

# 大模型识别谣言不同来源效能研究

何静<sup>1</sup> 陈逸然<sup>2</sup>

1 北京航空航天大学人文与社会科学高等研究院 北京 100191

2 北京航空航天大学人工智能学院 北京 100191

**摘要** 针对网络谣言识别面临的新挑战,探索大模型在识别谣言不同来源中的效能。研究构建国内外人为谣言与 AI 谣言数据集,据此在零样本设置情况下,对 4 种大模型的谣言来源辨识能力进行测试。研究发现,单一大模型识别谣言的精确度较低,存在明显错误倾向。为提高识别性能,采用预训练、微调 and 集成学习等方法,使得大模型性能得到显著提升。进一步,提出基于模型碰撞的集成学习方法,利用多模型反馈改善谣言来源识别效能。实验结果显示,集成学习框架能够整合各模型优势,显著提高识别准确性。通过实证研究验证了大型语言模型在谣言识别中的潜力和改进方向,有助于应对当前复杂的网络谣言环境,维护网络空间的清明。

**关键词**: AIGC; 大模型; 谣言识别; 谣言治理

**中图分类号** G206; TP18

## Study on Efficiency of Large Model in Recognizing Rumors from Different Sources

HE Jing<sup>1</sup> and CHEN Yiran<sup>2</sup>

1 Institute for Advanced Studied in Humanities and Social Sciences, Beihang University, Beijing 100191, China

2 School of Artificial Intelligence, Beihang University, Beijing 100191, China

**Abstract** This study aims to address the new challenges faced by online rumor recognition and explore the effectiveness of large models in recognizing different sources of rumors. Constructing domestic and foreign rumor and AI rumor datasets, and testing the rumor source identification ability of four large models under zero sample settings. Research has found that a single large model has low accuracy in identifying rumors and has a clear tendency towards errors. To improve recognition performance, methods such as pre-training, fine-tuning, and ensemble learning are adopted to significantly enhance the performance of the large model. Furthermore, a model collision based ensemble learning method is proposed to improve the effectiveness of rumor source recognition by utilizing multi model feedback. Experimental results show that the ensemble learning framework can integrate the advantages of various models and significantly improve recognition accuracy. This study verifies the potential and improvement direction of large-scale language models in rumor recognition through empirical research, which helps to cope with the current complex online rumor environment and maintain the clarity of cyberspace.

**Keywords** AIGC, Large model, Rumor recognition, Rumor control

## 1 引言

随着人工智能生成内容 (Artificial Intelligence Generated Content, AIGC) 的快速发展,特别是以大型语言模型 (Large Language Model, LLM, 以下简称大模型) 为代表的大模型的涌现,社会进入了多模态大模型创生阶段。这一新技术的出现,在 2023 年已经引起广泛的关注,而在 2024 年初,OpenAI 的 Sora 则如同一颗炸弹再次引爆了全球的关注。

AIGC 技术极大地降低了制造谣言的门槛和成本,从文本、图像、视频、时序等多个维度进行造假,从而引发了“数字野火”,导致后真相时代的到来。这种技术具有自动生成、广泛传播、高度真实等特性,使得谣言的生成变得简易化、批量化和拟真化,因此网络谣言乱象急剧恶化,严重危害社会安定<sup>[1]</sup>。在信息传播方面,AIGC 技术使得谣言不仅可以实现

网络全面渗透,还能够进行精准定向攻击。随着大型预训练语言模型如 OpenAI 的 GPT-4、Meta 的 OPT 以及 Big Science 的 BLOOM 的迅速发展,它们在 AI 辅助写作<sup>[2]</sup>、医疗问题回答、金融服务以及法律领域等多个方面的应用,展现了生成复杂、连贯内容的能力。这些模型仅需简单的文本提示,就能生成难以区分其来源是人类还是 AI 的文本,为社交媒体上的谣言检测和鉴别带来了新的挑战<sup>[3]</sup>。

目前,已有大量研究聚焦于如何使用大模型识别谣言内容,但现有研究更加关注谣言辨别能力,以及 AIGC 内容的辨别,但对于谣言的来源——是由人工智能 (AI) 生成还是由人类创造,目前研究较为缺乏。而在 AIGC 生成谣言泛滥的今天,识别谣言来源对于开发精准工具检测控制谣言,并补充法律责任框架,具有重要意义。本文将聚焦于大模型对谣言来源的识别能力展开研究和验证,并深入探讨提升其识别谣言

基金项目:北京市社会科学基金(23XCC020);北京市教育科学“十四五”规划课题(CGCA23128)

This work was supported by the Beijing Social Science Fund (23XCC020) and Beijing Education Science 14th Five Year Plan Project (CGCA23128).

通信作者:何静(bhhejing@buaa.edu.cn)

精确度的有效路径,为构建更加健全的网络信息生态并维护社会稳定和公共利益提供支持。

## 2 相关研究

传统上,谣言检测依赖于手工提取的特征和简单的分类技术,但这些方法在处理复杂的信息传播行为时常显不足。近年来,大量研究开始探索结合先进的机器学习技术和 LLM 来提升谣言检测的性能<sup>[4]</sup>。

### 2.1 谣言识别理论更新

首先,Kwon 等<sup>[5]</sup>的研究揭示了在线社交媒体中谣言传播的显著特征,特别是谣言内容中积极情感词的稀缺性。这一发现为开发更加精确的谣言识别方法提供了重要线索。为了充分挖掘和利用这些特征,Yang 等<sup>[6]</sup>提出了一种集群智能和 ChatGPT 辅助的网络(CICAN),该网络通过群体智能学习语义特征并结合 ChatGPT 的知识增强,有效提升了谣言检测的性能。此外,LLM 的应用不仅限于知识提取,Chen 和 Shu 以及其他学者探索了 LLM 在验证信息准确性方面的潜力,尽管他们也指出了单纯依赖文本特征的局限性。

在此基础上,其他研究通过采用图卷积网络(Graph Convolutional Network,GCN)和评估不同的假新闻检测方法,进一步拓展了谣言识别的方法论<sup>[7]</sup>。Shahid 等<sup>[8]</sup>对检测恶意用户和机器人的最新方法进行了全面调查。D'Ulizia 等<sup>[9]</sup>提供了一个关于假新闻检测评估数据集的概述,而 Alsaif 等<sup>[10]</sup>则回顾了利用立场检测来识别谣言的最新方法,此类方法为理解和应对社交媒体上的谣言传播提供了多维度的视角。

### 2.2 谣言识别技术适配

在技术适应性方面,Slimi 等<sup>[11]</sup>通过将 RoBERTa 语言模型适配到 Twitter 上的谣言检测任务,展现了预训练语言模型超越传统手工特征方法的潜力。他们的工作表明,通过微调,模型对谣言识别的精度可以得到显著提升。最近,一些研究工作试图将 LLM 应用于识别信息的准确性<sup>[12]</sup>,譬如,Liu 等<sup>[13]</sup>使用 ChatGPT 辨别文本真实性。此外,Yan 等<sup>[14]</sup>采用知识驱动的提示方法来增强 LLM 在谣言检测领域的性能,证明了领域知识的重要性和有效性。

### 2.3 本文主要创新贡献

据调研,识别信息的来源——是 AI 生成还是人为制造的谣言,尚未引起足够的关注。LLM 以模仿人类写作的方式轻松创建内容,放大了错误信息传播的风险。随着 LLM 变得更加普及,其越来越多地用于更广泛的业务,假叙述、假新闻和垃圾信息的广泛传播潜力增加<sup>[15]</sup>。因此,了解内容的来源,对于在下游应用中的适当使用和监督至关重要<sup>[16]</sup>。识别 AI 生成内容的来源,不仅可以帮助标记信息的准确性,还能提醒用户注意该模型的特定偏见和限制。本文的主要贡献点如下:

1)构建了一个综合性的谣言数据库,并对现阶段多种大型模型在识别谣言来源上的效能进行了深入分析。研究发现了大型模型在识别中的一些集中共性问题。

2)基于对单个大型模型的测试结果,提出了一种基于数

据集预训练和微调的方法,旨在提高大型模型在识别谣言来源上的准确率。通过对比分析不同预训练策略的效果,证实了这种改善思路的可行性。

3)研究了多模型碰撞情况,并引入了集成学习框架,以整合多个大型模型的优势,进一步提升谣言来源识别的效率和准确性。

## 3 大模型识别谣言不同来源效能研究方法与实验

### 3.1 构建综合性谣言数据集

根据谣言不同来源,可构建包含 AIGC 谣言和人为谣言的综合性谣言数据集,其中数据准备的整体流程如图 1 所示。

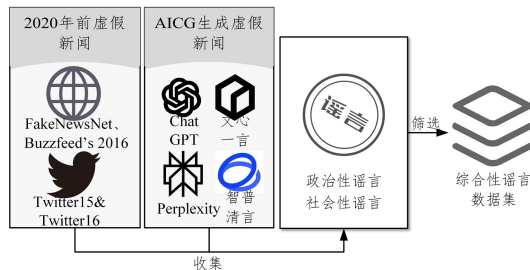


图 1 数据集构建流程

Fig. 1 Dataset construction process

按照谣言内容划分,谣言总体可分为政治谣言、经济谣言、军事谣言、社会生活谣言和自然现象谣言;按照后果划分,谣言可分为有害谣言和无害谣言<sup>[17]</sup>。其中,政治与社会谣言传播广泛,辨识困难,后果严峻;相对地,自然与经济谣言由于其背后的知识基础,其识别与检测相对容易。因此,构建数据集特别关注政治与社会谣言两大领域。

现有谣言数据集并未对谣言来源加以区分,无法辨别谣言生成来源。为保证数据集中数据特征的准确性,从 FakeNewsNet<sup>1)</sup>,Twitter15 & Twitter16 和 Buzzfeed's 2016 数据集中收集 6407 条数据,涵盖 2020 年以前国内外知名社交平台的真实人为谣言数据,在保障数据代表性的基础上,避免了 AIGC 生成导致的数据误判问题。

1)FakeNewsNet 是一个收集、分析和可视化假新闻以及社交媒体上的相关传播信息的数据库。该数据集提供了来自 PolitiFact<sup>2)</sup> 和 Gossipcop 的新闻信息。其中,PolitiFact 是一个美国的政治事实检查网站,由坦帕湾时报(Tampa Bay Times)在 2007 年创建;GossipCop 是一个专门用于核实娱乐和名人八卦新闻真实性的网站,由记者 Michael Lewittes 和前警察 Turner 派出所所长 Dan Abrams 在 2009 年共同创立。

2)Twitter15&Twitter16 数据库收集了来自 Twitter15-16 的贴文,并对真实性进行标注,内容以社会性新闻为主。

3)Buzzfeed's 2016 数据库包含了作者收集的 2016 年期间的选举假新闻、讽刺新闻等,内容以政治性事件为主。

在生成 AIGC 谣言数据部分,基于 SimilarWeb<sup>3)</sup> 统计的各生成式人工智能(Generative Artificial Intelligence,GAI)平台月度访问量数据,选取访问量排名靠前的平台与国内自研平台,即 ChatGPT-4、Perplexity、文心一言,分别生成不同类别谣言 120 条,作为 AI 谣言样本。使用提示词“构造”虚拟社

<sup>1)</sup> <https://github.com/KaiDMML/FakeNewsNet>

<sup>2)</sup> <https://www.politifact.com/>

<sup>3)</sup> <https://github.com/similarweb>

会新闻和政治新闻,构造内容共分为两种形式:

1)选取真实新闻作为背景框架,通过使用 AI 修改新闻部分关键词,并对新闻事实进行改编重写;

2)虚拟事件新闻,并由 AI 完全自生成内容和细节。

为保证统一性,人为谣言和 AIGC 谣言均包含不同语言的新闻数据;为确保用户隐私,新闻中所有关键人名均进行匿名处理。由此初步获得构建完成包含人为谣言和 AIGC 谣言的综合性数据集,数据集具体信息特征如表 1 所列。

表 1 谣言数据集各项信息的特征统计

Table 1 Statistics of various information features in rumor dataset

谣言类别	来源	社会生活谣言	政治谣言	平均文本长度/byte
AIGC 谣言	FakeNewsNet	474	5336	3926.7
	Twitter15 & Twitter16	532	17	1239.5
	Buzzfeed's 2016	0	48	2394.5
人为谣言	ChatGPT-4	20	20	3299.0
	Perplexity	20	20	2546.5
	文心一言	20	20	2365.0

### 3.2 大模型识别谣言不同来源性能评估

为平衡实验数据结构,对初识数据集进行抽样,分别抽取 60 个 AIGC 谣言和 60 个人造谣言作为测试数据集,其中政治新闻和社会新闻比例为 1:1,并使用以下关键指标对谣言来源识别性能进行评估。

1)准确度(Accuracy):准确度表示模型在所有样本中正确分类的比例。

$$Accuracy = \frac{TP + TN}{TP + FP + FP + FN}$$

2)精确度(Precision):精确度表示模型判断为 AI 生成谣言的样本中,实际为 AI 生成的比例。

$$Precision = \frac{TP}{TP + FP}$$

3)召回率(Recall):召回率表示模型能够正确识别 AI 生成谣言的比例。

$$Recall = \frac{TP}{TP + FN}$$

4)F1 分数:F1 分数是精确度和召回率的调和平均数,综合考虑模型的准确率和完整性。

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

其中,TP 表示真正例(模型将 AI 生成的谣言正确判断为 AI 生成),TN 表示真负例(模型将人类生成的谣言正确判断为人类生成),FP 表示假正例(模型将人类生成的谣言错误判断为 AI 生成),FN 表示假负例(模型将 AI 生成的谣言错误判断为人类生成)。

在零样本设置(Zero-shot Setting)即不训练数据而直接对数据集样本进行评估的情况下,对 4 种大模型(ChatGPT-4、Perplexity、文心一言和智普清言)的辨别能力进行了测试,具体提示词和实验流程如表 2 所列。

表 2 谣言识别测试流程

Table 2 Rumor recognition testing process

Prompt	Title: Innovative Urban Development Plan to Transform Cities An innovative urban development plan has been unveiled, aiming to transform cities into more sustainable, livable, and technologically advanced environments. The plan includes initiatives for green infrastructure, smart technology integration, and community engagement programs. By reimagining urban spaces, the project seeks to address challenges such as pollution, housing shortages, and traffic congestion, setting a new standard for city living in the 21st century. 请你告诉我这条新闻谣言是 AI 生成的还是人工编造的,如果是人工编写请回答 1, AI 生成请回答 2。
	Output

- 普遍性与缺乏细节:这个描述非常通用,没有提供特定的实施城市、负责的组织、资金来源或项目启动时间等细节。这种普遍性既可能是为了保持信息的简洁,也可能是因为缺乏具体的背景信息。
  - 创新性与趋势:描述中提到的概念,如“绿色基础设施”“智能技术集成”和“社区参与计划”,都是当前城市规划和发展的热门话题。这些概念的使用显示了一定程度的行业知识,但也很常见,并不特定于任何一个具体的计划。
  - 目标与挑战:文中提及的目标(解决污染、住房短缺和交通拥堵等挑战)是许多城市发展计划共有的。这些目标的普遍性和广泛性使得新闻听起来更像是一个理想化的计划概述,而不是一个具体的、即将实施的项目。
  - 语言风格:描述的语言风格既可以是人工编写的,也可以是 AI 生成的。没有明显的语言或语法错误,也没有深入的技术细节或特定数据支持其声明。
- 综合来看,我认为是 AI 生成的,我的回答是 2。

在实验中,首先向大模型发送新闻标题和完整文本内容,并根据固定 prompt 向 AI 发送指令,关注 AI 给出的最终答案;其次,对 sample 进行手动测试并逐一记录结果;最后,在 4 个平台上,分别进行完全一致的重复测试。通过实验,即可得到各模型在谣言识别任务上的性能指标,如表 3 所列。

表 3 性能比较及各模型表现结果

Table 3 Performance comparison of various models

模型	准确率	精确度	召回率	F1 分数
ChatGPT-4	64.2	96.7	58.6	73.0
Perplexity	40.7	75.0	44.6	55.9
文心一言	55.0	93.3	52.9	67.5
智普清言	50.5	83.3	50.5	62.9

辨别要求时,识别准确率整体较差,并且错误倾向性较为明显,大部分错误集中在将 AIGC 文本识别为人工造谣。在实验过程中发现,随着文本长度的增长,大模型会出现注意力不集中的状况,具体表现为给出模棱两可的答案。例如,模型 A 回复如下:“基于上述分析,这个新闻条目看起来像是基于真实研究的报道,而不是完全由 AI 生成或人工编造的。然而,没有具体的研究链接或出版日期,无法完全验证所有细节的真实性。”这种情况在 2000 字节以上文本格外常见。针对大模型在实际识别中出现的问题,需要进行精确度层面的进一步完善。

### 3.3 基于数据集预训练和微调提升谣言来源识别精确度

单个大模型在零样本设置条件下对谣言来源识别过程中,表现出注意力涣散与准确率低下的问题,对此,需要深化

从表 3 中可以看到,单个模型在提示词直接指示说明

模型对谣言不同来源的特征理解来提高识别精确度。

Zaeem 等<sup>[18]</sup>观察到负面情绪和假新闻之间存在显著的统计学关系。这些情绪反过来又会产生特定的行为或行动,如传播谣言的动机<sup>[19]</sup>。以此为例, AI 生成谣言与人造谣言可能在情绪、表述和语法逻辑上产生细致差异,将数据集特征进行进一步划分并对大模型进行数据集预训练,将有效提升模型对不同类别谣言的感知能力。

表 4 使用提示词进行特征提醒

Table 4 Use prompt words for feature reminders

Prompt	Title: Groundbreaking Study Reveals Surprising Insights into the Impact of Artificial Intelligence on the Job Market Researchers at the University of Oxford have published a landmark study that sheds new light on..... 这是一条新闻谣言, 请你根据新闻撰写的风格一致性、分析复杂性和深度、细节的真实性和准确性、情感表达和态度, 告诉我这条新闻谣言是 AI 生成的还是人工编造的, 如果是人工编写请回答 1, AI 生成请回答 2。
Output	1. 风格一致性: 这段描述保持了正式和信息性的风格, 这与政府声明和新闻报道的常规风格相吻合。 2. 分析复杂性和深度: 文本提到了政府的几项具体措施, 如改善基础设施、提供教育资源和支 持技术创新, 这显示了一定程度的复杂性。然而, 缺乏具体实施计划的详细信息, 比如投资额、具体项目或时间表。 3. 细节的真实性和准确性: 虽然提及了政府投资计划的大致方向, 但没有提供足够的细节来验证这些计划的真实性。真实的政策声明通常会包含更多具体信息, 如财政预算、政策实施的具体地区以及预期的效果评估。 4. 情感表达和态度: 文本传达了一种积极向上的态度, 这是政府宣传中常见的情绪。 综上所述, 这条消息虽然在结构和表达上符合一般新闻报道的标准, 但缺乏具体细节和可验证性, 可能表明它是人工编造的, 我认为是 AI 生成的。

在此基础上, 采用每次将一个样本集用于训练, 另一个样本集用于测试的方法, 即两倍交叉验证的方法, 在 4 个不同的大型模型平台上进行重复测试。实验结果如表 5 所列。

表 5 基于预训练与微调后识别效果的比较

Table 5 Comparison of recognition performance based on pre-training and fine-tuning

模型	准确率	精确度	召回率	F1 分数	精确度提升率 (%)
ChatGPT-4	70.8	96.7	63.8	76.9	10.28
Perplexity	45.8	75.0	47.4	58.0	12.53
文心一言	60.0	91.7	56.1	69.7	9.09
智普清言	58.3	93.3	54.9	69.1	15.45

上述实验结果验证了研究提出的假设与猜想, 即引入预训练并使用提示词对信息特征进行强调后, 大模型的识别准确率将有不同程度的提升。在对足够多的数据样本进行学习后, 大模型能够更加精准地把握谣言不同来源的差异特征, 通过对特征进行提示加强, 大模型对长文本的处理更加精确, 能够按照提示反馈具体答案。

### 3.4 基于模型碰撞的集成学习框架

为进一步提升谣言来源识别的效率和准确性, 研究引入一种基于模型碰撞的集成学习框架。核心思想在于利用不同大模型平台在处理同一问题时的差异性和统一性, 通过综合各模型的判断结果, 达到更高的识别准确率和可靠性。

模型碰撞是指在集成学习过程中, 不同模型对于同一输入数据给出不同输出结果的现象。在谣言识别任务中, 这种差异性可以用来评估和提升最终决策的准确性。如果多个模型对同一任务的识别都给出了错误答案, 则可能意味着模型对这种类型的谣言识别能力有所欠缺或存在值得注意的模型风险。

选取 ChatGPT-4、Perplexity、文心一言和智普清言 4 个

在测试前, 需要对 4 个大模型平台进行预训练。从综合数据集中筛选测试样本后, 再次抽取 60 个 AIGC 谣言和 60 个人造谣言作为训练数据集, 对每条谣言来源进行标注, 并将这 120 条新闻谣言作为预训练数据提供给大模型进行学习。

在测试中, 对测试样本的每条新闻的数据特征进行标注, 并在提示词中着重强调文本特征。具体操作如表 4 所列。

大模型, 在第 3.2 节中, 对于每个模型  $i$ , 已分别测出了其样本数据的准确率  $a_i$ , 则错误率  $e_i = 1 - a_i$ 。

使用 Pearson 相关系数评估两个模型错误预测之间的相关性  $\rho_{ij}$ , 具体计算公式如下:

$$\rho_{ij} = \frac{\sum (e_{i,k} - \bar{e}_i)(e_{j,k} - \bar{e}_j)}{\sqrt{(e_{i,k} - \bar{e}_i)^2} \sqrt{(e_{j,k} - \bar{e}_j)^2}}$$

其中,  $e_{i,k}$  和  $e_{j,k}$  分别代表模型  $i$  和  $j$  在样本  $k$  上的错误预测 (错误为 1, 正确为 0),  $\bar{e}_i$  和  $\bar{e}_j$  分别代表模型  $i$  和  $j$  的平均错误率。

对于任意两个模型  $i$  和  $j$ , 其碰撞率  $C_{ij}$  可以基于它们的错误率和错误预测的相关性来估算:

$$C_{ij} = e_i \times e_j \times \rho_{ij}$$

其中,  $C_{ij}$  衡量了两个模型同时给出错误预测的概率。最后, 整体模型碰撞率可以通过考虑所有可能的模型对组合来计算:

$$C = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N C_{ij}$$

经计算, 整体碰撞率  $C$  为 3.08%, 两两模型之间的碰撞率如图 2 所示。

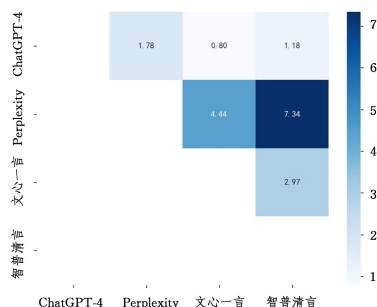


图 2 不同大模型之间两两的碰撞率

Fig. 2 Collision rate between different large models in pairs

根据不同大模型在谣言识别的碰撞情况,使用简单多数投票法(Majority Voting)设计集成学习框架。简单多数投票是一种基本的集成方法,其核心可以表示为:

$$V(y) = \text{mode}\{c_1(x), c_2(x), \dots, c_n(x)\}$$

其中, $V(y)$ 代表最终的投票结果; $C_i(x)$ 代表第*i*个模型对样本*x*的分类结果; $\text{mode}$ 函数返回一组数中出现次数最多的数,即多数模型选择的类别。

基于以上框架,可以重新计算出在4个大模型集成学习后,系统整体的识别情况,如表6所列。

表6 基于预训练与微调后识别效果的比较

Table 6 Comparison of recognition performance based on pre-training and fine-tuning

(%)				
模型	准确率	精确度	召回率	F1分数
集成学习模型	78.3	93.3	71.8	81.2

从实验结果可知,在使用集成学习框架后,系统整体准确率有所提升,且对正反例判断更加平衡,增强了对AIGC谣言识别的准确性。

**结束语** 本研究围绕使用大型语言模型对AIGC谣言和人为谣言进行识别的问题展开。实证研究和实验验证发现,通过在测试前对模型进行预训练,测试中使用提示词(prompts)对信息特征进行强调,以及测试后采用多模型集成学习方法,能够显著提升谣言来源的识别准确率。具体而言,通过对大模型进行定向的预训练和微调,增强模型对谣言特征的识别能力,尤其是在区分AIGC生成的谣言与人为创造的谣言方面。实验结果显示,集成学习框架能够整合不同模型的优势,提高系统整体的识别准确性,相较于单一模型在零样本设置下的性能,展现了显著的改进效果。

未来,考虑到信息传播的多模态性,需要进一步探索跨模态(文本、图像、视频等)的谣言识别,从而持续性地提高谣言识别的准确性,并增强模型对复杂信息环境的适应能力,以期提升信息安全水平,完善法律和政策框架及保护公共讨论空间提供有效帮助。

## 参考文献

- [1] YAN Y X, LI B, FENG J Y, et al. Research on the impact of trends related to ChatGPT [J]. *Procedia Computer Science*, 2023, 221: 1281-1294.
- [2] HUTSON M. Robo-writers; the rise and risks of language-generating AI [J]. *Nature*, 2021, 591(7848): 22-25.
- [3] LI H, MOON J, PURKAYASTHA S, et al. Ethics of large language models in medicine and medical research [J]. *The Lancet Digital Health*, 2023, 5(6): 333-335.
- [4] LIU Y, SHEN H, SHI L. A review of rumor detection techniques in social networks [J]. *Journal of Intelligent & Fuzzy Systems*, 2023, 44(3): 3561-3578.
- [5] KWON S, CHA M, JUNG K, et al. Prominent Features of Rumor Propagation in Online Social Media [C] // *Proceedings of the 13th International Conference on Data Mining*. New York: IEEE Press, 2013: 1103-1108.
- [6] YANG C, ZHANG P, QIAO W, et al. Rumor Detection on Social Media with Crowd Intelligence and ChatGPT-Assisted Networks [C] // *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 2023: 570-5717.
- [7] VARLAMIS I, MICHAEL D, GLYKOU F, et al. A survey on the use of graph convolutional networks for combating fake news [J]. *Future Internet*, 2022, 14: 70.
- [8] SHAHID W, LI Y R, STAPLES D, et al. Are You a Cyborg, Bot or Human? - A Survey on Detecting Fake News Spreaders [J]. *IEEE Access*, 2022, 10: 27069-27083.
- [9] D'ULIZIA A, CASCHERA C M, FERRI F, et al. Fake news detection: a survey of evaluation datasets [J]. *PeerJ Computer Science*, 2021, 7: 518.
- [10] ALSAIF H F, ALDOSSARI H D. Review of stance detection for rumor verification in social media [J]. *Engineering Applications of Artificial Intelligence*, 2023, 119: 105801.
- [11] HAMDA S, IBRAHIM B, YAHYA S. Adapting Pre-trained Language Models to Rumor Detection on Twitter [J]. *Journal of Universal Computer Science*, 2021, 27(10): 1128-1148.
- [12] CHEN C Y, SHU K. 2023. Combating misinformation in the age of LLMs: Opportunities and challenges [J]. *AI Magazine*, 2024, 45(3): 1-15.
- [13] LIU Q, TAO X, WU J F, et al. Can Large Language Models Detect Rumors on Social Media? [J]. *arXiv:2402.03916*, 2024.
- [14] YAN Y Q, ZHENG P, WANG Y J. Enhancing large language model capabilities for rumor detection with Knowledge-Powered Prompting [J]. *Engineering Applications of Artificial Intelligence*, 2024, 133(C): 108259.
- [15] BANG Y J, CAHYAWIJAYA S, LEE N, et al. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity [J]. *arXiv:2302.04023*, 2023.
- [16] ALI H, QADIR J, SHAH Z, et al. ChatGPT and large language models (LLMs) in healthcare: Opportunities and risks [C] // *Proceedings of the 2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things*. New York: IEEE Press, 2023: 1-4.
- [17] ZHANG Y Y, YUAN J J. A review of rumor detection research based on social media [J]. *Data Communication*, 2024(1): 28-33.
- [18] ZAEEM R N, LI C J, BARBER K S, et al. On sentiment of online fake news [C] // *Proceedings of 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. New York: IEEE Press, 2020: 760-767.
- [19] ZHANG N, SONG J H, CHEN K, et al. Emotional contagion in the propagation of online rumors [J]. *Issues in Information Systems*, 2022, 23(2): 1-19.



**HE Jing**, born in 1989, Ph.D, assistant professor. Her main research interests include AI and big data.