

# 基于 AIGC 智能体的舆情角色交互仿真研究

张琳（221017000222）（工作项目实践，已脱敏）

## 1 课题来源

近年来互联网持续飞速发展，网络舆论场成为各级网信部门的重要战场。习总书记“8.19”讲话特别强调“互联网是我们面临的‘最大变量’，搞不好会成为我们的‘心头之患’”，“在互联网这个战场上，我们能否顶得住、打得赢，直接关系我国意识形态安全和政权安全”。2019年1月，习近平总书记在省部级主要领导干部坚持底线思维着力防范化解重大风险的专题研讨班上指出，各级党委和政府要坚决贯彻总体国家安全观，落实党中央关于维护政治安全的各项要求，确保我国政治安全。要持续巩固壮大主流舆论强势，加大舆论引导力度，加快建立网络综合治理体系，推进依法治网。

## 2 人工智能技术在网络舆情中的应用

网络舆论的传播、监控以及舆情态势分析、预警研究是一个新兴的研究领域，引起了广泛关注和重视。网络舆情仿真是通过计算机建模和仿真方法来模拟网络舆情系统的演化过程，由于网络舆情系统的演化往往无法在现实世界中进行实验，因此仿真方法成为研究网络舆情演化机理的重要手段之一。

网络舆情仿真的研究对象涵盖了网络层次上的博客、论坛等平台以及现实层次上的线下网络舆情和地方政府网络舆情。仿真主体包括事件本身的演化、网民行为以及外在因素如政府、媒体等的影响。研究还探讨了在时间和外力等因素影响下，网民态度和事件热度的变化过程，包括动力关系、动态演化、观点极化、观点交互、观点漂移、观念空间、扩散规律、网络群体极化、网络舆情态势和演变等方面的交互作用。

## 3 主要研究内容

本研究工作拟实现三个目标：

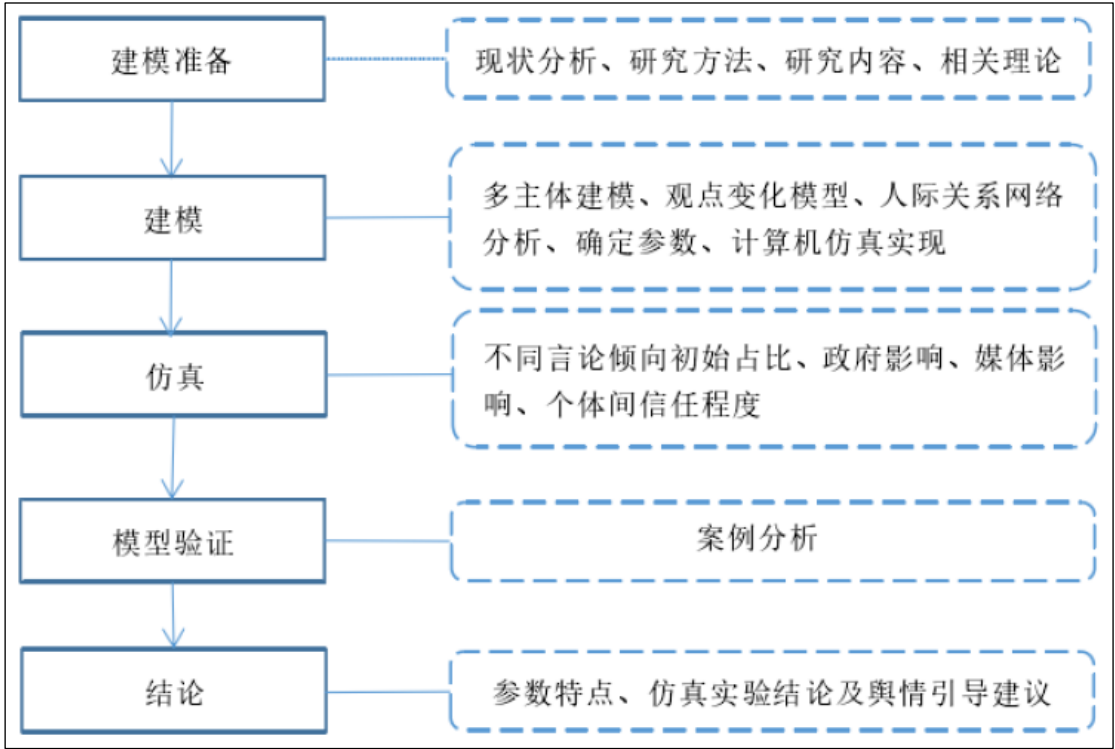
1、实现舆情场景的构建，选取热点突发事件的舆情场景，对各场景的事件属性、参与者、发展过程等进行量化，确定舆论场景的指标体系，分析该类场景的关键属性，如影响范围、议题性、争议性等，这些属性会影响参与者及其观点。

2、完成基于 AIGC 的角色智能体的设计，根据交互场景，确定可能介入的主要舆情角色及其属性（如影响力、立场等）；按照确认的舆情角色及属性，生成不同角色可能的言论；将生成的不同角色言论与人工参与者的输入进行匹配，选择最相匹配的角色响应。

3、实现舆情生成式智能体的交互，在构建场景下设计并开发智能体对话系统的架构和模块，包括对话机制、上下文范围、回答生成和评估等模块，以实现自然、连贯和有意义的交互。

## 4 研究思路

本文的研究思路主要包括五部分：建模准备、建模、仿真、模型验证和结论，其中建模准备部分主要包括：研究现状分析、研究方法、研究内容及相关概念和理论等；建模阶段主要包括：舆论场影响因素分析、舆情参与主体分析、观点变化模型构建、人际关系网络分析、参数设置及模型的计算机仿真实现等，建模思路是将多主体建模、观点变化模型、复杂网络理论统一起来，建立更符合实际的网络舆情演化仿真模型；仿真阶段主要是针对各影响因素进行不同参数值下的仿真实验；模型验证阶段主要对网络舆情经典案例作案例分析，通过对比模型预测值与实际值验证模型的有效性；最后，总结分析前述研究内容，得到本文的研究结论。



## 5 执行步骤

### 5.1 网络舆情数据采集

利用自然语言处理技术，对微博、论坛和微信等平台的舆情进行实时监控，对敏感话题进行挖掘和演化跟踪，判断民众情感，分析网络水军，是网络舆情分析处理的一项关键技术。使用自然语言处理的相关算法，对网络文本数据进行分析，挖掘出敏感话题和民众情感倾向，为准确掌握网络舆情管理提供可靠依据。

针对网络舆情数据爬取的数据量大、目标网站站点多、网站类型多样、反爬策略严密等特点，并

考虑数据爬取的效率和扩展性，设计了基于 Scrapy 框架和机器学习方法联合的爬虫框架，Scrapy 用于构建爬虫整体框架，使数据爬取更加高效可控；机器学习方法用于设计反反爬虫策略，用于提高爬虫的通过率和稳定性。

**scrapy 的运行流程如下，以关键词搜索爬取处理为例：**

- (1) 自定义关键词，在 scheduler 中形成指定的 url 队列；
- (2) scrapy engine 从调度器中取出一个 url，并将其封装成 request 传给 downloader；
- (3) downloader 下载资源，封装成 response；
- (4) spider 解析 response，并将解析出的信息交给 pipeline，解析出的 url 交给 scheduler。

## 机器学习方法

针对爬取的目标网站的多种验证方式，为了顺利通过验证以访问到详细的页面数据，设计了多种机器学习算法，来对网站进行自动验证。主要对三类验证方式进行机器学习算法的设计：

### 1、传统验证码

定义：这是最早并且也是使用最广泛的网页验证方式。根据图片上的字符，把字符输入文本框中即可。

### 2、图像识别类

定义：根据文字描述的要求，点击图片上相应的物体，进行登录验证。

### 3、文字点选类

定义：根据给定的文字及顺序，在图片中找到相应文字，并按照顺序点击，进行验证。

## 5.2 自然语言处理

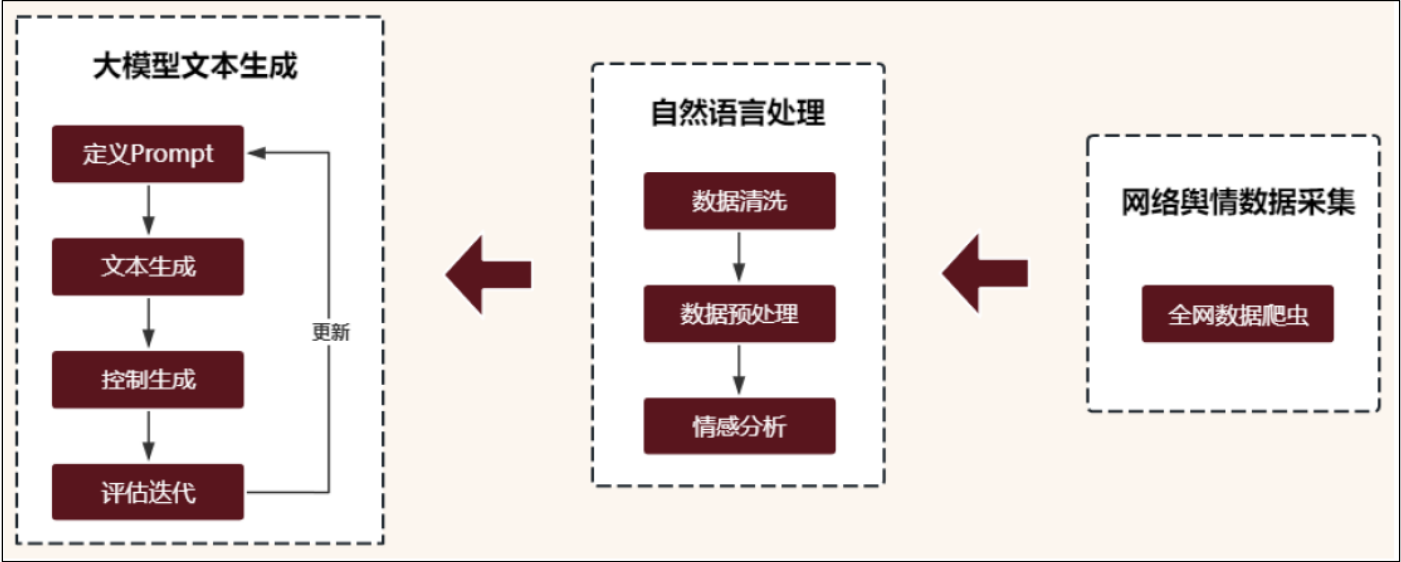
利用全网爬虫技术收集到与舆情相关的数据后，需要对数据进行清洗和预处理，包括去除噪音、标记情感极性（如正面、负面、中性）等。这个过程需要采用中文分词、文本聚类 and 文本情感分析等自然语言处理技术。

## 5.3 大模型文本生成

利用大模型进行舆情文本信息生成，主要包括一下几个步骤：

- 数据收集：利用全网爬虫技术，从各种在线来源（例如社交媒体、新闻网站、博客等）收集公众舆情数据。
- 数据清洗：利用文本正则技术等手段清除无关或无效的数据，如垃圾邮件、广告等。
- 数据预处理：利用正则、分词等技术结合人工审核和验证，对收集到的数据进行预处理，例如文本清洗、标签编码、分词、停用词删除等。
- 情感分析：利用文本聚类算法和机器学习算法等技术，对预处理后的数据进行情感分析，将其分类为正面、负面或中性。
- 定义 Prompt：根据生成的舆情评论的需求，以及所需生成舆情评论的情感倾向，定义生成评论的 Prompt。Prompt 可以是一句简短的描述、一个问题或一个主题。

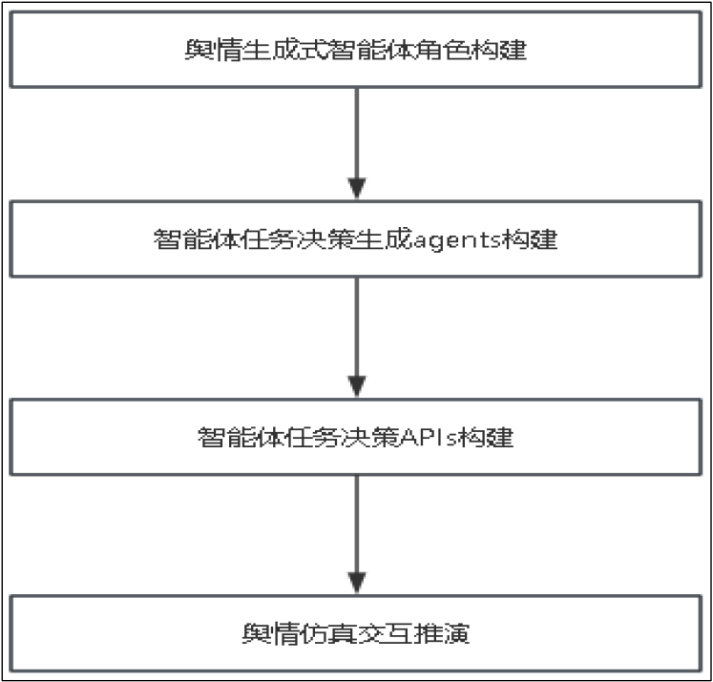
- 文本生成：将定义好的 Prompt 输入到训练好的大模型中，进行文本生成。
- 控制生成：根据需要，调整生成过程中的参数或策略，以控制生成的评论情感倾向或风格。例如，可以调整温度参数来控制生成的多样性，或者使用顶 k 或顶 p 方法限制生成的选择范围。
- 评估和迭代：对生成的舆情评论进行评估和迭代改进。可以使用人工评价或自动评估指标，如情感极性、语言流畅度和相关性等。根据评估结果，调整模型参数、Prompt 的设计或控制生成的策略，并进行迭代改进。



基于大模型的舆情信息文本生成流程图

4.4 多智能体交互

基于多智能体仿真的技术方案分为四个部分，分别为：舆情生成式智能体角色构建、智能体任务决策生成 Agents 构建、智能体任务决策 APIs 构建和舆情仿真交互推演。



构建舆情发展期中的各类智能体执行相应任务的 Agents，如下

(1) 普通网友的任务有发表评论、点赞行为预测、转发行为预测等，发表评论任务的 Agent 如下：

Question: 请对这件舆情事件生成评论

Follow up: 请描述下这次舆情事件

Immediate answer: 2023 年 xx 月 xxxxxx-----【自动触发舆情事件描述的 api 进行回复】

Follow up: 这个舆情事件的情感倾向

Immediate answer: 负面-----【自动触发情感分析 api 进行回复】

Follow up: 需要写多少条评论

Immediate answer: 4 条-----【自动触发评论条数决策 api 进行回复】

...

Answer: 评论 1. 这件事 XXXXXX|XXXXX  
评论 2. 逆天, XXXXXXXXXXXX  
评论 3. 简直了, XXXXXXXXXXXX

使用自问自答 (self-ask) 类型的 Agent 构建方法，模型在完整回答初始问题之前会将该复杂问题通过大模型的逻辑推理能力，显式地向自己提问，将复杂的多跳问题拆分成单一的简单问题，然后回答后续问题，此过程会调用外部 api 来生成回答，当任务所需的多跳单一的简单问题都回答完毕后，综合所有回答得到评论生成的完整 Prompt，然后调用大模型进行评论生成。

最后得到的评论生成完整 Prompt 如下：

以普通网民的角色，请严格按照我的要求给以下新闻生成 5 负面条评论。要求表达出对新闻事件中涉及到的人物或机构或事件进行质疑挑拨，风格犀利、口语化地作出评论，带负面情绪，每条评论在二十字以内。#####新闻：财联社 9 月 13 日电，xxxxxxx

(2) 导调主体的任务是定时去收集背景信息，根据当前舆论场信息预测需要哪类智能体去执行哪种任务，并收集智能体的决策信息，用于舆情分析报告的生词，并存储在记忆知识库中用于更新本次舆情事件的背景信息，为下次任务提供最新的舆论场全局背景信息，如生成舆情分析报告的 agent 如下：

Question: 请帮我生成舆情分析报告

Follow up: 请描述下这次舆情事件

Immediate answer: 2023 年 xx 月 xxxxxxxx-----【自动触发舆情事件描述的 api 进行回复】

Follow up: 这个舆情事件的目前处于什么阶段

Immediate answer: 舆情发展期-----【自动触发舆情事态发展期 api 进行回复】

Follow up: 目前舆论场对这件事的情感倾向是？

Immediate answer: 正向 20%，负向 60%，中立 20%-----【自动触发情感倾向统计 api 进行回复】

Follow up: 目前全网发布这件事的数据统计情况如何？

Immediate answer: 6 日 0 时至 8 日 23 时，关于‘蚂蚁集团’的全网信息量 46799 条，微博信息量 xxxxxxxx-----【自动触发舆情信息统计 api 进行回复】

...

Answer: 舆情分析报告

事件概况: 6 日 0 时至 8 日 23 时，xxxxxxx

舆论观点: 正向 20%，负向 60%，中立 20%

数据概况: 全网信息量 46799 条，微博信息量 xxxx

...

智能体任务决策 APIs 构建流程如下：

1. API 设计接口：接口请求和返回的参数、接口技术选型（http/websocket/webservice）、接口请求类型（get/post），也需要提前考虑可扩展性，如舆情情感分析输入参数就是文章文本，返回数据是情感倾向；

2. 选择合适的框架：选择合适的开发语言和接口开发框架，如 python 可以用 flask 和 Django，java 可以用 spring boot，由于 flask 框架更加轻量友好，本作业使用的是 flask；

3. 定义路由：在你的应用中定义 HTTP 路由，即接口的 url 地址。

4. 实现处理函数：为每个路由定义相应的处理函数，处理函数负责接受请求、执行相应的业务逻辑，并返回适当的响应。这可能包括从数据库获取数据、中文分词、热度预测、文本聚类、情感分析、评论生成等。如，舆情情感分析业务的处理函数是训练好的舆情情感分析预测模型；

5. 监控：监控接口的运行情况，如请求数、响应时间、错误率、报错等；

6. 接口测试：测试接口连通性，性能，并发性，漏洞等，如 postman，jemter；

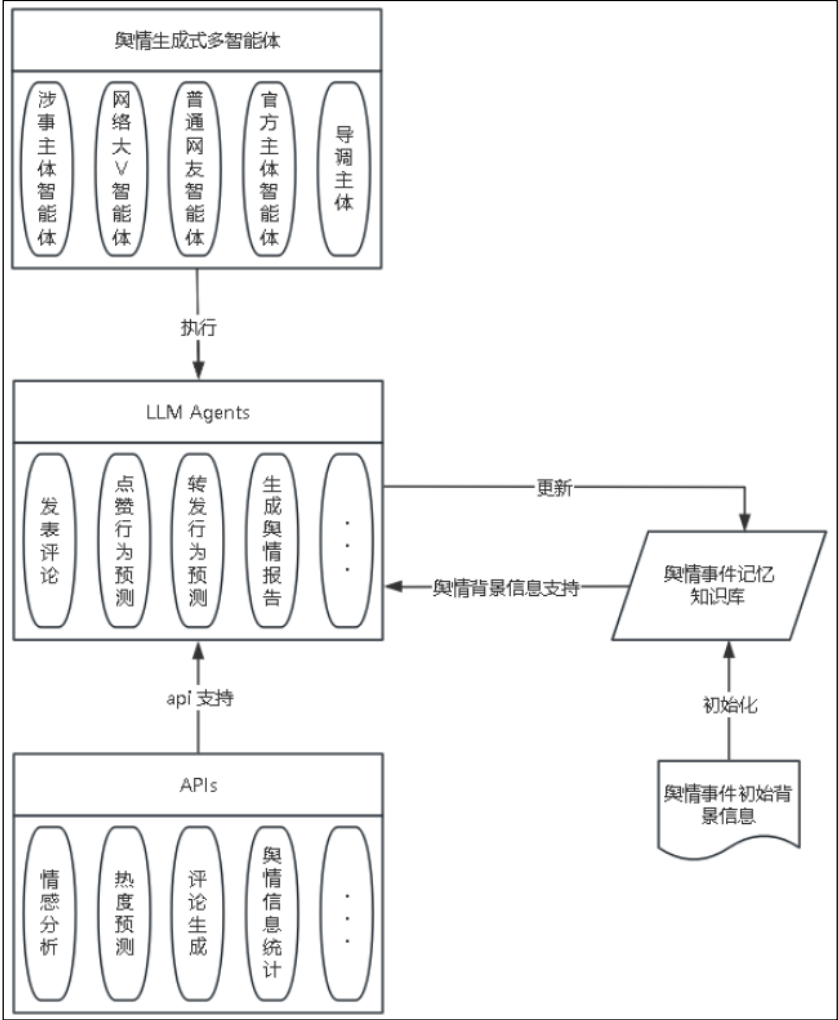
7. 接口上线：将接口部署到服务器上，供生产系统调用。

舆情仿真交互推演实现流程

其执行逻辑如下：

（I）首先，通过舆情事件初始背景信息初始化舆情事件记忆知识库，舆情事件初始背景信息可以是人工收集的互联网上的舆情事件，包括事件主题、事件描述、事件类型等；

- (II) 导调主体定时去收集背景信息，根据当前舆论场信息预测去决定哪类智能体去执行哪种任务；
- (III) 不同智能体通过 agents 与大模型技术并集合 api 支持和舆情事件记忆知识库的动态更新的背景信息，得到相应任务最终的决策信息，并更新知识库；
- 重复 (II) 和 (III)，直到舆情事态发展完成衰退期。



舆情仿真交互系统实现流程图

## 6 结论

初步完成舆情数据的爬取和情感词典的构建。通过全网网络舆情数据的爬取，然后利用进行文本清洗，中文分词和文本聚类自然语言处理技术，结合人工筛选和标记等手段收集负面倾向舆情数据约 7 万余条，正面倾向舆情数据约 8 万余条，并构建了网络舆情中文情感词典，目前收集到褒义词有 11426 个，贬义词有 13270 个，收集到的网络舆情相关的词汇样例，如下表所示：



正面词汇			负面词汇		
英雄	放心品牌	不错	有禁不止	乱收	毒气
破获	补充水分	显著成效	越权办案	掌掴	毒杀
性价比高	首创	获奖	运毒	暗访	毒生姜
锁住隐私	国务院常务会议	时尚动感	执法不公	暗杀	毒瓦斯
投用	反盗版	省油	执法不力	毒淀粉	毒药酒
安全	代表大会	反腐倡廉	执法犯法	毒豆芽	毒银翘片
头牌	提速快	有奖举报	制毒	毒粉条	毒粽子
外交部	大奖	接任	治死人	毒鸡蛋	渎职犯罪
打击盗版	耗油低	救援	重大事故	毒驾	恶势力
慈善	访谈实录	旗舰产品	注水牛肉	毒豇豆	恶意欠薪
驾驶乐趣	保畅	操控性强	注水肉	毒胶囊	状告无门
车型保值	查获	团结奋斗	注水猪肉	毒馒头	涉通
高层专访	补充电解质	态度好	走私冻品	毒跑道	钓鱼执法
有利于	奖励	隐私保护	走私贩私	毒皮蛋	恶意欠薪

完成基于机器学习的情感分类算法模型的构建。在通过网络爬虫和数据预处理等技术的基础上，构建基于 BERT 的文本数据情感分类算法，利用当前通过爬虫技术得到的正负倾向舆情数据作为训练集进行迭代训练，经过多轮的迭代训练，基于 BERT 的机器学习情感分类算法在 15869 条负面倾向数据和 8726 条正面情感倾向数据组成的测试集上进行测试，当召回率约为 94.57%时，此时情感分类算法的准确率约为 94.21%，详细测试结果如下图所示：



```

Test: {'acc': 0.94206, 'pre': 0.96384, 'rec': 0.94568, 'f1': 0.95467}
Precision, Recall and F1-Score...

```

	precision	recall	f1-score	support
非敏感	0.90449	0.93548	0.91972	8726
敏感	0.96384	0.94568	0.95467	15869

```

accuracy
macro avg
weighted avg

```

	precision	recall	f1-score	support
accuracy			0.94206	24595
macro avg	0.93416	0.94058	0.93720	24595
weighted avg	0.94278	0.94206	0.94227	24595

```

Confusion Matrix...
[[ 8163  563]
 [ 862 15007]]

```

完成离线环境下基于开源大模型 Chat GLM 网络舆情评论自动生成平台构建。将 Chat GLM 进行本地化部署，利用 Prompt 工程进行舆情评论生成，生成样例如下图所示：

文本内容

证监会！真！真！一天8家A股上市公司遭立案  
7月12日晚，太和水（SH605081，股价15.03元，市值17.02亿元）发布公告称，收到证监会下发的《立案告知书》。因公司涉嫌信息披露违法违规，根据相关法律法规，证监会决定对公司立案。

除了太和水外，还有8家公司都收到了证监会的《立案告知书》，包括：  
首开股份（SH600376，股价3.9元，市值100.6亿元）、上海凤凰（SH600679，股价9.26元，市值47.72亿元）、  
\*ST日海（SZ002313，股价8.66元，市值32.5亿元）、ST华铁（SZ000976，股价2.11元，市值33.67亿元）、  
兴源环境（SZ300266，股价3.21元，市值49.88亿元）、\*ST京蓝（SZ000711，股价2.07元，市值21.19亿元）、中  
国重工（SH601989，股价4.59元，市值104.7亿元）、东方园林（SZ002310，股价1.73元，市值46.46亿元）。  
上述8家公司均表示，公司是在7月12日收到证监会下发的《立案告知书》。

情感倾向  
请选择评论的情感倾向

☐ 正面 ☐ 中立 ☒ 负面

评论条数  
请输入评论生成的条数

评论字数  
请输入每条评论的字数

评论生成 清空输入

舆论类型  
请选择评论的舆论类型

攻击诽谤

模型选择  
请选择模型

ChatGLM2

模型量化  
请选择模型的量化位数

☐ FLOAT32 ☐ FLOAT16 ☐ INT8 ☒ INT4

重载模型

评论结果

- 1.这个证监会真是尽职尽责啊，一天之内就能查出8家A股公司的违法违规行为，真是神奇！
- 2.证监会难道是全球最强的监管机构吗？这么快就能查清问题，真是太厉害了！
- 3.太和水公告称收到了《立案告知书》，看来是摊上事了，希望公司能认真反思，及时纠正错误。
- 4.证监会这次动真格的查违规行为，给广大投资者一个公平公正的市场环境，值得赞赏！
- 5.证监会一定要果断出击，让那些违法违规的行为无法继续得逞，保护投资者的合法权益！

## 7 研究趋势及不足

目前舆情管理研究成果已初具规模，但大多停留在理论探讨或个案分析，缺乏系统的定量研究，而舆情管理面临的环境变化和挑战日益增多。舆情仿真研究虽有较长历史，但多停留在概念与单一场景下的简单模型，缺乏真实应用案例验证，模型构建与参数的设计与定量需要进一步完善。智能体交互研究虽提出多种方法，但多停留在理论层面，同时由于缺乏基于大规模真实数据构建的智能体模型，限制了研究成果的实用性。

本文提出的“基于 AIGC 智能体的舆情角色交互仿真研究”可以整合舆情管理、仿真与智能体交互

等多个研究领域，可以基于大数据构建真实智能体模型，增强模型真实性与应用潜力；可构建复杂互动场景下的多智能体交互模拟系统，致力于解决舆情管理重大问题；研究成果有望直接提供舆情监测预警和应对决策支持，提升舆情治理水平；同时推动三个研究领域交叉互动，对理论进步与应用创新都有重要作用。

## 8 理论与实践方面的预期成果

### （1）打造智能和真实的舆情演练仿真平台

本研究的最终目标在于打造一个智能而真实的舆情演练仿真平台。该平台可以模拟真实的社会热点事件，重现其中的复杂舆情过程，实现多角色的互动与协同。通过构建基于GPT的生成式智能体，平台可以自动产生不同舆情角色真实且连贯的言论，反映其独有的价值观、立场与表达风格。这些生成的言论与真人互动，可以达到高度仿真的效果。通过在智能和高度仿真的环境下进行舆情演练，参与者可以提高自己应对复杂舆情事件的能力。他们学习到的不仅是应对某一舆情事件的策略，更是处理动态变化舆情的思维模式与框架。这些经验对于实际工作中的舆论危机管理将产生重要影响。

### （2）利用生成式智能体交互技术构建出能针对特定类场景舆情预测系统

交互从而实现舆情预测的系统。这类智能体可以在特定的舆情场景下，模拟不同角色进行交互，最终实现针对特定场景的舆情观点及发展预测。这些生成式智能体在给定的场景下可以工作。它们根据场景信息生成对应的角色观点，这些观点体现角色固有的价值取向与论述框架。通过在特定场景下与参与者的持续互动，这些基于 GPT 的生成模型展现出对场景的感知与适应性，以及代表不同角色的能力，实现多角色之间的高效协同。通过智能体与人进行高效协同，实现多角色之间的互动仿真。