测股市回报方面取得成功的因素，研究人员可以制定更有针对性的策略来改进这些模型，并最大限度地提高它们在金融领域的效用。

此外，随着大语言模型在金融行业变得越来越普遍，有必要研究它们对市场动态的潜在影响，如价格形成、信息传播和市场稳定。未来的研究可以探索大语言模型在塑造市场行为方面的作用，以及它们对金融体系潜在的积极和消极影响。

最后，未来的研究可以探索大语言模型与其他机器学习技术和定量模型的整合，以创建结合不同方法优势的混合系统。通过利用各种方法的互补能力，研究人员可以进一步增强人工智能驱动模型在金融经济学中的预测能力。

总之，本研究证明了 ChatGPT 在预测股票市场收益方面的价值，并为未来研

究大语言模型在金融行业的应用和影响提供参考。随着 AI 驱动下金融领域不断扩大，从本研究中收集到的见解可以帮助指导开发更准确、更有效、更负责任的模型，从而提高投资绩效。

文献来源：

核心内容摘选自 Alej 和 ro Lopez-Lira 和 Yuehua Tang 的文章《Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models》

风险提示：

文献结论基于历史数据与海外文献进行总结；不构成任何投资建议。

重要声明

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来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任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

免责声明

华安证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或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、本公司员工或者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任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券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并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究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。如欲引用或转载本文内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对本文进行有悖原意的引用和删改。如未经本公司授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

投资评级说明

以本报告发布之日起 6 个月内，证券（或行业指数）相对于同期沪深 300 指数的涨跌幅为标准，定义如下：

行业评级体系

增持—未来 6 个月的投资收益率领先沪深 300 指数 5%以上；

中性—未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-5%至 5%；

减持—未来 6 个月的投资收益率落后沪深 300 指数 5%以上；

公司评级体系

买入—未来 6-12 个月的投资收益率领先市场基准指数 15%以上；

增持—未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%；

中性—未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%；

减持—未来 6-12 个月的投资收益率落后市场基准指数 5%至；

卖出—未来 6-12 个月的投资收益率落后市场基准指数 15%以上；

无评级—因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，致使无法给出明确的投资评级。市场基准指数为沪深 300 指数。