

厦门大学“大学生创新创业训练计划” 创新训练项目申报书

学院:	信息学院
项目名称:	基于大语言模型的网络舆情分析与应对
负责人:	宋松原
联系方式:	19859305108
指导教师:	曾鸣
职称:	副教授
学历:	博士研究生毕业
E-mail:	974400809@qq.com
项目申请日期:	2023 年 12 月 15 日
项目起止年月:	2023 年 12 月 15 日至 2025 年 7 月 1 日

厦门大学教务处
二〇一三年十一月

填 写 说 明

1、本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要，申报书请按顺序逐项填写，填写内容必须实事求是，空缺项要填“无”。

2、申请人可以是个人，也可为创新团队，首页只填主持人。

3、本申请书为大 16 开本（A4），左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。填写完后用 A4 纸张双面打印，不得随意涂改。

4、主持人所在学院认真审核，经初评和答辩，签署意见后，将申请书（一式两份）报送教务处。

一、基本情况

项目名称	基于大语言模型的网络舆情分析与应对						
所属学科	计算机类						
申请金额	30000.00 元		起止年月		2023 年 12 月 15 日至 2025 年 7 月 1 日		
主持人姓名	宋松原	性别	男	民族	汉族	出生年月	2003 年 6 月
学号	30320212200061	联系电话	负责人联系方式				
指导教师	曾鸣	联系电话	15959236379				
主持人曾经参与科研的情况		参加 2022 年度第一批次大创实验 参加 2023 年度 c 语言积分赛并且进入决赛 参加 2023 年度海洋学院竞赛获得一等奖 参加 2022 年度挑战杯					
指导教师承担科研课题情况		指导参加 2020 国际 DeepFake 人脸视频检测挑战赛金牌（2000+队伍参赛，金牌取前 14 名）2020 年 参加国际计算机视觉大会 ICCV2019 三维人脸重建挑战赛，冠军，2019 年 指导参加首届中国人工智能-多媒体信息识别技术竞赛“创新之星”（全国仅 16 项，高校仅 1 项），2019 年 指导参加十二届英特尔杯软件创新大赛特等奖（全国仅一项），2019 年（为厦大在该赛事最好成绩）参加第十二届英特尔杯软件创新大赛最具创业潜力奖（全国仅两项），2019 年 获得“基于个性化定制的三维人体智能测量系统”，中国发明协会发明创业成果一等奖，2018 年 “人体智能测量及大数据应用服务”，中国产学研合作创新成果二等奖，2018 年					
指导教师对本项目的支持情况		提供本项目的选题并且全程进行指导，提供实验室为本项目的实验场所，并且提供进行搜索的渠道以及相关软件的提供以及选取，并且配备一名博士生进行指导。 提出所存在的问题并且引导团队解决					
项目主要成员	姓名	学号	专业班级		所在学院		项目中的分工
	宋松原	30320212200061	软件工程		信息学院		组长
	寇怀佳	30320212200028	人工智能		信息学院		组员

	张泽楷	3032021 2200093	软件工程	信息学院	组员
	周冠旭	3032021 2200099	计算机科学与技术	信息学院	组员
	蔡欣琪	3722023 2203599	软件工程	信息学院	组员

二、项目简介（限 200 字以内）

本项目利用大语言模型进行网络舆情分析与应对。通过自动抓取网络信息，理解并识别负面舆情，借助 AI 技术生成应对策略，为用户提供精准、高效的舆情管理方案。

三、立项依据（可加页）

（一）研究目的

随着互联网的普及和发展，网络舆情已成为反映社会民意、引导社会舆论的重要途径。但是，由于网络信息量巨大，传统的人工处理方式难以及时、准确地掌握舆情动态，这就需要借助科技力量，开发出能够自动获取、理解和分析网络舆情，并提出应对策略的智能系统。因此本项目贴合时代并且符合社会需求。

本团队一共 5 名成员，涵盖了计算机与科学专业，软件工程专业以及人工智能专业，与本次做的大创所需要的专业领域息息相关，所有成员都对该大创项目具有浓厚的兴趣，具有完善完成该项目的专业能力，并且自主学习能力强，可安排的自主学习时间

多，可以投入充分的时间进行大创项目开展。并且该项目的核心与我们生活息息相关，具有高度的关联性，团队成员都对舆情的产生以及辨识具有一定的经验并且可以提出相应的措施，并对于大语言模型也有着充分的了解，可以较好的将两者结合为一体并且提出相应的解决措施。

于此同时本团队由曾鸣副教授进行项目指导，曾鸣老师有着在此领域极其丰富的经验，并且对于该项目的语言模型和项目内容有着相关的指导经验。同时本项目还配备有一名经验丰富的博士生学长进行指导，具有充分的前期准备。因此开展此大创项目

（二） 研究内容

本项目旨在研发一套基于大语言模型的网络舆情分析与应对系统，该系统应具备以下功能：

自动抓取网络信息：从新闻网站、社交媒体、论坛等平台收集信息，保证数据来源的全面性和实时性。

舆情理解与分析：通过大语言模型理解文本内容，识别出其中的正面、中性或负面情绪，以及关键主题和观点。

应对策略生成：根据舆情的具体情况，运用 AI 技术生成有效的应对策略，包括信息发布、危机公关、舆情引导等。并且可以结合实时的法律系统提出合理的公关措施。

实时监控与预警：对重点事件和热点话题进行持续关注，及时发现潜在风险，发出预警。并设置好提预案。

（三） 国、内外研究现状和发展动态

国内研究现状：

近年来，我国在网络舆情分析与应对方面的研究取得了显著进展。许多高校和研究机构都投入了大量的精力进行相关的研究工作。在技术层面上，研究人员主要集中在以下几个方向：大数据处理、自然语言处理、人工智能等。特别是在大语言模型的应用方面，已有一些优秀的研究成果出现。然而，目前的网络舆情分析系统仍存在一些不足，如数据源不够丰富、分析深度不够、应对策略单一等，这些问题都有待进一步解决。

国际研究现状：在国外，网络舆情分析的研究同样受到广泛关注。美国、欧洲等地的一些顶级大学和公司都在这方面进行了深入的研究。在技术层面，国外的研究更注重前沿技术的应用，如深度学习、神经网络等。在实际应用方面，一些国际知名的舆情分析公司已经推出了成熟的产品和服务，但这些产品往往价格昂贵，不适合大规模普及。

发展动态：

随着人工智能技术的快速发展，大语言模型在舆情分析中的应用越来越广泛。未来，我们可以期待更多的新技术被应用于网络舆情分析，如无监督学习、强化学习等，这将进一步提高舆情分析的准确性和有效性。并且在全球范围内，各国政府开始重视网络舆情的监管和治理。相关的法规政策也在逐步出台和完善，这给网络舆情分析带来了新的挑战 and 机遇。例如，如何在保护个人信息和言论自由的同时，有效地进行舆情分析和应对，将成为一个重要的研究课题。随着社交媒体的普及，网络舆情的影响越来越大。因此企业和政府部门对舆情管理的需求也越来越强烈。这为网络舆情分析与应对系统的开发提供了广阔的市场空间。在未来，我们预计会有更多针对特定行业的舆情分析产品出现。

总结来说，网络舆情分析与应对是一个充满挑战和机遇的研究领域。尽管已经取得了一些成果，但还有许多问题需要解决。在这个背景下，我们希望通过大语言模型的应用，开发出一个高效、准确、实用的网络舆情分析与应对系统，为社会提供更好的服务。

（四） 创新点与项目特色

（1）项目特色

全面的数据来源：本项目不仅关注新闻媒体等权威发布渠道，还将社交媒体、论坛等非正式讨论场所纳入数据采集范围，充分体现了全网舆情的特点。

强大的理解能力：利用最新的大语言模型，如 BERT、GPT-3 等，

以及目前最贴近生活的文心一言以及通义等模型，对文本内容进行深入理解，准确捕捉舆情的关键信息和情感倾向。

高效的应对策略：结合 AI 算法，根据舆情的具体情况，快速生成有效的应对策略，包括信息发布、危机公关、舆情引导等。

实时监控与预警：实时跟踪热点事件和话题，及时发现潜在风险，发出预警信号。

用户友好的界面：提供简洁明了的操作界面，方便用户查看舆情报告，调整参数设置，以及接收预警通知。

（2）创新点

基于大语言模型的舆情理解：不同于传统的基于规则或统计的方法，本项目采用大语言模型来理解文本内容，能更好地捕捉语义特征，提高理解的准确性。

动态的舆情应对策略：传统的舆情应对往往是静态的，缺乏灵活性。本项目通过 AI 算法生成动态的应对策略，可以根据舆情的变化做出适时调整。

综合的舆情评估体系：除了情感倾向外，本项目还可以充分考虑舆情的传播速度、影响力等因素，构建了一个综合的舆情评估体系。

深度的跨学科融合：本项目不仅仅是计算机科学在其中参与，还设计到信息科学、社会学、新闻学等多方面的学科，具有一定深

度的跨学科融合。

综上所述，本项目的特色和创新点主要体现多个方面，有望为网络舆情分析与应对提供一种全新的解决方案。

（五） 技术路线、拟解决的问题及预期成果

（1） 技术路线

数据采集：使用爬虫技术从网络上抓取公开的文本信息，涵盖新闻、微博、论坛等多种来源。为了提高数据质量，我们将采用多线程和分布式爬虫，并设置一定的更新频率。

数据预处理：对抓取的数据进行清洗，去除无关信息，如广告、重复内容等。然后进行分词、去停用词等操作，以便后续的文本分析。

大语言模型应用：使用大语言模型（如 BERT、GPT-3 等）对预处理后的文本进行深入理解，提取关键词、主题和情感倾向。这部分工作涉及到大量的机器学习和自然语言处理技术。

AI 算法设计：针对不同的舆情场景，设计相应的 AI 算法来生成应对策略。这可能包括聚类分析、情感分析、主题模型等方法。

系统集成与优化：将上述模块整合到一个统一的系统中，并进行性能优化，确保系统的稳定运行和高效处理能力。

APP 或网站设置：可以创建相应的 APP 和网站来使项目具有实体化，并且充分考虑用户的交互体验，对于相应的部分进行优化。

(2) 人员分工

项目总负责人：负责整个项目的策划、组织和协调工作，确保项目按照预定的时间表顺利推进。

在数据采集和处理方面可以分成两个部分

a. 数据采集小组：了解网络爬虫的基本原理和使用方法，负责从网络上抓取公开的文本信息，涵盖新闻、微博、论坛等多种来源。

要

b. 数据预处理小组：需要熟悉自然语言处理的基本方法，负责对抓取的数据进行清洗，去除无关信息，如广告、重复内容等。然后进行文本分析。

c. 大语言模型小组：了解深度学习和自然语言处理的基本理论，负责大语言模型的应用，对预处理后的文本进行深入理解，提取关键词、主题和情感倾向。

d. AI 算法小组：负责设计和实现针对不同舆情场景的 AI 算法，包括聚类分析、情感分析、主题模型，并且可以做到对于多场合的 AI 算法的运用。

e. APP 开发小组：负责系统的集成和优化工作，确保系统的稳定运行和高效处理能力。并且进行系统的测试和调试，确保产品的

质量和稳定性以及进行产品的界面设计和用户体验优化，使产品更加易用和美观。

d.项目编写小组：负责编写项目文档，包括需求分析、设计文档、用户手册等。

以上分工为互相融合的小组，各个成员都可以参与不同的小组进行合作，发挥专业交融的优势，从多方面多角度对大创项目进行完善。

（3）拟解决问题

1. 考虑如何获取大量的平台账号？
2. 实时列出不限范围主要舆情（舆情榜单）
3. 实时列出关注点舆情
4. 给出舆情应对策略
5. 根据应对策略，自动处理舆情

（4）预期成果

a.技术成果：开发一套基于大语言模型的网络舆情分析与应对系统，该系统应具备自动获取、理解和分析网络舆情的能力，并能生成相应的应对策略。

利用大语言模型对舆情数据进行深入挖掘，提取关键信息和情感倾向，提高舆情分析的准确性和全面性。

结合 AI 算法，生成动态的舆情应对策略，可根据舆情的变化做出适时调整。

构建一个综合的舆情评估体系，除了情感倾向外，还考虑了舆情的传播速度、影响力等因素。

b.学术成果: 完成一篇或多篇高质量的研究论文，在国内外知名期刊或会议上发表，展示本项目的创新性和实用性。

参加相关的学术会议和研讨会，与其他研究者交流思想，提高自身的学术水平和影响力。

争取获得属于该项目自身的专利。

c.社会效益: 提高舆情管理的效率和效果，有助于维护社会稳定、促进信息透明。

（六） 项目研究进度安排

本项目预期完成时间为一年半左右，从 2024 年 1 月开始，具体进度安排如下：

- 1.（1-3 个月）：需求调研和技术选型，制定详细的项目计划和技术路线。并且采购相关所需的资源。
- 2.（4-6 个月）：数据采集和预处理，搭建基础的大语言模型环境。
- 3.（7-9 个月）：大语言模型的应用和 AI 算法的设计，初步实现舆情分析功能。
- 4.（10-12 个月）：系统集成和优化，完成产品原型的开发。

5. (13-15 个月): 产品测试和调试, 修复问题, 完善功能。
6. (16-18 个月): 产品上线和推广, 收集用户反馈, 进行持续改进以及由导师带领完成学术论文。
7. (第 18 月): 项目结题并且准备参与答辩。

(七) 已有基础

1. 与本项目有关的研究积累和已取得的成绩

我们团队在过去的研究中积累了丰富的经验和技能。我们在大数据处理、自然语言处理、机器学习等领域进行了深入的学习和实践, 具备了扎实的基础知识。此外我们团队成员对于舆情的研究都有着密切接触, 对此项目有着强烈兴趣。

2. 已具备的条件, 尚缺少的条件及解决方法

a. 已具备的条件:

(1) 技术储备: 我们拥有较强的技术实力以及自学能力, 掌握了大数据处理、自然语言处理等相关领域的核心技术。 (2) 人才资源: 我们有一支高素质的研发团队, 成员具有充沛的可支配时间。 (3) 硬件设施: 我们有进行本项目研究的大语言模型, 可以进行实验。

b.尚缺少的条件:

(1) 数据资源: 虽然我们已经有了部分舆情数据, 但仍需收集更多的样本, 以保证模型的泛化能力和实用性。 (2) 行业知识: 由于舆情涉及各行各业, 我们需要深入了解各个行业的特点和需求, 以便更好地开发针对性的产品和服务。

四、经费预算

开支科目	预算经费 (元)	主要用途	预计使用 时间
1. 业务费			
(1) 计算、分析、测试费	8000.00	购买服务器并且进行相关的计算测试以及分析	2023 年 12 月 15 日
(2) 能源动力费	1000.00	进行实验研究或者参加会议所需要花费	2023 年 12 月 15 日
(3) 会议、差旅费	3000.00	参与国内学术会议费用	2023 年 12 月 15 日
(4) 文献检索费	2000.00	用于检索相关的与该项目有关的文件并且获得授权	2023 年 12 月 15 日
(5) 论文出版费	6000.00	发表论文、软件著作权、专利所需费用	2023 年 12 月 15 日

2. 仪器设备购置费	2000.00	购买办公用品及必要的软件授权	2023 年 12 月 15 日
3. 实验装置试剂费	0.00		2023 年 12 月 15 日
4. 材料费	8000.00	购买并配备低值计算机配件等	2023 年 12 月 15 日
预算经费总额	30000.00		
学校批准经费	4000.00		

五、指导教师意见

<p>项目需求分析明确、技术方案详实新颖，有望做出高水平研究成果。建议予以优先支持。</p> <p>导师（签章）：曾鸣</p> <p>2023 年 12 月 15 日</p>

六、院系大学生创新训练计划专家组意见

<p>专家组组长（签章）：</p> <p>2024 年 1 月 1 日</p>

七、校大学生创新训练计划专家组意见

<p>负责人（签章）：</p> <p>年 月 日</p>

八、大学生创新创业训练计划领导小组审批意见

负责人（签章）：

年 月 日