AI智创·青春力量

首届AIGC与计算传播创新大赛

传播数据分析赛道

参赛作品：

突发公共卫生事件下“不良信息”型网络暴力演化机制与管控策略研究

AIGC与计算传播创新大赛组织委员会

2025 年 3 月 10 日

摘要：[目的/意义]近年来，突发公共卫生事件滋生的网络暴力问题对网络生态环境以及社会安全都造成了严重的不良影响，其中，“不良信息”型网络暴力信息以其庞大的数量基础以及治理的高难度加剧了此类信息对社会的负面影响。本研究通过考察突发公共卫生事件下“不良信息”型网络暴力的演化机制，为有效预防和管控“不良信息”型网络暴力提供理论与实践参考。[方法/过程]首先，利用 Python 在微博爬取相关数据形成数据集后，再训练 BERT 模型识别和预测微博评论中的“不良信息”型网络暴力信息。其次，结合时间周期理论将“不良信息”型网络暴力信息划分为扩散期、爆发期、衰退期、波动期四个阶段。最后，建立 LDA 主题模型，分析不同阶段“不良信息”型网络暴力信息主题演化趋势，并利用 K-means 聚类将用户评论分成三类，揭示不同类型用户的行为模式。[结果/结论]验通过 LDA 主题演化分析展示公众关注点从不易被聚焦的宏观突发公共卫生事件转移到易被攻击的微观个体的全过程，并揭示了展现立场型、宣泄情绪型以及聚焦事实型三类用户的行为模式，未来可以从增强主流意识形态的引领作用、扩大网络场域不良信息治理范围、有效控制网络暴力传播风险因素三个方面对“不良信息”型网络暴力进行有效管控。

关键词：突发公共卫生事件，“不良信息”型网络暴力，BERT 模型，LDA 模型，演化机制

目 录

1 引言

1.1 研究背景与问题提出

根据我国 2003 年国务院通过的《突发公共卫生事件应急条例》，突发公共卫生事件是指突然发生，造成或者可能造成社会公众健康严重损害的重大传染病疫情、群体性不明原因疾病、重大食物和职业中毒以及其他严重影响公众健康的事件。这一定义不仅涵盖了广泛的健康危机场景，也凸显了此类事件对社会稳定与公众福祉构成的紧迫挑战。随着互联网的蓬勃发展，特别是微博、微信等社交媒体平台的兴起，信息传播的方式发生了翻天覆地的变化。这些平台不仅成为人们获取信息的关键渠道，还赋予用户前所未有的双重角色——既是信息的接收者，也是信息的创造者与传播者。社交媒体的多元化特性，使得信息传播的主体变得更为复杂多样，信息的流通速度也随之加快，往往能在极短的时间内形成舆论的焦点，犹如风暴中的“台风眼”，吸引着全社会的目光。然而，这种即时性和广泛性的信息传播机制，虽然为人们提供了表达观点的自由空间，但同时也为突发公共卫生事件相关的谣言和错误信息提供了温床，增加了信息甄别的难度，容易引发不必要的恐慌和误解。

网络舆情，作为互联网时代的一种特殊现象，其核心在于以网络为媒介，围绕特定事件，汇聚广大网民的情感、态度、意见和观点的表达、传播与互动，以及这些表达后续产生的社会影响力。网络舆情带有鲜明的网民主观色彩，往往未经过传统媒体的严格验证和包装，便以文字、图片、视频等多种形式直接发布于网络。在突发公共卫生事件的背景下，网络舆情往往伴随着强烈的情感倾向，既有对事件正面应对的肯定和支持，也不乏对问题处理不当的批评与质疑，正负向情绪交织，形成了复杂多变的舆论场。

近年来，一系列突发公共卫生事件，如 2002 年的 SARS 疫情、2017 年的 H7N9 禽流感、2018 年的长春疫苗事件，以及 2019 年爆发并持续影响全球的新冠肺炎疫情，无不证明网络舆情在事件发展过程中的重要性。这些事件不仅考验着政府的应急管理能力，也深刻揭示了社交媒体平台在信息传播中的角色与影响。在社交媒体平台的推动下，网络空间成为公众情绪与意见交锋的前沿阵地。人们倾向于迅速识别并加入与自己观点相近的群体，形成不同的舆论阵营。在网络话语权的加持下，群体内的非理性情绪被进一步放大，不同群体间的相互影响与作用，

不仅加速了舆论的传播速度，也增加了舆论内容的复杂性和多样性。值得注意的是，社交媒体平台的群际化特点在促进信息快速传播的同时，也加剧了网络暴力事件和不良信息的产生。不同群体间的意见分歧，往往容易引发攻击性言论和负面情绪的累积，进而演变为网络空间的冲突与对立。这种趋势不仅损害了网络环境的健康生态，也对社会稳定构成了潜在威胁。

突发公共卫生事件的突发性和信息的不确定性等特征易使人们倾向于第一时间去传播而不是去证实信息[1]。这一背景下，伴随重大突发公共卫生事件暴发而产生的海量舆情信息使得以网络谣言为表现形态的舆情危机一触即发，谣言和不实信息快速扩散，社会压力加剧，公众的恐慌与焦虑情绪蔓延，极易滋生网络暴力问题[2]，这一现象的发生，不仅会对个人名誉、隐私造成严重侵害，还会对网络生态环境造成不良影响，甚至危害社会安全。目前学界对于网络暴力的相关研究多聚焦法律层面的政策制订和完善[3]，但网络暴力常以道德指责为起点，并伴随群体的无意识规模化攻击行为[4]，尤其是“不良信息”型网络暴力，一些具有争议性质的舆论，更加容易引发群体的过度关注和讨论，从而形成社会热点，在一定程度上放大了事件本身，造成舆论的漩涡甚至更加恶劣的传播效果。

2021 年 4 月 13 日，日本政府做出了一个在全球范围内引起广泛关注和争议的决策：正式决定将福岛第一核电站的核污水排放入海。这一决策不仅标志着日本在处理福岛核灾难后遗症上的一个新阶段，也引发了全球范围内的环境担忧和地缘政治紧张。2023 年 8 月 24 日中午 12 点（当地时间 13 点），日本按计划强行启动了福岛核污染水排海项目，首批 7788 吨核污染水在随后的两周内，即从 8 月 24 日至 9 月 11 日，被陆续排入大海。根据东京电力公司的规划，整个 2023 年内，将分四次排放总计 3.12 万吨的核污染水，而彻底完成这一排放计划预计将耗时超过 30 年。尽管日本政府声称这些核污染水在排放前已经经过严格的处理，但官方同时承认，处理后的水中仍含有一定量的放射性元素，这些元素无疑将对海洋生态系统以及周边国家的生态环境构成潜在威胁。中国，作为日本的近邻，由于地理位置的邻近性，其海洋环境无疑将受到这一决策的直接或间接影响。因此，中国公众对于此事的反应尤为强烈，网络上涌现了大量表达担忧和反对的声音。百度指数作为衡量互联网用户搜索关注度的工具，通过计算关键词在百度网页搜索中的频次并进行加权处理，能够直观地反映公众对于某一事件的关注度。

通过检索该时段内百度指数（如图 1 所示）的数值不难发现，网民搜索“日本排放核污水”关键词在 8 月 25 日达到峰值，总体指数表现达到 39165，在本研究周期内，该关键词的平均指数达到 4670，由此可见，该事件引发了一定的公众关注，同时在微博平台也引发了一定的热度和讨论度。

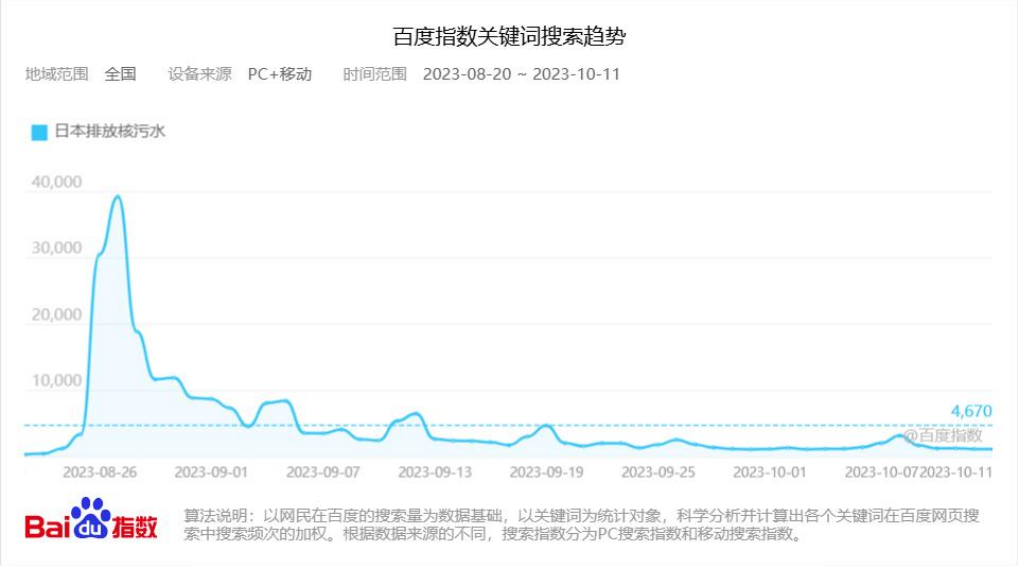

图 1 关于〝日本排放核污水〞关键词搜索趋势 (2023 年 8 月 20 日至 2023 年 10 月 11 日)

与此同时，微博等社交媒体平台也成为公众表达意见和情绪的重要场所。关于日本排放核污水的讨论在微博上持续发酵，引发广泛的关注和热议。这些讨论中，不乏对日本政府决策的质疑和批评，以及对可能受到影响的生态环境和人类健康的深切担忧。在此背景下，本研究聚焦于由这一突发公共卫生事件衍生的“不良信息”型网络暴力，旨在通过深入分析其演化机制，为有效预防和精准控制此类网络暴力提供参考。借助危机生命周期理论，本文将探讨在不同阶段，该舆论事件的演化特征，以及这些特征如何影响公众的认知、情绪和行为。本研究聚焦于 “不良信息” 型网络暴力这一现象，深入剖析其起源及形成机制，细致探究其在潜伏期、爆发期、持续期和衰退期等不同阶段所呈现的演化特征，并依据这些特征制定切实有效的预防与应对策略，旨在肃清网络环境，构建良好的网络生态，进而维护社会的稳定与和谐。通过本研究，期望能够为相关部门和机构提供有益的参考，帮助相关主体在面对类似事件时，能够更加有效地管理和引导网络舆论，减少不良信息的传播，保护公众免受误导和伤害。同时，也希望借此机会呼吁社会各界共同努力，共同维护一个健康、积极、向上的网络环境。

1.2 研究目的

本研究聚焦于深入探讨突发公共卫生事件背景下“不良信息”型网络暴力的演化机制，旨在为有效预防和控制此类网络暴力行为提供科学参考和实用策略。在理论层面，本研究有助于补充突发公共卫生事件与“不良信息”型网络暴力交叉领域的相关研究，推动相关理论的丰富和完善。在实践层面，本研究将助力社会各界，包括政府机构、网络平台、社会组织及广大网民，更加有效地关注和治理“不良信息”型网络暴力，共同营造一个健康、积极的网络生态，净化互联网舆论环境，从而维护社会的和谐稳定。

为了实现上述目标，本研究设定了四个具体的研究方向：

第一，在数据采集与处理方面，本研究采用 Python 爬虫技术，从微博这一具有广泛影响力的社交媒体平台爬取相关数据，形成包含大量用户言论、评论及转发信息的原始数据集。然后，利用这些数据对 BERT（Bidirectional Encoder Representations from Transformers）模型进行训练。BERT 模型作为一种先进的自然语言处理模型，能够高效理解文本语义，识别出蕴含“不良信息”的网络暴力言论。结合危机生命周期理论，本研究进一步分析“不良信息”型网络暴力的演化机制，从危机事件的酝酿、爆发、持续、衰退等各个阶段，详细剖析“不良信息”的波动变化，构建出一个科学且系统的理论模型。这一模型不仅有助于深入理解网络暴力的动态演变过程，也为后续研究提供了坚实的理论指导。

第二，为了更深入地揭示“不良信息”型网络暴力背后的公众情绪与关注度，本研究利用 Python 进行 LDA（Latent Dirichlet Allocation）主题模型分析。LDA 模型能够从大量文本数据中挖掘出潜在的主题结构，有助于理解公众在不同阶段关注的焦点和情绪倾向。结合生命周期理论，本研究对各阶段的主题进行文本分析，揭示公众对于突发公共卫生事件的情绪反馈及关注度变化，为“不良信息”类舆情危机的细分研究提供了坚实的理论支撑。

第三，鉴于不良信息的界定模糊且数量庞大，单纯依赖法律手段难以实现全面有效的控制。因此，本研究从信息传播的角度出发，对“不良信息”型网络暴力的演化机制进行深入解构。通过拓展并深化网络暴力研究的细分领域，本研究尝试探索一种从源头控制网络暴力传播的新路径。具体而言，本研究分析了不良信息的传播路径、扩散机制及其对社会心理的影响，旨在提出针对性的防控策略，以期从根本上减少不良信息的产生和传播，为有效控制“不良信息”型网络暴力

提供切实可行的理论指导。

第四，为了深入探索突发公共卫生事件下用户的评论模式，本研究对爬取的微博文本评论数据进行了 K-means 聚类分析，对用户的评论模式进行聚类，从而进一步了解不同类型的用户的评论行为路径以及扩散动机，为相关部门对网络环境的规范管理提供理论参考，促进网络空间的和谐发展。

综上所述，本研究综合运用文献分析、大数据挖掘、自然语言处理、主题模型分析等方法，系统探究突发公共卫生事件下“不良信息”型网络暴力的演化机制，不仅丰富了相关理论成果，也为实践操作提供了科学依据和实用策略。未来，随着技术的不断进步和社会环境的持续变化，本研究将继续深化和完善，为构建更加健康、和谐的网络空间贡献力量。

1.3 研究意义

本研究以“不良信息”型网络暴力为切入点，对其舆情演化机制展开深入探究，包括不良信息的精准识别、其生命生长周期的详细剖析、不良信息所引发情感的动态演化过程以及网络暴力的管控策略等方面。通过这一系列研究，为相关主体治理网络环境、提升网民个人素质等提供了有益的思路，对于维护社会稳定、促进网络健康发展具有重要的现实意义和深远的历史意义。

1.3.1 理论意义

第一，本研究基于危机生命周期理论，探究了突发公共卫生事件“不良信息”型网络暴力的演化生长过程，并阐述不同阶段的特征。现有的研究多基于不同突发公共卫生事件进行“问题—对策”等进行文本分析，本研究揭示了突发公共卫生事件在扩散期、爆发期、衰退期、波动期的不同表现，同时折射出人们对于公共事件的情绪反应及关注度，从而丰富了突发公共卫生事件网络暴力的相关研究。

第二，本研究利用 Python 爬虫爬取微博数据，通过人工文本标注后利用数据集训练 BERT 模型，再利用训练好的 BERT 模型对数据进行识别和分类，识别出符合研究要求的“不良信息”型网络暴力的文本内容，为研究“不良信息”型网络暴力的演化机制奠定数据基础。最后结合 LDA 主题演化模型分阶段分析“不良信息”型网络暴力的主题词分布特征和词频数量，有利于相关主体挖掘突发公共卫生事件的不同阶段情况特征和表现，有助于帮助相关主体理解舆论的发酵情况。

第三，本研究对现有文献的回顾发现：现有研究主要集中于不良信息的监测方法和治理对策两个层面，对于国内突发公共卫生事件相关研究主要聚焦于网络舆情的成因、传播机制及治理对策等三个层次进行研究，整体而言，对于突发公共卫生事件的网络舆情相关研究较为丰富，但是网络暴力分为“侮辱谩骂”“造谣诽谤”“侵犯隐私”“违法和不良信息”四种类型，关于网络暴力细分类型的相关研究较为有限，因此，本研究补充了突发公共卫生事件一定的相关研究缺失的现状，并提供了一定的理论支撑。

第四，本研究关注社交媒体平台的用户对于突发公共卫生事件的敏感度和讨论情况，探究其中“不良信息”型网络暴力的演化机制和演化趋势，并基于此提出“不良信息”型网络暴力的分阶段治理政策，不仅丰富了舆情发酵的不同阶段表现，同时也为相关主体如何应对拓宽了思路。

1.3.2 实践意义

第一，本研究提出的分阶段治理政策，有助于相关主体更加精准地把握网络暴力的演化趋势，制定有效的治理策略。政府可以根据不同阶段的特征，采取针对性的措施，如加强信息监管、打击谣言传播、引导舆论走向等，从而有效遏制网络暴力的蔓延，维护社会稳定和公共利益。

第二，本研究提供的数据处理方法和模型训练技术，有助于其提高不良信息的识别和过滤能力。网络平台可以利用这些技术，建立更加完善的信息审核机制，及时发现并处理不良信息，维护良好的网络环境。同时，网络平台还可以根据研究提出的舆情演化机制，优化信息推荐算法，减少不良信息的传播机会。

第三，本研究通过揭示网络暴力的危害性和演化机制，有助于提升网民的网络素养和自我保护意识。网民可以更加理性地看待网络信息，避免被虚假信息、谣言等误导。同时，网民还可以学会如何举报不良信息、维护自己的合法权益，共同营造一个健康、积极、向上的网络生态。

第四，本研究对于网络暴力的深入探究，有助于推动网络空间法治建设。政府可以根据研究提出的网络暴力特征和演化机制，制定更加完善的法律法规，明确网络暴力的界定和处罚标准。同时，政府部门还可以加强与网络平台的合作，共同打击网络暴力行为，维护网络空间的秩序和安全。

1.4 研究内容与思路

首先，运用 Python 爬虫从微博爬取“日料店”、“中国日料店会大批量倒闭吗”等话题下相关微博正文及评论，经数据清洗、人工标注后形成数据集，去除无评论内容及广告评论，用数据集训练 BERT 模型，再利用训练好的模型对微博评论数据进行预测，准确识别网络暴力信息；其次，结合危机生命周期理论对“不良信息”型网络暴力进行生命周期划分，通过对各阶段的特征分析，深入探讨其在不同阶段的演化规律；同时，运用 LDA 主题模型对各阶段的“不良信息”型网络暴力文本进行主题分析，识别各阶段的主题词分布特征和词频数量，揭示网络暴力信息在不同阶段的主题演化趋势。通过分析各阶段的主题分布，进一步理解公众关注点的变化以及网络暴力信息的内容特征，为制定针对性的治理策略提供依据。除此以外，采用 K-means 聚类算法对用户的评论模式进行聚类分析，通过对不同类型用户的行为模式进行深入分析，了解用户在网络暴力事件中的参与程度、情感倾向和关注焦点，为相关主体对网络环境的规范管理提供理论参考，促进网络空间的和谐发展。最后，提出“不良信息”型网络暴力的分阶段治理对策。本研究的具体思路如图 2 所示。

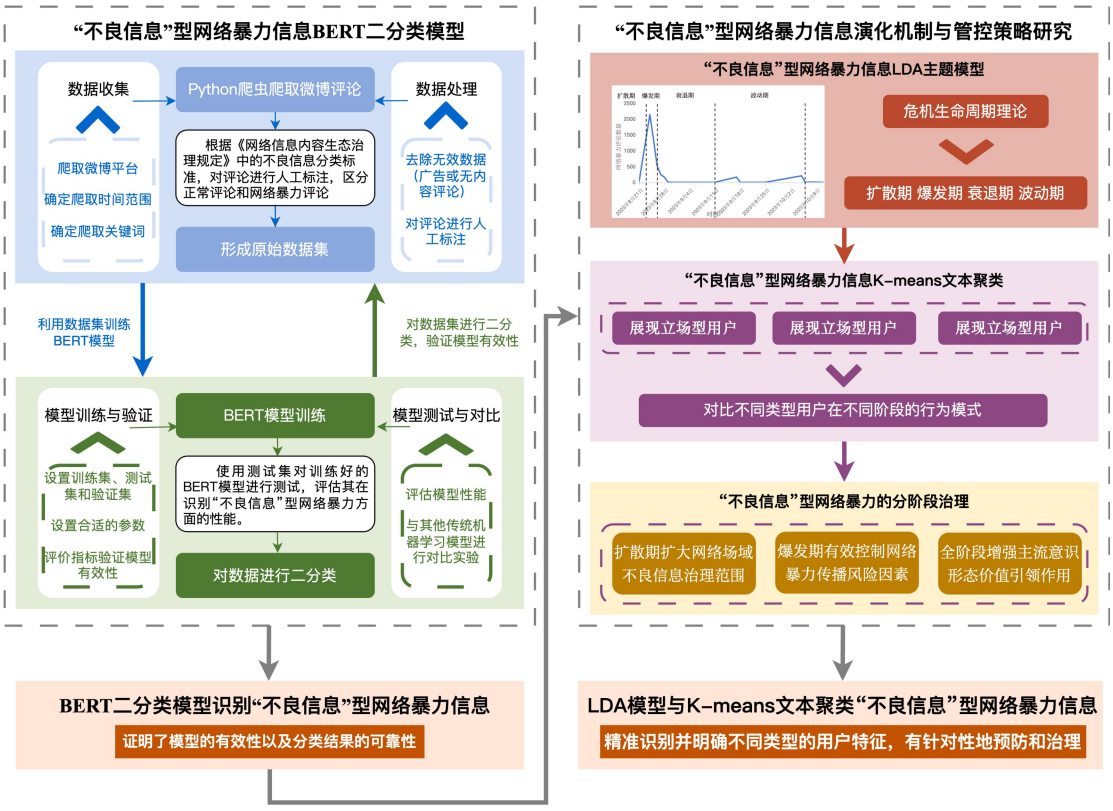

图 2 研究思路

1.5 研究方法

本文通过三种方法对突发公共卫生事件背景下“不良信息”型网络暴力的舆情演化机制进行研究：

1.5.1 BERT 模型

在微博平台上查找关键词相关的博文及评论，并用 Python 爬取相关数据，对数据进行清洗后形成数据集，对 BERT 模型进行预训练，然后用训练好的模型对数据集进行二分类，识别数据集中的“不良信息”型网络暴力信息。

1.5.2 LDA 主题模型

按时间顺序画博文数量的时序分布图，结合生命周期理论将“不良信息”型网络暴力划分为不同阶段，对每一个阶段分别用 Python 进行 LDA 主题分析，根据困惑度（Perplexity）和一致性得分（Coherence Score）确定每一阶段主题数，对每一阶段进行主题分析。

1.5.3 K-means 文本聚类

本文用肘部法确定最优聚类树，用 Python 进行 K-means 文本聚类，对用户行为模式进行分类，对每一类用户的“不良信息”型网络暴力的舆情演化机制进行分析并对不不同用户之间的区别。

1.6 创新点

本研究聚焦于突发公共卫生事件背景下“不良信息”型网络暴力的舆情演化机制，通过综合运用多种研究方法和技术手段，揭示了该类型网络暴力的演化规律和特征，为相关主体提供了有效的治理策略和提升网民个人素质的思路。本研究的主要创新点体现在以下四个方面：

第一，本研究在理论框架的构建上实现了创新，将危机生命周期理论与网络暴力研究相结合，提出了适用于突发公共卫生事件背景下“不良信息”型网络暴力研究的理论框架。该框架不仅涵盖了网络暴力的演化过程，还深入分析了不同阶段的特征、影响因素以及舆情演化机制。在此基础上，本研究进一步构建了基于 BERT 模型和 LDA 主题演化模型的舆情分析模型，用于识别和分类“不良信息”型网络暴力文本，并分阶段分析网络暴力的主题词分布特征和词频数量。这一理论框架与模型的构建，为深入研究网络暴力的演化机制提供了新的视角和方法。

第二，在数据处理与分析技术方面，本研究实现了多项创新。首先，本研究

利用 Python 爬虫技术从微博等社交媒体平台上爬取了大量数据，并通过人工文本标注和数据集训练，构建了高质量的 BERT 模型，实现了对“不良信息”型网络暴力文本的自动识别和分类。这一技术不仅提高了数据处理的效率和准确性，还为后续研究提供了可靠的数据基础。其次，本研究结合 LDA 主题演化模型，分阶段分析了网络暴力的主题词分布特征和词频数量，揭示了网络暴力的舆情演化规律和趋势。这一技术的运用，为深入挖掘网络暴力的舆情信息、理解舆论的发酵情况提供了有力的支持。

第三，在治理策略方面，本研究提出了分阶段治理“不良信息”型网络暴力的新思路。根据网络暴力的演化规律和特征，本研究将网络暴力的治理过程分为预警阶段、应对阶段和恢复阶段，并针对不同阶段的特点提出了相应的治理策略。在预警阶段，本研究建议通过加强信息监测、建立预警机制等方式，及时发现并预防网络暴力的发生。在应对阶段，本研究提出了加强信息监管、打击谣言传播、引导舆论走向等具体措施，以有效遏制网络暴力的蔓延。在恢复阶段，本研究则强调了加强网络素养教育、提升网民自我保护意识等长期治理策略的重要性。这一分阶段治理策略的创新，为相关主体提供了更加精准、有效的治理思路和方法。

第四，本研究在研究方法上实现了跨学科融合与创新。本研究不仅运用了传播学、心理学、社会学等学科的理论和方法，还结合了计算机科学、数据科学等技术手段，形成了跨学科的研究范式。这一跨学科研究方法的融合与创新，不仅丰富了网络暴力研究的理论和方法体系，也为深入理解网络暴力的本质和特征提供了新的视角和思路。例如，本研究通过运用计算机科学中的数据挖掘和文本分析技术，揭示了网络暴力的舆情演化规律和趋势；通过运用心理学中的情绪分析和认知理论，探讨了网络暴力对网民心理和行为的影响；通过运用社会学中的社会网络分析和群体动力学理论，分析了网络暴力的社会结构和传播机制。这一跨学科研究方法的融合与创新，为深入研究网络暴力提供了更加全面、深入的视角和方法。

2 文献回顾与理论基础

2.1 “不良信息”型网络暴力

网络暴力是指通过言语攻击等网络舆论手段，致使当事人人格权益受损的一系列网络失范行为[5]，其具有一定的重复性（例如单个信息可能被发送到数百个不同的个体），并且发生在关系权利不平衡的个体之间[6]。近年来网络暴力事件频发，对个人乃至社会造成了严重的不良影响[7]。2023 年 7 月 7 日，中央网信办发布《网络暴力信息治理规定（征求意见稿）》，将网络暴力信息定义为通过网络对个人集中发布的“侮辱谩骂”“造谣诽谤”“侵犯隐私”“违法和不良信息”四类信息，“不良信息”属于网络暴力其中一种。

“不良信息”由于其界定难度较大，介于合法与违法、道德与不道德之间，又被称为“灰色信息”[8]，庞大的数量基础以及治理的高难度加剧了此类信息对社会的负面影响。2020 年国家互联网信息办公室颁布的《网络信息内容生态治理规定》[9]第二章第七条将不良信息分为以下九类：（1）使用夸张标题，内容与标题严重不符的；（2）炒作绯闻、丑闻、劣迹等的；（3）不当评述自然灾害、重大事故等灾难的；（4）带有性暗示、性挑逗等易使人产生性联想的；（5）展现血腥、惊悚、残忍等致人身心不适的；（6）煽动人群歧视、地域歧视等的；（7）宣扬低俗、庸俗、媚俗内容的；（8）可能引发未成年人模仿不安全行为和违反社会公德行为、诱导未成年人不良嗜好等的；（9）其他对网络生态造成不良影响的内容。本研究中的不良信息即以此为标准。

目前对于不良信息的研究主要集中于不良信息的监测方法和治理对策两个方面。不良信息监测方法方面，主要结合数据挖掘方法[10]、人工神经网络[11]、情感分析[12]、自然语言处理[13]等方法建立不良信息监测模型；网络不良信息治理对策方面，多从法律立法角度提出通过完善立法的方式治理网络不良信息，例如，李铭轩[14]认为目前网络信息内容治理主要从信息生成和信息传播角度介入，并指出未来应该根据不同技术特点等进一步完善私法领域规则。张凌寒[15]从立法执法、信息管理制度、施暴主体惩戒三个角度详细探讨了如何构建基于场域特性的网络暴力治理制度。林爱珺等[16]提出可以从治理主体权利、不良信息鉴定、信息技术治理三方面促进多元主体参与网络治理。

本研究聚焦“不良信息”型网络暴力，通过收集微博评论数据并进行清洗形

成数据集，再根据《网络信息内容生态治理规定》中规定的九类不良信息，对收集到的训练集数据中的不良信息进行人工标注，得到训练后的 BERT 模型，从而识别出测试集数据中的不良信息，为后续研究“不良信息”型网络暴力的演化机制奠定基础。

2.2 突发公共卫生事件背景下的网络暴力

2003 年国务院颁布的《突发公共卫生事件应急条例》将突发公共卫生事件定义为：突然发生，造成或者可能造成社会公众健康严重损害的重大传染病疫情、群体性不明原因疾病、重大食物和职业中毒以及其他严重影响公众健康的事件[17]。突发公共卫生事件具有不确定性、复杂性、波动性的特征[18]，公众在面对突发公共卫生事件时，易出现恐慌混乱，产生惧怕、愤怒、厌恶等情绪[19]，因而，网络舆情是突发公共卫生事件的关注焦点之一。网络暴力是网络舆情的一种负面体现，是一种特殊类型的网络舆情。网络暴力不仅对个体的公众权力造成损害，还扰乱网络秩序、破坏网络生态，一旦发生，可能产生“社会性死亡”、精神失常、自杀等严重后果，对公众安全感造成极大危害[20]。因此，突发公共卫生事件网络舆情的健康发展离不开对网络暴力的有效治理。目前，国内突发公共卫生事件相关研究大部分聚焦于突发公共卫生事件背景下的网络舆情，主要从网络舆情成因、传播机理以及治理对策三方面进行研究。

舆情成因方面，现有研究从宏观、微观等不同维度进行分析，如张筱荣[21]认为社会矛盾和现实风险是网络舆情产生的根本动因，事件走向和主体行为是直接动因，新媒体平台的快速发展产生了放大效应进一步推动了网络舆情的爆发。也有从信息对象角度切入分析舆情成因，如胡象明等[22]研究认为网络舆情高影响力生成的核心条件是信息人和信息技术；Zhao 等[23]认为“旁观者”角色在网络暴力中起到了不可忽视的作用；吕鲲等[24]研究表明权威主体发布的、涉及范围较广的信息更容易被广泛讨论并传播。传播机理方面，研究人员大多采用定量研究的方式，结合模型和大数据构建研究框架，如曾子明[25]、马腾等[26]利用 BERT-BiLSTM-Attention 模型、LDA 模型等进行研究，研究表明舆情主题与情感演化与公共卫生事件演化关联，且各阶段主题内容侧重点不同，情感倾向分布演化趋势波动也较大；Raskhodchikov[27]结合神经网络和语言学分析方法提出了一种检测社交网络用户文本交流中的社会压力的模型，认为可以通过确定社会压力来

帮助相关部门及时调整计划方案，改善舆论压力。也有学者结合区块链技术对网络舆情进行预测分析，对网络舆情进行溯源[28]、监管[29]以及防控[30]等。治理对策方面，研究者主要从宏观和微观的角度切入，如朱广生等[31]从治理思维、治理原则、治理机制三个方面，提出从宏观到微观突发公共卫生事件的治理对策；也有从政府视角出发，为政府治理网络舆情提供思路，如温海涤[32]等认为需要重视网络意见领袖的作用，梳理引导正向信息的传播；Ferguson 等[33]研究表明负责公共卫生信息的政府部门并没有将信息传递给大多数公众，并指出相关部门应该努力提高公共卫生信息相关文件的可读性。

综上所述，当前关于突发公共卫生事件网络舆情的研究较为丰富，但对于网络暴力的研究还有待细化，对“不良信息”型网络暴力的研究更有待开展。本研究聚焦突发公共卫生事件下的“不良信息”型网络暴力，构建突发公共卫生事件背景下“不良信息”型网络暴力演化模型，并深入分析不同阶段的主题演化特征。

2.3 危机生命周期理论

危机生命周期理论最早由 Fink[34]提出，该理论认为危机从诞生、成长、成熟到结束的过程中具有不同的生命特征，并将危机划分为酝酿期、爆发期、扩散期、处理期、处理结果以及后遗症期等五个阶段。后续很多研究对其进行了拓展，现阶段的研究多以四阶段划分为主，根据实际研究情境对危机生命周期的五个阶段进行调整和完善，如王旭[35]将突发事件网络舆情分为萌芽期、生长期、成熟期、衰退期四个阶段，韩小伟等[36]将突发公共卫生事件网络舆情分为潜伏期、爆发期、扩散期、平复期四个阶段，张文杰等[37]将公共事件分为酝酿期、爆发期、持续期和消减期四个阶段。信息时代，信息传播速度大幅提高，舆情热点转移加快，信息更新周期变短，热点信息数据以天甚至小时为单位在短期内指数级增长，因此，从事件发生到讨论的爆发期之间时间很短，即整个生命周期内的生长期并不明显[38]，又由于突发公共卫生事件的特殊性，相较于其他事件更易引起人们的关注和讨论，进一步缩短了信息的生长期，基于此，本研究将突发公共卫生事件中的“不良信息”型网络暴力划分为扩散期、爆发期、衰退期、波动期四个阶段。其中，扩散期是指从突发公共卫生事件发生后到快速传播爆发的时期，此时相关讨论较少，还未引起广泛关注。爆发期指相关博文和评论信息数量指数级增长的阶段，此时网民关注度急剧上升，直至达到峰值。衰退期指相关博文和评论数量逐渐下

降，关注度也渐渐降低，此时热点已经转移。波动期指衰退期后，由于突发公共卫生事件发生后带来的其他影响，又引起了网民的小范围讨论。

2.4 LDA 主题模型

LDA（Latent Dirichlet Allocation）主题模型的发展历史可以追溯到自然语言处理（NLP）和文本挖掘领域的不断探索和进步。LDA 模型作为概率主题建模方法的代表，其诞生和发展标志着文本处理从传统的基于规则的方法逐渐过渡到基于统计和机器学习的范式。随着信息技术的飞速发展，文本数据呈现出爆炸式的增长。如何从这些海量的文本数据中提取有价值的信息，成为自然语言处理 (NLP) 领域的一个重要研究方向。主题模型作为文本挖掘中最为关键的技术之一，因其出色的降维能力和灵活的易扩展性，逐渐成为了研究的热点。其中，LDA 主题模型以其独特的概率生成方式，为文本数据的处理和分析开辟了新的道路。

目前对 LDA 等主题模型研究关注的领域主要包括利用主题模型进行大数据分析和挖掘以及在线评论分析两个方面。利用主题模型进行大数据分析和挖掘方面，研究如何改进传统的主题模型（如 LDA）以适应大数据和短文本的特点，包括利用词嵌入、神经网络等技术增强模型的表示能力。探索主题模型在社交媒体、新闻、医疗、图书馆与信息科学等多个领域的应用，如分析社交媒体话题的演变、新闻报道的主题分类、临床文本的分析等。也有部分研究分析主题随时间的演化过程，如通过时间序列分析、动态主题模型等方法研究主题的变化趋势，探索如何应用主题演化分析来揭示某一领域（如科研、政策、社会事件等）的发展趋势和动态变化。其中，Jang 等[1]探讨了如何利用专家对未来科技的看法来识别新兴技术。与以往基于过去和现在数据的研究不同，该方法收集了未来导向的专家意见，并使用主题建模和模糊聚类分析以发现未来的关注点和技术。Schillebeeckx 等[2]探讨在动态知识环境中创新活动与财务绩效之间的关系，并使用主题建模方法对创新活动进行映射和分析，通过对一家跨国公司的专利数据进行主题建模，将专利文本划分为不同的主题，并根据主题与公司财务指标的相关性分析了不同主题对公司财务绩效的影响。Altamirano 等[3]提出了两种不同的方法来获取课程聚类：基于文本教科书的方法和基于 BERT 模型的方法。在基于文本教科书的方法中，研究人员使用 LDA 模型来训练并提取与课程相关的主题，并将每个课程表示为一个时间序列，其中包含主题的概率分布。在基于 BERT 模

型的方法中，研究人员首先使用 BERT 模型对教师讲义进行编码，然后使用 UMAP 技术将其降维到三维空间中，最后使用高斯混合模型（GMM）来识别和描述教师讲义中的主题。

主题建模于在线评论中的应用的实验研究方面，部分研究使用主题建模技术（如 LDA、Bertopic 等）从在线评论中提取出核心主题或话题，从而了解消费者对产品或服务的关注点、意见和情感态度。Liu 等[4]通过对第三方平台和自有 B2C 在线药店的超过 13 万条非处方药评论数据进行 LDA 主题模型分析，发现物流、药品价格、客户服务和药品效果等 12 个影响消费者满意度的因素，并通过 SnowNLP 情感分析工具对评论进行了情感分析。Shang 等[5]探讨徒步旅行者对于长城旅游体验的看法和感受，并通过话题建模技术对这些看法进行分析。作者采用了大型评论数据集，并使用 LDA 技术构建了游客们讨论的话题列表。Saura 等[6]讨论内容营销对在线用户行为的影响，并提出了一个基于数据驱动的预测分析方法。该方法包括三个步骤：首先使用机器学习进行情感分析，以提高对用户生成内容（如 Twitter 上的推文）的情感分类准确性；其次使用潜在狄利克雷分配算法将数据库分为主题；最后使用 Python 编程语言进行文本分析。

总体而言，国外学者主要研究 LDA 等主题模型用于大数据分析与挖掘以及在线评论分析中的应用。国内 LDA 主题模型研究主要涉及政策量化研究，文本主题演化分析、用户需求与市场分析等。本研究基于微博评论中的“不良信息”型网络暴力信息新数据建立 LDA 主题模型，对各阶段主题词进行识别统计，分析不同阶段“不良信息”型网络暴力信息主题演化趋势。

3 “不良信息”型网络暴力信息 BERT 二分类模型

3.1 方法介绍

BERT（Bidirectional Encoder Representation from Transformers）是由 GoogleAI 研究院提出的一种预训练模型，原理是通过双向预训练捕获句子的上下文信息。BERT 模型使用了 Vaswani 等[39]提出的多层 Transformer 结构，即一种基于 Attention 机制的可以捕获句子中上下文信息的深度学习模型。本研究采用 Python 爬虫从微博爬取数据并形成数据集，利用数据集对 BERT 模型进行训练，再通过训练好的 BERT 模型识别数据中符合网络暴力的“不良信息”型。BERT 训练过程如图 3 所示，将人工标注过的微博评论数据作为训练集和测试集输入，根据损失函数和评价指标对模型进行训练和评价优化，微调时用到的交叉熵损失函数公式如下：

$$Loss = \frac{1}{N}\sum_{i} -\left[y_i \times \log(p_i) + (1-y_i) \times \log(1-p_i)\right] \quad (1)$$

其中，y_i 表示人工标注的标签，正常评论标签为 0（负类），网络暴力评论标签为 1（正类）；p_i 表示样本 i 预测为正的频率。

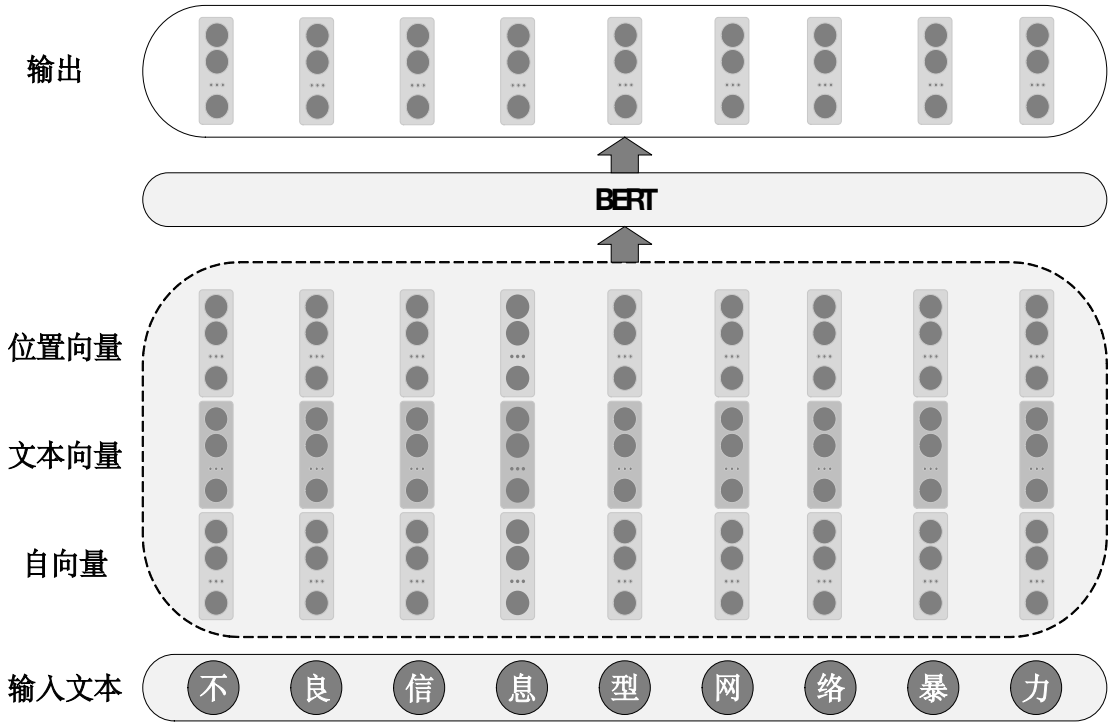

图 3 BERT 流程图

3.2 数据采集与预处理

根据《突发公共卫生事件应急条例》，排放核污水属于突发公共卫生事件，并且国内公众对于此事件反应较为激烈。日料店作为与日本文化及食材紧密相关的行业，在此事件背景下，其发展前景受到公众质疑，相关话题在微博等社交媒体平台上引发了热烈讨论，成为网络舆情的焦点之一。这一事件不仅涉及生态环境、食品安全等公共利益问题，还触及了公众的民族情感与爱国情怀，极易引发情绪化表达和网络暴力行为，具有典型性和代表性，为研究 “不良信息” 型网络暴力提供了丰富的数据基础。在日料店相关话题的讨论中，出现了大量 “不良信息” 型网络暴力言论，如对日料店店主的人身攻击、地域歧视、恶意揣测等。这些言论不仅对个人名誉和隐私造成侵害，还对网络生态环境产生不良影响，甚至可能引发社会不稳定因素。本研究选择该事件为研究情境，利用 Python 爬虫爬取“日料店”“中国日料店会大批量倒闭吗”等话题下的用户微博正文及评论共 11540 条，爬取时间为 2023 年 8 月 20 日至 2023 年 10 月 11 日，爬取的数据包括用户名称、评论内容、发布时间，去除无评论内容的评论和广告评论后，共得到 9833 条数据作为数据集。其次，按照《网络信息内容生态治理规定》中不良信息的九个分类对评论进行人工标注，使用布尔值区分，其中正常评论标签为 0，网络暴力评论标签为 1，人工标注后得到数据集中网络暴力评论数据 4077 条，取 80%数据作为训练集和验证集，剩下的数据作为测试集，训练集和验证集数量比为 8：2，数据样本示例见表 1，数据集构成见表 2。

表 1 数据样本示例

微博评论	评论类型
吃高端日料是钓到中国女大学生的必要手段之一	网络暴力评论（1）
在中国开日料的 90%都骗，常去吃日料的 80%是装	网络暴力评论（1）
不得不说陕西的面是真的好吃，又香又筋道。	正常评论（0）

表 2 数据集构成

类型	训练集和验证集	测试集	总计
正常评论	4006	982	4988
网络暴力评论	4077	768	4845
总计	8083	1750	9833

3.3 实验过程

3.3.1 实验环境与参数设置

本研究采用的操作系统为 Windows10，深度学习框架为 PyTorch，Python 版本为 3.9.7，GPU 为 NVIDIA GeForce MX330，选择“BERT-base-Chinese”模块，设置 Transformer 层数为 12，优化器采用 adam，模型参数设置见表 3。

表 3 模型参数设置表

参数名称	参数说明	值
loss	交叉熵损失函数	CrossEntropyLoss
max_length	最大文本长度	50
batch_size	迭代选取的数据量大小	64
lr	学习率	2e-5
epoch_num	迭代次数	10
Hidden_dim	隐藏层向量维度	768

3.3.2 评价指标

为了评估模型在不良信息分类模型中的效果，本研究使用 Accuracy（准确率）、Precision（精确率）、Recall（召回率）和 F1 值四个指标作为评价指标[40]。指标的定义和计算公式如下：

Accuracy（准确率）：预测正确的结果占总样本的百分比。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Precision（精确率）：又叫查准率，指所有被预测为正的样本中实际为正的样本的概率，精确率越高，模型对于负样本区分能力越强。

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Recall（召回率）：又叫查全率，指实际为正的样本中被预测为正样本的概率，召回率越高，模型对于正样本的区分能力越强。

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

F1 指数：权衡 Precision（精确率）和 Recall（召回率），F1 越接近 1 模型质量越好。

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \tag{7}$$

其中，TP（True-Positive）为正确预测属于网络暴力的不良信息的样本数量，TN（True-Negative）为正确预测不属于网络暴力的不良信息的样本数量，FP（False-Positive）为错误预测属于网络暴力的不良信息的样本数量，FN

（False-Negative）为错误预测不属于网络暴力的不良信息的样本数量。

3.4 实验结果

经过 10 次迭代，训练模型得到 Accuracy（准确率）为 0.86，Precision（精确率）为 0.81，Recall（召回率）为 0.91，FI 指数为 0.85，利用训练好的模型对测试集进行预测，得出测试集中正常评论有 982 个，网络暴力评论有 768 个。准确率和迭代次数如图 4 所示，训练集准确率随着迭代次数增加而上升，最终训练集准确率为 0.98；测试集准确率随迭代次数增加略有波动，保持在 0.85 左右，最终测试集准确率为 0.86，说明 BERT 模型在识别符合网络暴力的不良信息识别上有较好的表现，证明了模型的有效性以及分类结果的可靠性。

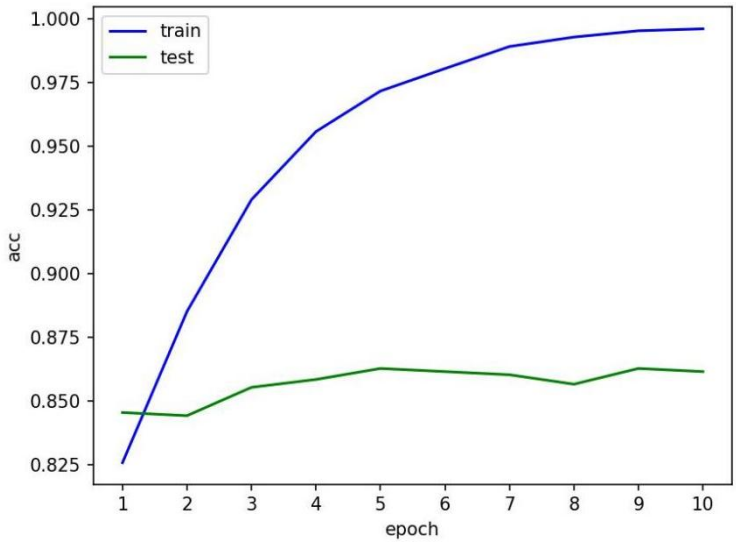

图 4 迭代次数与准确率

3.5 对比实验设置

为了验证 BERT 模型的分类性能，本研究在相同实验环境下做了与传统机器学习模型的分类效果对比，包括 XGBoost、随机森林、逻辑回归、支持向量机、朴素贝叶斯等。对比实验结果见表 4。

表 4 传统机器学习二分类结果指标对比

模型	Accuracy	Precision	Recall	FI
XGBoost	0.83	0.84	0.83	0.83
随机森林	0.82	0.85	0.82	0.81
逻辑回归	0.86	0.87	0.86	0.86
支持向量机	0.86	0.87	0.86	0.86

朴素贝叶斯	0.74	0.74	0.74	0.74

从上表可以看出，和传统机器学习分类方法相比，BERT 在 Accuracy、Recall 等指标的数据表现均为最优，BERT 相较于 XGBoost 在“不良信息”型网络暴力信息数据集上提升了 3%，相较于随机森林提升了 4%，相较于朴素贝叶斯提升了 14%，相较于逻辑回归和支持向量机，BERT 的 Recall 指标提升了 5%，验证了 BERT 在“不良信息”型网络暴力信息文本二分类实验中的有效性。

4 “不良信息”型网络暴力信息 LDA 主题模型

4.1 方法介绍

LDA（Latent Dirichlet Allocation，隐含狄利克雷分布）是由 Bleietal[41]在 2003 年提出的生成式主题模型，又被称为三层贝叶斯概率模型，包含文档（d）、主题（z）、词（w）三层结构，能够有效对文本进行建模，通过挖掘数据集中的潜在主题,分析数据集的核心主题及其相关特征词。LDA 采用词袋特征(bag-of-word feature）代表文档，将每一篇文档视为一个词频向量，从而将文本信息转化为易于建模的数字信息。由于 LDA 是无监督学习算法，不需要人工标注数据集，因此，广泛应用于文本主题识别、文本分类等研究中。本研究用 LDA 模型分析“不良信息”型网络暴力各阶段的主题分布以及同一阶段的主题分布，进而分析其主题演化过程。利用困惑度（Perplexity）和一致性得分（Coherence Score）确定主题数，困惑度与一致性得分的计算公式如下：

$$Perplexity(D) = erp\left\{-\frac{\sum_{d=1}^{M}\log(p(W_d))}{\sum_{d=0}^{M}N_d}\right\} \tag{2}$$

$$C(Z^{\cdot}, S^{z}) = \sum_{n=1}^{N}\sum_{t-1}^{n-1}\log\frac{D_2(W_n^2, W_1^2)+1}{D_1(W)} \tag{3}$$

其中，M 表示文档数量，D 表示文档中全部词的集合，W_d 表示文档中的词，$P(W_d)$ 表示词频，N_d 表示词的数量，$D_1(W)$ 是单词 W 的词频，$D_2(W_1, W_2)$ 是单词 W_1、W_2 的共现文档频率，为了避免模型过拟合，本研究选择困惑度最低且一致性得分最高的参数值作为最终参数。

4.2 “不良信息”型网络暴力生命周期划分

突发公共卫生事件下的“不良信息”型网络暴力评论量时间分布特征如图 5 所示。从核污水排放消息公布后到人们开始关注日料店并未有明显的生长期，而是快速进入扩散期，信息传播速度急剧上升，并且人们对日料店产生了剧烈的负向情感反馈，“不良信息”型网络暴力数量短时间内指数级上涨，进入爆发期，并于 8 月 25 日达到峰值。8 月 25 日后讨论热度逐渐冷却，“不良信息”型网络

暴力数量有明显衰退迹象，进入衰退期，并于 28 日达到谷底，但是仍保有一定讨论度，进入波动期，在 9 月和 10 月数据量产生了一定波动，于 9 月 17 日和 10 月 5 日上升到了小高峰。

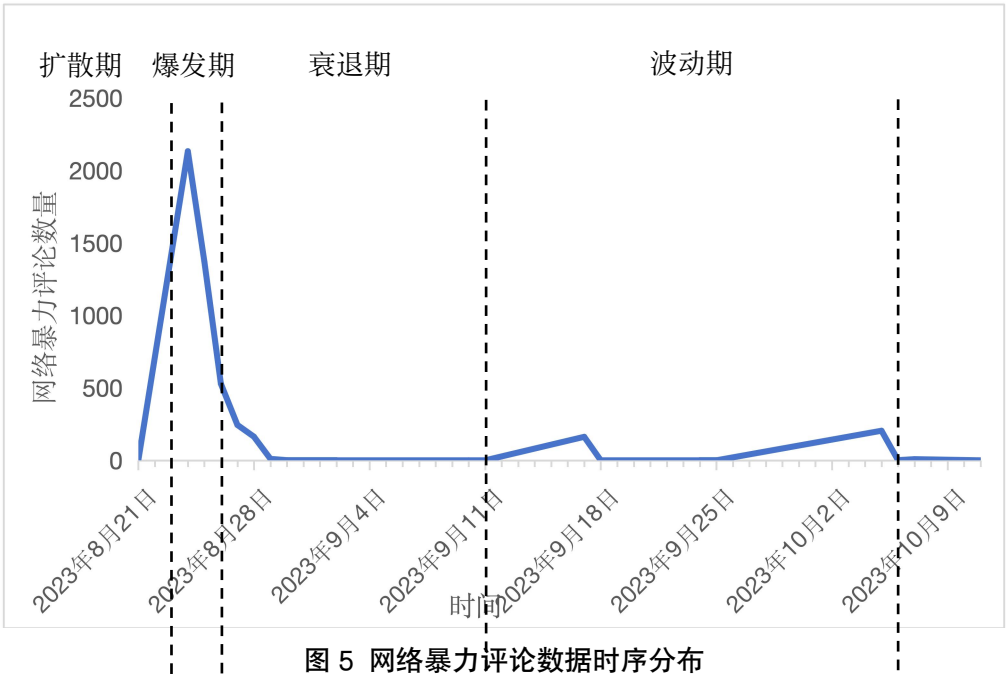

图 5 网络暴力评论数据时序分布

利用 Python 进行 LDA 模型分析，得到“不良信息”型网络暴力生命周期各阶段的主题数(见图 6-图 9)及对应的主题词(见表 5)。根据利用困惑度(Perplexity)和一致性得分（Coherence Score）确定主题数，可知扩散期主题数为 2，主要涉及“核污水排放”，对应的主题包括核污水、生物等；以及“民族情绪”，对应的主题包括日本人、素质等。在扩散期，排放核污水的消息刚刚放出，公众关注点主要集中于排放行为本身，从道德层面对该行为进行谴责，表达对生态环境和海洋生物的担忧，讨论度较低但扩散速度极快。

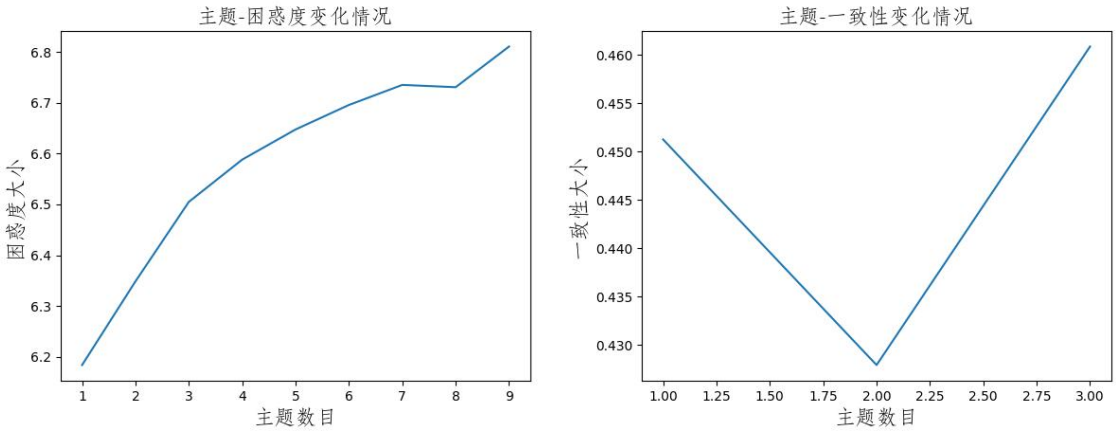

图 6 扩散期主题数-困惑度/一致性变化情况

爆发期主题数为 2，主要涉及“排污影响日料店”，对应的主题包括福岛、日本料理等；以及“日料店食材”，对应的主题包括进口、三文鱼等。爆发期，随着日料店店主提出日料店亏损严重，甚至有日料店因为排放核污水事件倒闭等，公众将目光从环境污染转移到国内日料店发展问题上，公众开始担忧食材安全，加剧了对该行为的谴责，并且开始出现泛意识形态化解读现象，部分极端情绪的公众开始将矛头指向日料店店主，发表言语辱骂、阴阳怪气等偏激评论。也有因为日料店价格过于高昂而开始地域攻击、引起地域歧视的现象，从关键词中可以看到此阶段“江苏”“上海”等地名频繁出现。

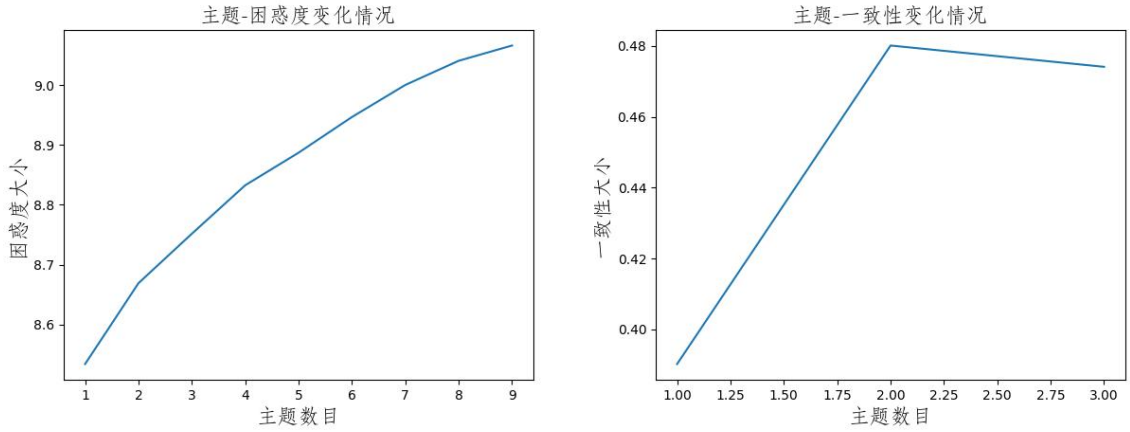

图 7 爆发期主题数-困惑度/一致性变化情况

衰退期主题数为 2，主要涉及“民族情怀”，对应的主题包括情怀、民族等；以及“爱国情怀”，对应的主题包括人格、爱国等。衰退期随着公众情绪逐渐平复，以及其他热点事件吸引注意力，因此，相关评论量逐渐减少，但是泛意识形态化解读现象仍然存在，部分网友将“开日料店”与“不爱国”联系起来，发表偏激评论，并进行人身攻击。也有公众理性指出排放核污水对经济的影响，并提出对日料店店主的网暴可能会导致部分从业人员失业等问题。除此以外，有公众对比此次事件和 2012 年“反日游行”事件，认为这两次事件中公众反应有较大差别，指出此次并没有“怒砸日产车”等情况发生，相比之前国民素质有了一定提升。

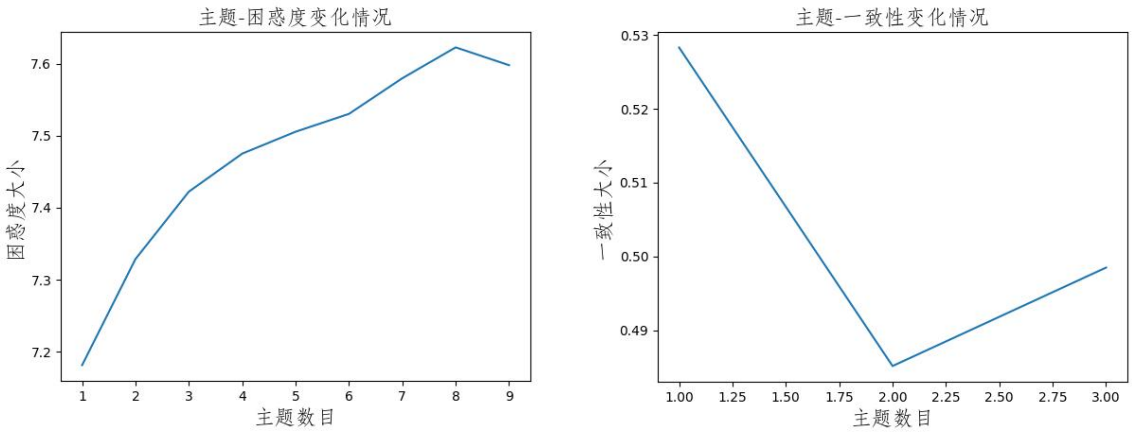

图 8 衰退期主题数-困惑度/一致性变化情况

波动期主题数为 2，主要涉及“日料店未来”，对应的主题包括吃日料、消费等；以及“反对进口食材”，对应的主题包括反对、邋遢等。波动期公众对于该事件本身的讨论热情慢慢消褪，主要讨论日料店的未来发展以及表达对进口食材安全的担忧。但是由于突发公共卫生事件本身的复杂性、不确定性等特性，易引起人们的恐慌情绪，人们对于该事件的担忧一直存在，因此，当有人发布相关言论时，仍会引起小范围的讨论，数据量在小幅波动。

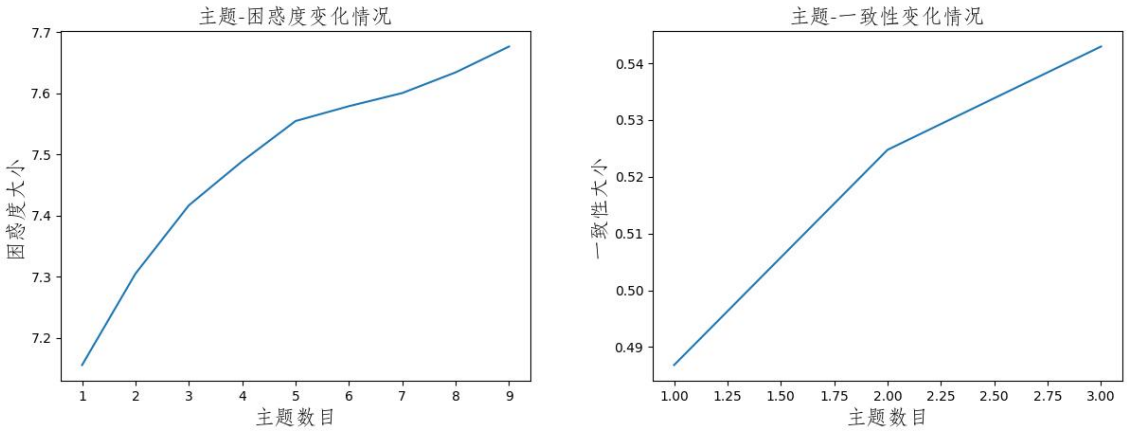

图 9 波动期主题数-困惑度/一致性变化情况

表 5 各阶段主题分布

阶段	主题	主题词及权重
扩散期	主题 1：核污水排放	核污水（0.021）、干净（0.018）、日本（0.018）、处理（0.013）、排放（0.012）、叶文洁（0.008）、人类（0.008）、生物（0.006）、大义（0.006）、污染（0.006）
	主题 2：民族情绪	核污水（0.057）、灭霸（0.022）、日本（0.020）、核污染（0.011）、日本人（0.009）、素质（0.008）、夏天（0.007）、质疑（0.007）、核废料（0.006）、喜欢（0.006）

爆发期	主题 1：排污影响日料店	中国（0.010）、日料店（0.006）、福岛（0.004）、料理（0.003）、氢弹（0.003）、江苏（0.003）、喜欢（0.003）、日本料理（0.003）、玫瑰花（0.003）、核污染（0.002）
	主题 2：日料店食材	日本（0.030）、日料店（0.017）、日料（0.016）、进口（0.013）、海鲜（0.010）、上海（0.006）、食材（0.005）、国内（0.005）、三文鱼（0.004）、吃日料（0.004）、日本人（0.004）
衰退期	主题 1：民族情怀	日料店（0.017）、情怀（0.014）、民族（0.013）、进口（0.010）、海鲜（0.008）、韩国（0.004）、呵呵（0.004）、流量（0.003）、中国（0.003）、食材（0.003）、核辐射（0.003）
	主题 2：爱国情怀	日本（0.014）、表演（0.012）、海水（0.006）、民族（0.005）、核污染（0.005）、人格（0.005）、爱国（0.005）、媒体（0.005）、情怀（0.005）、砸掉（0.004）
波动期	主题 1：日料店未来	吃日料（0.007）、核污染（0.006）、进口（0.006）、商家（0.004）、智商（0.004）、好吃（0.004）、高端（0.003）、食品（0.003）、装逼（0.003）、消费（0.003）
	主题 2：反对进口食材	日本（0.051）、日料店（0.020）、核污水（0.012）、进口（0.008）、反对（0.007）、鳗鱼（0.005）、海鲜（0.005）、食材（0.005）、邋遢（0.005）、排放（0.004）

5 “不良信息”型网络暴力信息 K-means 文本聚类

5.1 方法介绍

随着网络技术不断进步和信息快速发展，如何从大量的数据中获取对决策有价值的知识成为人们当前面临的主要问题与挑战，数据挖掘技术为解决这一矛盾提供了有效的方法。其中，对文本文件实施自动聚类是许多检测与控制过程的重要环节，目的在于使得类内的文档尽量相似，类间尽可能相异。在文本聚类中，k-means 算法被广泛应用于检测相关数据的不同集群，对此国内外许多研究者已在人工智能领域与信息获取技术上做了相关工作。k-means 文本聚类算法是一种基于形心的技术，簇的形心是它的中心点。文本聚类是一种无监督的文档分类，把一个文本集分成若干成为簇（Cluster）的子集，每个簇的文本之间具有较大的相似性。K-means 聚类算法属于动态聚类，该算法通过迭代操作，每次对数据样本的分类进行测试与调整。

5.2 聚类分析与结果

本研究采用肘部法来选择 K 值，肘部法通过观察不同 K 值下聚类结果的 SSE（总平方误差）变化，来找到一个 “肘部” 点，说明增加 K 值对聚类效果的提升在某个临界点后会显著减小，本研究通过多次实验和计算，最终得出此数据集下微博评论文本数据的最优聚类数为 3，聚类后三类中的数据量如表 6 所示，通过对聚类结果进行进一步分析，将用户行为模式分为展现立场型、宣泄情绪型、聚焦事实型三类。

表 6 k-means 聚类结果

分类	数量	典型例子
0（展现立场型）	7294	还敢吃寿司？赶紧戒了！我大华夏那么多美食你不吃吃小日子的?
1（宣泄情绪型）	1199	完全就是坑人嘛
2（聚焦事实型）	1015	呵，这文稿前后矛盾，既然是爱国情怀为什么之前要搞成日料店直接搞中餐店风格不就好了？自己砸明明就是为了省个拆卸费，还夹带蹭一波流量

5.2.1 展现立场型用户

展现立场型用户的评论主题较为广泛，涵盖对日料店经营、日本核污染水排放、民族情怀、消费观念等多方面内容的讨论。从情感倾向来看，此类用户评论呈现出复杂多样的情绪，部分评论表现出对日料店的不满与抵制，如“不用调整

了，给我倒闭吧，要么你就转中餐，要么就滚出中国”“全部倒闭！抵制一切日货！！！”等，反映了用户对日料店食材来源及品质的担忧，以及对日本核污染水排放可能带来的食品安全问题的关切，情绪上带有明显的焦虑与愤怒。此外，一些评论则更倾向于表达对整个事件的理性思考，如“很无奈，日本明明可以有很多选择，却最终选择了排海，我们的渔民和相关从业者都收到了牵连”、“感觉这件事都搞魔怔了，变得非黑即白。国内很多家日料店都是中国人开的，食材也大部分是用中国本地食材，这也是别人赖以生计的工作，突然被自己国人捅刀子，受伤的依旧是那些中国商家和中国食材供应商；而且现在污染还没到我们这，大家也是趁着最后机会再去尝一尝”等，用户试图从客观角度分析问题，情感相对平和。在扩散期，这类用户的评论主要集中在对事件的初步反应和态度表达上，信息扩散速度极快，公众关注度逐渐上升。在爆发期，随着公众情绪的激化，这类用户的评论数量急剧增加，极端情绪开始激化舆论，地域歧视等严重影响网络环境的不良信息快速产生并扩散。在衰退期，随着公众情绪逐渐平复和其他热点事件的出现，这类用户的评论热度逐渐下降，但仍可能在特定情况下再次引发关注。在波动期，由于事件的复杂性和不确定性，这类用户的评论仍会引发小范围的讨论和情绪波动。

此类评论往往具有较强的主观性与直接性。用户不避讳表达自己的观点与情感，如“爱吃”“不坑穷人”“抵制”等表述，直截了当地表明对日料店、日本文化以及相关事件的态度。从心理动因角度来看，用户发表此类评论可能是出于对自身健康与安全的担忧、对公共事件的参与感与责任感、以及对社会公平与正义的追求。他们希望通过评论来传递自己的声音，引起更多人的关注与重视，从而推动社会问题的解决。在信息传播方面，由于此类用户评论语言风格通俗易懂，情感表达直接，能够迅速吸引其他用户的关注与回应，形成话题热点。在微博这一社交媒体平台上，此类评论的转发、点赞与回复数量往往较高，进一步扩大了信息的传播范围。

5.2.2 宣泄情绪型用户

宣泄情绪型用户评论普遍流露出较为强烈的负面情感色彩，用户在提及相关话题时，频繁使用如“讨厌”“可恶”“恶心”等词汇，反映出对该事件的不满与反感。这种情绪的表达并非孤立，而是与一些历史事件紧密关联，激发了用户

的民族情感与爱国情绪，进而促使他们在评论中宣泄情绪。在扩散期，这类用户的情绪宣泄可能较为温和，主要表现为对事件的初步反应和情绪表达；在爆发期，其情绪宣泄可能达到高峰，评论内容更加激烈和极端；在衰退期和波动期，其情绪宣泄可能逐渐减弱，但仍可能在特定情况下再次爆发。从认知层面来看，该类用户通过分享自己的观点和经历，试图揭示事件的“真实面目”，并警示他人。例如，有用户指出核污染水排放对海洋生态和人类健康的潜在威胁，显示出他们对这一问题的关注与担忧，以及对此决策的不信任。在行为倾向上，此类用户表现出明显的抵制行为。他们不仅在言语上表达对自己的不满，还积极倡导抵制相关产品、文化等。用户认为通过减少对相关产品和文化的消费，可以在一定程度上对决策者施加压力，促使其改变不当行为。此外，部分用户还表现出对国产替代品的支持，强调国内产品在质量、价格等方面的优势，鼓励消费者选择国产。

在社交互动方面，该类带有强烈个人情绪的用户评论往往能够引发其他用户的共鸣与讨论。他们的观点鲜明、情感真挚，容易吸引志同道合者的关注与回应，具有较强的传播性，通过转发、点赞等行为，使得相关信息能够在网络空间迅速扩散，扩大了影响力，从而形成一定的舆论氛围。除此以外，此类用户在评论中提及的信息来源较为广泛，包括新闻报道、个人经历、网络资料等。他们通过对这些信息的整合与分析，形成了自己对问题的看法，并在评论中加以阐述。这表明此类用户具有一定的信息获取能力和分析能力，能够从多渠道收集信息，并基于此构建自己的观点体系。

5.2.3 聚焦事实型用户

此类用户评论主要围绕“日料”这一主题展开，聚焦于核污水排放后国内日料店的发展前景，从对日料的喜爱到对日料店经营现状的讨论，再到对日料食材安全性的担忧，涵盖了日料店的经营、食材来源、消费者态度等多个方面。在扩散期，这类用户的评论主要集中在对事件的初步反应和对日料店发展的关注上；在爆发期，随着公众情绪的激化，这类用户的评论可能更加关注日料店的经营状况和食材安全性；在衰退期和波动期，这类用户的评论可能逐渐趋于理性和平和，但仍可能在特定情况下再次引发关注。从情感倾向来看，该类用户的评论呈现出较为复杂的情感态度。一方面，有部分用户表达了对日料的喜爱之情，如“其实，我很喜欢吃日料韩国料理也喜欢……东亚三国的饮食我都觉得很美味。牛排汉堡

披萨我倒是一般般，不是很爱，当然，偶尔吃一次也可以啦”等，显示出日料在一些消费者心中的吸引力。另一方面，也有不少用户对日料店的经营行为、食材来源等表示质疑和担忧，如“赶紧转行吧！日料贵就算了，到时因为食材污染然后重新问题就麻烦了”“希望中国日料行业能够注重食品安全和质量”，反映出用户对日料行业的不满和对食品安全的关注。

从行为倾向上分析，此类用户表现出一定的消费行为和决策倾向。部分用户表示愿意尝试和消费日料，如“吃日料没事”“这日料我非吃不可”，显示出对日料的消费热情。然而，也有用户因对日料食材安全性的担忧而选择抵制日料，如“国内美食不多吗，干嘛非要吃日料”“日料贵的吓人，利润高的离谱，还不怎么好吃”，反映出用户在消费决策中的谨慎和理性。在社交互动方面，此类用户评论具有较强的互动性和传播性，用户之间的评论相互呼应，形成了热烈的讨论氛围。例如一些用户对日料店的经营行为进行质疑，引发了其他用户的共鸣和回应，这种互动不仅加深了用户之间的情感联结，也进一步强化了他们对日料行业的认知和态度。同时，一些用户的评论具有较强的传播性，能够引发更多的关注和讨论，扩大了日料话题的影响力。从价值取向来看，此类用户评论体现了用户对食品安全、消费权益和文化认知的重视。用户对日料店的食材来源和真实性表示关注，显示出用户对食品安全和消费权益的维护意识。同时，一些用户对日料文化的认知和评价也反映了他们对文化多样性的尊重和对本土文化的自信，体现了用户在文化认知上的理性和批判精神。

5.3 不同类型用户“不良信息”型网络暴力信息演化机制

对不同类型用户在不同时期的评论数量进行汇总，得到数据表格如下,由表可知，扩散期展现立场型用户的评论数量相对较少，但已经开始表达对事件的初步反应和态度。宣泄情绪型用户的评论数量最少，表明在事件初期，情绪宣泄的需求较低。聚焦事实型用户的评论数量略高于宣泄情绪型用户，显示出对事件事实的关注。爆发期是评论数量的高峰期，展现立场型用户的评论数量急剧增加，表明在事件引发广泛关注后，用户开始积极表达自己的立场和观点。宣泄情绪型用户的评论数量也显著增加，但远低于展现立场型用户，显示出情绪宣泄的需求在这一阶段有所上升。聚焦事实型用户的评论数量相对较少，但也有明显增加，表明对事实的关注度提高。衰退期展现立场型用户的评论数量迅速减少，表明随

着事件热度的下降，用户的参与度降低。宣泄情绪型用户的评论数量也有所减少，但相对稳定，显示出情绪宣泄的需求逐渐减弱。聚焦事实型用户的评论数量最少，表明对事实的关注度进一步降低。波动期展现立场型用户的评论数量继续减少，几乎达到最低点。宣泄情绪型用户的评论数量有所回升，表明在事件的后续发展中，用户的情绪宣泄需求再次上升。聚焦事实型用户的评论数量有所增加，显示出对事件后续发展的持续关注。

表 7 三类用户不同阶段的评论数量

阶段	展现立场型	宣泄情绪型	聚焦事实型
扩散期	51	11	15
爆发期	7173	1059	910
衰退期	57	48	12
波动期	12	81	78

展现立场型用户在扩散期（51 条）、爆发期（7173 条）、衰退期（57 条）和波动期（12 条）的评论数量呈现出先急剧增加后迅速减少的趋势。特别是在爆发期，评论数量达到峰值，远高于其他阶段，可能是因为展现立场型用户希望通过表达自己的观点来获得更多的关注和认同。在事件热度下降后，这类用户的参与度迅速降低，表明他们的行为更多是受事件热度和公众关注度的影响。宣泄情绪型用户在扩散期（11 条）、爆发期（1059 条）、衰退期（48 条）和波动期（81 条）的评论数量也呈现出先增加后减少的趋势，但在波动期有小幅回升。总体而言，宣泄情绪型用户的评论数量在爆发期达到次高峰，但远低于展现立场型用户，可能是因为宣泄情绪型用户需要一个情绪宣泄的出口，而在事件热度较高时，这种需求更为强烈。在事件热度下降后，这类用户的情绪宣泄需求逐渐减弱，但在波动期可能会因为事件的后续发展而再次上升。聚焦事实型用户在扩散期（15 条）、爆发期（910 条）、衰退期（12 条）和波动期（78 条）的评论数量变化相对平稳，没有出现像展现立场型用户那样的急剧波动。聚焦事实型用户的评论数量在爆发期达到次高峰，但在波动期有所增加，可能是因为聚焦事实型用户更关注事件的事实和真相，而不是情绪宣泄或立场表达。在事件热度较高时，这类用户会更加积极地参与讨论，以获取更多的信息和事实。

6 研究启示

研究结果表明，通过训练 BERT 模型可以较准确地识别出包含“不良信息”型网络暴力信息的微博评论，证明了模型的有效性以及分类结果的可靠性。本研究通过 LDA 主题演化分析展示了公众关注点从不易被聚焦的宏观突发公共卫生事件转移到易被攻击的微观个体的全过程，在扩散期，人们更关注事件本身对环境等的影响，信息扩散速度极快。这一阶段，公众的注意力主要集中在事件的规模和影响上，对事件的细节和个体的关注较少。在爆发期，公众的极端情绪开始激化舆论，地域歧视等严重影响网络环境的不良信息开始快速产生并扩散。这一阶段，网络暴力问题最为严重，不良信息的传播速度极快，对受害者造成了极大的伤害。在衰退期，随着公众情绪平复以及有其他热点事件吸引关注，讨论热度逐渐下降。这一阶段，网络暴力问题得到了一定的缓解。在波动期，由于突发公共卫生事件的复杂性、不确定性等特性，公众的担忧仍然存在，仍会随机引起小范围的讨论，网络暴力问题可能会再次出现。针对“不良信息”型网络暴力不同阶段主题的演化和发展，本研究提出以下几点措施，为“不良信息”型网络暴力的分阶段治理提供一定的参考。

6.1 在扩散期扩大网络场域不良信息治理范围，强化不良信息监管机制

网络暴力信息并非是大量违法信息的简单集合，而是由少数诽谤信息辅以大量不良信息形成，是由“首发者”+“跟风者”、“起哄者”+“微小参与者”+“吃瓜者”的行为共同造成的网络暴力[42]，因此传统法律的二分模式，即“违法——合法”模式很难有效治理网络暴力。一方面，受害者受到的损害与网络暴力之间的因果关系难以认定。网络暴力受害者遭受网络暴力后，容易进入一种无力摆脱的心理瘫痪状态，类似于家庭暴力受害者的“受虐者综合征”，极易导致受害者心理抑郁甚至自杀[19]。然而网络暴力间接导致的这种严重损害，法律上难以明确相关主体的法律责任，尤其是对于“不良信息”型网络暴力来说，其因果关系更难认定。在此基础上，受害者及其家属更加难以获取法律救助。另一方面，即使平台识别出了网络暴力并阻止了网络暴力信息的发送，虽然可以一定程度上阻止网络暴力的扩大和进一步加重，但是受害者已经受到的伤害和已经产生的心

理阴影并没有得到有效的救济，对于受害者的处境并没有明显改善，因此从受害者的角度目前传统的解决方式收效甚微。

党的十九大报告明确提出，要加强互联网内容建设，建立网络综合治理体系，营造清朗的网络空间。近年来，国家相继出台了一系列法律法规和政策措施，加强对网络信息的管理和监督。例如，《网络安全法》《互联网信息服务管理办法》等法律法规，明确了网络运营者的责任和义务，规范了网络信息的传播秩序。同时，国家还积极推进网络文明建设，倡导文明上网、理性发言，营造良好的网络文化氛围。

遵守法律法规是网络场域不良信息治理的基本要求。网络不是法外之地，任何在网络上的行为都必须遵守法律法规。加强对网络法律法规的宣传和教育，提高公众的法律意识和法治观念的同时，要加大对网络违法犯罪行为的打击力度，维护网络空间的安全和秩序。尊重社会公德和伦理道德也是营造良好网络生态的重要保障。社会公德是全体公众在社会交往和公共生活中应该遵循的行为准则，伦理道德是人们在处理人际关系和社会事务中应该遵循的道德规范。树立正确的价值观和道德观，自觉抵制不良网络文化的影响。

要加强网络监管，建立健全网络监管机制。通过加强对网络信息的审核和管理，及时发现和处理不良信息，维护网络空间的良好秩序。新时代网络监管机制模式有别于传统的监管模式，更需要充分考虑互联网的特殊性及无边界性特点，要探索结合传统法律法规、制度体系和现代互联网模式详见和的信息治理制度，构建治理主体明晰、治理手段的治理模式。要充分认清互联网信息治理的风险点和难点，建立跨部门、跨行业、多部门联动的信息共享和沟通协调机制，建立健全互联网信息治理相关法律法规和行政制度，完善监督检查机制。强化互联网企业参与互联网信息治理，提供必要的技术支撑，强化行业自律的作用。

要建立互联网舆情分类管控制度，对于涉及国家安全和意识形态的舆情，要坚决打击控制；对涉及网民和社会群体的舆情，应加强行业自律和公众个人自律发挥作用，提倡自我消化解决，要利用网络媒体积极引导舆论，积极发挥社会正能量的作用，杜绝网络社会热点上升为网络舆情。全面加强网络生态综合治理。通过技术、行政、法律等多种手段的综合运用，要加强网络生态治理，注重基本规范、注重基础管理，强化主管部门的属地管理责任，强化网站的主体责任，要

提升管理效能，加强网络许可管理工作统筹协调，全面加强网络新闻从业人员管理工作，加强教育培训，推进分级分类管理。要督促企业切实落实企业主体责任，开展专项检查行动，完善内容审核机制，全面排查清理违法违规信息。

对于“不良信息”型网络暴力的治理，必须首先正视网络场域不良信息的累积和聚合效应对社会的危害，基于网络这一特殊场域建立并完善网络暴力治理制度，扩大治理范围，才能让网络暴力，尤其是“不良信息”型网络暴力得到有效治理。可以重点从“不良信息”型网络暴力扩散期切入，利用人工智能技术对网络信息进行及时监控和阈值预警，对扩散期的恶意辱骂、诽谤等不良信息精准有效打击，鼓励人们发表积极、客观的观点和言论，从源头打击网络暴力。此外，应明确界定网络暴力的不良信息的界定标准，并以此要求平台对网络评论和私信严恪标准管理，阻止“不良信息”型网络暴力朝下一阶段转化。

6.2 在爆发期重点控制网络暴力传播风险因素，精准施策降低传播风险

突发公共卫生事件背景下的“不良信息”型网络暴力产生的过程大致如下：由特定突发公共卫生事件引起人们的关注和讨论，自媒体等各类新兴媒体开始进行报道，经过广泛传播后网络舆情环境呈一致性或群体对立性两种趋势，在此传播过程中存在各类风险因素或促进、或推动了网络暴力的传播和扩散[2]。首先是匿名性心理导致的去抑制化效应。互联网重塑了信息传播和人与人之间交流的形式，使人们的自我认同和社会系统认同之间出现张力，然而互联网信息的多元化和碎片化又导致了信息透支、信息泛化等问题，网民往往容易根据网上只言片语的片面信息轻易对事件下结论，加之互联网的匿名机制为网民进行道德审判提供的便利，秉着“法不责众”的心理，网络场域的匿名性弱化了人们的道德责任感，社会矛盾和生活压力又让人们不断从互联网寻找发泄口，更催生了网络暴力的发展。其次是“沉默的螺旋”效应导致的网络舆论单一化。大数据时代平台会按照用户的喜好推送给用户感兴趣的信息，因此对于特定突发公共卫生事件，人们无法同时获取差异化的信息，加剧了网络舆论的单一化现象，同时对于与主流观点持不同意见的人来说，一旦发表相左的观点，可能会被视为与受害者同一阵营从而遭受网暴，因此为了避免这种情况，人们往往会选择沉默，进一步加剧了网络舆论单一化，这种单一化会导致网暴行为对个人的损害更加严重。最后是信息透

支、信息极化导致的信息信任问题。互联网的海量信息会让人们产生信息透支、信息泛化等问题，信息时代知识的更新速度极快，知识不断被推翻、重塑，专家权威被挑战，人们对于权威的不信任导致其对于信息的判断多基于个体经验，也更容易被互联网上夸张化的、未经证实的信息吸引注意力，并由于网络的匿名性发表激烈的观点，再形成“沉默的螺旋”效应，推动了网络暴力的扩散和发展。

实名制进程正在逐步推进，目前的实名制为账号发言显示 ip 省份地址，现实地域荣誉感已经在很大程度上对人们的言行起到了规范作用。所以我们要不断加强网络账号与现实身份的联系，加大其对网络用户的约束作用。目前部分平台网络账户采用手机号绑定登录，然而手机号存在二次放号的可能性，因为实名效果不理想。故平台要推进网络账户与居民唯一身份证明——身份证的绑定登录，并进行人脸识别鉴定，来保证实名制的效果；同时将相关数据录入数据库，以便及时有效对应身份。对该数据库，平台也需要采取相应的保护措施，即采用应用防火墙对网络用户的隐私信息进行保护，防止出现大量用户信息被盗取及泄露传播，以至于酿造网络暴力事件。针对网络暴力事件侵权责任人难以确定的问题，平台具有一定的操作权限，即可利用先进的技术推动实名进程。第一，建立黑名单与信誉数据库。通过数据分析若该网络账户存在相关网络暴力行为，便锁定封禁其账号，禁止其进行网络活动。第二，采用 ip 地址定位服务。网络用户可能拥有多个网络账户，但登录账户的设备是有限的，电子设备对应的 ip 地址是唯一的。所以通过 ip 地址定位可及时锁定网络暴力侵权人的地理位置。第三，设立识别被盗账号的装置。通过异常登录检测，如果用户用不常用地理位置和新设备登录，可发送警告邮件和信息并要求额外验证，降低利用他人账号进行网络暴力行为的可能性。

网络安全和信息化是一体之两翼，驱动之双轮，互联网时代，网络信息治理和信息化发展是密不可分的。应建立以政府治理为主导，行业组织、社会群体等多个治理主体协同参与的体制机制，充分发挥市委网信办统筹协调作用。建立多部门多行业信息共享研判机制，解决多头管理、互相推诿的问题。要加强党对互联网的管理，建立健全网联网信息治理体制机制。控制网络暴力传播风险因素，有助于从各个阶段尤其是爆发期对网络暴力的产生和扩散进行有效遏制，打破“法不责众”的认知和网络舆论单一化的局面，改善人们的社会信任感，打造健

康清朗的网络舆论环境。对于匿名性心理导致的去抑制化效应，可以通过在扩散期阶段对于少数的非法评论精准打击，屏蔽相关评论，并采取手段对当事人进行批评教育，加大惩罚力度，打破人们对于“法不责众”的错误认知。对于“沉默的螺旋”效应导致的网络舆论单一化，平台要不断优化推荐算法，避免完全以用户喜好为核心的算法推荐内容，减少信息茧房和舆论单一化的现象出现，并且将关闭个性化推荐的选项更清晰地呈现给用户，让用户根据自身需求选择是否关闭个性化推荐。对于信息透支、信息极化导致的信息信任问题，相关部门可以开通社交账号以便在应对重大突发公共卫生事件时能够及时、准确地发表权威消息，并鼓励有影响力的意见领袖评论转发，禁止不良媒体的恶意营销和炒作。让谣言和不实消息不攻自破，加强人们对于权威消息的信任感。

6.3 全阶段增强主流意识形态价值引领作用，合理规制用户信息权力

习近平总书记在准确把握互联网时代新形势的前提下，不仅高度重视意识形态工作，还积极应对互联网意识形态新挑战，提出了一系列关于网络意识形态的新观点新论断。首先，网络意识形态治理要始终坚持以马克思主义为指导的主流意识形态，尤其在多元思潮的冲击下，更要巩固其指导地位不动摇。“宣传思想工作就是要巩固马克思主义在意识形态领域的指导地位，巩固全党全国人民团结奋斗的共同思想基础”。其次，网络意识形态治理要坚持中心工作与意识形态工作相适应，促进我国物质文明和精神文明协调发展，坚持“硬实力”与“软实力”协同进行。“经济工作搞不好要出大问题，意识形态工作搞不好更要出大问题”。再次，网络意识形态治理要坚持以人民为主体，坚持党性与人民性的统一，全心全意为人民服务。“网信事业要发展，必须贯彻以人民为中心的发展思想”。最后，网络意识形态治理要依法治理与科学治理，根据具体问题，完善互联网相关立法，加强互联网行为规范。“网络空间同现实社会一样，既要提倡自由，也要保持秩序。要坚持依法治网、依法办网、依法上网，让互联网在法治轨道上健康运行。”

互联网深刻改变了人们的生产生活方式，互联网也成为人们交往的主要空间。然而网络具有开放性与包容性，匿名性与隐蔽性等特征，无门槛的内容创作及以秒计的传播速度使不良信息得以快速占领舆论场地，错误意识形态得以逐渐实现隐性渗透，不仅影响网络风气，还严重危及社会公共安全，故营造“风清气正”

的网络空间刻不容缓。要牢牢掌握网络意识形态工作领导权，坚持破立并举，维护网络意识形态安全。要加大网络意识形态工作统筹力度，形成以重点网站为主力，各类社会力量广泛参与的工作格局。提升网络意识形态工作水平，积极管理和反击错误思想，强化正面宣传，让广大网民充分认识到应该旗帜鲜明的反对什么，支持什么，使全体网民在理想信念、价值理念、道德观念更加紧密的团结在一起。要采取线上线下相结合的联动机制，通过深入实践、深入基层、深入生活的方式，拓展网络意识形态工作空间，弘扬社会主义核心价值观。政府应密切关注网民诉求，有效准确回应，用事实说话，提升网民信任感。要加强互联网企业党建工作，夯实互联网意识形态工作基础。要推动移动互联网更好的服务社会、服务公众。要引导网民养成崇德向善的网络行为习惯，筑牢文明守法的网络行为底线。要坚持“以人民为中心”为工作导向，在信息公开、舆论引导、辟谣科普等方面服务好广大网民。

当下网络空间中较易出现泛意识形态化解读现象，而泛意识形态化会一定程度上加剧网络暴力，从而进一步对现实社会产生危害[43]。网络空间的匿名性塑造了一种碎片化、扁平化的交往形式，使得人们可以脱离现实社会个体身份的束缚，个体表达更加主体化，如从个体经验解读政治问题等。但与此同时，存在部分个体为了博取关注，故意发表过激观点，将一般问题上升到意识形态层面的情况等。同时作为社会弱关系的匿名空间，网络的虚拟性也会弱化人们的道德责任意识，使其在发表言论时带有很强的个人主观想法，网络空间不断流动的海量信息也会“掩护”人们的真实信息，因此更助长了网络空间的泛意识形态化解读现象。核污水对日料店冲击的讨论中，就有人将“开日料店”等同于“不爱国”，对店主进行人身攻击和辱骂等，对店主个人以及网络环境均产生了严重不良影响。现实生活中出于政治文化和社会道德制约，泛意识形态化现象即使出现也很难传播，然而在网络空间中，由于海量信息出现导致的信息透支和信息泛化问题，专家知识的权威性被打破[44]，网民对事件的判断更多地基于个人的现实经验、某种个人权威等，从而更容易信任挑战权威、挑战传统的个性化解读，这种网络环境也进一步加剧了泛意识形态化现象。同时，由于网络传播的聚合效应，网民个体的泛意识形态化解读会通过“回声室效应”不断扩大传播，进而从个体的解读转变为群体的共识，不断发酵成为现象级事件。

在突发公共卫生事件中的“不良信息”型网络暴力爆发期这种个体的泛意识形态化解读和传播对于事件走向和网络空间都有着严重不良影响，因此要增强主流意识形态在网络空间的价值引领作用，引导人们正确解读和使用自己的信息权力。增强主流意识形态在网络空间的价值引领作用，一方面要加强对主流意识形态内涵的阐释，让普罗公众从根本上理解意识形态的内涵，避免公众因为对主流意识形态理解不透彻而错用、滥用意识形态批判的情况。此外，还可以引导公众将其对主流意识形态的理解和日常生活紧密联系，充分利用现有的网络工具，灵活、生动地诠释主流意识形态的核心要义，提高全社会的认同度。另一方面需要加强对主流意识形态分析方法的把握，提高利用马克思主义立场、观点、方法解决问题的能力，坚持实践第一的观点，强调调查研究的重要性，在历史的观点下解读信息，用客观的判断代替主观的臆想。

6.4 精准识别并明确不同类型的用户特征，有针对性地预防和治理

本研究通过对 “不良信息” 型网络暴力信息进行 K-means 文本聚类分析，将用户评论分为表明态度型、宣泄情绪型和聚焦事实型三类。表明态度型用户评论数量最多，这类用户在评论中直接表达对日料店、日本文化及相关事件的态度，其评论主题广泛，情感倾向复杂多样。部分用户因对日料店食材来源及品质的担忧而表现出不满与抵制，也有部分用户从客观角度进行理性思考。这表明在网络暴力事件中，用户的态度和情感表达具有多样性和复杂性。针对这类用户，网络平台和相关部门应加强对评论内容的审核和引导，鼓励用户理性表达观点，避免使用过激言语和不实信息。同时，提供权威信息和专业解读，帮助用户形成客观、理性的态度，减少因误解和偏见导致的网络暴力行为。

宣泄情绪型用户评论数量相对较少，但情感色彩强烈。这类用户在评论中宣泄对事件的不满与反感，频繁使用负面词汇，且其情绪表达与历史事件紧密关联，激发了民族情感与爱国情绪。这表明网络暴力事件中，用户的情绪宣泄可能受到多种因素的影响，包括个人情感、社会文化背景等。针对这类用户，网络平台应加强对情绪化评论的监测和管理，及时发现和处理过激言论，防止其引发更大的网络暴力行为。同时，相关部门应关注用户的情绪变化，积极回应用户的关切和诉求，通过加强信息公开和沟通交流，缓解用户的焦虑和不满情绪，引导用户以更加理性的方式表达自己的情感。

聚焦事实型用户评论数量最少，但其评论内容具有较高的信息价值。这类用户关注日料店的经营、食材来源、消费者态度等具体事实，对日料店的发展前景进行讨论和分析。这表明在网络暴力事件中，部分用户能够保持理性和客观，关注事件的本质和核心问题。针对这类用户，网络平台和相关部门应提供准确、权威的信息，满足用户对事实真相的追求，同时加强对虚假信息和谣言的打击力度，维护网络空间的信息真实性和可信度。此外，还应加强对用户的教育和引导，提高用户的媒介素养和信息辨别能力，使用户能够更加理性地看待网络信息，避免被虚假信息误导。

7 结论与展望

本研究首先通过 BERT 模型识别突发公共卫生事件下符合网络暴力的“不良信息”，其次结合危机生命周期理论，利用 LDA 模型对突发公共卫生事件下“不良信息”型网络暴力演化机制进行深入剖析，并利用 K-means 聚类将用户评论分为“展现立场型”、“宣泄情绪型”、“聚焦事实型”三类，并对每一类用户的行为模式进行分析；最后针对“不良信息”型网络暴力不同阶段的主题演化特征，提出相应的治理策略，如在扩散期扩大网络场域不良信息治理范围，爆发期重点控制网络暴力传播风险因素，及持续增强主流意识形态价值引领作用多种途径等。本研究的结论成果可助力“不良信息”型网络暴力的分阶段治理，进而维护社会稳定与和谐。然而本研究仍存在一定的局限性，一方面，“不良信息”型网络暴力的有效识别率有待进一步提升；另一方面，本研究中样本数据局限于文本数据。后续研究可以进一步优化 BERT 模型，提高模型识别准确率，同时扩大数据集来源，将音频、视频等多模态数据也纳入样本数据范围，从而减弱“不良信息”型网络暴力对网络环境和社会生活的影响，维护法律和伦理秩序，保障社会稳定。

本文围绕突发公共卫生事件下“不良信息”型网络暴力的演化机制与管控策略展开了深入研究，但鉴于该领域的复杂性与动态性，仍存在诸多方面值得后续进一步探究。具体而言，我们认为未来研究还可以从以下几个方面进行：

第一，在研究方法层面存在可拓展空间。当前主要采用的研究方法可能在全面性上有所欠缺，例如对网络暴力信息传播路径的分析多基于特定时间段内的数据收集，存在一定局限性。未来可引入时间序列分析等创新型方法，对网络暴力信息在不同阶段的传播动态进行长期追踪，更精准地剖析其演变规律，进一步验证和完善现有研究结论。同时，积极采用网络爬虫技术与大数据分析平台相结合，获取更广泛的网络数据，包括不同社交平台、小众网络社区等信息，使研究数据来源更加丰富全面，增强研究结果的可靠性。

第二，研究结果的跨地区普适性有待检验。本研究主要聚焦于特定地区或国家在突发公共卫生事件下的网络暴力情况，然而不同地区的社会文化、法律制度、网络生态等存在显著差异，这些因素可能对网络暴力的演化及管控产生不同影响。后续研究可针对多个具有代表性的地区进行案例分析与对比研究，将本研究成果

置于不同文化与社会背景下进行验证, 从而确定研究模型与管控策略在全球范围内的普适性与局限性，为制定更具通用性的网络暴力治理方案提供依据。

第三，样本选取的多样性可进一步提升。目前研究的调查对象可能集中于单一平台特定网络使用习惯的人群，限制了研究结果的普遍代表性。未来可通过与专业网络调研机构合作，利用其庞大的用户样本库，广泛收集不同社会阶层、年龄段、教育背景及地域人群的数据。同时，在网络平台上采用分层随机抽样的方法，确保涵盖活跃用户与普通用户、不同网络社交圈的参与者等各类群体，使研究样本能更真实地反映整个网络社会的情况, 增强研究结论的说服力与应用价值。

参考文献

[1]吕途, 陈昊, 林欢, 等. 突发公共事件下网络谣言治理策略对谣言传播意愿的影响研究[J]. 情报杂志, 2020, 39(7): 87-93.

[2]熊尧. 网络舆论引导关键技术研究[D]. 华侨大学, 2023.

[3]王文华. 论反仇恨言论视阈下网络暴力的法律治理[J]. 中国应用法学, 2023(5): 63-75.

[4]赵精武. 异化的网络评论——再论网络暴力信息的阶段化治理[J]. 北方法学, 2023, 17(5): 21-36.

[5]姜方炳. “网络暴力”: 概念、根源及其应对——基于风险社会的分析视角[J]. 浙江学刊, 2011(6): 181-187.

[6]Kowalski R M, Giumetti G W, Schroeder A N, et al. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth[J]. Psychological bulletin, 2014, 140(4): 1073.

[7]Giumetti G W, Kowalski R M. Cyberbullying via social media and well-being[J]. Current Opinion in Psychology, 2022, 45: 101314.

[8]程睿.治理观视域下网络不良信息内容的法律认定标准[J].江西社会科学,2022,42(6):159-167.

[9]网络信息内容生态治理规定[J].中华人民共和国国务院公报,2020,(8):46-50.

[10]张玉峰,张婧.基于数据挖掘的 Web 文本不良信息监测模型研究[J].情报理论与实践,2009,32(11):89-92.

[11]黄辉宇,李从东,任家东,等.基于人工神经网络的不良信息实时监测原型系统[J].计算机工程,2006,(2):254-256,265.

[12]明弋洋,刘晓洁.基于短语级情感分析的不良信息检测方法[J].四川大学学报(自然科学版),2019,56(6):1042-1048.

[13] Paul S, Saha S. CyberBERT: BERT for cyberbullying identification: BERT for cyberbullying identification[J]. Multimedia Systems, 2022, 28(6): 1897-1904.

[14]李铭轩,文继荣.AIGC 时代网络信息内容的法律治理——以大语言模型为例[J].北京理工大学学报(社会科学版),2023,25(6):83-92.

[15]张凌寒.“不良信息”型网络暴力何以治理——基于场域理论的分析[J].探索与争鸣,2023,(7):96-107+178-179.

[16]林爱珺,章梦天.网络内容生态治理的多元主体责任规制[J].新闻爱好者,2021,(4):14-16.

[17]中华人民共和国中央人民政府.突发公共卫生事件应急条例（2003 年 5 月 9 日中华人民共和国国务院令第 376 号公布根据 2011 年 1 月 8 日《国务院关于废止和修改部分行政法规的决定》修订）[EB/OL].https://www.gov.cn/zhengce/202203/content_3338257.htm,2011-01-08.

[18]刘宇桐,王晰巍,王楠阿雪,等.重大突发公共卫生事件下社交媒体信息传播的算法抵抗行为研究[J].图书情报工作,2024,68(9):98-109.DOI:10.13266/j.issn.0252-3116.2024.09.010.

[19]宋慎铭,王琛,詹东远.突发公共卫生事件下的在线社交媒体公众情绪挖掘[J].管理评论,2024,36(3):246-257.

[20]林威宇.网络暴力中自媒体行为归责之类型化配置[J].学术探索,2024(4):125-137.

[21]张筱荣,郭圳凝.突发公共卫生事件网络舆情危机及其治理[J].北京交通大学学报(社会科学版),2023,22(2):133-140.

[22]胡象明,刘腾.突发公共卫生事件网络舆情影响力生成机理及风险规避[J].山东社会科学,2023,(2):96-107.

[23] Zhao Y, Chu X, Rong K. Cyberbullying experience and bystander behavior in cyberbullying incidents: The serial mediating roles of perceived incident severity and empathy[J]. Computers in Human Behavior, 2023, 138: 107484.

[24]吕鲲,施涵一,靖继鹏.突发公共卫生事件网络舆情热点话题形成组态路径研究——基于微博热搜数据的模糊集定性比较分析[J].情报理论与实践,2022,45(9):148-156.

[25]曾子明,陈思语.基于 LDA 与 BERT-BiLSTM-Attention 模型的突发公共卫生事件网络舆情演化分析[J].情报理论与实践,2023,46(9):158-166.

[26]马腾,殷跃,赵树宽等.多维数据融合的突发公共卫生事件网络舆情演化特征研究[J].情报理论与实践,2022,45(12):170-177.

[27]Raskhodchikov A N, Pilgun M. COVID-19 and Public Health: Analysis of Opinions in Social Media[J]. International Journal of Environmental Research and Public Health, 2023, 20(2): 971.

[28]曾子明,江新林.突发公共卫生事件中基于区块链的网络舆情溯源体系研究[J].现代情报,2023,43(6):149-157.

[29]郑佳悦,王亮.基于区块链的网络舆情监管机制研究[J].中国传媒科技,2022(6):10-13.

[30]周婕.基于区块链技术的大连高校舆情防控机制研究[J].图书馆学刊,2022,44(1):13-20.

[31]朱广生,刘阳.突发公共卫生事件网络舆情的治理思维、原则及方略[J].广西社会科学,2022,(8):122-127.

[32]温海涤,邱振博.新媒体时代突发公共危机事件网络舆情治理能力研究[J].情报科学,2022,40(8):38-43,49.

[33]Ferguson C, Merga M, Winn S. Communications in the time of a pandemic: the readability of documents for public consumption[J]. Australian and New Zealand Journal of Public Health, 2021, 45(2): 116-121.

[34]Fink S, American Management Association. Crisis management: Planning for the inevitable[M]. Amacom, 1986.

[35]王旭,孙瑞英.基于 SNA 的突发事件网络舆情传播研究——以“魏则西事件”为例[J].情报科学,2017,35(3):87-92.

[36]韩小伟,张传洋,张起超,等.大数据背景下突发公共事件网络舆情情感演化及舆情引导策略研究[J/OL].情报科学:1-20[2024-05-22].http://kns.cnki.net/kcms/detail/22.1264.G2.20240129.0941.008.html.

[37]张文杰,许门友.主流媒体引导下公共事件社会舆情泛化特征分析[J].情报科学,2022,40(1):25-30.

[38]毛太田,蒋冠文,李勇,等.新媒体时代下网络热点事件情感传播特征研究[J].情报科学,2019,37(4):29-35+96.

[39]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[40]Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation[C]//European conference on information retrieval. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 345-359.

[41]Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 3(Jan): 993-1022.

[42]于冲.网络“聚量性”侮辱诽谤行为的刑法评价[J].中国法律评论,2023,(3):87-98.

[43]赵宴群.网络空间泛意识形态化解读现象分析及其引导[J].江苏社会科学,2022(1):96-103.

[44]闫臻.数字化时代网络集体非理性惩罚现象的结构逻辑与个体特征[J].西安交通大学学报(社会科学版),2022,42(5):107-114.

AI智创·青春力量

AIGC与计算传播创新大赛组织委员会

2025 年 3 月 10 日