

中国科学技术大学

专业硕士学位论文



HPV 疫苗相关议题在多模态社交媒体上 的内容研究

作者姓名： 何浏

专业领域： 新闻与传播硕士

导师姓名： 朱孟潇 教授 邵祁丰 博士

完成时间： 二〇二五年五月二十七日

University of Science and Technology of China
A dissertation for master's degree



A Content Study on HPV Vaccine-Related Topics on Multimodal Social Media

Author: He Liu

Speciality: Master of Journalism and Communication

Supervisors: Prof. Zhu Mengxiao, Dr. Shao Qifeng

Finished time: May 27, 2025

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：_____

签字日期：_____

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

☒ 公开 ☐ 控阅（____年）

作者签名：_____

导师签名：_____

签字日期：_____

签字日期：_____

摘 要

近年来,关于 HPV 疫苗的健康话题在中国引发了广泛讨论,并引起了大量社交媒体用户的关注。以前关于 HPV 疫苗和社交媒体的研究,往往仅关注以文本内容为主的社交媒体平台。随着技术的发展,社交媒体的内容形式变得多样化,不再仅限于文本内容,还包括音频和视频。社交媒体平台主要呈现的内容可以分为三种类型:文本、音频和视频。在此基础上,本研究应用健康信念模型(Health Belief Model, HBM)的四个维度——易感性、严重性、效益和障碍,对三种类型的社交媒体上的内容进行分类,并从以上四个维度来分析 HPV 疫苗相关议题在社交媒体平台上的多模态内容。

本文收集了 2018 年 7 月至 2023 年 3 月微博、喜马拉雅、抖音三个平台上关于 HPV 疫苗讨论的数据作为文本、音频、视频三类模态内容的例子,其中包含 273,357 条微博帖子、1,228 个喜马拉雅音频和 1,225 个抖音视频。由于不同模态的内容无法进行直接对比,且音频和视频中的关键内容多集中在人声部分,因此本研究对音频和视频中的人声进行提取和转录,并将不同模态内容的对比聚焦在文字层面的比较上。通过对长文本数据进行健康信念模型四个维度的人工编码和对短文本数据先人工编码后应用自然语言处理技术进行小样本分类后发现,文本平台主要关注障碍维度,音频平台关注维度较为均匀,而视频平台主要关注严重性、效益和障碍维度。将维度分类后通过主题建模发现,文本平台主要关注国家政策,音频平台主要关注知识普及,视频平台主要关注个体经验和叙述。通过情绪分析发现,音频和视频平台相较于文本平台更具有情绪,且情绪表达更积极。本文也发现文本平台和视频平台的讨论多从“女性”视角展开,音频平台的讨论则表现出“迟滞性”和“长期性”的特征。

综上所述,本文揭示了不同内容类型的平台在讨论 HPV 疫苗话题时的差异,并从健康信念模型的四个维度出发解释了这些差异。研究结果有助于了解 HPV 疫苗在我国公众认知中的现状,并为社交媒体平台的治理和公众政策宣传提供参考。未来研究议题可以扩展到其他健康传播、政治传播等领域。

关键词: HPV 疫苗 健康信念模型 社交媒体 自然语言处理 文本分类 主题建模

ABSTRACT

In recent years, the health topic of the HPV vaccine has sparked widespread discussion in China, attracting significant attention from social media users. Previous research on the HPV vaccine and social media has often focused primarily on text-based platforms. With the advancement of technology, the content forms of social media have diversified to include text, audio, and video. This study applies the four dimensions of the Health Belief Model (HBM)—susceptibility, severity, benefits, and barriers—to categorize content across these three types of social media, aiming to study the multi-modal content of the HPV vaccine issue on social media platforms.

Data were collected from Weibo (text), Ximalaya (audio), and Douyin (video) from July 2018 to March 2023, totaling 273,357 Weibo posts, 1,228 Ximalaya audios, and 1,225 Douyin videos. Since content in different modalities cannot be directly compared and the key content in audio and video is often concentrated in human voice, this study extracts and transcribes the voice content from audio and video, focusing the comparison on the textual level. Through manual coding of the four dimensions of the Health Belief Model for long text data and initial manual coding followed by small sample classification using natural language processing for short text data, the study found that text platforms mainly focus on the barrier dimension, audio platforms have a more balanced focus across dimensions, and video platforms mainly focus on the dimensions of severity, benefits, and barriers. Topic modeling after dimension classification revealed that text platforms mainly focus on national policies, audio platforms on knowledge dissemination, and video platforms on individual experiences and narratives. Sentiment analysis showed that audio and video platforms exhibit more emotional and positive content compared to text platforms. The study also found that discussions on text and video platforms often unfold from a “feminine” perspective, while discussions on audio platforms exhibit characteristics of “latency” and “long-term.”

In conclusion, this study reveals the differences in how different types of content platforms discuss the topic of the HPV vaccine and explains these differences from the perspective of the four dimensions of the Health Belief Model. The findings help to understand the current state of public perception of the HPV vaccine in China and provide references for social media platform governance and public policy promotion. Future research topics can extend to other fields such as health communication and political communication.

Key Words: HPV Vaccine, Health Belief Model, Social Media, Nature Language Processing, Text Classification, Topic Model

目 录

第 1 章 绪论	1
1.1 研究背景	1
1.2 研究概况和研究意义	3
1.3 技术路线	4
第 2 章 国内外研究现状	6
2.1 公众健康视角下的 HPV 疫苗议题	6
2.1.1 人乳头瘤病毒 (HPV)	6
2.1.2 人乳头瘤病毒疫苗 (HPV 疫苗)	7
2.1.3 人乳头瘤病毒疫苗 (HPV 疫苗) 在中国的现状	8
2.2 HPV 疫苗议题与社交媒体	9
2.2.1 国外文献综述	9
2.2.2 国内文献综述	13
2.3 不同模态内容的社交媒体平台	14
2.3.1 文本社交媒体平台	15
2.3.2 音频社交媒体平台	17
2.3.3 视频社交媒体平台	18
2.4 多模态社交媒体平台的跨平台比较	19
2.5 健康信念模型 (Health Belief Model, HBM)	20
2.5.1 健康信念模型理论	20
2.5.2 有关健康信念模型的研究	20
2.5.3 HPV 疫苗和健康信念模型在社交媒体上的研究	21
2.6 文本分类任务、小样本学习	21
2.6.1 文本分类任务	21
2.6.2 小样本学习	22
第 3 章 研究设计	23
3.1 研究方法	23
3.2 数据采集	23
3.2.1 文本平台数据采集	24
3.2.2 音频平台数据采集	24
3.2.3 视频平台数据采集	25

3.3 数据预处理	25
3.3.1 文本平台数据预处理	25
3.3.2 音频平台数据预处理	26
3.3.3 视频平台数据预处理	27
3.4 健康信念模型维度分类	29
3.4.1 抽样数据建立编码表	29
3.4.2 训练深度学习模型进行小样本分类	29
3.5 基于深度学习的主题建模技术	30
3.5.1 主题建模数据预处理	32
3.5.2 BERTopic 短文本主题建模	33
3.5.3 BERTopic 长文本主题建模	33
3.6 基于深度学习模型的情绪分类	34
第 4 章 结果分析	36
4.1 三类平台的健康信念模型分析	36
4.1.1 文本平台的健康信念模型分析	36
4.1.2 音频平台的健康信念模型分析	37
4.1.3 视频平台的健康信念模型分析	38
4.2 三种平台在健康信念模型维度下的比较分析	40
4.2.1 易感性维度比较	40
4.2.2 严重性维度比较	42
4.2.3 效益维度比较	44
4.2.4 障碍维度比较	46
4.2.5 小结	47
4.3 三类社交平台的情绪分析	48
4.3.1 总占比分析	48
4.3.2 时间序列分析	49
4.3.3 小结	52
第 5 章 结论与讨论	54
5.1 研究结论	54
5.1.1 健康信念模型维度在三种平台上分布不均	54
5.1.2 不同平台在健康信念模型四个维度具有内容上的共性	55
5.1.3 文本、视频平台的内容多与女性相关	55
5.1.4 音频平台表现出“迟滞性”和“长期性”	56
5.1.5 音频和视频平台内容更有“情绪”和更“积极”	57

5.1.6 音频平台的“国际视野”和视频平台的“支持国产”	58
5.2 关于内容差异形成的机理的探究	59
5.2.1 模态的差异	59
5.2.2 内容发布者的差异	59
5.2.3 内容消费者的差异	60
5.3 研究的意义	60
5.3.1 学术贡献	60
5.3.2 实践启示	62
5.4 研究讨论	63
5.4.1 研究局限性	63
5.4.2 未来展望	64
参考文献	66
附录 A 补充材料	77
A.1 文本预处理无关词表	77
A.1.1 微博无关词表	77
A.1.2 微博无关用户	77
致谢	78
在读期间发表的学术论文与取得的研究成果	81

第1章 绪 论

1.1 研究背景

随着科技的飞速发展,公众在社交媒体上对健康信息的讨论越来越多。健康机构和专家利用社交媒体平台传播健康信息,使公众可以以较低的成本轻松获取健康信息。此外,公众愿意根据自己的经验在社交媒体上分享健康信息^[1]。例如,在 COVID-19 大流行期间,研究人员观察到公众讨论了许多健康话题,如食品安全和健康不平等^[2]。然而,有学者发现社交媒体也可能传播带有误解和偏见的信息,从而对人们的健康行为产生负面影响^[3]。例如,研究人员发现个体接种 COVID-19 疫苗的意愿与在社交媒体上接触该疫苗相关信息之间存在负相关关系^[4]。因此,对社交媒体健康话题进行内容分析,可以帮助我们了解公众参与健康传播活动的方式。

在健康信息的讨论中,宫颈癌讨论的热度非常高。宫颈癌主要由感染人乳头瘤病毒 (HPV) 引起,这种癌症在全世界女性恶性肿瘤中发病率位居前列^[5]。2003 年,香港艺人梅艳芳之死将宫颈癌带入公众视野,媒体开始注意到这种高发疾病。在中国每年约有 13 万宫颈癌新增病例,占全球新发病例的三分之一,其中约有 8 万女性去世,宫颈癌也是中国 15 岁至 44 岁女性中的第二大高发癌症。放眼全球,平均每一分钟即确诊一例新发病例,每两分钟就有一名女性死于宫颈癌^[6]。

宫颈癌是众多癌症中唯一一种能够通过接种疫苗进行预防的癌症^[7]。这种疫苗,即人乳头瘤病毒 (HPV) 疫苗,专用于预防宫颈癌及其相关的性传播疾病,在公众中被广泛知晓,通常被称为“宫颈癌疫苗”。从医学角度看,人乳头瘤病毒包含多种类型,例如 HPV1、2、6 等。其中,HPV16 和 18 型病毒是宫颈癌的主要致病因素。因此,女性接种 HPV 疫苗能有效预防宫颈癌、生殖器疣、肛门癌等由高危型 HPV 引起的重大疾病^[6]。目前,HPV 疫苗有三种类型:二价疫苗、四价疫苗和九价疫苗。中国的 HPV 疫苗于 2006 年香港上市,2017 年在中国内地上市。内地上市后,HPV 疫苗需求旺盛,导致供不应求。2018 年上半年,接种疫苗常常需要提前预约并长时间等待。到了下半年,尽管上海、广州等地的二价和四价疫苗供应已经足够,九价疫苗依然需预约并数量有限。

在中国社交媒体上对于健康信息的讨论中,HPV 疫苗是一个非常热门的话题。中国 HPV 疫苗的接种率比较低。2020 年,中国仅有不到 1% 的年轻女孩接种了疫苗^[8]。为了有效降低宫颈癌的发病率,中国政府已经启动了一项全国性的 HPV 疫苗接种计划,旨在到 2025 年使 90% 符合条件的 15 岁以下女孩接种疫苗^[9]。

目前, 社交媒体平台的内容形式愈发丰富多样, HPV 疫苗这一健康话题在各类平台上被广泛讨论。在文本内容主导的社交媒体如微博、推特上, HPV 疫苗的话题讨论频频出现; 而在音频内容为主的播客平台上, 如 Spotify、喜马拉雅等, 也能听到关于 HPV 疫苗的深入解析和讨论; 此外, 以视频内容为主的社交媒体平台, 如抖音、快手、B 站等, 更是以直观、生动的方式呈现 HPV 疫苗的相关知识和公众观点。这些多元化的内容形式不仅拓宽了信息传播的渠道, 也增强了公众对于 HPV 疫苗的了解和认识。

HPV 疫苗在社交媒体上的研究起步于 2012 年, Fowler 等人在《International Journal Of Press-Politics》上首次探讨了媒体对 HPV 疫苗的报道^[10]。自那时起, 相关研究逐渐增多, 主要分为定性研究、混合研究和定量研究, 研究方法包括文本数据采集、问卷调查和眼动实验。在定性研究方面, 研究者通过问卷调查和在线实验分析了社交媒体上关于 HPV 疫苗的讨论。例如, Lee 等人探讨了博客上个人健康故事对社会规范认知的影响^[11]。Jiang 等人则通过在线调查研究了社交媒体使用与 HPV 知识之间的关系^[12]。Chen 等人研究了社交媒体上的阴谋论信息对 HPV 疫苗认知的影响^[13]。混合研究方法结合了定量和定性分析。例如, Himelboim 等人使用 Twitter 数据分析 HPV 疫苗争议的对话内容, 并结合网络分析、计算机辅助内容分析和人工编码^[14]。Jiang 等人则通过微博数据爬取和在线调查研究了 HPV 疫苗相关主题的传播模式^[15]。定量研究主要集中在文本和视频平台上。Himelboim 等人和 Agergaard 等人研究了 Twitter 和 Facebook 上的 HPV 疫苗讨论^[14,16]。Chen 等人使用健康信念模型和解释阴谋论文献分析中国网络空间中的 HPV 疫苗话语^[13]。Li 等人通过分析微博平台上的帖子, 探讨了 HPV 疫苗相关信息的性别不平等问题^[17]。目前的研究存在几大不足: 样本量有限, 理论框架研究较少, 公共卫生领域深度学习模型应用不足, 多平台对比研究较少。中文文献研究主要集中在微博和知乎平台, 样本数据量小, 研究方法以内容分析和语义网络分析为主, 存在较大改进空间。未来的研究可以通过扩大样本量, 采用更加多样化的理论框架和研究方法, 以及进行多平台对比分析, 来弥补当前研究的不足。

当前中国社会正处于转型期, 在经济发展与性观念开放等外界因素共同作用下, 宫颈癌的发病率持续上升, 对女性身心健康构成了严重威胁。为了有效应对这一挑战, 我们首先需要深入了解 HPV 疫苗在我国公众中的认知现状, 这不仅能为社交媒体平台的治理策略提供参考, 还能为公众政策的宣传提供有力参考。社交媒体作为信息传播的重要渠道, 有责任宣传宫颈癌的危害, 推广 HPV 疫苗的健康传播知识。无论民众的接种意愿如何, 媒介都应致力于制作内容全面、价值观正确的健康知识普及报道, 从传播角度助力降低宫颈癌等相关疾病的发病率。同时, 对于社交媒体上可能出现的错误信息, 我们也需要积极治理, 确

保公众能够获取到准确、可靠的健康信息，从而提高对 HPV 疫苗的认知度和接受度。

1.2 研究概况和研究意义

为了研究 HPV 疫苗议题在社交媒体上的多模态内容差异，并探索在内容层面的情绪差异，本研究选取微博、喜马拉雅、抖音三个平台上的数据作为文本、音频和视频三类模态内容的例子，其中包含 273,357 条微博帖子、1,228 个喜马拉雅音频和 1,225 个抖音视频。目前学界在直接对比不同模态的内容时面临挑战，因此通常需要将它们转化为同一模态的信息（如文本）后，再进行深入的比较和分析。分别对不同模态的内容数据分析后发现，音频和视频内容关于 HPV 疫苗议题讨论的内容多集中在人声部分。因此，对音频和视频模态的人声部分进行了提取，并通过语音识别技术将其转化为长文本。由于数据量较大，为了发现内容差异的多样性，本研究引入了健康信念模型（Health Belief Model, HBM）的四个维度——易感性、严重性、效益和障碍。利用自然语言处理技术和人工编码，将三种平台的文本数据按照这四个维度进行了分类，并通过主题建模对比了各平台间的差异。同时，还比较了三种平台在文本内容层面的情绪差异。

本研究在理论、研究方法和实践上均有重要的意义。在理论上，本文在文本、音频、视频的社交媒体平台上应用了健康信念模型。先前的研究多集中于应用健康信念模型理论于单一文本或视频平台，而本研究则进一步将其拓展到跨模态和多平台领域，特别地，本研究包括了音频平台喜马拉雅，这是关于中国 HPV 疫苗讨论的首次研究涉及该平台，从而扩大了研究的视角和深度。本文从健康信念模型的各个维度出发，对比了文本、音频和视频平台上的多模态内容差异，这是健康信念模型的理论发展上的创新。当对内容进行健康信念模型四个维度的分类时，内容的表述可以同时涵盖健康信念的多个维度，而非仅限于单一维度，这种方法为健康信念模型的应用提供了新的范式。

此外，在研究方法上面有四个意义。第一，本研究对比了文本、音频、视频三种社交媒体平台上关于 HPV 话题内容的差异。先前关于 HPV 疫苗在社交媒体上的内容研究主要局限于文本平台，而本研究扩展到了音频和视频平台。鉴于不同模态的内容在直接比较上存在困难，且音频和视频中的关键内容多集中在人声部分，因此本研究对音频和视频中的人声进行提取和转录，将三种模态的内容转化为文字层面（即短文本和长文本），以便进行内容分析和建模。这种跨模态的研究方法有助于更全面地理解 HPV 疫苗在社交媒体上的讨论态势，并揭示不同模态内容之间的差异。第二，本研究利用深度学习技术对大数据进行了全面分析。从文本、音频、视频等平台收集多模态数据，并通过人声提取和文本转录将

其聚焦在文字层面的比较上，最终应用深度学习技术对全部内容数据进行分析。相比以往研究主要采用抽样的方式分析几百条样本，本研究实现了大规模数据的深度学习分析。第三，本研究创新性地在健康信念模型的维度上对比了短文本（如微博、推特的 200 字以内内容）和长文本（音频、视频转录后的人声文本内容，达几千至上万字）。在以往基于健康信念模型的分类研究中，长文本和短文本因长度和数量的差异难以直接比较，而本研究通过在长文本和短文本中包含多个健康信念模型维度的内容，进行了多次二分类任务，从而实现了长文本和短文本在各维度上的对比分析。第四，在主题建模方面，由于长文本和短文本使用的建模方法不同，本研究采用了基于深度学习的 BERTopic 模型。在对比三种模态内容在健康信念模型各维度上的差异时，为了更深入探究长文本的潜在主题，本研究将长文本拆分为短文本，并基于这些短文本进行主题建模，从而在主题建模内容对比层面进行了创新。

在实践上，本研究探究了不同模态内容在社交媒体上所呈现的特性，有助于根据特性进行针对性的平台治理和政策宣传。此外，本研究还探讨了公众在中国多种社交媒体上对 HPV 疫苗等议题的认知现状，研究结果有助于国家和企业制定相应的政策和措施。

1.3 技术路线

本文研究 HPV 疫苗相关议题在社交媒体上的多模态内容。整体的研究路线如图 1.1 所示。具体内容如下所示：

第一步，通过关键词收集 2018 年 7 月至 2023 年 3 月微博、喜马拉雅、抖音三个平台上关于 HPV 疫苗讨论的数据作为文本、音频、视频三类模态内容的例子，其中包含 273,357 条微博帖子、1,228 个喜马拉雅音频和 1,225 个抖音视频。

第二步，对这些来自不同模态的数据进行精细的预处理。鉴于不同模态的内容难以直接对比，本研究采取了针对性的处理策略，特别是音频和视频，这些模态中的关键信息多集中于人声部分。预处理方法是先将音频中的人声数据转换为文本形式，以便进行后续分析。具体的处理方法是：1）对于视频数据，利用 FFmpeg 软件提取出音频数据，并构建了相应的音频数据集。2）通过 Spleeter 包对音频中的人声部分进行精准分割，再借助科大讯飞的语音转写（ASR）API，将音频数据转录为简体中文文本。3）对文本数据集进行了细致的清洗，过滤掉了与 HPV 疫苗无关的内容，如商业广告、非中文数据等。

第三步，为了分析不同模态内容对 HPV 疫苗讨论的影响，本研究根据健康信念模型的四个核心维度——易感性、严重性、效益和障碍，对三个平台的文本数据进行了分类。由于一条内容会涉及到健康信念模型的多个维度。对于微博平

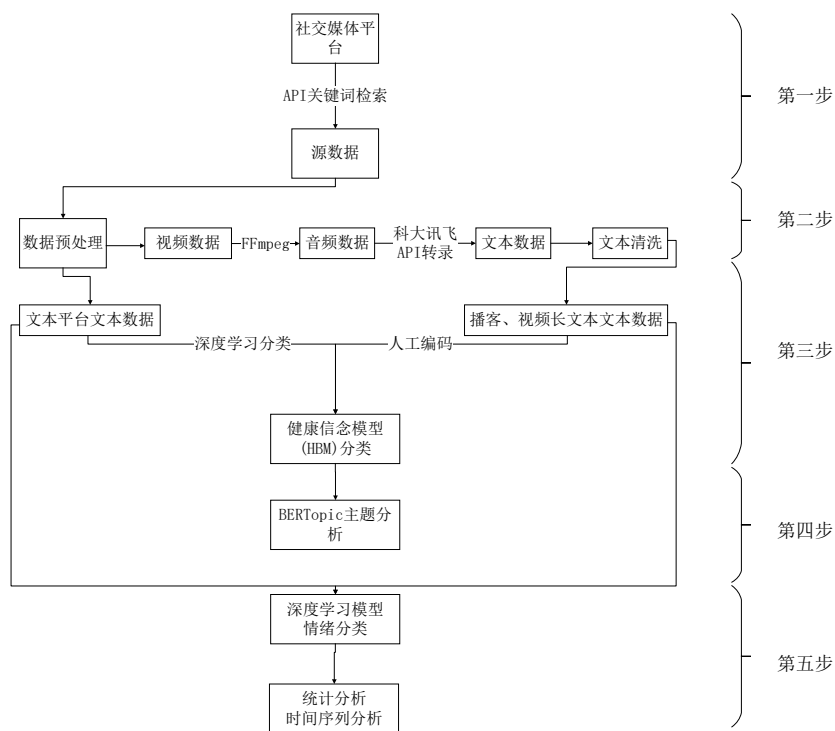


图 1.1 技术路线

台的短文本数据，本研究训练了四个深度学习模型，每个模型针对健康信念模型的一个维度进行分类，从而全面捕捉文本中涉及易感性、严重性、效益和障碍的信息。对于喜马拉雅音频和抖音视频中的长文本数据，采用人工编码的方式，本研究采用人工编码的方式，根据健康信念模型的四个维度进行分类。

第四步，在内容分析方面，本研究采用了 BERTopic 主题建模技术，该技术基于预训练的 Transformer 语言模型 BERT 生成文档嵌入，进而对这些嵌入进行聚类，最终通过基于类的 TF-IDF 过程生成主题表示。对三个社交媒体平台的数据在健康信念模型的四个维度下进行了分类，并对生成的十二个小数据集进行了深入的内容分析。BERTopic 模型将语料库聚类为不同的主题簇，本研究进一步分析了每个平台在健康信念模型的四个维度下前三个最具代表性的主题簇，以揭示不同模态内容对 HPV 疫苗讨论的差异。

第五步，为了探究不同模态内容在情感表达上的差异，本研究采用了基于 BERT 中文预训练版本的深度学习模型，并训练了情绪分类模型。该模型能够准确识别文本中的积极情绪、消极情绪和无情绪状态，从而帮助本文深入探究来自不同模态的内容在 HPV 疫苗讨论中的情感倾向差异。

第2章 国内外研究现状

2.1 公众健康视角下的 HPV 疫苗议题

健康机构和专家利用社交媒体平台传播健康信息，使公众可以以较低的成本轻松获取健康信息。此外，公众愿意根据自己的经验在社交媒体上生成和分享健康信息^[1]。研究人员观察到公众讨论了许多健康话题，如食品安全问题等^[2]。学者们发现社交媒体并不总是发挥积极作用，信息也可能存在误解和偏见，从而对人们的健康行为产生负面影响^[3]。例如，研究人员通过对英国居民中进行与 COVID-19 相关的社交媒体使用问卷调查，发现 COVID-19 健康保护行为与将社交媒体作为信息来源之间存在负相关关系^[4]。因此，对社交媒体健康话题进行内容分析，可以帮助我们了解公众参与健康传播活动的方式。

本研究关注中国社交媒体上与 HPV 疫苗相关的健康话题。人乳头瘤病毒 (Human papillomavirus, HPV) 是一种常见的病毒，它与宫颈癌息息相关。宫颈癌在中国仍然是一个重大的公共卫生威胁。中国是全球宫颈癌发病率和死亡率第二高的国家，每年约有 110,000 例新病例和 59,000 例死亡病例。宫颈癌在中国的发病率排在癌症发病率的第六位，死亡人数排在癌症死亡人数的第七位^[18]。然而，在中国，HPV 疫苗的接种率比较低。2020 年，中国仅有不到 1% 的年轻女孩接种了疫苗^[8]。为了消除宫颈癌，中国政府已经启动了一项全国性的 HPV 疫苗接种计划，旨在到 2025 年使 90% 符合条件的 15 岁以下女孩接种疫苗^[9]。自从 2017 年首个 HPV 疫苗在中国获得批准以来，中国公众一直在社交媒体上讨论 HPV 疫苗的各种话题。2023 年 8 月，HPV 疫苗接种相关话题标签成为中国微博平台微博上最热门的话题之一^[15]。因此，本研究选取了 HPV 疫苗作为研究的话题。

2.1.1 人乳头瘤病毒 (HPV)

人乳头瘤病毒 (Human papillomavirus, HPV) 感染是极为普遍的病毒感染，其传播途径多样，包括皮肤接触和黏膜接触，通过皮肤或黏膜的微小损伤侵入体内。尽管大部分 HPV 感染能够被人体免疫系统自然清除，但它依然是最常见的性传播疾病之一。全球范围内，男性和女性在其一生中至少有 50% 的几率感染 HPV^[19-20]。

HPV 感染不仅会导致生殖器疣等皮肤问题，还与一系列非皮肤疾病相关。当前，HPV 主要通过性接触传播。^[19-20]。据美国疾病控制和预防中心 (Centers for Disease Control and Prevention, CDC) 的数据，HPV 感染在男性和女性中均极为常见，每年约有 0.13 亿美国人 (包括青少年) 感染 HPV 病毒^[21]。在中国，HPV

感染率同样居高不下。据统计，中国约有 0.5582 亿 15 岁及以上的女性感染了 HPV，面临着罹患宫颈癌的风险。每年，有高达 109,741 名女性被诊断出患有宫颈癌，其中近 59,060 人因此离世，使宫颈癌成为 15 至 44 岁女性中第三大最常见的癌症^[6]。

值得注意的是，世界卫生组织（World Health Organization, WHO）的数据显示，每年超过 85% 的宫颈癌死亡病例发生在低收入和中等收入国家，这凸显了发展中国家妇女在宫颈癌防治方面所面临的严峻挑战。中国的数据也反映了这一趋势，即发展中国家妇女罹患宫颈癌的风险更大。然而，这些地区的公众对 HPV 预防的认识普遍较低，这进一步凸显了针对发展中国家妇女实施有效干预措施的紧迫性^[22]。

2.1.2 人乳头瘤病毒疫苗 (HPV 疫苗)

HPV 疫苗接种已被广泛证实为预防 HPV 感染的有效手段。2006 年，第一代 HPV 疫苗 Gardasil 经美国食品和药物管理局（Food and Drug Administration, FDA）批准，用于预防 HPV 6 型、11 型、16 型和 18 型的感染。而到了 2009 年，FDA 又批准了专门用于预防 HPV 16 型和 18 型感染的 Cervarix 疫苗，这两种病毒类型被认为是最具危险性的。HPV 疫苗的使用极大地保护了妇女免受宫颈癌的威胁，为消除这一恶性肿瘤带来了希望。鉴于宫颈癌对全球女性健康的严重威胁，世界卫生组织（WHO）在 2020 年 8 月通过了一项具有里程碑意义的文件——《消除宫颈癌的全球战略》。该战略提出了三大核心措施，旨在推动实现“90-70-90”目标，进而在 2030 年前全面消除宫颈癌。这三大核心措施是：确保 90% 的女孩在 15 岁前完成 HPV 疫苗接种；保证 70% 的女性在 35 岁前接受高效检测筛查，并在 45 岁时再次进行筛查；以及确保 90% 的患有早期癌症和 90% 的患有侵袭性癌症的妇女得到适当的治疗和管理。各国应在 2030 年前达到这一“90-70-90”的目标，以期在未来几十年内彻底消除宫颈癌^[23]。

截至 2023 年 4 月，全球已批准了六种预防性 HPV 疫苗，且在世界卫生组织 194 个成员国中，有 134 个（即 69%）已将 HPV 疫苗接种纳入其国家免疫计划（National Immunization Program, NIP）^[24]。然而，尽管有这一积极进展，全球范围内 HPV 疫苗接种的覆盖率仍然远低于 2030 年消除宫颈癌的目标。具体而言，截至 2022 年，15 岁女性接种 HPV 疫苗的第一剂全球覆盖率仅为 21%，远低于 90% 的目标。回顾近年来的数据，可以发现这一覆盖率呈逐年上升趋势，但增速显然不足以实现既定目标。例如，2021 年的覆盖率为 20%，2020 年为 17%，2019 年为 15%，2018 年为 13%，而到了 2010 年，这一数字仅为 3%。因此，面对这样的增长趋势，要实现 2030 年 90% 的消除宫颈癌目标，发展中国家面临着巨大的挑战^[25]。未来应注重确保更多国家将 HPV 疫苗引入其免疫计划，特别是对于

宫颈癌负担较重的国家，以进一步提高全球免疫计划的执行效果。

2.1.3 人乳头瘤病毒疫苗 (HPV 疫苗) 在中国的现状

中国作为宫颈癌的高发地区，其病例数占据了全球总数的约五分之一，其中2020年新增病例高达109,741例，这一数据不容忽视^[26]。然而，令人遗憾的是，由于中国食品药品监督管理局（China Food and Drug Administration, CFDA）对相关药品的审批延迟，葛兰素史克公司的Cervarix疫苗直至2017年7月31日才正式登陆中国市场，而Gardasil9更是在2018年4月28日才获得有条件上市许可。在这两款疫苗问世之前，中国大陆仅依靠薄层细胞学检测试剂盒（ThinPrep Cytology Test, TCT）作为预防宫颈癌的手段。这意味着，首款人乳头状瘤病毒（HPV）疫苗的上市时间被推迟了近十年，导致约1.14亿名9-14岁的女孩错过了预防宫颈癌及其癌前病变的最佳时机^[27-28]。

有研究显示，若在2006年至2015年间，中国实施面向9-14岁女孩的国家性HPV疫苗接种计划（采用二价HPV疫苗，接种率达到90%），则可预防该年龄段内783,000例宫颈癌病例和435,000例相关死亡^[28-29]。然而，中国目前尚未将HPV疫苗纳入国家免疫规划（National Immunization Program, NIP）。据预测，若NIP的启动继续推迟，在2022年至2030年期间，每年可能新增超过11.9万例宫颈癌病例和4.2万例相关死亡^[30]。为了弥补这一昂贵的延误并加速HPV疫苗接种的全面推广，中国采取了积极的应对措施。这些措施包括自主研发国产HPV疫苗、探索个性化策略、启动免费疫苗接种计划以及将HPV疫苗纳入国家免疫规划等。

当前，中国市场上的HPV疫苗主要由私人购买，且主要面向成年女性，而非9-14岁的年轻女孩（2020年疫苗接种覆盖率不足1%）。因此，将HPV疫苗接种纳入国家免疫规划，确保到2030年前90%的女孩在15岁前接种疫苗，显得尤为重要。在HPV疫苗接种计划全面铺开之前，鼓励资源充足的地方政府先行试点，以提高接种率，并推动其在全国的普及^[31]。

2021年3月，中国国家卫生委员会启动了健康城市创新试点计划，旨在全面预防和控制宫颈癌。该计划设定了到2025年的目标：实现90%的女孩在15岁前接种疫苗，提高35-64岁妇女的筛查覆盖率（特别是超过70%的妇女在35岁和45岁前接受筛查），以及确保90%的妇女接受宫颈癌和癌前病变的治疗^[32]。在这一框架下，多个资源丰富的试点城市率先启动了地方政府资助的青少年HPV疫苗接种计划，为中国HPV疫苗免费接种计划的全面铺开奠定了坚实基础。

在两年的试点期间，大多数地区针对13-14岁的女孩采取了从学校通知到指定医疗机构或社区卫生服务中心的疫苗接种策略。同时，试点地区为HPV疫苗接种提供了多种灵活的融资方式，包括特定类型HPV疫苗的免费接种、任意类

型 HPV 疫苗的固定经济补偿,或二者结合的方式。在此过程中,所有相关利益方,如卫生委员会、教育部门、财政部门、疾病预防控制中心以及医疗产品管理局等,均积极参与并推动了 HPV 疫苗接种工作的进展。截至 2022 年底,所有实施 HPV 疫苗免费或经济补偿政策的试点城市均实现了目标女孩群体的高接种率,接近或超过 80% (具体比例因城市而异)。这一成果得益于多方面的共同努力,如地方政府的强势推进、教育部门的积极动员与组织、多部门的协同合作以及政府的财政支持。

中国的 HPV 疫苗免费接种试点可以被视为一个“边做边学”的过程。在试点成功的基础上,广东、海南、福建、江苏、西藏、重庆和江西等多个省级行政区已向所有 13 或 14 岁的目标女孩推出了免费或补贴 HPV 疫苗接种政策。截至目前,已有数十个地区实施了类似政策,覆盖了中国 17 个省级行政区。

中国对加速消除宫颈癌的坚定承诺,为中国推出 HPV 疫苗接种提供了前所未有的机遇和动力。这一决心在政策层面得到了充分体现。首先,在 2021 年 9 月,中国妇女发展计划(2021-2030)明确指示在特定试点地区推广 HPV 疫苗接种,为后续的全面推广奠定了坚实的基础^[33]。紧接着,到了 2022 年 1 月,政府进一步展现了其积极的态度和决心。从试点地区开始,政府逐步在全国范围内推广免费的 HPV 疫苗接种计划,这一举措无疑将极大地提高 HPV 疫苗的接种率,为更多女性提供保护^[31]。到了 2023 年 1 月,这一进程得到了进一步的加速。10 个中央政府部门联合发布了加速消除宫颈癌行动计划(2022-2030 年),其中明确提出了在全国范围内不断推广并提供女孩 HPV 疫苗试点服务的计划^[32]。这一政策的出台,不仅彰显了政府对于 HPV 疫苗接种工作的高度重视,也为中国在全球范围内实现宫颈癌的消除目标注入了强大的动力。

2.2 HPV 疫苗议题与社交媒体

2.2.1 国外文献综述

为了全面地掌握国内外对 HPV 疫苗和社交媒体既有研究,本文分别通过 Web of Science 和知网数据库,对 HPV 疫苗和社交媒体的研究进行了检索。在 Web of Science 上的文献检索流程为:选取 Web of Science 核心合集数据库中的 Social Sciences Citation Index (SSCI) 引文索引,检索所有字段中包含“HPV Vaccine”“Social Media”词段的文献,剔除掉会议摘要、修订、社论材料等类型的文献,一共得到 171 篇文章,其发表年份及文献类型见下图 2-1 所示。

这 171 篇文献的主要发表期刊如表 2-1 所示。在传播学的相关领域,HPV 疫苗和社交媒体的研究主要集中发表在期刊 Health Communication 和 Journal of Health Communication 上。传播学期刊或会议上最早关于 HPV 疫苗与媒体的研究

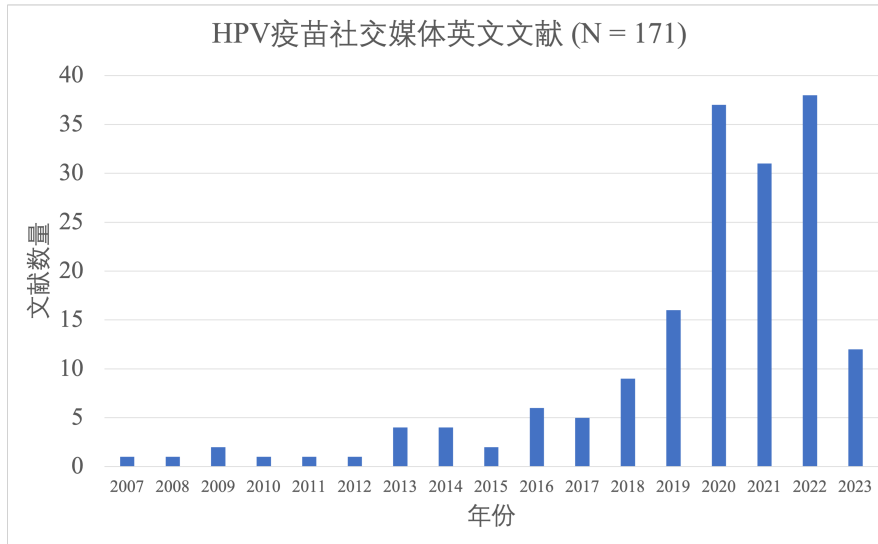


图 2.1 HPV 疫苗社交媒体英文文献发表年份与文献类型

数据截止至 2024 年 1 月 1 日。

于 2012 年发表在期刊 *International Journal Of Press-Politics* 上，国外学者 Fowler 等人^[10]在探讨竞争框架时，特别选取了 HPV 疫苗作为研究议题。他们深入分析了美国 50 个州内 101 家报纸对 HPV 疫苗的报道，通过对比分析内容的方法，旨在揭示在媒体引发的争议中如何更有效地影响公众舆论。此后，关于 HPV 疫苗和社交媒体的研究文献在传播学期刊和会议中逐渐增多。这些研究文献可以划分为定性研究、混合研究和定量探究三种类型。从数据采集的角度来看，研究者们主要采用了三种方法。一是直接采集社交媒体上的客观数据，这种方法能够获取大量的用户行为和信息传播数据；二是通过在线或线下实验的方式发放问卷，以收集用户对 HPV 疫苗的看法和态度；三是进行线下实验，利用眼动追踪技术收集用户在浏览 HPV 疫苗相关信息时的视觉数据。在定量研究方面，内容分析法是常用的研究方法之一。该方法主要分为三个类别：研究者抓取社交媒体上的客观数据，运用计算机技术进行建模分析；直接对客观数据进行人工编码，以识别和分析文本中的主题、情感倾向等关键信息；对文本信息进行语义网络分析，通过构建知识图谱来探究不同信息点之间的关系。

关于 HPV 疫苗和社交媒体的定性研究主要是通过问卷调查和在线实验来收集数据。例如，Lee 等人^[11]探讨了博客上的个人健康故事如何影响个体的社会规范感知，采用了社会规范是“基于群体身份的行为准则”的概念。通过测试故事框架和个人经历对感知相似性的影响，研究发现相似性与社交媒体指标互动，从而形成感知的社会规范。在一项在线实验中，220 名大学生阅读了博主的增益框架（接种 HPV 疫苗预防癌症）或损失框架（未接种疫苗患癌）故事。结果显示，参与者在增益框架的故事中觉得与博主更相似，特别是那些接种了 HPV 疫苗的人。这种感知相似性与描述性规范和规定性规范均正相关；但当博客的社交

表 2.1 HPV 疫苗社交媒体英文文献的主要发表期刊

期刊	文献数量
Vaccine	17
Human Vaccine Immunotherapeutics	14
Journal of Medical Internet Research	9
Health Communication	7
International Journal of Environmental Research and Public Health	7
Frontiers in Public Health	6
JMIR Public Health and Surveillance	5
Journal of Behavioral Medicine	5
Journal of Health Communication	5
Vaccines	5
BMC Public Health	4
Frontiers in Psychology	4
Preventive Medicine	3
Preventive Medicine Reports	3
Telematics and Informatics	3
BMC Women's Health	2
BMJ Open	2
Ethnicity Health	2
Health Promotion International	2
Healthcare	2
Journal of Adolescent Health	2
Journal of Community Health	2
Journal of the American Medical Informatics Association	2
PLOS One	2
Preventing Chronic Disease	2
Qualitative Health Research	2
Social Science Medicine	2

注: 仅列举发表了两篇及以上文献的期刊

媒体指标较低时, 相似性与描述性规范之间的正相关性消失。在 Jiang 等人^[12]于 2022 年 8 月的研究中, 他们在中国进行了两波在线调查, 他们基于认知中介模型, 考察了影响社交媒体用于人乳头瘤病毒 (HPV) 信息的动机, 并探讨了社交媒体使用与 HPV 知识之间的潜在路径。Chen 等人^[13]在 2020 年通过在线实验使用使用计划行为理论研究接触社交媒体上的阴谋论信息对中国年轻人 HPV 疫苗认知的影响。

混合方法研究结合了定量的大数据内容分析和访谈、在线实验数据的定性分析。例如, 在 2020 年国外学者 Himelboim 等人^[14]通过 Twitter 数据来了解 HPV 疫苗争议的对话内容特征和内容流动模式。结合网络分析、计算机辅助内容分析和人工编码内容分析, 发现积极情绪与群集密度呈正相关, 其次发现对 HPV 疫苗有把握的人和正在寻找答案的人聚集在一起。也有国外学者 Jiang 等人^[15]采用混合方法, 爬取了微博上关于 HPV 疫苗的 67773 篇帖子, 并进行了主题建模, 以确定在微博上流行的 HPV 疫苗相关主题。结果显示, 关于 HPV 疫苗的几个主要主题分别是政策、指导信息、广告、丑闻、个人经验分享和 HPV 风险。其次,

他们进行了一项样本量为 1982 的在线调查，以研究如何扫描、寻找和讨论从大数据分析中确定的六个 HPV 疫苗主题，可以影响 HPV 疫苗知识、安全关注和接种意愿。发现社交媒体健康传播显著影响用户健康知识、态度和行为意图。Xiao 等人^[34]设计实验和问卷考察了基于规范的信息（积极描述性、消极描述性、规定性和基本信息）对 18-29 岁的年轻人接种人乳头瘤病毒（HPV）疫苗的意愿和风险认知的影响。实验结果显示，与基本信息条件相比，额外接触到规定性规范信息增加了寻求更多关于 HPV 疫苗的信息的意愿，进而增强了接种疫苗的意愿。应谨慎使用消极描述性规范信息，因为它比积极描述性规范和基本信息更容易导致对 HPV 疫苗的高风险认知。

定量研究是直接对客观数据进行分析，目前的文献主要集中在两类社交媒体平台，文本平台和视频平台。在文本平台上的研究比较丰富。例如，国外学者 Himelboim 等人^[14]在 2020 年通过 Twitter 数据研究了 HPV 疫苗争议的对话内容特征和内容流动模式。学者 Agergaard 等人^[16]在 2020 年研究了在丹麦的 Facebook 页面上的 2012 年到 2019 年关于人类乳头瘤病毒（HPV）疫苗错误信息的讨论。Chen 等人^[13]在 2020 年的研究中通过两个理论视角，健康信念模型（HBM）和解释阴谋论的文献，研究了中国网络空间中关于人乳头瘤病毒（HPV）疫苗的媒体话语。Li 等人^[17]在 2022 年的研究关注了微博平台上 1,221 篇帖子中的 HPV 疫苗相关信息，他们通过定量内容分析发现，大部分内容来自非健康专家，并且强调了 HPV 疫苗的效果，带有推荐的语气；从性别角度看，女性预防性传播疾病的责任比男性更被强调，这引发了对应对 HPV 感染时性别不平等的担忧。也存在着一些关于 HPV 疫苗和社交媒体的研究基于视频平台。例如，Kim 等人^[35]在 2021 年 12 月的研究使用了内容分析法，检查了在韩国推广人乳头瘤病毒（HPV）疫苗的 201 个 YouTube 视频中几种消息策略的效果。他们发现，YouTube 视频的创作者经常使用的消息策略有时间框架、叙事格式、收益-损失框架和消息感知价值，而不使用情感诉求。

定量研究也有对眼动实验数据进行分析。例如，Kim 等人^[36]通过眼动实验研究了人们在社交媒体上接触到错误信息和更正信息时的注意力和可信度。通过对比幽默和非幽默的更正策略，研究发现幽默更正增加了对更正推文中图片部分的关注，降低了错误信息推文的可信度，间接减少了对 HPV 疫苗的误解。然而，非幽默更正由于其较高的可信度评分，在减少误解方面表现更佳。文章讨论了在社交媒体上更正错误信息的实际应用。

定量研究也涉及两种社交媒体平台的对话。例如，Coman 等人^[37]通过理论视角探讨了 2019 年 HPV 宣传日背景下在 Twitter 和 Instagram 上的风险沟通对话，聚焦于与 HPV、HPV 相关癌症以及 HPV 疫苗相关的污名化表达。研究结果揭示了自我污名和外污名均有体现，涉及非营利组织、官方代言人以及普通用

户。此外，其他与污名相关的类别也出现了，这些类别可以被视为对污名的反应以及对改善个人和社会态度的呼吁（如消除刻板印象），这些内容来自官方和非官方来源，涵盖支持和反对疫苗的言论。尽管这两个平台上出现了相同的类别，但在叙事和信息传递方面存在差异。研究讨论了这些发现的实际影响。

经过文献梳理，目前英文的文献还存在着以下几点不足。第一，定量研究数据样本较少，仅仅是从海量的大数据里面选取了一部分来进行分析，而不是对全部的大数据进行分析。第二，部分定量研究的文献缺少理论框架的支撑，这些文献直接通过客观数据人工定性总结框架。第三，目前的文献对深度学习技术使用较少。第四，目前的文献研究对象只涉及单个或者两个平台的客观数据，没有研究三个或三个以上平台的客观数据。

2.2.2 国内文献综述

在知网上的文献检索流程为：检索主题包含“HPV 疫苗”“社交媒体”的所有期刊文章，选择北大核心和 CSSCI 索引的期刊，一共得到 5 篇文献，其发表年份及文献类型见图 2.2 所示，发表期刊见表 2.2 所示。与英文文献相比，中文文献的研究开始时间较晚，数量较少。

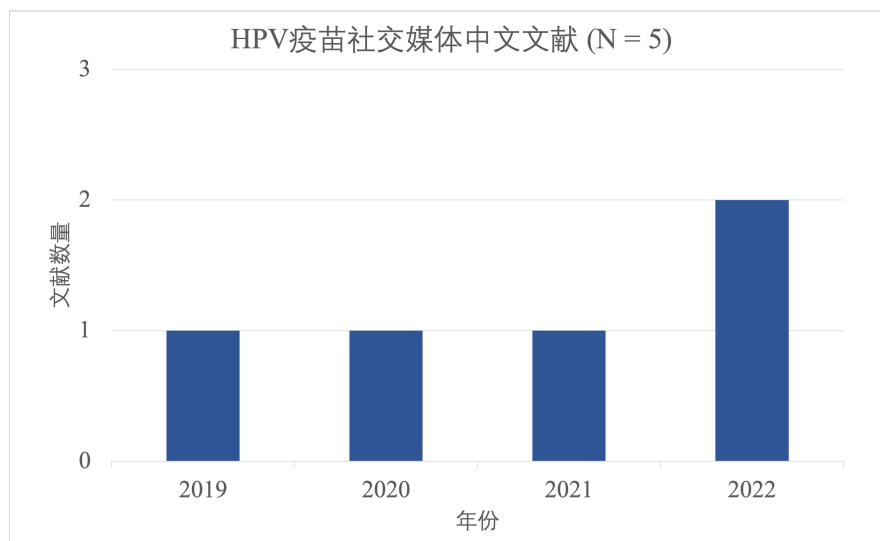


图 2.2 HPV 疫苗社交媒体中文文献发表年份与文献类型

数据截止至 2024 年 1 月 1 日。

与 HPV 疫苗和社交媒体相关最早的中文研究是在 2019 年由向菲和冯思佳撰写的基于 HPV 疫苗的新浪微博数据的量化研究，经过爬虫抓取 27715 条微博数据为研究对象，根据对微博数据转赞评的数量分类为高关注度微博和低关注度微博，对高关注度微博进行人工编码进行主题划分，对低关注度的微博进行 LDA(Latent Dirichlet Allocation) 主题建模^[38]。在这之后，2020 年，郭小安和王天翊通过问卷调查法，引入健康信念模型，尝试以健康信念为中介变量，在分析女

大学生群体对 HPV 病毒/疫苗认知状况基础上,考察新媒体接触对女大学生健康信念和接种意向的影响^[39]。在这之后,姚志臻、张斌和司湘云在 2021 年通过对知乎社交媒体平台上的 519 篇优质回答进行分析,通过内容分析法从疫苗情绪、疫苗类型、目标人群和信息来源四个方面讨论问答内容特征;其次,构建疫苗情绪语义网络,分析疫苗情绪网络特征和核心概念,了解不同情绪网络下的主题差异^[40]。在这之后,孙鹿童在 2022 年以微博 HPV 疫苗相关视频为研究对象,对选取样本 132 个进行内容分析,引入扩展的平行过程模式 (expended parallel process model, EPPM) 理论,通过人工编码分析 EPPM 中的“威胁”和“效能”,探究有可能影响公众健康态度和疫苗健康决策的社交媒体内容如何呈现 HPV 疫苗相关议题^[41]。在几个月之后,商建辉和赵昱也在 2022 年通过对新浪微博官方媒体账号发布的 HPV 疫苗 200 条博文数据及 1608 条微博评论数据进行内容分析,考察使用的不同主题框架与用户在线疫苗情绪表达的关系^[42]。

经过文献梳理,中文文献关于 HPV 疫苗在社交媒体上的研究基本上集中于微博和知乎这两个社交媒体平台,既有定量研究也有定性研究,数据来源有客观数据和问卷调查,但是样本数据量都比较小。分析的数据形式有文本数据和视频数据,但是都是以人工编码为主。研究方法以内容分析和语义网络分析为主,其中内容分析主要用到 LDA 主题建模等的机器学习的方法。基于此研究现状,本文在数据样本,理论框架,研究方法等方面都有很大的改进。

表 2.2 HPV 疫苗社交媒体中文文献的主要发表期刊

期刊	文献数量
青年记者	1
医学与哲学	1
信息资源管理学报	1
新闻与传播研究	1
中国公共卫生	1

2.3 不同模态内容的社交媒体平台

“模态”被定义为“通过媒介递送或呈现信息的方式”,涵盖了文本、音频、视频等多种形式^[43]。其中,音频模态相较于文本,展现出其独特的特征,这些特征包括不同说话人声音的多样性^[44]、音乐的巧妙融入、以及制作效果或创意声音的强调^[45]、音效的精细运用以及沉默的恰当留白。值得注意的是,相较于视频内容,复杂的纯音频内容在提升注意力和增强记忆能力方面展现出了显著的优势^[46]。

此外,视频模态在信息传递中同样占有举足轻重的地位。研究显示,人们往往更倾向于信任视频内容,因为它与现实世界的相似度更高^[47]。先前的研究进一步支持了视频在健康信息传播中的有效性。例如,Huang 等人^[48]在 2009 年的

研究中发现, 视频不仅具有高度的视觉吸引力, 而且在解释和阐述各种健康概念和过程时发挥着重要作用。通过使用视频, 可以有效提升患者的健康知识水平和理解能力^[49], 增强医患之间的信任^[48], 进而促进患者采取更健康的行为方式^[50-51]。这些发现均彰显了视频模态在健康信息传播领域的独特价值和潜力。

在社交媒体这一领域, 信息通过多元化的模态(如文本、音频和视频)进行传递。随着多媒体技术的飞速发展, 用户能够在不同的社交媒体平台上接触到形式多样的信息源^[52]。例如, 在文本内容的社交媒体平台(如 Twitter 和 Reddit)上, 用户可以接收和讨论基于文字的信息; 在音频内容的社交媒体平台(如 Spotify 播客)上, 用户则可以通过音频内容接收信息; 而在视频内容的社交媒体平台(如 TikTok 和 YouTube)上, 用户则能直观地观看并参与到视频内容的讨论中。先前的实证研究进一步揭示了, 在健康传播的语境下, 不同模态的特征可能对受众感知健康信息的方式产生深远影响^[53]。因此, 本研究旨在深入探讨和分析不同类型社交媒体平台上 HPV 疫苗多模态内容。

2.3.1 文本社交媒体平台

1. 文本社交媒体平台上的 HPV 疫苗研究

基于文本的社交媒体平台, 其核心在于通过文本内容传递信息, 已然成为当今社会公共健康话题的重要讨论场所。HPV 疫苗, 作为此类平台上的热点议题之一, 其在全球范围内的讨论尤为热烈。在美国, 自从 2006 年第一款 HPV 疫苗问世以来, 有关 HPV 疫苗的信息就开始在 Twitter 上广泛讨论, Dunn 等美国学者深入分析了 Twitter 上的 HPV 疫苗信息接触情况, 以探究这一信息接触能否解释美国各州疫苗接种覆盖率的差异。他们选取了 2013 年 10 月 1 日至 2015 年 10 月 30 日期间的 258,418 条推文, 累计信息接触量高达 273.8 百万次, 并运用机器学习方法对这些推文进行了细致的主题分类。通过构建多变量模型, Dunn 等人利用各主题的比例性信息接触预测了美国各州的 HPV 疫苗接种覆盖率, 并将其与基于社会经济因素(如贫困、教育和保险)构建的模型进行了对比。研究结果显示, 与疫苗接种覆盖率最为密切相关的主题, 主要围绕媒体争议展开, 无论是正面的还是负面的影响。此外, 教育和保险在社会经济因素中占据了举足轻重的地位。

值得注意的是, 信息接触度量解释了 2015 年女性 HPV 疫苗一剂覆盖率 68% 的差异(男性为 63%), 而基于社会经济因素的模型解释了女性 42% 的差异(男性为 40%)。这一发现表明, Twitter 上的信息接触对疫苗接种覆盖率具有显著影响, 尤其是那些社会经济因素无法完全解释的差异。进一步分析发现, 社交媒体上的负面言论, 尤其是关于疫苗安全性、错误信息和阴谋论的讨论信息接触量较高比例的州, 疫苗覆盖率较低。这一现象再次印证了社交媒体在公共健康议题中

的影响力不容忽视^[54]。

在丹麦，HPV 疫苗的宣传活动则主要依托于 Facebook 平台。Loft 等丹麦学者在 2020 年针对对 HPV 疫苗持怀疑态度的父母群体的研究中，评估了不同 Facebook 帖子对父母群体的共鸣程度，并探讨了如何在 Facebook 上与父母群体进行积极有效的沟通。他们利用一个名为“停止 HPV-停止宫颈癌”的 Facebook 页面，结合特定的社交媒体策略，吸引了大量目标受众——那些对给女儿接种 HPV 疫苗心存疑虑的母亲。通过深入分析 2017 年 5 月至 12 月间该页面上发布的各类帖子，他们评估了这些帖子的影响力、参与度以及评论的情感倾向（积极、中性或消极）。此外，他们还通过问卷调查收集了受众对这些帖子以及 FB 管理团队回应的态度。研究结果表明，相较于纯事实性的帖子，个人故事更能引起受众的共鸣，促进更积极的对话。然而，Loft 等人也强调，为父母提供准确、全面的事实信息，依然是推动他们对 HPV 疫苗做出决策的关键^[55]。

在中国，HPV 疫苗在文本平台上的讨论引起了广泛关注，比较多的是微博^[15]和知乎^[56]。Jiang 等人于 2023 年进行的研究深入探讨了社交媒体上 HPV 疫苗信息的传播以及这些信息如何影响中国的健康相关行为。他们通过混合方法的研究设计，首先收集了微博平台上发布的 67,773 条关于 HPV 疫苗的帖子，并运用主题建模技术，识别出六个核心主题，这些主题涵盖了政策、指导信息、广告、丑闻、个人经验分享以及 HPV 风险等方面。紧接着，他们通过一项在线调查（样本量为 1,982），进一步探讨了这些主题如何影响用户的 HPV 疫苗知识、安全关切以及接种意愿。研究结果表明，社交媒体健康沟通对用户健康知识、态度和行为意向具有显著影响^[15]。另一方面，J. Li 和 Zheng 在 2020 年的研究中，聚焦于知乎这一社交媒体平台上 HPV 相关信息的覆盖情况。他们运用 Python 网络爬虫技术，通过关键字搜索，识别和收集了 115 篇相关文章，这些文章的时间跨度为 2017 年至 2018 年。通过对这些样本进行多维度的编码分析，他们发现知乎上的文章主要聚焦于 HPV 与宫颈癌的关系、HPV 的严重性，以及 HPV 疫苗的益处。这些文章更倾向于使用主题框架来传播 HPV 相关知识，并且研究发现表述策略与信息覆盖之间存在显著关联。此外，这一研究还揭示了中国社交媒体平台上 HPV 疫苗讨论的独特性，与西方世界的 HPV 疫苗话语体系形成了鲜明对比^[56]。Zhang 等人于 2013 年的研究揭示了一个有趣的现象，即与美国公民相比，中国公民对 HPV 疫苗的知识相对缺乏，但对这一话题却表现出浓厚的兴趣。特别是那些对接种疫苗感兴趣的人，对疫苗的相关信息（如副作用和商业保险保障）表现出了强烈的好奇心^[57]。而 Y. Li 等人在 2022 年的研究中，专注于微博平台上关于 HPV 疫苗的讨论。他们发现，尽管微博上对 HPV 疫苗的功效给予了更多的强调，但对感染 HPV 的易感性和严重性的关注却相对较少。此外，研究还发现微博上的讨论存在性别差异，女性更倾向于参与关于性传播疾病预防的

讨论^[17]。

2. 文本社交媒体平台的 HPV 疫苗情绪研究

关于文本社交媒体平台上对 HPV 疫苗的情绪研究呈现出了多样化的观点。在 Luisi 的 2020 年研究中,对 Facebook 上与 HPV 疫苗相关的 6506 条帖子进行了深入分析,结果表明,高达 47.1% 的帖子表达了负面或反对的态度,这一比例显著超过了表达支持立场的 19.8% 的帖子。社交媒体上用户对 HPV 疫苗的态度并非一成不变^[58]。此外, Himelboim 等人在 2020 年的 Twitter 研究中则发现了积极情绪与集群密度之间的正相关关系。这里的集群密度是强连接和快速信息流的指标,这进一步揭示了社交媒体上情绪传播的动态性和集群效应^[14]。Zhang 等人在 2021 年的研究中,针对 HPV 疫苗的讨论进行了情绪分析,发现评论中包含了广泛的积极情绪、负面情绪和愤怒。值得注意的是,在接种率高和低的地区,帖子的情绪表达呈现出了一定的相似性,这为后续的研究者理解公众对 HPV 疫苗的情绪反应提供了新的视角^[59]。同样, Kearney 等人在 2019 年的关于 Instagram 研究中发现,虽然关于 HPV 疫苗基本效果的信息大多是正面的,但来自普通人的描述,尤其是疫苗消费者的描述,往往更加负面^[60]。在社交媒体平台上,公众对于 HPV 疫苗的情绪态度呈现出多元化和复杂化的特点,这需要本文进一步深入研究和分析。

2.3.2 音频社交媒体平台

音频社交媒体平台是以音频内容为主要形式的平台,例如播客平台。播客平台通过其去中心化的架构和逐步平台化的转型,凸显了其作为社交媒体平台的独特性质。播客平台不仅通过整合 RSS 和平台服务,促进了用户间的交流与互动,推动了社区的形成,还极大地助力了内容的广泛传播与分享^[61]。在过去的十年里,播客不仅得到了显著的发展和巩固^[62-63],而且其数量与听众基数均呈现稳步增长的趋势。这种增长不仅推动了播客行业的专业化进程,更使其成为了一种广受欢迎的媒体内容消费形式,展现出持续发展的市场潜力。播客不仅仅是传统广播的延伸,更是一种新兴的、独立的社交媒体形态,为公众提供了更加多元、丰富的信息获取和娱乐方式。

1. 音频社交媒体平台上的研究

在西方国家,特别是欧洲,关于播客的研究更加普遍。学者们利用内容分析和在线民族志的结合研究了播客粉丝社群。他们探索了这些受众的独特特点以及他们如何通过在线工具与数字媒体互动^[64]。其他学者将播客视为一种媒体消费形式,使用使用与满足理论作为框架,突出其多面性及其对个人效率的影响^[65]。播客是妇女健康护理实习生的有效学习工具,解决了提供者在治疗阴道疾病方面对知识的缺乏问题^[66]。播客还成为传播意识和促进疫苗接种的有力工

具，为青年作家提供了一个平台，引发对话并向同龄人教育了他们接种 HPV 疫苗的重要性^[67]。然而，目前关于播客平台在 HPV 疫苗方面的研究还很有限，尤其是在中国。

2. 音频社交媒体平台上的情绪研究

在播客平台上，关于情绪的研究日益受到关注，其影响力不容忽视。Lindgren 在近期（2023 年）的深入研究中探讨了播客制作人如何利用情感技巧，与听众建立起亲密的聆听体验^[68]。这一发现不仅彰显了播客在情感交流方面的独特作用，还为后续研究提供了宝贵的参考。此外，Y. Kim 等人（2016 年）的研究则聚焦于播客内容对个体政治情绪的影响^[69]。他们的调研揭示了这样一个现象：倾向于收听自由派播客的听众对保守派候选人的态度普遍呈现出较强的负面情绪，而对自由派候选人的情绪则呈现负相关关系。这一发现不仅揭示了播客内容在塑造个体政治情感方面的潜在作用，还为理解音频社交媒体在政治传播中的影响力提供了新的见解。综上所述，对音频社交媒体平台上的情绪研究具有显著的学术价值和实践意义。

2.3.3 视频社交媒体平台

基于视频的视频社交媒体平台是以音频内容为主要形式的平台，例如 Youtube、Tiktok。在本研究中，研究了抖音，即中国版 Tiktok，也是全球增长最快的视频平台之一^[70]。

1. 视频社交媒体平台上 HPV 疫苗的研究

针对 HPV 疫苗议题在视频社交媒体平台上的讨论情况，已有诸多研究展开探讨。深入理解抖音等平台上 HPV 疫苗信息的传播态势，对于构建高效的推广策略、吸引年轻群体并提高疫苗接受度具有举足轻重的意义。在这一过程中，支持疫苗的内容与健康专家的角色尤为关键，他们通过传递准确信息，有效影响着公众的疫苗认知^[71]。2023 年，Kirkpatrick 和 Lawrie 的研究进一步揭示了短视频社交媒体平台如 TikTok 在健康信息传递中的巨大潜力。他们发现，医生发布的自主支持性视频在提升年轻女性对宫颈癌刮片检查的态度和参与度方面起到了积极作用^[72]。然而，Basch 等人在 2022 年的研究中指出，尽管 TikTok 上的专业人士提及了与 HPV 相关的癌症，但对疫苗本身的讨论却相对有限。这凸显了在构建健康信息时，信息提供者需要警惕潜在的误导性信息和反对意见^[73]。为了更全面地了解 HPV 疫苗在社交媒体上的传播效果，Hopfer 等人在 2022 年进行了一项研究，对比了 12 种策略在 Instagram、TikTok 和 Twitter 上推广 HPV 疫苗视频的潜力。研究结果显示，Instagram 和 TikTok 在 HPV 疫苗推广方面显著优于 Twitter，这为公共卫生部门在社交媒体环境中规划和实施 HPV 疫苗接种干预提供了宝贵的参考框架^[74]。这些研究成果不仅揭示了社交媒体在疫苗推广中的

重要作用，也为未来相关策略的制定提供了重要依据。

2. 视频社交媒体平台上关于情绪的研究

视频社交媒体平台上的情绪研究，尤其是与疫苗相关的话题，已引发了广泛的学术关注。Y. Li 等人于 2021 年的研究，通过内容分析揭示了 TikTok 平台上最常表达的情感为希望^[75]。紧接着，Wang 和 He 在 2022 年的研究中，深入探讨了 TikTok 上通过“# 疫苗”标签表达的疫苗支持视频。他们不仅分析了这些视频中与疫苗相关的信念和情感类别，还揭示了信念、情感与支持性评论之间的关联。研究表明，信心成为最普遍的信念，而喜悦则是最频繁表达的情感^[76]。同样，Lewis 和 Grantham 在 2022 年的研究也指出，TikTok 上的内容在对待 COVID-19 疫苗时普遍呈现积极态度，中立态度相对较少^[77]。这一发现与疫苗支持性视频的流行趋势相呼应，凸显了社交媒体在传播积极疫苗信息方面的作用。转向 HPV 疫苗的情感问题，YouTube 平台上的讨论同样丰富。J. Kim 在 2021 年的研究中，聚焦于韩国 YouTube 上推广人乳头瘤病毒疫苗的视频。他发现，以当前为焦点的消息策略能有效引发观众的积极反应，而避免得失框架则能更直接地激发反应。此外，研究还指出，信息感受值对于观众的积极和消极反应都具有显著影响^[35]。然而，值得注意的是，Brionesdeng 在 2012 年的研究提出了不同的观点。他发现，YouTube 上的大部分视频在表达 HPV 疫苗相关情感时呈现负面态度，且观众更倾向于观看负面视频，而非积极或中立视频^[78]。这一发现与前述研究形成鲜明对比，揭示了不同时间和地域背景下，社交媒体平台上疫苗情绪问题的复杂性和多样性。

2.4 多模态社交媒体平台的跨平台比较

人们同时在多种社交平台上接收和讨论 HPV 疫苗信息，但在不同模态内容的社交媒体平台上，这些讨论各不相同。为了了解这些多模数据如何构建人们的信念和行为，学者们提议分析和比较多模社交媒体数据内容，而不是孤立地研究单一模态的数据^[79-80]。现有研究已经揭示，不同模态的平台在同一主题下往往提供了多元化的内容信息。例如，对 Twitter、Facebook 和 YouTube 的比较研究显示，在品牌营销的语境下，品牌更容易在 Twitter 上占据主导地位，而在 Facebook 和 YouTube 上则显得较为边缘^[81]。类似地，另一项针对 Twitter 和 YouTube 上某一社会运动信息的研究也发现，这两个平台在信息来源和质量上存在着显著差异^[80]。基于这些发现，本研究旨在通过对比分析多模态社交媒体内容数据，进一步增进对不同形式社交媒体特征的理解，并为此提供如何在各个平台上有效推动健康行动的见解。具体而言，针对当前 HPV 疫苗相关社交媒体信息的主题和情感研究，本研究提出了以下两个研究问题：

- 研究问题 1: 在三类平台上的文本、音频和视频内容, *HPV* 疫苗相关讨论的主题是什么?
- 研究问题 2: 在三类平台上的文本、音频和视频内容, *HPV* 疫苗相关讨论在文本层面上所表达的情感是什么?

2.5 健康信念模型 (Health Belief Model, HBM)

2.5.1 健康信念模型理论

为了系统研究社交媒体上关于 HPV 疫苗的话语, 本研究运用了健康信念模型 (HBM)。健康信念模型是健康促进和沟通中最常用的理论之一, 它解释了健康信息如何促进积极的健康行为, 特别是预防行为^[82]。健康信念模型理论认为, 在与健康相关的背景下, 人们的健康行为受他们避免疾病的愿望和相信健康行动将预防疾病的影响^[83]。具体而言, 健康信念模型包括以下四个维度。第一个维度是易感性, 它反映了患病的主观风险^[84]。第二个维度是严重性, 它反映了对给定健康问题严重程度的信念^[84]。第三个维度是健康行动的益处, 它反映了对推荐健康行动有效性的信念, 以减少健康威胁^[83]。第四个维度是健康行动的障碍, 它反映了人们权衡健康行动的不利因素, 例如可能有危险、昂贵、不便等^[83]。

2.5.2 有关健康信念模型的研究

鉴于健康信念模型 (HBM) 提供了一个全面且系统的理论框架, 以深入剖析影响人们健康信念和健康行为的社会心理因素, 因此, 这一理论在媒体话语中得到了广泛的运用, 用以分析众多健康主题。学者纷纷借助健康信念模型理论, 对公众的健康观念和行为模式进行了深入的剖析和研究。在一项研究中, 研究人员研究了 1990 年至 2008 年间关于体育中类固醇的新闻报道, 发现新闻机构强调了使用类固醇的非法性, 而对于类固醇的易感性和严重性报道较少^[85]。也有研究人员使用健康信念模型来检视 58 种提前护理规划教育材料的内容, 发现这些材料缺乏对益处和障碍的具体描述^[86]。研究人员比较了在 Twitter 和 Pinterest 上有关埃博拉病毒的信息, 发现 Pinterest 上关于埃博拉的易感性和严重性更频繁^[87]。研究人员采用健康信念模型理论对 TikTok 上关于 COVID-19 疫苗的内容进行了内容分析。研究结果表明, COVID-19 疫苗的益处 (15.7%) 和障碍 (34.7%) 比 COVID-19 的易感性 (2.8%) 和严重性 (5.6%) 更频繁^[88]。

2.5.3 HPV 疫苗和健康信念模型在社交媒体上的研究

健康信念模型在使不同类型的社交媒体的可比较方面展示了强大的潜力,可以针对 HPV 相关和 HPV 疫苗相关的话题进行比较。在基于文本内容的平台方面,有国外研究人员在 2018 分析了 2000 年至 2016 年中国报纸上的 HPV 相关文章,发现中国报纸更加关注 HPV 预防方法(例如疫苗)的益处,而其他健康信念模型维度则较少关注^[89]。在中国的另一项研究中,研究人员调查了中国类似 Quora 的问答社交媒体平台知乎上的 HPV 报道。他们发现,知乎上大多数相关文章包括易感性、严重性和益处。此外,易感性的存在与严重性的存在呈正相关^[56]。研究人员 Du 等人在 2020 年应用深度学习方法分析了 2013 年至 2018 年 Twitter 上关于 HPV 疫苗的讨论。他们发现随着时间的推移易感性和严重性维度的出现频率普遍增加,而益处和障碍维度的出现频率普遍降低。结果表明 Twitter 用户越来越倾向于讨论 HPV 的风险^[90]。至于图像内容平台上,学者 Guidry 等人在 2020 年研究了 Pinterest 上关于 HPV 疫苗信息的搜索结果在平台开启不实信息审查实前后的变化。研究发现,两个健康信念模型维度的频率变化最大——益处从 13.3% 增加到 40.4%,而障碍则从 65.6% 减少到 11.3%^[91]。在视频内容平台上,Briones 等人在 2012 年通过健康信念模型模型考察了 YouTube 上有关 HPV 疫苗的报道。他们发现,在 172 个样本视频中,37.8% 提及了 HPV 的易感性,53.4% 提到了 HPV 的严重性,41.9% 提到了 HPV 疫苗的益处,而 53.5% 提到了 HPV 疫苗的障碍^[78]。

先前的研究表明健康信念模型是一种可靠的理论框架,可用于在各种社交媒体环境下进行 HPV 相关信息的内容分析。因此,本研究借鉴健康信念模型理论,比较了以文本、音频和视频为主要呈现内容的平台微博、喜马拉雅和抖音上 HPV 疫苗讨论的情况。本研究希望探索这三个平台如何包含有关 HPV 疫苗不同维度的信息,以此了解社交媒体如何向用户提供关于 HPV 疫苗的信息并且影响用户对于 HPV 疫苗的观点。因此,本研究提出研究问题 3:

研究问题 3: 在三类平台上的文本、音频和视频内容如何涵盖健康信念模型 (Health Belief Model, HBM) 的易感性、严重性、益处和障碍?

2.6 文本分类任务、小样本学习

2.6.1 文本分类任务

文本分类是自然语言处理领域中的一项重要任务,在情感分析信息检索、舆情分析、话题跟踪等领域有广泛的应用。具体来说,文本分类就是在预先定义类别的情况下,根据文本的主题和内容,将文本按照一定的分类体系或标准进行自

动划归类别的过程。针对于文本分类的具体任务，预先训练的文本特征表示被广泛用于提高任务处理的性能。

针对于文本分类的具体任务，预先训练的文本特征表示被广泛用于提高任务处理的性能^[92]。这些表征是在大型语料库上训练的（通常是无监督的数据），然后将其作为特征输入到处理下游任务的模型之中。Turian 在 2018 年的研究表明，预先训练好的词嵌入是优于那些从随机初始化开始训练的^[93]。自从深度学习开始流行以来，词嵌入已经被广泛使用，并且许多不同的训练策略已经出现^[94-96]。在深度学习中，这些特征也可以在下游任务的数据上进行微调，可以说预训练微调范例已成为解决自然语言处理的一种标准方法，并在多个基准测试中取得了优异的结果^[97-99]。

有关文本分类方面的优秀研究成果，一般都是通过微调 Transformer 网络^[100]结合完型填空 MLM 损失^[97]实现的。Howard 在 2018 年的研究中表明，微调方法通常效果优于之前提出的基于特征的方法，另一方面微调方法也不需要按照之前的方法一样根据特定任务进行模型设计。然而微调仍然有其问题，对于每一项新的下游任务而言，都需要训练新的网络权重集来处理任务，而且每次训练都需要大量的数据支撑^[101]。

2.6.2 小样本学习

小样本学习旨在利用有限的监督信息实现在目标任务上较为泛化性能，这在许多现实场景中是迫切需要的。最近的小样本学习方法大致是基于三个方面设计的：数据、模型和算法。第一类小样本学习方法主要是通过数据增强方法防止模型过拟合^[102]。这类方法或者采用一些标准的数据增强操作直接增加样本，或者从辅助数据集中学习先验知识，继而用以生成新的样本。第二类小样本学习方法尝试用辅助数据集来约束假设空间，从而使少量的样本足以优化模型^[103-104]。最后一类小样本学习方法是“模型无关的”，它专注于从辅助数据集学习好的通用初始点或优化步骤。由于只有目标任务的几张标签样本不足以进行泛化，所以大多数仍需要辅助数据集来获取先验知识^[105]。

第3章 研究设计

3.1 研究方法

本文探讨了文本、音频和视频作为主要内容的社交媒体平台关于 HPV 疫苗议题的讨论差异。基于研究问题选取了研究样本，通过 Python 脚本采集了微博平台 273,357 条帖子，喜马拉雅平台 1,228 条音频，抖音平台 1,225 条视频作为文本、音频和视频三类模态内容的例子。

由于文本、音频和视频三种模态对于 HPV 疫苗议题的内容差异无法直接进行跨模态对比，且音频和视频模态关于 HPV 疫苗议题的讨论主要集中在人声部分。因此，对音频和视频模态的人声部分进行了提取，并通过语音识别技术将其转化为长文本，从而用文字形式承载音频和视频平台上的讨论。具体步骤如下：首先使用 FFmpeg 工具将视频数据转换为音频数据，接着通过 Spleeter 工具包分离出人声部分，最后对音频和视频模态中的人声讨论应用语音转写（ASR）技术，将其转录为文本，以便与文本模态内容进行统一对比分析。

由于三种社交媒体平台的数据量差异较大，直接主题建模进行内容分析难以进行有效对比，本研究引入健康传播领域经典的理论—健康信念模型（Health Belief Model, HBM）四个维度——易感性、严重性、效益、障碍，来对数据进行分类，从四个维度的角度进行差异化的对比。在进行分类任务时，由于一条文本数据在内容上涵盖了多个健康信念模型维度，对于文本平台的数据，通过自然语言处理技术训练四个深度学习模型 RoBERTa 进行分类，对于音频平台和视频平台的长文本数据，通过人工编码进行分类。对于分类后的数据，采用主题建模技术 BERTopic 进行建模，对每个平台每个维度选取最大的三个代表性的簇，最后进行内容的差异对比。同时，本文也训练了基于情绪的深度学习语言模型，对三种平台的内容数据进行情绪分析。

3.2 数据采集

为了深入探究人乳头瘤病毒（HPV）疫苗在社交媒体上的讨论趋势，本研究系统性地收集了来自多个平台的公开数据。具体而言，数据的采集范围涵盖了文本平台微博、音频平台喜马拉雅以及视频平台抖音，时间跨度自 2018 年 10 月至 2023 年 3 月，且数据类型包括文本、音频和视频三种。这些数据集在时间上进行了对齐处理，以确保分析的一致性和准确性。

数据集的具体构成涵盖以下三部分：首先是文本、音频和视频的原始数据；其次是相关的互动信息，如文本的点赞数、评论数等，音频的播放数、标题、点

赞数、评论数等，以及视频的点赞数、播放量、评论量、收藏量等；最后是文本、音频和视频发布者的用户信息。这些数据为分析提供了全面且详尽的基础。

在数据采集的过程中，有几点需要特别说明。首先，微博平台的检索机制较为完善，能够获取到全量的相关数据，因此本研究使用了关键词“HPV”、“人乳头瘤病毒”和“宫颈癌”进行了全面的检索。然而，喜马拉雅平台和抖音平台的检索机制存在一定的限制，为了获取足够多的 HPV 疫苗相关的音频和视频数据，本研究采用了特定的方法进行检索，以确保数据采集的严格性和完整性。

其次，HPV 疫苗相关的讨论在微博平台、喜马拉雅平台、抖音平台的起始时间并不一致。尽管采集了三种社交媒体上截至 2023 年 3 月之前所有的关于 HPV 疫苗的文本、音频、视频的社交媒体数据，但为了保证研究的严谨性和避免时间不一致对研究结果的影响，本研究对三个平台的数据进行了时间对齐处理，最终选取了 2018 年 10 月到 2023 年 3 月期间的文本、音频和视频数据进行分析。

3.2.1 文本平台数据采集

为了在微博平台上获取关于 HPV 的文本数据，本研究采用了三个搜索关键词：“HPV”、“人乳头瘤病毒”和“宫颈癌”，进行了全面的检索。经过这一步骤，共采集到了 70 余万条相关数据。随后，为了聚焦于 HPV 疫苗的讨论，本研究在微博文本数据中进一步检索了关键词“疫苗”，并成功筛选出了与 HPV 疫苗直接相关的文本数据，共计 318,855 条。

这些数据不仅包含了与 HPV 疫苗相关的帖子内容，还涵盖了发布这些帖子的用户信息，这些信息是通过微博 ID 进行关联的。为了高效存储这些数据，本研究采用了 JSON 文件格式。在后续的分析阶段，从 JSON 存储格式中提取了 HPV 疫苗相关帖子的数据和发帖人的用户信息，并将这两部分数据分别保存为 CSV 格式，同时采用了 UTF-8 编码以确保数据的可读性和兼容性。

此外，值得注意的是，这些数据的时间跨度为 2011 年 03 月至 2023 年 03 月，共计包含了 318,855 条文本数据。这一庞大的数据集为深入研究 HPV 疫苗在微博平台上的讨论趋势提供了坚实的基础。

3.2.2 音频平台数据采集

为了从喜马拉雅获取音频数据，本研究采取了以下步骤。第一，使用关键词“HPV 疫苗”、“人乳头瘤病毒疫苗”和“宫颈癌疫苗”进行检索，收集了数据集的一部分。第二，使用关键词“HPV”、“人乳头瘤病毒”和“宫颈癌”进行检索，并且在标题、描述和标签中通过关键词“疫苗”筛选数据，生成了数据集的另一部分。第三，移除前两个步骤中收集的数据集中的重复内容，生成了最终的喜马拉雅数据集。根据音频的 ID，采集音频文件的元数据，该音频的详细页面数据。

第四,对这些音频数据发布者的用户信息也进行采集。时间跨度为2016年12月到2023年3月,共1,228条音频数据。

3.2.3 视频平台数据采集

为了从抖音平台获取关于 HPV 疫苗的音频数据,本研究采取了以下系统性步骤。

首先,本研究利用关键词“HPV 疫苗”、“人乳头瘤病毒疫苗”和“宫颈癌疫苗”进行了初步检索,从而收集到了数据集的第一部分。随后,为了扩大数据覆盖范围,本研究又使用了关键词“HPV”、“人乳头瘤病毒”和“宫颈癌”进行检索,并在标题、描述以及标签中通过关键词“疫苗”进行了进一步的筛选,从而生成了数据集的第二部分。

其次,为了确保数据的唯一性和准确性,本研究对前两个步骤中收集到的数据集进行了去重处理,最终生成了喜马拉雅平台的最终数据集。在此过程中,本研究还根据视频的 ID,采集了视频文件的元数据以及详细页面数据,这些数据为深入分析提供了丰富的信息。

此外,本研究还采集了这些视频数据发布者的用户信息,这些信息对于理解用户特征和讨论背景具有重要意义。值得注意的是,本研究收集的数据时间跨度为2018年12月至2023年3月,共计获得了1,225条音频数据。

综上所述,本研究通过一系列严谨的步骤,成功地从喜马拉雅平台获取了关于 HPV 疫苗的音频数据,为后续的分析 and 研究奠定了坚实的基础。

3.3 数据预处理

3.3.1 文本平台数据预处理

为了提升微博平台文本数据集的质量,本研究对其进行了严格的清洗过程,旨在过滤掉与 HPV 疫苗无关的数据。以下是数据清洗的具体流程:

第一步,本研究对数据集进行了语言筛选,去除了英文、韩文、日文等非中文数据,仅保留了中文数据,以确保数据的语言一致性。

第二步,为了初步判断数据的相关性,本研究按月分层抽取了每个月总数据的 0.5% 进行人工筛选。通过人工判定关键词与 HPV 疫苗的相关性,对于无关数据,本研究记录了关键词,并利用这些关键词检索相关微博文本。随后,通过正则匹配每个关键词对应的数据,进一步验证其相关性。对于不相关的数据,本研究将其关键词记录在词表中,并批量删除了相关的文本数据。该词表详见附录 A1.1。

第三步,为了识别并去除由特定新闻事件引发的、与 HPV 疫苗主题无关的

讨论峰，本研究以天为单位绘制了微博总数量的时序统计图。通过人工判定，本研究识别出了与 HPV 疫苗主题无关的讨论峰，如明星广告打榜、微博热搜广告等。针对这些无关讨论峰，本研究采取了以下措施：A. 总结并记录了相关关键词，形成了新的词表，并批量删除了与这些关键词相关的文本数据。随后，将新生成的关键词词表与第二步处理的关键词词表进行了合并。B. 根据发帖人的身份和内容，本研究识别出了频繁发布无关内容的用户，如明星打榜用户和广告用户。对于这些用户，本研究直接删除了其发布的所有文本数据，并将这些用户记录在无关用户词表中。该词表详见附录 A1.2。

第四步，为了进一步识别并去除广告用户，本研究统计了所有用户的发布数量，并按照数量从高到低进行了排序。通过查看排序靠前的用户所发布的内容，本研究识别出了广告用户，并将其记录在无关用户词表中。该词表同样详见附录 A1.2。

综上所述，本研究通过一系列严谨的数据清洗步骤，成功地去除了与 HPV 疫苗无关的数据，为后续的数据分析提供了高质量的数据集。

3.3.2 音频平台数据预处理

鉴于不同模态的内容难以直接进行比较，且音频和视频中的关键信息往往集中在人声部分，本研究采取了针对性的策略。首先，对音频和视频中的人声进行了精确的提取和转录，以确保信息的完整性和准确性。接着，将对比的焦点集中在了文字层面，通过将不同模态的内容转化为文字形式，使得原本难以直接对比的音频和视频信息能够在同一维度上进行对比分析。这一方法不仅有效解决了模态间比较的难题，还提高了对比分析的准确性和效率。对于喜马拉雅音频数据集，进行以下预处理：

1. 音频人声抽取

使用 Spleeter 工具从音频背景音中抽取人声。Spleeter 是 Deezer Research 于 2019 年提出的带有预训练模型的音乐源分离工具。旨在帮助音乐信息检索 (MIR) 社区利用源分离的能力来完成各种 MIR 任务，例如从音频中分析歌唱歌词（音频/歌词对齐，歌词转录...），音乐转录（和弦转录，鼓声转录，低音转录，和弦估计，节奏跟踪），歌手识别，任何类型的多标签分类（情绪/风格...）或歌唱旋律提取。他们认为源分离已经达到了一定的成熟度，使其对这些任务值得考虑，并且从孤立的人声、鼓声或低音中计算的特定特征可能有助于提高性能，尤其是在数据匮乏的情况下（小数据集，有限的注释可用性）。它可以基于 Tensorflow，可以通过命令行预训练模型将音频文件分离为 2 个、4 个或 5 个音轨。^[106]它是目前比较成熟的音频模态中人声和背景音分离的成熟工具。本研究应用 Spleeter 中的预训练模型，编写脚本，对喜马拉雅数据集的音频数据中的人声部分进行提取，

从而生成人声的喜马拉雅数据的源文件。

2. 人声转录文本

语音转写（Long Form ASR）基于深度全序列卷积神经网络，将长段音频数据转换成文本数据，为信息处理和数据挖掘提供基础。本文采用科大讯飞的语音转写 API 将音频数据中的人声讨论内容转录成为长文本内容^[107]。转录的步骤如下所示：

第一步，首先调用预处理接口，上传待转写音频文件的基本信息（文件名、大小）和分片信息（分片大小设置为 10M，若无需分片，`slice_num=1`）和相关的可配置参数。调用成功，返回任务 ID（`task_id`，转写任务的唯一标识），是后续接口的必传参数。

第二步，预处理成功，调用文件上传接口；按预处理设置的分片信（`slice_num`）依次上传音频切片（文件以二进制方式读取上传），直到全部切片上传成功（如预处理时 `slice_num=2`，则需将音频切分成两部分，`slice_id=aa` 和 `ab`，并按顺序调用该接口）。

第三步，全部文件切片上传成功后，调用该接口，通知服务端进行文件合并与转写操作。

第四步，在调用方发出合并文件请求后，服务端已将任务列入计划。在获取结果前，调用方需轮询该接口查询任务当前状态。

经过语音转写 API 转写人声音频，使用 JSON 文件格式来存储这些人声文本数据。由于喜马拉雅平台的是播客，声音讨论 HPV 疫苗的时间非常的长，所以转录之后的是非常长的文本。本文对非中文的转录长文本和跟 HPV 疫苗无关的长文本数据进行删除。

3.3.3 视频平台数据预处理

鉴于不同模态的内容难以直接进行比较，且音频和视频中的关键信息往往集中在人声部分，本研究采取了针对性的策略。首先，对音频和视频中的人声进行了精确的提取和转录，以确保信息的完整性和准确性。接着，将对比的焦点集中在了文字层面，通过将不同模态的内容转化为文字形式，使得原本难以直接对比的音频和视频信息能够在同一维度上进行对比分析。这一方法不仅有效解决了模态间比较的难题，还提高了对比分析的准确性和效率。对于抖音视频数据集，本研究进行如下预处理流程：

1. 视频抽取音频

本研究使用软件 FFmpeg 从视频中提取出音频数据，创建了相应的音频数据集。FFmpeg 是可以在多种操作系统环境下运行的开源软件，提供了音视频数据处理完整的技术框架，在音视频处理领域中占有了不可忽视的地位。FFmpeg 处

理音视频的能力非常强,并且支持大部分的音视频编码标准^[108]。使用命令行工具,将抖音数据集的视频数据通过 FFmpeg 抽取了音频。

2. 音频人声抽取

本研究使用 Spleeter 工具作为从音频背景音中抽取人声。Spleeter 是 Deezer Research 于 2019 年提出的带有预训练模型的音乐源分离工具^[106]。该工具旨在帮助音乐信息检索 (MIR) 社区利用源分离的能力来完成各种 MIR 任务,例如从音频中分析歌唱歌词 (音频/歌词对齐,歌词转录...),音乐转录 (和弦转录,鼓声转录,低音转录,和弦估计,节奏跟踪),歌手识别,任何类型的多标签分类 (情绪/风格...) 或歌唱旋律提取。他们认为源分离已经达到了一定的成熟度,使其对这些任务值得考虑,并且从孤立的人声、鼓声或低音中计算的特定特征可能有助于提高性能,尤其是在数据匮乏的情况下 (小数据集,有限的注释可用性)。它可以基于 Tensorflow,可以通过命令行预训练模型将音频文件分离为 2 个、4 个或 5 个音轨。它是目前比较成熟的音频模态中人声和背景音分离的成熟工具。应用 Spleeter 中的预训练模型,对抖音数据集的音频数据中的人声部分进行提取,从而生成人声的抖音数据的源文件。

3. 人声转录文本

语音转写 (Long Form ASR) 基于深度全序列卷积神经网络,将长段音频数据转换成文本数据,为信息处理和数据挖掘提供基础。本研究采用科大讯飞的语音转写 API 将音频数据中的人声讨论内容转录成为长文本内容^[107]。转录的步骤如下所示:

第一步,首先调用预处理接口,上传待转写音频文件的基本信息 (文件名、大小) 和分片信息 (分片大小设置为 10M,若无需分片, slice_num=1) 和相关的可配置参数。调用成功,返回任务 ID (task_id, 转写任务的唯一标识),是后续接口的必传参数。

第二步,预处理成功,调用文件上传接口;按预处理设置的分片信 (slice_num) 依次上传音频切片 (文件以二进制方式读取上传),直到全部切片上传成功 (如预处理时 slice_num=2,则需将音频切分成两部分, slice_id=aa 和 ab,并按顺序调用该接口)。

第三步,全部文件切片上传成功后,调用该接口,通知服务端进行文件合并与转写操作。

第四步,在调用方发出合并文件请求后,服务端已将任务列入计划。在获取结果前,调用方需轮询该接口查询任务当前状态。

经过语音转写 API 转写人声音频,用 JSON 文件来存储这些人声文本数据。由于喜马拉雅平台的是播客,声音讨论 HPV 疫苗的时间非常的长,所以转录之后的是非常长的文本。本文删除了非中文的转录长文本和跟 HPV 疫苗无关的长

文本数据。

3.4 健康信念模型维度分类

本研究使用健康信念模型（Heath Belief Model, HBM）对社交媒体平台（微博、喜马拉雅和抖音）上关于 HPV 疫苗的讨论进行系统分析。健康信念模型假设几个认知结构能够影响个人的健康和行为^[82-83]。由于数据集中大部分数据采用第三人称视角且没有反映个人行为，健康信念模型只有四个维度涵盖这些文本数据，即易感性、严重性、收益和障碍。在这里注意的是，一条文本内容可能涵盖健康信念模型的多个维度，这意味着用户在同一时间可能会感知不同的健康信念模型维度。因此，本文将数据集分为四个互相包含的类别。

3.4.1 抽样数据建立编码表

为了对微博数据集进行分类，首先手动编码了样本数据，并建立了编码规则。在建立编码规则时，随机抽取微博数据集的 1% 作为样本（2,733 条文本数据），而喜马拉雅和抖音数据集则对全部数据进行人工编码。人工专家一起对样本数据集进行手动的人工编码。编码规则和示例详见表 3.1 和表 3.2。

3.4.2 训练深度学习模型进行小样本分类

经过手动编码，通过黄金标签的样本数据训练四个深度学习模型来对剩余数据集进行分类。深度学习模型使用 RoBERTa 模型，一种基于 Transform 语言模型，具有多层自注意机制。该模型通过训练包括新闻、交互讨论和百科全书在内的各种英语语料库，总计 160GB 以上的未压缩文本。模型还涉及近 1000 亿个汉字^[109]。在这里，预训练模型使用 RoBERTa 中文预训练的版本^[110]，使用健康信念模型四个维度黄金标记的微博数据集对该预训练模型进行了微调。对于每个健康信念模型维度，本文微调独立的二元分类器，输出值为 1（分类到关注维度）或 0（未分类到关注维度）。首先，由于社交媒体数据噪声较大，特征不明显，带有黄金标签的样本数据较少等原因，使用数据增强技术对小样本的数据进行数据增强。其次，对分类模型更换多种适合小样本分类的二分类损失函数，并且通过 Gamma 参数调整易于分类样本和难分类样本的权重。通过大量的模型训练的参数调整，训练的分类器的表现 F1 得分均大于 0.7。对微博数据集，训练四个独立的健康信念模型维度分类器，分别是易感性维度分类器、严重性维度分类器、效益维度分类器、障碍维度分类器，经过深度学习模型的训练，选择了分类效果表现最优的网络权重参数和损失函数。模型的性能如表 3.3 所示，表明其具有可接受的精确度和质量。

表 3.1 健康信念模型编码规则和示例

类型	描述	一级说明	二级示例
易感性	HPV 感染风险的评估。	1. 感染几率; 有可能感染; 是否感染 (HPV) 2. 某种行为可以/会导致 (HPV) 感染 3. HPV 感染的原因和感染的症状 4. 男性女性应该打/需要打/必须打/可以打 5. HPV 疫苗的适用人群	1. 在国外, 适龄的男性也可以接种 HPV 疫苗。 2. 这种疾病是只属于女孩子的疾病, 因此我鼓励身边的女性去接种 HPV 疫苗, 万一中招, 我会后悔死的。 3. 绝大部分 (80%) 的女性将在一生中感染 HPV 病毒。
严重性	HPV 感染是否构成足够的健康问题的评估。	1. 死于, 死亡率, 感染率, 发病率 2. 潜伏期长 3. 疫苗作用效果有限, 无法完全预防所有病毒; 还需要检查 (筛查); 一级预防、二级预防、三级预防。 4. 反复感染 5. 传播方式/传播途径 6. 癌症 7. 打疫苗后需要筛查	1. 宫颈癌是一种恶性肿瘤, 在中国女性生殖系统恶性肿瘤中的发病率和死亡率居首位。 2. HPV 疫苗不能预防能引发宫颈癌的所有 HPV 类型, 已经接种疫苗的人仍需进行健康筛查。 3. 世界卫生组织制定了针对宫颈癌的一级预防、二级预防和三级预防。 4. 治病需要花很多钱, 对一个家庭是毁灭性的打击 5. 九价能预防 9 个 HPV 亚型病毒、四价能预防 4 个 HPV 亚型病毒、二价能预防 2 个 HPV 亚型病毒

同时, 对于音频平台 (喜马拉雅) 和视频平台 (抖音) 的长文本数据, 经过人工编码对数据进行分类。

3.5 基于深度学习的主题建模技术

主题建模兴起于计算机科学领域, 通过归纳文本语料库中重复出现的单词集群的主题数量来发现文本信息中潜藏的语义信息, 从海量网络文本中提炼主题并跟踪某一主题的发展, 在文本挖掘领域有着广泛的应用。主题建模基本上是一组方法和算法的统称, 这些方法和算法通过揭示共同出现的单词的递归环簇来揭示文档集合中隐藏的主题结构。虽然针对每个成形主题建模有几种不同的算法, 最常使用的算法是 LDA (Latent Dirichlet Allocation), 但是 LDA 仅基于词概率没有考虑语义, 所以选择更加先进的基于预训练的语言模型 BERT 的主题建

表 3.2 健康信念模型编码规则和示例 (续表)

类型	描述	一级说明	二级示例
效益	HPV 疫苗在预防 HPV 感染、HPV 感染引起的癌症等方面的好处。	1. 疫苗本身的好处：保护作用、预防作用 2. 有必要接种	1. 九价 HPV 疫苗可以预防 92.1% 的宫颈癌。 2. 接种 HPV 疫苗将全球受益，并为世界卫生组织关于加速消除宫颈癌的全球战略做出贡献。
障碍	HPV 疫苗的不良反应；接种 HPV 疫苗的成本；关于 HPV 疫苗的年龄限制、供应和性别限制等等。	1. HPV 疫苗的价格（便宜/贵）；检查费用 2. HPV 疫苗与其他疫苗的冲突 3. 年龄限制 4. 接种条件限制：在感染之前接种、限制性别（男性不易接种） 5. 打疫苗是否有效果/HPV 是否会反复感染 6. 接种条件的增加或者减少	1. 过敏；HPV 疫苗将会更加昂贵。 2. TCT 检查费用为 180 元人民币。 3. 感染病毒前接种效果最佳 4. HPV 疫苗上市并已开始预约。 5. 我自己是洁身自好的，我的伴侣也是洁身自好的，那么这辈子我们都不太可能感染 HPV 病毒，那打了 HPV 疫苗之后会不会反而感染呢？ 6. 理论上推测 HPV 疫苗最后一针之后是肯定有效果的

表 3.3 健康信念模型分类模型

健康信念模型维度	准确率	召回率	F1-得分
易感性	0.78	0.79	0.78
严重性	0.89	0.83	0.85
效益	0.76	0.77	0.77
障碍	0.77	0.71	0.74

模技术 BERTopic。

BERTopic 主题建模技术^[11]，它是用预训练的基于 transformer 的语言模型生成文档嵌入，对这些嵌入进行聚类，最后用基于类的 TF-IDF 过程生成主题表示。BERTopic 通过三个步骤生成主题表示。首先，使用预训练的语言将每个文档转换为其嵌入表示，这里使用了 Sentence-BERT (SBERT) 框架^[12]。该框架允许用户使用预训练的语言模型将句子和段落转换为密集向量表示。它在各种句子

嵌入任务上取得了最先进的性能^[112-113]。然后,对得到的嵌入进行降维以优化聚类过程,这里采用 UMAP 中来降低之前生成的文档嵌入的维度^[114]。UMAP 和 PCA, t-SNE 等降维方法相比,在较低的投影维度中保留了更多的高维数据的局部和全局特征^[114]。由于 UMAP 对嵌入维度没有计算限制,因此可以跨具有不同维度空间的语言模型使用。接下来对降维的嵌入进行聚类,是使用 HDBSCAN 进行聚类^[115]。它是 DBSCAN 的扩展,通过将 DBSCAN 转换为层次聚类算法来查找不同密度的簇。HDBSCAN 使用软聚类方法对聚类进行建模,允许将噪声建模为离群点。这可以防止不相关的文档被分配到任何簇,并有望改进主题表示。最后,利用自定义的基于类别的 TF-IDF 变换从文档簇中提取主题表示。BERTopic 主题建模技术目前基于中文的研究用的比较少。

3.5.1 主题建模数据预处理

为了提升文本数据的有效性与准确性,要通过一定的处理方式完成对文本数据的清洗,比如 html 解析,去除非中文字符等处理。网络文本的预处理是实现将文本数据(非结构化数据)转化为结构化数据的必要准备工作,一般地要将清洗过文本语料数据进行中文分词、去除停用词等方法处理,才可以开展下一步文本表示模型构建工作。中文是一种复杂的语言,计算机难以识别。利用中文分词技术可以将整个语句拆分成独立的可被机器识别的词语。相较于英文而言,由于中文句中词语之间没有空格,以及存在大量一词多义的情况,所以造成中文文本在词语切分以及词性识别方面实现难度较大。现有的分词算法有基于字符串匹配,基于理解和基于统计的分词三种方法。随着专家学者对中文分词技术的不断深入研究,出现了很多中英文分词工具,目前常用的分词工具包括 python 中的 Snow NLP 函数库,清华大学自然语言处理实验室推出 THULAC 中文词法分析工具包,中科院分词工具包 NLPIR 和 jieba 分词等。去停用词是指在文本数据分词之前或之后会自动过滤掉的某些非中文字符或语气助词。文本数据进行分词处理后,语料中可能还会存在语气助词标点符号等干扰词汇,这些干扰词汇会影响对后续对文本特征的提取,从而影响到文本聚类的准确性。自然语言处理领域去停用词使用停用词表对分词后的数据进行过滤剔除,将包含在停用词表中的词汇直接删除。本文在文本预处理中使用哈工大停用词表进行停用词的剔除。继而,进行单个词汇的人工复核,剔除没有明确意涵以及不具备一定程度指代性的词汇。与此同时,进行词对频的分析,剔除没有明确意涵和信息价值的词对。如上所述的人工复核过程重复三次之后,得到最终的文本数据库。

在本文中,采用了在中文分词领域得到了广泛应用的 Python 第三方库 Jieba 库进行分词,使用词图和动态规划等技术找出基于词频的最大切分组合,并使用 HMM 模型预测未登录的新词。分词完成后,导入哈工大停用词表剔除文本

中的停用词，并完成繁简转换和去除标点等预处理操作。通过去除 URL、提及、Hashtag 和数字符号来预处理数据集。基于中文 Jieba 库分词工具^[116]和哈尔滨工业大学中文停用词表，定义一个专门针对本研究的数据集的分词表和自定义停用词表。对文本数据进行中文分析后，对文本、音频、视频平台上四个健康信念模型维度易感性、严重性、效益和障碍进行 BERTopic 主题建模。

3.5.2 BERTopic 短文本主题建模

本文使用 BERTopic 模型^[111]来分析微博短文本内容主题。这种主题建模的方法可以在文本语料库中发现潜在的主题。BERTopic 模型基于 BERT 嵌入、降维和聚类算法。它提供了可解释的人类主题以及它们随时间的重要性。对每种平台上健康信念模型的四个维度进行 BERTopic 主题建模。通过这种视角，本文可以区分基于健康信念模型的四个维度视角不同平台之间的独特特征。BERTopic 微博短文本主题建模的步骤如下所示：

第一步，对基于深度学习模型进行健康信念模型四个维度分类的易感性、严重性、效益和障碍的四个微博数据集从总的微博数据集里面提取。

第二步，对微博易感性数据集、微博严重性数据集、微博效益数据集和微博障碍数据集进行 BERTopic 主题建模。本文在主题建模第一层选择了 Sentence-BERT (SBERT) 框架，语言选择了中文版本，第二层选择了 UMAP 降维算法从高维降到低维度，第三层对降维的嵌入进行聚类，是使用 HDBSCAN^[115]进行聚类，第四层使用词袋使用词袋对聚类簇的无意义词和主题词进行去除，第五层使用基于类别的 TF-IDF 变换从文档簇中提取主题表示。

第三步，在对每个数据集经过多次调整主题建模的参数和人工判定主题建模达到了较好的效果。本文对主题建模的各个分散主题进行主题合并。

第四步，选取每个维度主题建模后最大的三个代表性簇进行代表主题分析。

3.5.3 BERTopic 长文本主题建模

对于音频平台（喜马拉雅）和短视频平台（抖音）上的长文本，由于 BERTopic 主题建模模型的第一层 Sentence-BERT (SBERT) 的输入是 512token，对于长文本为了更好的主题建模，挖掘音频模态和视频模态用户发布内容所表达关于 HPV 疫苗的讨论，本研究在对每个平台每个维度的数据集主题建模的时候，会对句子按照句号进行分割，在单句层面上进行主题建模，以更好地探索主题和语义。BERTopic 音频视频平台长文本主题建模的步骤如下所示：

第一步，对基于深度学习模型进行健康信念模型四个维度分类的易感性、严重性、效益和障碍的四个喜马拉雅、抖音数据集从总喜马拉雅、总抖音数据集里面提取。

第二步，对长文本数据按照句号为单位，对长文本进行拆分。

第三步，对喜马拉雅易感性数据集、喜马拉雅严重性数据集、喜马拉雅效益数据集和喜马拉雅障碍数据集进行 BERTopic 主题建模。同时也对抖音易感性数据集、抖音严重性数据集、抖音视频数据集和抖音障碍数据集进行 BERTopic 主题建模。本文在主题建模第一层选择了 Sentence-BERT (SBERT) 框架，语言选择了中文版本，第二层选择了 UMAP 降维算法从高维降低维度，第三层对降维的嵌入进行聚类，是使用 HDBSCAN 进行聚类^[115]，第四层使用词袋使用词袋对聚类簇的无义词和主题词进行去除，第五层使用基于类别的 TF-IDF 变换从文档簇中提取主题表示。

第四步，在对每个数据集经过多次调整主题建模的参数和人工判定主题建模达到了较好的效果。对主题建模的各个分散主题进行主题合并。

第五步，选取喜马拉雅、抖音平台每个维度主题建模后最大的三个代表性簇进行代表主题分析。

3.6 基于深度学习模型的情绪分类

为了分析三种社交媒体上内容的情绪，本文还将数据分为两个情绪类别，即积极情绪和消极情绪。基于学者崔等人在 2021 年提出的深度学习模型 BERT^[117]，本文使用的预训练模型 Chinese-BERT-wwm，该模型是由哈工大讯飞联合实验室和旷视科技联合研发的中文预训练模型，具有更好的中文自然语言处理能力。本文使用了包含了 10 万条微博数据的 Weibo_Senti_100k 数据集^[118]对预训练模型进行微调，该数据集标注了积极和消极两种情绪。该模型输出一个二维向量，其中的值为零或一。数据可以被分类为消极情绪和积极情绪。微调后的模型的表现如表 3.4 所示。

表 3.4 情绪二分类模型

情绪类别	准确率	召回率	F1-得分
积极	0.90	0.93	0.90
消极	0.90	0.87	0.90

对于微博短文本数据集，本文直接用情绪模型对数据进行分类。而对于喜马拉雅和抖音长文本数据集，由于 Chinese-BERT-wwm 的输入是 512 字符 (token)，长文本难以全部输入模型进行情绪分类，本文进行以下步骤进行处理。首先将长文本拆分成分句。接着，计算每个分句的长度，并为每个情感标签分配数值：积极为 1，消极为-1，无情绪为 0。然后，通过将每个分句的长度与其情感分数相乘，计算出加权情感分数。随后，计算每个长文本的总长度和总加权情感分数，并通过将总加权情感分数除以总长度，计算出每个长文本的加权平均情感分数。

根据加权平均情感分数，将其映射回情感类别：分数大于 0 为积极，分数小于 0 为消极，分数等于 0 为无情绪。最后，将计算出的长文本的情感结果进行统计分析。

在情感分数等于 0 的长文本可以分为两类，一类是长文本的分句都是无情绪，所以长文本的情绪为 0。一类是长文本的分句积极情绪和消极情绪分句占比相等。通过对音频平台喜马拉雅和视频平台抖音的情感分数等于 0 的部分进行统计发现，在喜马拉雅平台上长文本的所有分句全部是无情绪的占比为 38.83%，积极消极抵消的占比为 0.31%；在抖音平台上长文本分句全部是无情绪的占比为 44.04%，没有积极消极情绪相抵消的部分。

第4章 结果分析

4.1 三类平台的健康信念模型分析

4.1.1 文本平台的健康信念模型分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三类社交媒体平台上的总体分布情况如图 4.1 所示。在文本平台微博上，健康信念模型的易感性维度占比为 4.58%，严重性维度占比为 11.84%，效益维度占比为 12.23%，障碍维度占比为 87.15%。其中文本平台微博障碍维度关于 HPV 疫苗的讨论最多，易感性维度关于 HPV 疫苗的讨论最少，严重性维度和效益维度关于 HPV 疫苗的讨论差不多一样多。障碍维度是易感性维度的 19 倍，障碍维度是严重性维度的 7 倍，障碍维度是效益维度的 7 倍。严重性维度是易感性维度的 2.50 倍，效益维度是易感性维度的 2.60 倍。而严重性维度和效益维度基本相同。健康信念模型的易感性、严重性、效益和障碍四个维度在文本平台上的月度数量分布情况如图 4.2 所示。由时间序列图可以发现，对于每个月的讨论，健康信念模型障碍维度的占比都在 80% 以上。关于 HPV 疫苗讨论集中在障碍维度的原因，会在下节进行内容分析讨论。

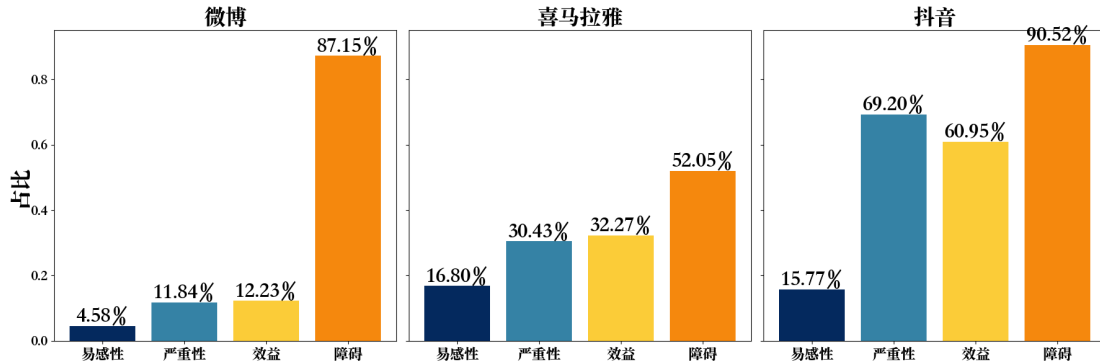


图 4.1 健康信念模型的四个维度在三类社交媒体平台上的分布

对微博平台上健康信念模型的四个维度（易感性、严重性、效益和障碍）的分布进行 Cochran's Q 检验分析。分析结果显示，四个维度的分布存在显著差异（ $Q = 190634.70$, $df = 3$, $p < 0.01$ ）。为了进一步确定具体哪些维度之间存在差异，进行了成对 McNemar 检验的事后分析，结果如下表所示。

表 4.1 Cochran's Q 检验结果

检验统计量	自由度	Q 值	P 值
Cochran's Q	3	190634.70	<0.01

事后检验结果表明，所有维度之间均存在显著差异（ $p < 0.01$ ）。其中，障碍维度（143,469 条）的文本数量显著高于其他三个维度；效益维度（20,133 条）

表 4.2 成对 McNemar 检验结果

组别比较	文本数量差异	P 值	显著性
障碍 vs 效益	123336	<0.01	显著差异
障碍 vs 严重性	123980	<0.01	显著差异
障碍 vs 易感性	135926	<0.01	显著差异
效益 vs 严重性	644	<0.01	显著差异
效益 vs 易感性	12590	<0.01	显著差异
严重性 vs 易感性	11946	<0.01	显著差异

和严重性维度（19,489 条）的文本数量相对接近但仍存在显著差异；易感性维度（7,543 条）的文本数量显著低于其他维度。

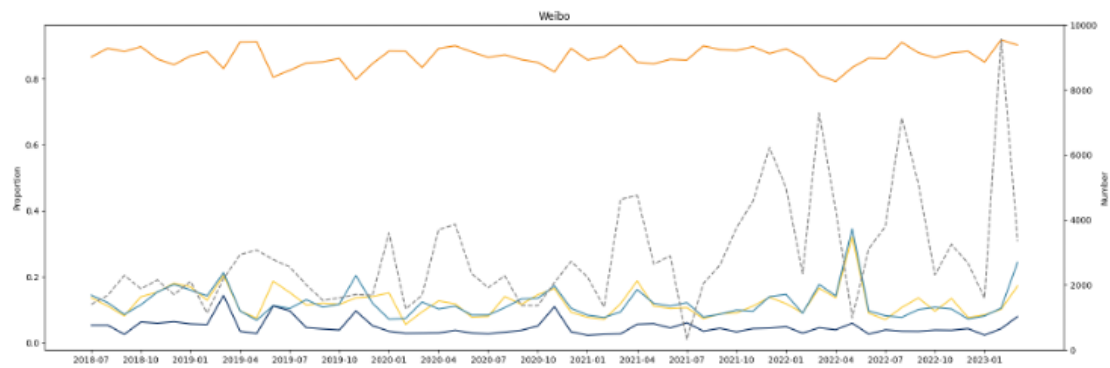


图 4.2 健康信念模型的四个维度在文本平台上的月度占比时序图

4.1.2 音频平台的健康信念模型分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三种社交媒体平台上的总体数量分布情况如图 4.1 所示。在音频平台喜马拉雅上，健康信念模型的易感性维度占比为 16.80%，严重性维度占比为 30.43%，效益维度占比为 32.27%，障碍维度占比为 52.05%。其中音频社交媒体平台喜马拉雅障碍维度关于 HPV 疫苗的讨论最多，易感性维度关于 HPV 疫苗的讨论最小，严重性维度和效益维度关于 HPV 疫苗的讨论一样多。但是障碍维度是易感性维度的 1.60 倍，障碍维度是严重性维度的 1.16 倍，障碍维度是效益维度的 1.14 倍。严重性维度是易感性维度的 1.36 倍，效益维度是易感性维度的 1.39 倍。而严重性维度和效益维度基本相同。关于 HPV 疫苗主要讨论了什么，会在下小节进行内容分析讨论。健康信念模型的易感性、严重性、效益和障碍四个维度在音频平台上的月度数量分布情况如图 4.3 所示。

对喜马拉雅平台上健康信念模型的四个维度（易感性、严重性、效益和障碍）的进行 Cochran's Q 检验分析。基于总文本数 1227 条，其中易感性维度 206 条（16.80%），严重性维度 373 条（30.43%），效益维度 396 条（32.27%），障碍维度 639 条（52.05%）。分析结果显示，四个维度的分布存在显著差异（ $Q = 428.90$, $df = 3$, $p < 0.01$ ）。为了进一步确定具体哪些维度之间存在差异，进行了

成对 McNemar 检验的事后分析，结果如下表所示。

表 4.3 Cochran's Q 检验结果

检验统计量	自由度	Q 值	P 值
Cochran's Q	3	428.90	<0.01

表 4.4 成对 McNemar 检验结果

组别比较	文本数量差异	P 值	显著性
障碍 vs 效益	243	<0.01	显著差异
障碍 vs 严重性	266	<0.01	显著差异
障碍 vs 易感性	433	<0.01	显著差异
效益 vs 严重性	23	<0.01	显著差异
效益 vs 易感性	190	<0.01	显著差异
严重性 vs 易感性	167	<0.01	显著差异

事后检验结果表明，所有维度之间均存在显著差异 ($p < 0.01$)。其中，障碍维度（639 条，52.05%）的文本数量显著高于其他三个维度；效益维度（396 条，32.27%）和严重性维度（373 条，30.43%）的文本数量相对接近但仍存在显著差异；易感性维度（206 条，16.80%）的文本数量显著低于其他维度。



图 4.3 健康信念模型的四个维度在声音平台上的月度占比时序图

4.1.3 视频平台的健康信念模型分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三种社交媒体平台上的总体数量分布情况如图 4.1 所示。在视频平台抖音上，健康信念模型的易感性维度占比为 15.77%，严重性维度占比为 69.20%，效益维度占比为 60.95%，障碍维度占比为 90.52%。其中，视频社交媒体平台抖音上，障碍维度关于 HPV 疫苗的讨论最多，易感性维度关于 HPV 疫苗的讨论最少，严重性维度比效益维度的讨论多。障碍维度是易感性维度的 5.70 倍，严重性维度是易感性维度的 4.40 倍，效益维度是易感性维度的 3.90 倍。严重性维度占比超出 69%，效益维度占比超出 60%，障碍维度占比高达 90%。在视频平台抖音中，关于 HPV 疫苗的讨论严重性维度、效益维度、障碍维度占比非常高，易感性维度稍微比其他维度讨论较少。关于 HPV 疫苗主要讨论了什么，会在下小节进行内容分析讨论。健康

信念模型的易感性、严重性、效益和障碍四个维度在视频平台抖音上的月度数量分布情况如图 4.4 所示。

对抖音平台上健康信念模型的四个维度（易感性、严重性、效益和障碍）的分布进行 Cochran's Q 检验分析。基于总文本数 1224 条，其中易感性维度 193 条（15.77%），严重性维度 847 条（69.20%），效益维度 746 条（60.95%），障碍维度 1108 条（90.52%）。分析结果显示，四个维度的分布存在显著差异（ $Q = 1247.3$, $df = 3$, $p < 0.01$ ）。为了进一步确定具体哪些维度之间存在差异，进行了成对 McNemar 检验的事后分析，结果如下表所示。

表 4.5 Cochran's Q 检验结果

检验统计量	自由度	Q 值	P 值
Cochran's Q	3	1247.30	<0.01

表 4.6 成对 McNemar 检验结果

组别比较	文本数量差异	P 值	显著性
障碍 vs 效益	362	<0.01	显著差异
障碍 vs 严重性	261	<0.01	显著差异
障碍 vs 易感性	915	<0.01	显著差异
效益 vs 严重性	-101	<0.01	显著差异
效益 vs 易感性	553	<0.01	显著差异
严重性 vs 易感性	654	<0.01	显著差异

事后检验结果表明，所有维度之间均存在显著差异（ $p < 0.01$ ）。其中，障碍维度（1108 条，90.52%）的文本数量显著高于其他三个维度；严重性维度（847 条，69.20%）和效益维度（746 条，60.95%）的文本数量相对接近但仍存在显著差异；易感性维度（193 条，15.77%）的文本数量显著低于其他维度。

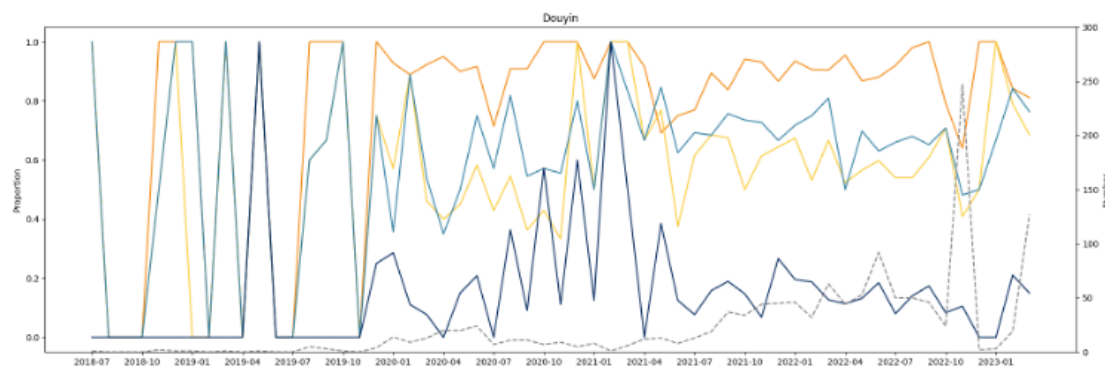


图 4.4 健康信念模型的四个维度在视频平台上的月度占比时序图

总结三种平台的内容，发现以下的平台特点。

第一，从四个维度来总结平台讨论内容的差异发现，文本平台微博内容集中于国际国家政策和相关实践新闻；音频平台喜马拉雅内容集中于 HPV 疫苗，从客观的科学视角出发科普疫苗的各种知识和疫苗公司的发展；音频平台抖音内容则以个人叙述和个人的视角用生活中的例子和观点讨论。

第二，音频平台喜马拉雅热衷于把中国大陆的 HPV 疫苗的效果、接种政策进行中外对比。视频平台抖音则支持国产疫苗，比较拥护“国产”。

第三，文本平台微博、视频平台抖音在谈论内容时给疫苗打上“女性”标签或以女性的角度进行叙事，具有性别化的视角。

第四，文本平台微博和音频平台喜马拉雅对接种对象内容讨论应该扩展到全民，不仅仅局限于女性。

4.2 三种平台在健康信念模型维度下的比较分析

4.2.1 易感性维度比较

1. 描述性统计分析

三种社交媒体平台在健康信念模型的易感性、严重性、效益和障碍四个维度上的总体数量分布情况如图 4.5 所示。在健康信念模型的易感性维度上，文本平台微博占比为 4.58%，音频平台喜马拉雅占比为 16.80%，视频平台抖音占比为 15.77%。在易感性维度上，喜马拉雅平台讨论的内容占比最多，微博平台讨论的内容占比最少。喜马拉雅平台的占比是微博平台的 3.67 倍，抖音平台的占比是微博平台的 3.44 倍。在健康信念模型的易感性维度上，喜马拉雅平台和抖音平台的讨论占比较多，而微博平台上关于易感性维度的讨论相对较少。音频平台和视频平台比文本平台更加关注易感性。内容分析将会具体分析文本平台、音频平台、视频平台关于易感性维度的具体关注内容。

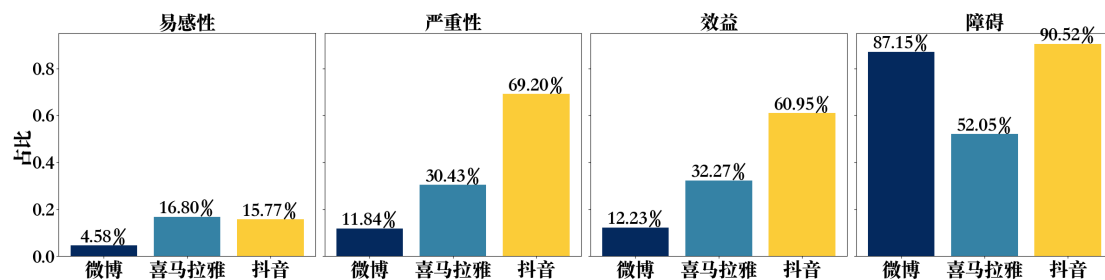


图 4.5 三类社交媒体平台在健康信念模型的四个维度的分布

为了分析微博、抖音和喜马拉雅在易感性维度上的显著差异，本文使用卡方检验（Chi-square test）来比较这三个平台的易感性维度比例。结果见表 4.7 所示，卡方检验结果显示三个平台在易感性维度上存在显著差异（ $\chi^2 = 892.34$, $df = 2$, $p < 0.01$ ）。为了进一步确定具体哪些平台之间存在差异，本文进行了 Bonferroni 校正的事后检验，结果见表 4.8 所示。事后检验结果表明，微博平台（4.58%）的易感性维度比例显著低于喜马拉雅（16.80%）和抖音（15.77%）平台（ $p < 0.01$ ），而喜马拉雅和抖音平台之间的差异不显著（ $p = 0.52$ ）。这表明在易感性维度的讨论上，音视频平台（喜马拉雅和抖音）比文本平台（微博）具有更高的占比。

表 4.7 易感性维度在三个平台的卡方检验结果

统计量	值	自由度	P 值
卡方值	892.34	2	<0.01

表 4.8 易感性维度事后检验结果

平台对比	比例差异	P 值 (校正后)	显著性
微博 vs 喜马拉雅	-12.22%	<0.01	显著
微博 vs 抖音	-11.19%	<0.01	显著
喜马拉雅 vs 抖音	1.03%	0.52	不显著

2. 内容分析

在健康信念模型的易感性维度上，文本平台微博占比为 4.58%，音频平台喜马拉雅占比为 16.80%，视频平台占比为 15.77%。健康信念模型易感性维度的内容分析在表 4.9 显示。其中平台代表的是文本平台微博、音频平台喜马拉雅、视频平台抖音。簇号代表的是在易感性维度上，三种平台进行 BERTopic 主题建模之后，聚类的最大的三个代表性的主题。微博上的易感性维度讨论了 HPV 的传播途径、进行疫苗推广和疫苗接种的目标人群。在微博上，WS0 主题提到了女性感染 HPV 的各种途径。WS1 主题与 2022 年 HPV 疫苗供应短缺的消息对应，讨论了女性接种 HPV 疫苗的时间，进行 HPV 疫苗的推广。WS2 主题出现在 2022 年，指出 HPV 疫苗不应局限于适龄女性，而应该对所有人开放，包括男性和成人的男性与女性。喜马拉雅播客平台专注于 HPV 疫苗，对 HPV 疫苗的现状进行中外对比，目前国内的 HPV 疫苗落后于国外，强调疫苗的安全性，以及推广 HPV 疫苗到男性或者女性。XS0 主题提到中国目前使用的二价疫苗在国外已经不再使用，国外使用的是九价疫苗。XS1 主题强调疫苗接种的好处大于副作用。XS2 主题指出 HPV 疫苗在国外男性和女性都可以接种。抖音视频平台关注感染 HPV 从而引起的疾病症状。D0 主题强调 HPV 病毒会导致各种疾病，也可能导致癌症。D1 主题强调 HPV 疫苗对 HPV 病毒进行预防。D2 主题站在女性的视角，强调了女性在不知情的情况下，因为男友在外面乱搞而感染了 HPV 病毒，站在日常生活的视角，从当前事例出发展示了 HPV 病毒的易感性。

从健康信念模型的易感性维度视角来看三种平台内容的差异，结果表现出以下特点。

第一，文本平台微博讨论了 HPV 的传播途径、疫苗接种的年龄限制和目标人群。音频平台喜马拉雅更加客观强调 HPV 疫苗，讨论了 HPV 疫苗的科普和中外对比。而视频平台抖音讨论则关注了关注感染 HPV 从而引起的疾病症状。

第二，文本平台微博、音频平台喜马拉雅有推广 HPV 疫苗的内容讨论，但是视频抖音只是在讨论 HPV 病毒易于传播。

第三，文本平台微博、视频平台抖音具有女性化的表达，音频平台更加客观。

表 4.9 易感性维度的话题讨论

平台	簇号	数量	代表句子
微博	WS0	822	HPV 可以通过多种途径进行传播, 如果女性感染了 HPV 应该怎么办?
	WS1	627	女性在什么时候接种疫苗最好? 越快越好
	WS2	589	广泛接种 HPV 疫苗可以预防宫颈癌, 接种对象应该扩展到男孩和成人
喜马拉雅	XS0	794	中国用的是国外已经不再使用的二价 HPV 疫苗
	XS1	460	HPV 疫苗非常安全, 接种之后的好处远大于副作用
	XS2	446	HPV 疫苗男女都可以接种
抖音	DS0	751	HPV 不仅引起宫颈癌, 除了宫颈癌之外, 它还会导致皮肤和粘膜的多发性乳头状瘤、肛门癌、口咽癌, 并且可以传播给他人。
	DS1	127	我们生活的环境中有一百多种 HPV 病毒, 而四价疫苗可以预防 80% 的宫颈癌和 90% 的尖锐湿疣。
	DS2	85	那个女生毫不知情的情况下, 因为男友在外面乱搞而被传染了 HPV

第四, 文本平台微博、音频平台抖音对于 HPV 疫苗的接种对象来说, 不仅仅局限于适龄女性, 要扩展到男性和其他年龄段的女性。

第五, 视频平台抖音特别关注到了中国研发生产的二价和四价疫苗。

4.2.2 严重性维度比较

1. 描述性统计分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三种社交媒体平台上的总体数量分布情况如图 4.5 所示。在健康信念模型的严重性维度上, 文本平台微博占比为 11.84%, 音频平台喜马拉雅占比为 30.43%, 视频平台抖音占比为 69.20%。在严重性维度上, 抖音平台讨论的内容占比最多, 微博平台讨论的内容占比最少。喜马拉雅平台的占比是微博平台的 2.57 倍, 抖音平台的占比是微博平台的 5.84 倍。在健康信念模型的严重性维度上, 抖音平台的讨论占比最多, 喜马拉雅平台次之, 而微博平台上关于严重性维度的讨论相比于其他平台占比较少。音频平台和视频平台比文本平台更加关注严重性。内容分析将会具体分析文本平台、音频平台、视频平台关于严重性维度具体关注的内容。

为了分析微博、抖音和喜马拉雅在严重性维度上的显著差异, 本文使用卡方检验 (Chi-square test) 来比较这三个平台的严重性维度比例。结果见表 4.10 所示, 卡方检验结果显示三个平台在严重性维度上存在显著差异 ($\chi^2 = 3641.52$, $df = 2$, $p < 0.01$)。为了进一步确定具体哪些平台之间存在差异, 本文进行了 Bonferroni 校正的事后检验, 结果见表 4.11 所示。事后检验结果表明, 三个平台之间的严重性维度比例均存在显著差异 ($p < 0.01$): 抖音平台 (69.20%) 显著高于喜马拉雅 (30.43%) 和微博 (11.84%), 同时喜马拉雅的比例也显著高于微博。这表明

在严重性维度的讨论上，视频平台（抖音）的占比最高，其次是音频平台（喜马拉雅），文本平台（微博）的占比最低。

表 4.10 严重性维度在三个平台的卡方检验结果

统计量	值	自由度	P 值
卡方值	3641.52	2	<0.01

表 4.11 严重性维度事后检验结果

平台对比	比例差异	P 值 (校正后)	显著性
微博 vs 喜马拉雅	-18.59%	<0.01	显著
微博 vs 抖音	-57.36%	<0.01	显著
喜马拉雅 vs 抖音	-38.77%	<0.01	显著

2. 内容分析

在健康信念模型的严重性维度上，文本平台微博占比为 11.84%，音频平台喜马拉雅占比为 30.43%，视频平台抖音占比为 69.20%。严重性维度的话题在下表 4.12 中显示。在以文本平台微博上，最大的两个簇群 WS0 和 WS1 讨论了在中国女性中，宫颈癌是第二大致命癌症，适龄女性接种率不到 1%。而 WS3 话题声称 HPV 疫苗的有效性有限，无法提供 100% 的宫颈癌保护。从严重性来看，微博平台上存在两种相反的观点，一种是观点是 HPV 疫苗的接种率很低，另一种观点却在说明 HPV 疫苗效果有限。在音频平台喜马拉雅上，XS0 话题提到 HPV 疫苗无法提供完全保护。XS1 话题称在香港和新加坡地区没有年龄限制，在中国大陆疫苗供应短缺，受到年龄的限制。XS2 话题指出 LGBTQ+ 群体可能患有与 HPV 相关的癌症。总体而言，播客平台提供了对 HPV 政策的客观解读和中外对比。在视频平台抖音上，重点更多地放在个人叙述上。最大的簇群 DS0 话题是女性分享与 HPV 和 HPV 疫苗相关的自身经历的女性主义叙述。DS1 话题建议已感染 HPV 的女性需要进行宫颈涂片筛查以预防宫颈癌。DS2 话题提到如果无法接触进口疫苗，应该选择国产疫苗进行自我保护。在视频平台抖音上，用户从个人角度出发，当接种不到九价的 HPV 疫苗时，讨论应该采取的补救措施和行动。

从健康信念模型的严重性视角来看三种平台内容的差异，结果表现出以下特点。

第一，文本平台微博上的观点有两种倾向：一种认为癌症严重且 HPV 疫苗接种率很低，另一种则指出 HPV 疫苗效果有限。音频平台喜马拉雅主要聚焦于 HPV 疫苗的有效性及其供应的国际对比。视频平台抖音更关注个体感染 HPV 病毒后对生活的具体影响，以及九价 HPV 疫苗短缺时的补救措施。

第二，文本平台微博和视频平台抖音都涉及女性话题。微博将 HPV 疫苗与女性健康相关联，而抖音以女性视角讲述 HPV 话题。音频平台喜马拉雅则采取

表 4.12 严重性维度的话题讨论

平台	簇号	数量	代表句子
微博	WS0	1016	目前, 中国学龄女生的 HPV 疫苗接种率不到 1%。
	WS1	845	在中国, 宫颈癌是女性发病率第二高的癌症类型。
	WS2	660	接种 HPV 疫苗是否意味着不会患上宫颈癌? 不是的!
喜马拉雅	XS0	1023	目前, HPV 疫苗是一种预防性疫苗, 不能治愈 HPV 感染, 并且不能百分之百预防宫颈病变和宫颈癌。
	XS1	948	HPV 疫苗供应紧张, 并且接种受年龄限制, 但是在新加坡没有严格的限制
	XS2	617	HPV 还可以导致直肠癌, 这种癌症也存在于同性恋群体中。
抖音	DS1	691	自从我感染了 HPV 以来, 我从未与男性发生过性行为。并且我已经接种了 HPV 疫苗。
	DS2	423	超龄女性可以进行 TCT 检测来预防宫颈癌。
	DS3	373	为了避免在等待九价疫苗时感染了 HPV, 可以先接种国产的双价或四价疫苗。

相对中立的立场, 既未强调特定性别标签, 还涉及了性少数群体的健康需求。

4.2.3 效益维度比较

1. 描述性统计分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三种社交媒体平台上的总体数量分布情况如图 4.5 所示。在健康信念模型的效益维度上, 文本平台微博占比为 12.23%, 音频平台喜马拉雅占比为 32.27%, 视频平台抖音占比为 60.95%。在效益维度上, 抖音平台讨论的内容占比最多, 微博平台讨论的内容占比最少。喜马拉雅平台的占比是微博平台的 2.64 倍, 抖音平台的占比是微博平台的 4.99 倍。在健康信念模型的效益维度上, 抖音平台的讨论占比最多, 喜马拉雅平台次之, 而微博平台上关于效益维度的讨论相比于其他平台占比较少。音频平台和视频平台比文本平台更加关注效益。内容分析将会具体分析文本平台、音频平台、视频平台关于效益维度具体关注的内容。

为了分析微博、抖音和喜马拉雅在效益维度上的显著差异, 本文使用卡方检验 (Chi-square test) 来比较这三个平台的效益维度比例。结果见表 4.13 所示, 卡方检验结果显示三个平台在效益维度上存在显著差异 ($\chi^2 = 3128.76$, $df = 2$, $p < 0.01$)。为了进一步确定具体哪些平台之间存在差异, 本文进行了 Bonferroni 校正的事后检验, 结果见表 4.14 所示。事后检验结果表明, 三个平台之间的效益维度比例均存在显著差异 ($p < 0.01$): 抖音平台 (60.95%) 显著高于喜马拉雅 (32.27%) 和微博 (12.23%), 同时喜马拉雅的比例也显著高于微博。这表明在效益维度的讨论上, 视频平台 (抖音) 的占比最高, 其次是音频平台 (喜马拉雅), 文本平台 (微博) 的占比最低。

表 4.13 效益维度在三个平台的卡方检验结果

统计量	值	自由度	P 值
卡方值	3128.76	2	<0.01

表 4.14 效益维度事后检验结果

平台对比	比例差异	P 值 (校正后)	显著性
微博 vs 喜马拉雅	-20.04%	<0.01	显著
微博 vs 抖音	-48.72%	<0.01	显著
喜马拉雅 vs 抖音	-28.68%	<0.01	显著

2. 内容分析

在健康信念模型的效益维度上,文本平台微博占比为 12.23%,音频平台喜马拉雅占比为 32.27%,视频平台抖音占比为 60.95%。效益维度的话题在表格 4.15 中显示。在以文本体平台微博上, WB0 话题突出了世界卫生组织的政策。WB1 话题建议国产疫苗和进口疫苗对宫颈癌保护效果相同。WB2 话题表明 HPV 疫苗的有效性与价格呈正相关关系。在音频平台喜马拉雅上, XB0 话题讨论了默沙东九价疫苗的年龄限制扩大问题,这在 2022 年 9 月和 10 月是一个热门讨论。XB1 话题强调了早期筛查和治疗对 HPV 感染和宫颈癌的重要性。XB2 话题提出国产疫苗和进口疫苗具有类似的功能和安全性。在视频平台抖音上, DB0 话题认为九价疫苗应该对 45 岁及以上人群开放。DB1 话题对九价疫苗的高价格表示关切,并提倡低成本的国产二价疫苗。DB3 话题建议人们尽早接种疫苗。

表 4.15 效益维度的话题讨论

平台	簇号	数量	代表句子
微博	WB0	1676	疫苗接种、筛查和治疗是世界卫生组织加速消除宫颈癌的全球战略的一部分。
	WB1	1328	进口和国内产的 HPV 疫苗正在全国各地接种,来预防宫颈癌。
	WB2	661	HPV 疫苗的价数越高,能防护的 HPV 病毒种类越多。
喜马拉雅	XB0	1360	默沙东的九价 HPV 疫苗已获得中国国家药品监督管理局批准,新的适应症范围扩展至 9 岁至 45 岁的适龄女性
	XB1	1281	在感染和恶性癌症之间存在一个时间间隔期,本研究可以早发现早治疗。
	XB2	1021	国产和进口疫苗在效果和安全性方面可能没有太大的差异。
抖音	DB1	581	未成年人可以稍晚接种九价疫苗,接近 45 岁的年龄可以首次接种,接种与 TCT 检查无关。
	DB2	526	九价疫苗在成本效益上不具备优势,效果也不是很显著,而且存在被骗的风险;二价疫苗在预防宫颈病变和宫颈癌方面最有效,明显优于四价疫苗,并且提供交叉保护。
	DB3	373	疫苗接种越早,感染 HPV 的风险越低,疫苗提供的保护越大。

从健康信念模型的效益视角来看三种平台内容的差异,结果表现出以下特点。

第一，文本平台微博关注国际和国内的战略和政策，疫苗效果的比较。音频平台喜马拉雅则关注了九价扩龄这个特殊事件，除此之外则关注疾病和疫苗的科学知识。视频平台抖音则关注 HPV 疫苗可以接种的人群不受限制，推广国产疫苗。

第二，音频平台喜马拉雅在中国研发和国外研发的疫苗的安全性的角度进行中外对比。

4.2.4 障碍维度比较

1. 描述性统计分析

健康信念模型的易感性、严重性、效益和障碍四个维度在三种平台上的总体数量分布情况如图 4.5 所示。在健康信念模型的障碍维度上，文本平台微博占比为 87.15%，音频平台喜马拉雅占比为 52.05%，视频平台抖音占比为 90.52%。在障碍维度上，抖音平台讨论的内容占比最多，喜马拉雅平台讨论的内容占比最少。微博平台的占比是喜马拉雅平台的 1.67 倍，抖音平台的占比是喜马拉雅平台的 1.74 倍。在健康信念模型的障碍维度上，抖音平台和微博平台的讨论占比最多，而喜马拉雅平台上关于障碍维度的讨论相比于其他平台占比较少。文本平台和视频平台比音频平台更加关注障碍性。内容分析将会具体分析文本平台、音频平台、视频平台关于障碍维度具体关注的内容。

为了分析微博、抖音和喜马拉雅在障碍维度上的显著差异，本文使用卡方检验（Chi-square test）来比较这三个平台的障碍维度比例。结果见表 4.16 所示，卡方检验结果显示三个平台在障碍维度上存在显著差异（ $\chi^2 = 1247.89$, $df = 2$, $p < 0.01$ ）。为了进一步确定具体哪些平台之间存在差异，本文进行了 Bonferroni 校正的事后检验，结果见表 4.17 所示。事后检验结果表明，三个平台之间的障碍维度比例均存在显著差异（ $p < 0.01$ ）：抖音平台（90.52%）和微博平台（87.15%）的比例相对较高，且两者之间存在显著差异，而喜马拉雅平台（52.05%）的比例显著低于其他两个平台。这表明在障碍维度的讨论上，视频平台（抖音）和文本平台（微博）的占比较高，而音频平台（喜马拉雅）的占比明显较低。

表 4.16 障碍维度在三个平台的卡方检验结果

统计量	值	自由度	P 值
卡方值	1247.89	2	<0.01

表 4.17 障碍维度事后检验结果

平台对比	比例差异	P 值 (校正后)	显著性
微博 vs 喜马拉雅	35.10%	<0.01	显著
微博 vs 抖音	-3.37%	<0.01	显著
喜马拉雅 vs 抖音	-38.47%	<0.01	显著

2. 内容分析

在健康信念模型的障碍维度上,文本平台微博占比为 87.15%,音频平台喜马拉雅占比为 48.8%,视频平台占比为 89.24%。障碍维度的话题在表格 4.18 中显示。在文本平台微博上, WB0 主题关注 HPV 疫苗短缺。WB1 主题讨论了 HPV 疫苗的副作用。WB3 主题比较了不同 HPV 疫苗的价格。在音频平台喜马拉雅上, XB0 主题讨论接种疫苗的年龄限制。XB1 主题强调了昂贵疫苗的普及度。XB2 主题指出疫苗公司获得了巨大利润,其股价飙升。在视频平台抖音上, DB0 主题讨论了疫苗短缺。DB1 主题关注接种疫苗的资格。DB2 主题建议人们尽早接种疫苗,不盲目等待九价疫苗。

表 4.18 障碍维度的话题讨论

平台	簇号	数量	代表句子
微博	WB0	20,760	去支付宝预约 HPV 疫苗抢不上
	WB1	7,429	接种疫苗后,我的胳膊开始疼痛,并且出现不良反应,发烧和重感冒。
	WB2	6,585	国产二价疫苗比进口的九价疫苗便宜。
喜马拉雅	XB0	1,086	HPV 疫苗对于 26 岁以上的人的效果有限
	XB1	1,070	价格较高的 HPV 疫苗比价格较低的更受欢迎,且市场上更为稀缺。
	XB2	404	沃森生物和万泰生物医药的营收和利润实现了大幅增长。
抖音	DB1	970	抢到九价疫苗就像中彩票一样。
	DB2	732	如果已经感染 HPV,仍可以接种 HPV 疫苗。
	DB3	629	接种疫苗越早越好。如果有二价疫苗,就接受二价疫苗接种;如果有四价疫苗,就接受四价疫苗接种;不要盲目等待九价疫苗。

从健康信念模型的障碍视角来看三种平台内容的差异,结果表现出以下特点。

第一,文本平台微博集中关注 HPV 疫苗短缺、疫苗的不良反应和国产疫苗费用较低等民生话题。音频平台喜马拉雅更关注话题更客观,话题是疫苗效果有限、哪种疫苗更受欢迎,疫苗公司的经济和产业机遇,更加专注疫苗本身及相关话题。视频平台抖音则关注市场上疫苗短缺的个人叙述。

第二,文本平台微博、视频平台抖音叙事的角度是从普通用户的角度出发叙事的。文本平台微博提到了疫苗的不良反应,这在其他平台的讨论较少。视频平台抖音表达了支持接种 HPV 疫苗的表述。音频平台在客观讲述 HPV 疫苗知识,销售欢迎度,以及疫苗企业的利润增涨,从客观说明了 HPV 疫苗在中国接种使多方受益。

4.2.5 小结

从健康信念模型的四个维度来分析三种平台讨论内容的差异,结果如下。

第一，在健康信念模型的易感性维度，文本平台微博讨论了 HPV 的传播途径、疫苗接种的年龄限制和目标人群。音频平台喜马拉雅更加客观强调 HPV 疫苗，讨论了 HPV 疫苗的科普和中外对比。而视频平台抖音讨论则关注了关注感染 HPV 从而引起的疾病症状。

第二，在健康信念模型的严重性维度，文本平台微博一种观点谈论癌症严重，HPV 疫苗的接种率很低，另一种观点却在说明 HPV 疫苗效果有限。音频平台喜马拉雅聚焦于 HPV 疫苗，讨论了 HPV 疫苗效果有限和中外对比供应的对比。而视频平台抖音讨论则关注个体层面感染 HPV 病毒对生活造成的影响，九价 HPV 疫苗短缺时应该采取的补救措施。

第三，在健康信念模型的效益维度，文本平台微博关注国际和国内的战略和政策，疫苗效果的比较。音频平台喜马拉雅则关注了九价扩龄这个特殊事件，除此之外则关注疾病和疫苗的科学知识。视频平台抖音则关注 HPV 疫苗可以接种的人群不受限制，推广国产疫苗。

第四，在健康信念模型的障碍维度，文本平台微博集中关注 HPV 疫苗短缺、疫苗的不良反应和国产疫苗费用较低等民生话题。音频平台喜马拉雅更关注疫苗效果、哪种疫苗更受欢迎，疫苗公司的经济和产业机遇，更加专注疫苗本身及相关话题。视频平台抖音则关注市场上疫苗短缺的个人叙述。

总结三种平台的内容，发现以下的平台特点。

第一，从四个维度来总结平台讨论内容的差异发现，文本平台微博内容集中于国际国家政策和相关实践新闻；音频平台喜马拉雅内容集中于 HPV 疫苗，从客观的科学视角出发科普疫苗的各种知识和疫苗公司的发展；音频平台抖音内容则以个人叙述和个人的视角用生活中的例子和观点讨论。

第二，音频平台喜马拉雅热衷于把中国大陆的 HPV 疫苗的效果、接种政策进行中外对比。视频平台抖音则支持国产疫苗，比较拥护“国产”。

第三，文本平台微博、视频平台抖音在谈论内容时给疫苗打上“女性”标签或以女性的角度进行叙事，具有性别化的视角。

第四，文本平台微博和音频平台喜马拉雅对接种对象内容讨论应该扩展到全民，不仅仅局限于女性。

4.3 三类社交平台的情绪分析

4.3.1 总占比分析

图 4.6 显示了微博、喜马拉雅和抖音的情绪特征。在微博上，非情绪性内容的比例最高（50.10%），其次是喜马拉雅（39.14%）和抖音（44.04%）。本研究发现微博上只有 11.40% 的帖子是积极的，38.50% 的帖子是消极的。在喜马拉雅

上, 22.23% 的内容是积极的, 38.63% 是消极的。在抖音上, 21.64% 的内容是积极的, 34.33% 是消极的。

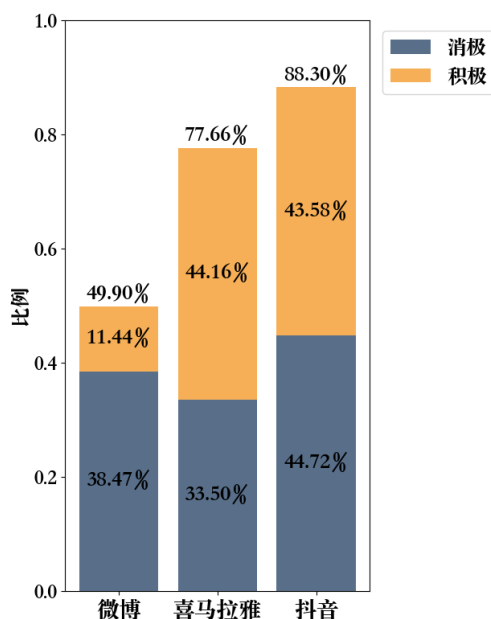


图 4.6 微博、喜马拉雅、抖音上的情绪分布

注：在图中，展示了两种情绪，其余为无情绪

总之，在文本平台微博，拥有更少的情绪化表达，更多的是客观的说明。在音频平台喜马拉雅和视频平台抖音上情绪化表达相比文本平台微博更多。音频平台情绪化表达相比于视频平台的表达的情绪更多。在社交媒体的情绪占比中，音频平台和视频平台相比较于文本平台更加积极。

4.3.2 时间序列分析

1. 文本平台

本研究还分析了情绪分布随时间的变化。图 4.7 表示了以文本平台微博从 2018 年 7 月到 2023 年 3 月每个月的积极消极情绪占比随月份变化的时间序列图。从图中可以看出，在 2018 年 7 月到 2023 年 3 月中每个月无情绪的占比超过情绪占比，而在情绪占比中，消极情绪占比超过积极情绪。消极情绪和无情绪占据了每个月关于 HPV 疫苗讨论的主流。

本文还分析了文本平台讨论高峰的月份和主题关键词，如表 4.19 所示。例如，在 2022 年 3 月 8 日，中国“两会”期间，多位全国人大代表提出为适龄女性免费接种 HPV 疫苗的建议。这一建议引发了较多的积极情绪讨论，与无情绪或消极情绪的讨论相比有所增加。这一讨论高峰持续了两个月，即 2022 年 3 月和 4 月，其中积极情绪的讨论较事件发生前后的月份明显增多，这显示出健康信念中障碍性的减少可能增加积极情绪的产生。讨论内容主要集中在适龄女性免

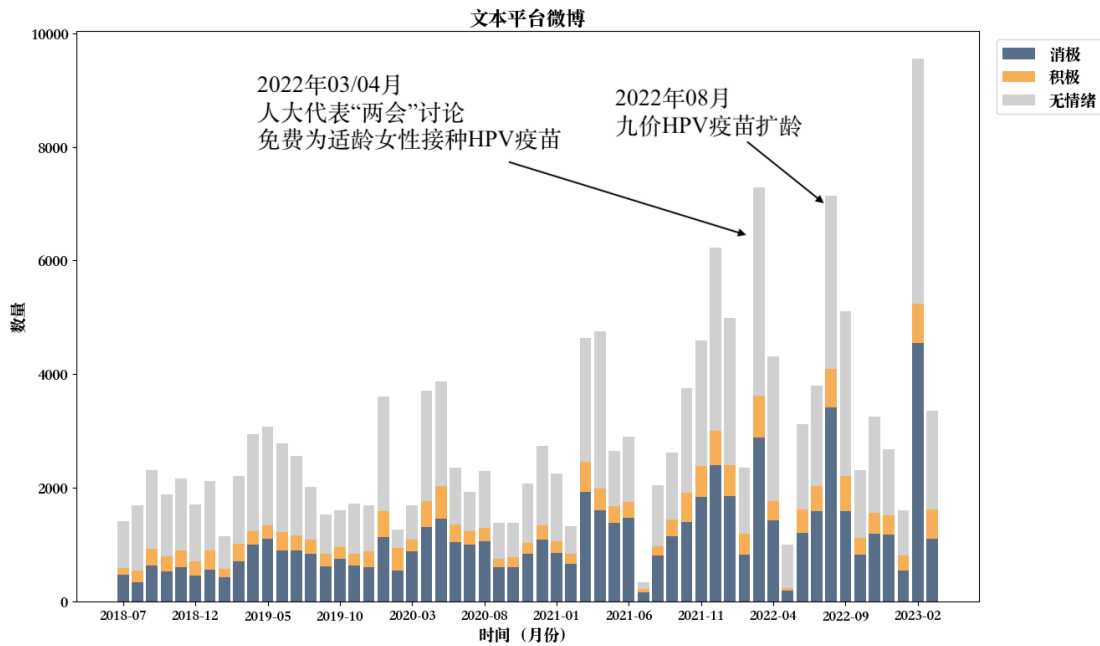


图 4.7 微博时间序列情绪分布

表 4.19 三种平台上讨论的高峰月份的主题关键词

平台	时间	关键词
微博	2022 年 03 月	适龄女性免费接种；全国人大代表；国产疫苗；纳入国家免疫规划
	2022 年 08 月	九价扩龄；国家药监局批准；9 到 45 周岁
喜马拉雅	2022 年 05 月	葛兰素史克；默沙东；感染；病毒
	2022 年 09 月	默沙东；九价疫苗扩龄；年龄 9 到 45 周岁；获批
	2022 年 10 月	适龄女性；免费接种；默沙东
抖音	2022 年 03 月	二价疫苗；免费接种
	2022 年 06 月	摇号预约；澳门；香港预约
	2022 年 11 月	国产二价疫苗；熬夜偷偷抢；国货之光；高价疫苗

费接种、全国人大代表、国产疫苗及纳入国家免疫规划等方面。在另一事件中，2022 年 8 月，九价 HPV 疫苗扩展至更广泛的年龄层，这一变化同样引发了讨论高峰。在这一时期，积极情绪的讨论较无情绪和消极情绪的讨论有所增加，并且较事件发生前后的月份显著增多，反映出减少健康信念障碍可以促进积极情绪的产生。讨论焦点包括九价疫苗年龄扩展、国家药监局批准、以及扩展的年龄范围——九岁至四十五岁。

2. 音频平台

图 4.8 表示了以音频平台喜马拉雅从 2018 年 07 月到 2023 年 03 月每个月的积极消极情绪占比随月份变化的时间序列图。表 4.19 表示了音频平台喜马拉雅讨论高峰的月份和主题关键词。从图和表中可以看出，在 2018 年 07 月到 2021 年 09 月中每个月关于 HPV 疫苗的讨论比较少，到 2021 年 9 月之后才关于 HPV 疫苗的讨论开始逐渐增多。在发生特定的事件 2022 年 5 月国家卫健委推广 HPV

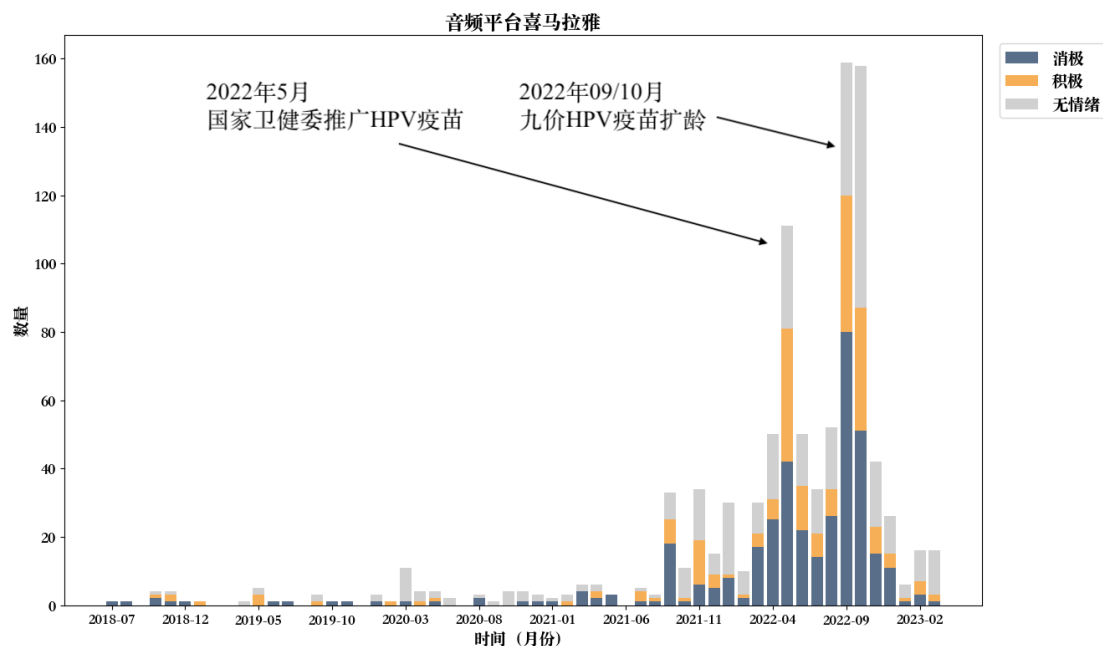


图 4.8 喜马拉雅时间序列情绪分布

疫苗，关于 HPV 疫苗的讨论没积极的情绪的讨论多于消极情绪的讨论。关于这次讨论的内容则是葛兰素史克、默沙东、感染、病毒等关键词，而不是事件本身，更多的关注 HPV 疫苗，包括医药企业的名称。而另一件特定的事件 2022 年 9 月 10 月讨论的 HPV 疫苗九价扩龄，消极情绪的讨论多于积极情绪的讨论。也有部分是无情绪。通过主题建模查看讨论的内容发现，讨论的内容主要是默沙东、九价疫苗扩龄、年龄 9 到 45 周岁、获批、适龄女性、免费接种、默沙东。相比于文本平台，音频平台关于这个事件讨论的峰值具有迟滞性并且讨论的周期较长，音频平台关于此事件的讨论更加专注于疫苗和疫苗企业，音频平台关于此事件的讨论更加积极。

3. 视频平台

图 4.9 表示了以视频平台抖音从 2018 年 07 月到 2023 年 03 月每个月的积极消极情绪占比随月份变化的时间序列图。表 4.19 表示了视频平台抖音讨论高峰的月份和主题关键词。从图和表中可以看出，在 2020 年 01 月到 2020 年 07 月相较于之前的月份讨论较多，但是都是以无情绪的讨论为主。从 2021 年 2 月开始讨论逐渐增多，并且情绪讨论占主导地位。在发生特定的事件 2022 年 6 月九价疫苗短缺，这是一个负面的事件，但是积极的讨论多于消极的讨论，通过内容分析发现关于这次讨论的内容则是摇号预约、澳门、香港预约等关键词，更加关注九价短缺之后用户的解决方案和视角。而另一件特定的事件 2022 年 11 月国产二价疫苗是国货之光。消极的情绪多余积极的情绪，无情绪的讨论增多了。这本来是个积极的事件，通过内容分析发现有国产二价疫苗、熬夜偷偷抢、国货之光、高价疫苗。在宣扬国产二价 HPV 疫苗的同时又表达了抢不到，价格高等负面情

绪，更加关注个人视角。并且视频平台的讨论相较于文本平台和音频平台关注不同的新闻事件和内容。

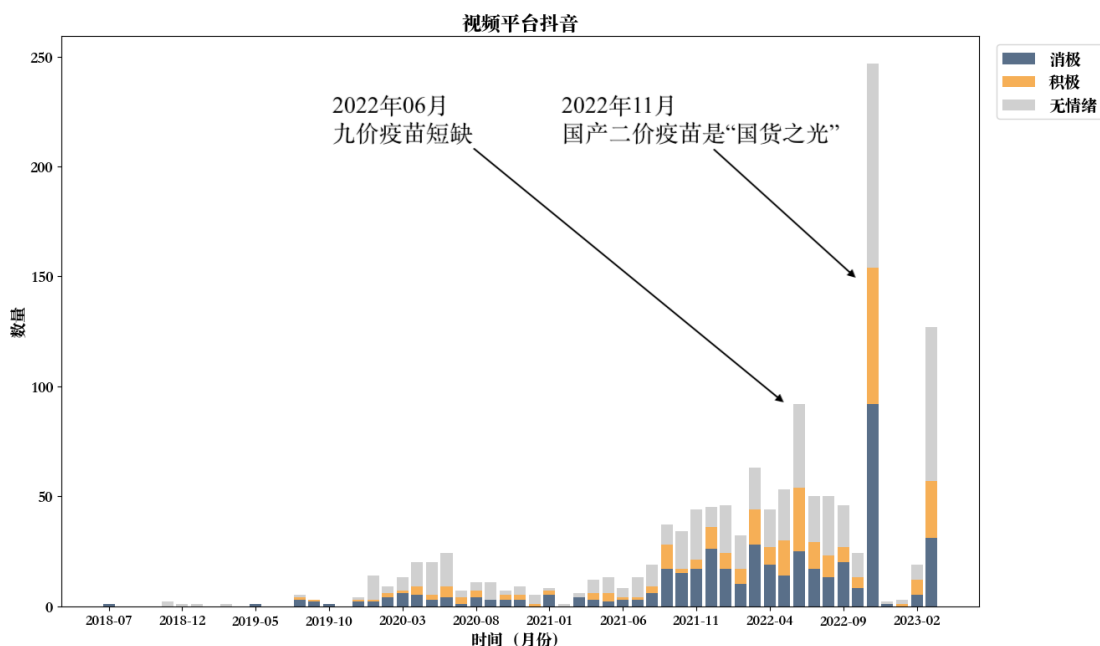


图 4.9 抖音时间序列情绪分布

4.3.3 小结

经过对文本、音频、视频三种平台（微博、喜马拉雅、抖音）的占比分析和时间序列分析，本研究有了以下的发现。

第一，在文本平台微博，拥有更少的情绪化表达，更多比较客观。在音频平台喜马拉雅上情绪化表达相比文本平台更多。音频平台喜马拉雅情绪化表达相比于视频平台抖音表达的情绪更多。在社交媒体的情绪占比中，文本平台微博和于音频平台喜马拉雅相比于视频平台抖音更加消极。

第二，对于特定的发生的事件，音频平台喜马拉雅关于这个事件讨论的峰值具有迟滞性并且讨论的周期较长，并且更加专注于讨论疫苗和疫苗企业。

第三，三种平台讨论峰关注不一样的事件。文本平台微博讨论的峰关注 2022 年 3 月的全国人大代表建议把 HPV 疫苗纳入国家免疫规划，2022 年 8 月九价扩龄这两个事件，这些事件的来源都属于新闻机构和中国政府官方，文本平台会迅速讨论，持续周期较短。音频平台喜马拉雅讨论峰关注 2022 年 5 月国家卫健委提倡打 HPV 疫苗，2022 年 9 月 10 月九价扩龄。2022 年 5 月国家卫健委提倡打 HPV 疫苗事件是 2022 年 3 月全国人大代表提倡 HPV 疫苗纳入国家免疫规划的发展事件，中国有关部门对人大代表事件作出的反应。视频平台抖音的讨论峰是 2022 年 6 月的九价短缺和 2022 年 11 月的国产二价疫苗是国货之光。其中 2022 年 6 月的九价短缺是 2022 年国家卫健委提倡打 HPV 疫苗的事态发展事件，而

2022 年 8-10 月九价扩龄之后由于疫苗短缺 2022 年 11 月的国产二价疫苗是国货之光事件又发生。文本平台微博关注的子事件是国家政策，音频平台喜马拉雅关注的子事件是疫苗和疫苗信息。视频平台抖音关注的子事件是更具有用户视角。

综上所述，文本平台微博更加无情绪，音频平台喜马拉雅和视频平台抖音更加积极。对于特定的事件，音频平台喜马拉雅关于这个事件讨论的峰值具有迟滞性并且讨论的周期较长，并且更加专注于讨论疫苗和疫苗企业。文本平台微博、音频平台喜马拉雅关注有相同的事件，但是视频平台抖音关注事件讨论峰是不同于文本和音频平台的。

第 5 章 结论与讨论

5.1 研究结论

本文深入探讨了 HPV 疫苗相关议题在社交媒体上的多模态内容差异。在研究中，社交媒体平台上的内容主要被划分为三种模态：文本、音频和视频。研究样本是通过收集 2018 年 7 月至 2023 年 3 月期间来自微博、喜马拉雅、抖音等社交媒体平台的数据而获得的。这些样本包括 273,357 条微博文本、1,228 条喜马拉雅音频以及 1,225 条抖音视频。鉴于不同模态内容在形式上的显著差异，直接对比存在困难。并且音频和视频数据的关键信息多集中于人声部分，故本文采取了人声提取和转录的方法，从而将对比焦点转移至文本层面，确保了跨模态内容的可比性。

由于三种社交媒体平台的数据量差异较大，直接主题建模进行内容分析难以进行有效对比，本研究引入了健康信念模型（Health Belief Model, HBM）的四个维度——易感性、严重性、效益和障碍，作为分类框架。这一策略旨在从这四个关键维度出发，系统地研究 HPV 相关议题在社交媒体上多模态内容的差异和特征。基于详细的数据分析和深入的探讨，本研究得出了以下六点主要结论。

5.1.1 健康信念模型维度在三种平台上分布不均

从平台内容差异性的角度来看，文本平台微博主要聚焦于健康信念模型的障碍维度。具体来说，障碍维度按照编码定义可以分为障碍增加与减少两类。在微博上，障碍增加主要体现在 HPV 疫苗短缺和年龄限制等方面，而障碍减少则更多地体现在年龄限制的放宽和免费接种等政策的推广。这些讨论通常与中国政策以及社会新闻紧密相连，以简短且引人注目的文本形式迅速传播。

相比之下，音频平台喜马拉雅在健康信念模型各维度上分布较为均衡。其中障碍维度的占比较高。当人们通过语音进行交流时，讨论的内容往往更为丰富，且观点深刻而客观。喜马拉雅平台常围绕社会事件和国家政策等新闻话题展开深入讨论，剖析背后的医学、科学、经济和科技知识。在探讨 HPV 疫苗相关新闻时，其讨论亦能涵盖健康信念模型的各个维度，从而使各维度分布更加均衡。

在视频平台抖音上，严重性、效益和障碍维度的占比相较于易感性更高，其中障碍维度尤为突出。抖音上的视频内容多为个人对某一话题的见解或生活感悟的记录。当涉及 HPV 疫苗这一关乎个人健康的议题时，用户更倾向于从个人角度出发，讨论与其相关的健康、疫苗短缺、接种的好处与坏处等话题。因此，在抖音平台上，严重性、效益和障碍维度的占比相对较高。

除了平台内容差异之外，差异的原因可能来源于它们所承载内容模态的固

有特性。文字消息短，传播迅速；音频内容长，谈话深入；视频内容较长，会表达观点。具体从文字层面来看，微博内容倾向于短小精悍，而喜马拉雅和抖音则呈现较长的内容长度。较长的内容长度更可能在一条文本上出现多个健康信念模型的维度，所以维度差异较小。

基于以上的两种原因，健康信念模型各维度在这三种平台上的分布存在显著的不均衡性。

5.1.2 不同平台在健康信念模型四个维度具有内容上的共性

相同平台在健康信念模型的四个维度上的内容表现出明显的共性，但在共性中又有子话题的差异。具体来说，文本平台微博在健康信念模型的四个维度上内容都是国家政策及衍生出来的子话题，音频平台喜马拉雅在健康信念模型的四个维度上内容是 HPV 疫苗的科学、科技和财经类话题，视频平台抖音在健康信念模型四个维度上内容是关于 HPV 疫苗的个体经历和个人观点。

在健康信念模型的易感性维度上，文本平台微博讨论了 HPV 的传播途径、疫苗接种的年龄限制和目标人群，音频平台喜马拉雅讨论了 HPV 疫苗的知识科普和中外对比，而视频平台抖音则关注了个人感染 HPV 的症状。

在健康信念模型的严重性维度上，文本平台微博讨论了宫颈癌在中国的严重程度、HPV 疫苗的接种率、疫苗的有效性等从国家的角度出发的宏观话题，音频平台喜马拉雅讨论了 HPV 疫苗的效果有限、进行中外疫苗对比等科学知识，而视频平台抖音则讨论了个人感染 HPV 所患的疾病和经历、在个人层面推广国产疫苗。

在健康信念模型的效益维度上，文本平台微博讨论了 HPV 疫苗的国家与国际战略、疫苗效果，音频平台喜马拉雅讨论了宫颈癌和 HPV 疫苗的科学知识，而视频平台抖音则呈现出明显的个人观点，鼓吹国产 HPV 疫苗。

在健康信念模型的障碍维度上，文本平台微博 HPV 疫苗短缺、疫苗的不良反应和国产疫苗费用较低等民生话题，音频平台喜马拉雅疫苗效果有限、哪种疫苗更受欢迎，疫苗公司的经济和产业机遇，更加专注疫苗本身及相关话题，而视频平台抖音则关注市场上疫苗短缺的个人叙述。

5.1.3 文本、视频平台的内容多与女性相关

在微博平台上，对于健康信念模型的易感性维度，讨论主要围绕女性视角展开，例如“女性感染了 HPV 应该如何处理？”以及“女性接种 HPV 疫苗的最佳时机是什么？”等问题。同样，在严重性维度的讨论中，目标群体也主要聚焦于女性，这些讨论强调了适龄女生接种率低下问题以及 HPV 感染可能导致的女性患癌风险。在障碍维度上，从女性个人角度出发，详细叙述了接种疫苗和抢疫

苗接种资格的经历。这些讨论使 HPV 疫苗和 HPV 病毒被贴上了性别的标签，被视为“女性疫苗”和“女性病毒”。

而在抖音平台上，关于健康信念模型的易感性维度的讨论则聚焦于女性在不知情的情况下，因其男友的高风险性行为而感染 HPV。此外，在严重性维度的讨论中，还涉及了女性感染后可能产生的心理变化，如“感染 HPV 后就不敢与男性接触”的叙述。这些讨论往往从女性受害者的角度出发，将男性视为主要的传播源，从而加剧了性别刻板印象。

值得注意的是，由于抖音平台推荐算法的影响，这些基于“女性视角”的 HPV 疫苗讨论被大量推荐给用户，形成了一种信息茧房。这在一定程度上影响了女性用户对 HPV 的客观认识，可能导致不科学的理解和恐慌情绪的产生，甚至可能加剧关于 HPV 疫苗的男女性别对立。

5.1.4 音频平台表现出“迟滞性”和“长期性”

音频平台喜马拉雅从而表现出“迟滞性”和“长期性”。

“迟滞性”指音频平台喜马拉雅关于 HPV 疫苗话题和事件的讨论要晚于文本平台微博。具体表现在九价 HPV 疫苗扩龄这一事件，文本平台微博讨论这一事件的时间是 2022 年 8 月，音频平台喜马拉雅讨论这一事件的时间是 2022 年 9 月到 10 月。音频平台喜马拉雅在文本平台微博讨论升起热度之后，然后才开始传播讨论。

探究“迟滞性”出现的原因，可以从以下方面来考虑。第一，从文本、音频和视频的制作成本和制作周期来说，文本制作成本很低，制作周期也很短，仅仅几分钟之类就可以完成。音频的制作成本更高，制作音频时需要音频发布者就 HPV 疫苗发表自己的观点和看法，搜寻相关资料并与其他人进行讨论，在录音之后还需要剪辑，配乐等后期制作，也需要专业的设备进行制作。这一过程需要持续几天甚至几个星期，音频制作周期长。第二，从文本和音频内容发布者发布内容的状态来分析。发布文本的时候仅仅需要思考几秒钟就可以用文字表达出观点。在用声音表达自己的观点的时候，讨论的内容更加深刻更长，需要思考输出内容，观点的思考和提炼需要在人脑接受信息和进行一定的知识储备和事件全貌之后进行输出。所以音频内容较文本内容更深刻也更具有思考性。

“长期性”指在音频平台喜马拉雅在事件讨论相较于文本平台的讨论周期更长。具体来说，具体表现在九价 HPV 疫苗扩龄这一事件，文本平台微博讨论了一个月，音频平台喜马拉雅讨论这一事件长达两个月。此外也在视频平台抖音上观测到了“长期性”的现象。

探究“长期性”出现的原因，可以从以下几个方面来考虑。第一，从音频、视频和文字传达内容的深刻程度角度来分析。音频平台喜马拉雅讨论的内容更

加深刻更长,需要思考输出内容,观点的思考和提炼需要在人脑接受信息和进行一定的知识储备和事件全貌之后进行输出,并且音频需要进行口述思考,对于人类来说口语表述比文本写作更加便捷。视频的制作成本更加高昂,虽然手机和网络的普及极大的降低了制作成本,但是当有画面出现的时候,当视频发布人出现在视频画面或者构思画面表达自己的想法,制作周期会更长,并且发布者会希望发布视频引起广泛的讨论。第二,从平台属性和用户特点的角度来看,不同平台之间的差异性也导致了“长期性”现象的出现。在视频平台如抖音上,当一个话题引起广泛关注时,往往会激发用户发布视频来表达自己的观点。这种表达方式通常围绕某个关键词或口号展开,用户创作内容会有对热点的“跟风效应”,而视频内容的制作成本相对较高,即使是随着手机和网络的普及降低了部分制作门槛,但视频的构思、拍摄和后期处理仍然需要一定时间。因此,当视频发布并引起讨论时,这种讨论往往会持续一段时间,形成了所谓的“长期性”。

“长期性”在内容传播上表现为“信息茧房”现象,即特定内容在一段时间内持续受到关注和讨论。这种现象在视频平台上尤为明显,因为视频内容的制作和发布往往需要较长的周期,而这期间用户对于相关话题的关注和讨论也会持续进行。这种长期性不仅反映了平台属性和用户特点对内容传播的影响,也揭示了信息在不同媒介中的传播规律和效应。

5.1.5 音频和视频平台内容更有“情绪”和更“积极”

音频平台喜马拉雅和视频平台抖音的内容比文本平台微博在表达上更具有“情绪”。三种平台的情绪中虽然负面情绪占据主流,但是音频平台喜马拉雅和视频平台抖音相较于文本平台更加“积极”。

音频平台喜马拉雅和视频平台抖音更具有“情绪”和更加“积极”的原因,可以从平台主要用户的角度分析。第一,在文本平台微博上,其信息发布主要由新闻机构和用户构成,因此新闻文本占据较大比重。这类内容往往注重信息传递,较少涉及情感表达,从而呈现出较少的情绪色彩。第二,在音频平台喜马拉雅上,其用户群体主要由专业主播和医生构成,他们普遍拥有良好的教育背景和丰富的知识面。这些用户在发布内容时,相较于文本平台,更倾向于以乐观、积极的态度进行阐述,从而营造出更加积极的氛围。第三,在视频平台抖音上,其用户群体广泛,主要是个体用户和医生等。在抖音上,个体用户往往表现出对国家和民族的自豪感,当面临如进口疫苗短缺等挑战时,他们倾向于从积极的角度解读,并传播正能量。

从三种平台模态的角度来分析,具有以下原因。第一,文本制作成本相对较低,使得消极情绪更容易通过文字快速传播。当接收到负面消息时,人们往往会出于恐慌而迅速扩散信息以消解情绪,这一过程极为迅速。例如,当 HPV 疫苗

的负面新闻和话题出现时,消极情绪和讨论会迅速蔓延。第二,语音交流相较于文本更能直接表达情感。尽管新闻主播在播报新闻时情绪表达较少,但在喜马拉雅这样的平台上,情感表达更为丰富和直接的主播占据了主导地位。第三,视频作为一种多媒体形式,更能全面、直观地展现个人用户的情绪。在抖音上,用户通过视频表达自己的情感和观点,使得情绪的传播更加生动和真实。

综上所述,喜马拉雅和抖音在展现“情绪”和“积极性”方面相较于微博更为突出,这既与平台用户群体的特点有关,也与平台模态的特性紧密相连。

5.1.6 音频平台的“国际视野”和视频平台的“支持国产”

在音频平台喜马拉雅上,“国际视野”指的是用户在讨论 HPV 疫苗时,不仅关注中国的情况,也会与欧美发达国家的情况进行对比,帮助受众更全面地了解中国 HPV 疫苗的现状与发展。

具体来说,这种“国际视野”在以下几个方面有所体现。第一,在健康信念模型的易感性维度上,有用户的讨论指出中国大陆目前使用的二价疫苗已逐渐在欧美国家淘汰,展现出在疫苗研发和选择方面的差距。第二,在严重性维度,有用户的讨论涉及中国大陆和其他国家在疫苗接种人群限制方面的不同。第三,在效益维度上,有用户的讨论对国产疫苗和进口疫苗的安全性进行了比较。最后,在障碍维度上,有用户的讨论分析了国产二价、四价疫苗与进口九价疫苗的价格差异。

探究音频平台出现“国际视野”的原因,可能包括以下几点。第一,喜马拉雅上的播客内容生产者中有许多专业主播和医生,部分在欧美接受过教育或有国际视野,他们在讨论 HPV 疫苗时自然会融入国际对比的内容。第二,音频平台的内容消费者往往具有较高的教育背景,且更加倾向于获取跨国信息,因而内容生产者会提供更多与国外疫苗情况相关的内容,满足这部分用户对多维信息的需求。

在视频平台抖音上,“支持国产”指用户在讨论中国生产的二价和四价 HPV 疫苗时,多数内容倾向于积极评价与推广,呈现出对“国货”的支持,同时反映出用户对国产疫苗的认同感和信任感。

这种“支持国产”在视频平台上的流行可以从以下几个角度来看。第一,中国政府通过新闻媒体和政策支持国产疫苗的推广,尤其是在九价疫苗供应紧张的情况下,推动国产疫苗的发展能够减少对进口疫苗的依赖。第二,抖音等视频平台的用户群体较为广泛,其中包括许多个人用户。许多用户成长在崇尚集体主义的社会氛围中,对“国货”天然抱有认同感。在分享自己接种国产疫苗的体验时,这些用户往往带有积极的情感表达,体现了他们对国家产业发展的支持与期待。这种社会氛围与集体意识在一定程度上强化了视频平台上的“支持国产”

内容。

5.2 关于内容差异形成的机理的探究

造成三种平台内容差异的机理有很多方面的原因，本小节从以下几个方面来分析。

5.2.1 模态的差异

从模态的角度来看，根据麦克卢汉“媒介即讯息”的观点，文本、音频和视频作为三种不同的传播介质，具有各自的特点，因此在 HPV 疫苗这个话题上展现出不同的讨论方式。

文本模态具有生产成本低、传播速度快的特性，因而在传播中往往偏向“吸引眼球”，更容易扩散迅速传播的内容。这种模态特性使得微博平台上的讨论更集中于国家政策和公共卫生事件，具体涉及疫苗接种计划、疫苗有效性、政策导向等层面的探讨。

音频模态因为内容制作和传播周期较长，受众在收听时会更关注内容的连贯性和表达的细致性。因此，喜马拉雅平台上的 HPV 疫苗讨论更多地涵盖科学、科技以及财经类话题，聚焦于疫苗的原理、技术背景和经济影响等方面，展现出一种系统性和逻辑性。

视频模态则因其具备视听结合的沉浸性，能够更强烈地影响受众的情绪和认知，因此抖音平台上更常见的内容是围绕个人经历和主观观点的分享，例如疫苗接种体验、个人感受和观点。这种模态使得视频内容的讨论更具情感冲击力，着重于个体化的表达。

5.2.2 内容发布者的差异

从社交媒体上内容发布者的差异角度来分析，在不同的平台上的内容发布者的占比是不同的。通过对文本平台微博的用户进行初步分析，用户组成为普通用户（72%）、名人（13%）、公司（3%）、媒体（2%）、媒体（2%）、学校（0.4%）。个人用户占比比较大，由于新闻政策发布的及时性，用户会对政策进行及时的讨论和反馈。通过对音频平台喜马拉雅的用户类别分析，用户组成为医学知识的医生或传播者（27%）、技术和金融相关（22%）、新闻媒体（11%）、未经验证的个人主播（35%）。用户占比较大的是专业的医生和健康博主，部分拥有国内外生活的经历，所以在内容上呈现出中外对比和客观知识。用户也有部分是金融科技相关的自媒体，所以在内容上出现了疫苗公司的讨论。通过对视频平台抖音的用户进行初步分析，以及未经验证的个人用户的抖音视频上传者（39%）、经过验

证的医生（22%）、政府机构（7%）、新闻媒体（24%）和其他验证用户（9%），用户占比较大的是个人用户和已经验证过的医生，这些医生大部分来自于中国城镇和县乡的医院，发布视频的目的在于科普跟 HPV 相关的疾病和 HPV 疫苗，会通过个人事例和个人观点在进行内容讨论。

5.2.3 内容消费者的差异

从社交媒体内容消费者差异的角度来看，不同模态的社交媒体平台吸引了各具特征的受众群体，而这些受众的需求和喜好反过来深刻地影响了内容的生产方向和风格。

在文本平台微博上，为迎合内容消费者的偏好，内容生产者往往会发布标题醒目、带有情绪张力的内容，以吸引快速消费的受众进行转发和传播。这种模式下，微博的内容更倾向于简洁、直击要点，尤其是围绕社会事件和热点话题。

音频平台喜马拉雅因制作成本较高且周期较长，其受众群体通常具备较高的知识和文化水平，更倾向于通过播客等形式获得专业知识和深入见解。这种内容需求促使生产者在内容策划上更注重深度和系统性，以满足听众的知识获取和思考需求。

在视频平台抖音上，由于制作门槛低，内容消费者不仅是观看者，许多人也能通过手机轻松制作和分享自己的观点和体验，成为内容生产者。抖音用户对直观的视觉体验和个人化内容表现出浓厚兴趣，这促使生产者偏向简短、生动且具有强烈个人风格的内容，以吸引关注并促进互动。

这种内容生产和消费的互动关系，体现了不同模态平台在满足各类受众需求时形成的特定内容风格和生产模式。

5.3 研究的意义

5.3.1 学术贡献

在学术研究方面，本文不仅在理论方面有贡献，在研究方法上也有一定的创新。首先，在理论方面，本文有以下贡献。第一，本文对比研究了文本平台微博、音频平台喜马拉雅、视频平台抖音上的健康信念模型。健康信念模型由 Rosenstock 在 1974 年提出^[82]，在与健康相关的背景下，人们的健康行为受他们避免疾病的愿望和相信健康行动将预防疾病的影响。在文本平台微博上，有研究人员研究了报纸上关于 HPV 疫苗讨论，发现更加关注健康信念模型的效益维度^[89]。还有研究人员研究了在线问答平台知乎上关于 HPV 疫苗的报道，他们发现包括易感性、严重性和益处。此外，易感性的存在与严重性的存在呈正相关^[56]。Du 等人还研究发现了 Twitter 上用户随着时间的推移更加倾向讨论 HPV

的风险^[90]。在图像平台上,有学者研究 Pinterest 上关于 HPV 疫苗审查不实信息前后前后讨论 HPV 疫苗的变化,发现效益维度增加,障碍维度减少^[91]。在视频平台上,有学者研究了健康信念模型维度的变化,发现 37.8% 提及了 HPV 的易感性,53.4% 提到了 HPV 的严重性,41.9% 提到了 HPV 疫苗的益处,而 53.5% 提到了 HPV 疫苗的障碍。先前的研究只是在某个单一的平台应用了健康信念模型理论,并且只应用在了文本平台和视频平台,本文不仅分析了来自文本平台和视频平台的数据,还分析了来自音频平台的数据。这个特别研究的播客平台是喜马拉雅,该平台在中国的 HPV 疫苗背景下从未进行过研究。第二,通过将人声讨论从音频和视频模态提取,通过 ASR 技术转写为文本。借助这些文本信息,可以系统地在健康信念模型的各个维度上进行对比文本平台微博、音频平台喜马拉雅、视频平台抖音上内容的差异,这是健康信念模型的理论发展上的创新。第三,在进行健康信念模型分类的时候,健康信念的各个维度可以出现在一条内容上面,并不是把健康信念的各个维度单独的分开,所以健康信念模型的各个维度应用被开创了新的应用方式。

其次,在研究方法上,本文具有以下贡献。第一,本文对比了文本、音频、视频三种社交媒体平台上关于 HPV 话题内容的差异。先前关于 HPV 疫苗在社交媒体上的内容研究主要局限于文本平台,而本研究则扩展到了音频和视频平台。鉴于三种模态在内容层面上难以直接比较,本研究将三种模态中关于 HPV 话题的内容统一转换为文字层面,即短文本和长文本,以便进行内容分析和建模。这种跨模态的研究方法有助于更全面地理解 HPV 疫苗在社交媒体上的讨论态势,并揭示不同模态内容之间的差异。

第二,本文利用深度学习的方法进行了大数据分析。从文本、音频、视频等不同平台收集多模态数据,并将其转换为文字层面上,通过深度学习技术分析了所有采集的内容数据。此前关于 HPV 疫苗在社交媒体上的研究多采用抽样,人工选出几百条样本进行分析,而本研究应用了深度学习技术,分析了全部数据。

第三,本文在健康信念模型的维度上开创性地对比了短文本和长文本。在文本平台微博和推特上,多为短文本,一条文本的长度在 200 个字以内,而在音频和视频平台中,人声内容转换为文字后形成了长文本,长度可达几千至上万字。以往在健康信念模型维度分类的研究中,长文本和短文本因数量级和文本长度差异难以直接比较,而本研究通过让单条内容包含多个健康信念模型维度,进行了多次二分类任务,实现了长文本和短文本在健康信念模型各维度上的对比分析。

第四,本文在长文本和短文本的内容对比上进行了方法上的创新。由于在主题建模时,长文本和短文本适用的主题建模方法不同,本文采用了基于深度学习的语言模型 BERT 的主题建模技术 BERTopic。在对比三种模态内容在健康信念

模型各个维度上的差异时,为更好地探究长文本的潜在主题,本文将长文本拆分成短文本,并基于这些短文本进行主题建模,从而在主题建模内容对比层面进行了创新。

5.3.2 实践启示

本文对公共卫生领域的实践有以下几点启示。

第一,本文作为对 HPV 疫苗议题在文本、音频、视频三种平台上内容差异的量化研究,发现了关于 HPV 疫苗议题在三个不同平台上的健康信息话语有各自的特点。在文本平台微博上,讨论更注重国家政策,并且对新闻事件作出迅速回应。在音频平台喜马拉雅上,健康专家制作的播客内容较多,并且对事件的回应较微博比较晚。在视频平台抖音上,发现更多女性分享了她们接种 HPV 疫苗的个人经历。基于这些发现,健康机构、国家疫苗政策制定机构可以根据不同模态的特性制定适当的健康信息宣传策略。在文本平台上,应着重打造简洁明了、易于理解的文字内容,并辅以引人入胜的标题和简短的描述,以迅速捕获用户的注意力。此类平台特别适合快速传播热点事件和敏感话题,因此,应密切关注新闻动态,结合健康信念模型的多个维度,设计相关内容,以激发用户的讨论和分享欲望。在音频平台上,可以邀请权威的健康专家进行访谈,就相关健康知识进行深度解析。这种方式可以为听众提供权威、专业的信息,增加他们对健康信息的了解和信任。在视频平台上,可以鼓励个人分享自己的经历和感受。这种真实、直观的方式可以大大增加信息的可信度和感染力,吸引更多人的关注。

第二,在社交媒体的公众健康领域内,音频平台喜马拉雅关于健康信息的科普更加客观且不带有个人主观的意见。在文本平台微博、视频平台抖音上呈现出“女性”的视角,进而在平台上出现一些关于 HPV 疫苗比较极端的认识。可以根据文本平台微博、视频平台抖音上内容的特点,对关于 HPV 疫苗讨论的内容进行更加客观的引导,从而提高全民的公众健康意识,而不仅仅是局限于“女性视角”。

第三,本文同样揭示了一些关于 HPV 疫苗的不实信息和假信息的传播现象。这些不实信息和假信息具有更高的传播效率,然而,由于社交媒体平台缺乏对 HPV 疫苗专业信息的有效审核机制,导致公众在传播和认知 HPV 疫苗相关信息时容易受到误导。因此,为了应对这一问题,本文建议在社交媒体平台上建立公众健康信息的审核机制。具体而言,这一机制应能在文本、音频、视频等多种信息形式中,从源头上遏制不实信息和假消息的扩散。通过这样的措施,我们有望在治理公共健康领域的信息传播方面取得显著成效,并进一步提升全民的基本健康素养。

第四,可以通过本文的研究结果了解中国社交媒体上 HPV 疫苗的讨论现状。

现阶段关于 HPV 疫苗健康信念模型的障碍维度占据主导地位,根据研究结果应建议中国政府应该把 HPV 疫苗的接种年龄,接种人群扩展,接种地区扩展到全国的各个省市自治区,不再限制接种的性别,预防 HPV 引起的男性女性的各种疾病,而不仅仅是预防女性有关的宫颈癌。

第五,本文观测到了在音频平台和视频平台上面的迟滞性和长期性,呈现出“信息茧房”的现象。之前有学者的研究^[119]指出在视频平台上,视频内容会在影响社交媒体上的用户形成信息茧房方面起着重要的作用,而个性化的推荐算法不一定会造成严重信息茧房。所以在视频平台上宣传健康议题的时候,需要根据宣扬的目的,采取合理的措施。例如,当需要宣扬科学和正向的政策的时候,可以在视频讨论的内容上制作更加吸引人关注的内容,从而利用“信息茧房”效应,让在视频平台上的用户更多次的触达健康信息,从而提高公众健康素养。当要对视频平台的信息进行平台治理的时候,信息的检查不应该局限于标题和描述,应该更加注重视频内容的讨论,防止有害的讨论通过“信息茧房”传播触达更多的平台用户。

5.4 研究讨论

5.4.1 研究局限性

本研究在以下几个方面存在一定的局限性。

第一,在数据的获取层面,在文本、音频、视频三种平台中,每种类型的平台只选取了最受欢迎的一个平台,而不是选取了多个平台进行对照实验。研究目标是对比不同模态作为传播媒介的平台对于健康话题内容传播的差异,所以并没有考虑同一个模态是否存在平台属性的差异,而只是从模态差异的角度出发去探究基于健康信念模型的四个维度内容传播的差异和情绪层面的差异。在音频、视频平台的层面,由于平台对于数据获取的限制,只获取到一部分关于 HPV 疫苗的数据,所以只能从用户使用社交媒体的视角出发,站在用户使用社交媒体所看到的内容的角度来分析不同模态的平台关于 HPV 疫苗的讨论。

第二,在数据分析的层面,从内容的角度进行文本、音频、视频平台上讨论健康信息的对比,由于目前的技术在不同的模态无法在内容角度上进行对比,所以把文本、音频、视频平台的内容转换到了文字层面,从文字层面研究不同模态讨论内容的差异,所以本文分析了三种模态中文本的信息,并没有探究模态的差异。探究模态之间的差异没有固定的分析框架,目前没有在现有的研究和文献中找到合适的方法进行比较,传播学界的多模态传播也是目前学界存在的难点问题。本文为之后进行模态差异的比较奠定了基础。

第三,在研究方法中,针对长文本和短文本的研究方法针对健康信念的维度

分析和情绪分析尝试了很多种方法。如果对长文本进行了拆分,拆分成短文本之后进行统计分析,但是本文认为拆分成短文本之后,长文本的句子语义结构遭到了破坏,在音频社交媒体播客上关于 HPV 疫苗的讨论需要联系前后语境才会判断出情绪和维度,如果按照单个句号进行拆分可能会改变原有的语义结构,进而结果完全不同。如果对长文本进行拆分,那么因为长文本比短文本在健康信念模型的四个维度层面和情绪层面出现的出现的概率比短文本的概率高,存在着局限性。由于考虑在社交媒体上面关于 HPV 疫苗的讨论的一条内容作为一个整体,无论是微博上 HPV 疫苗的讨论,喜马拉雅 HPV 疫苗的讨论,抖音上 HPV 疫苗的讨论,都是一个用户发布关于 HPV 疫苗的讨论,是一个整体,所以从每个用户发布的每条内容作为整体来出发来开展研究,但是在这个角度上有局限性。

第四,对于不同平台上观察到的内容和情感的差异,根据本文的结论,文本平台微博在健康信念模型的四个维度上讨论内容都是国家政策及衍生出来的子话题,音频平台喜马拉雅在健康信念模型的四个维度上讨论的 HPV 疫苗的科学、科技和财经类话题,视频平台抖音在四个维度上讨论的是关于 HPV 疫苗接种的是个体经历和个人观点。但是这些差异是什么造成的,本研究仅仅是从三种模态的平台看到了这些差异,没有探究差异是如何形成的,有哪些原因。传播模态的差异会造成现在的现象,但是在文本、音频、视频平台存在着不同内容生产者,以及内容消费者。从内容生产者的角度来说,文本传播模态社交媒体平台微博存在着扮演者各种各样社会角色的内容生产者,例如,政府机构、新闻媒体、疫苗推广机构等。而在音频平台喜马拉雅存在着内容生产者的社会角色是国内外科学家、医院的医生。在视频平台抖音存在的内容生产者的角色是中国各省市医院的医生等。从内容消费者的层面,文本平台微博、音频平台喜马拉雅、视频平台抖音视频消费者的年龄、受教育程度、知识背景、性别的差异等的差异。所以造成内容差异背后的原因可能是多种多样的。本研究只是在观测者的角度,发现了这些差异的存在,从文本、音频、视频三种模态的角度解释造成这些差异的原因,所以具有局限性,未来可以进一步深刻分析,不仅仅是初步分析。

第五,目前的研究只针对了“HPV 疫苗”这个单一的主题,从公共卫生的大背景来看,目前的研究方法和理论创新依然适用,不仅仅局限” HPV 疫苗“这个主题,还有其他关于公共卫生领域的话题,如果针对不同平台的特性,可以指定对应的策略。然而,也不仅仅局限于公共卫生领域,也可以扩展到其他领域,比如政治等。

5.4.2 未来展望

针对研究局限,未来研究可以从以下几个方面进行深入探索。

第一,在文本、音频和视频三类不同模态的社交媒体平台选取的层面,文本

只选取了三类社交媒体平台上最受欢迎的平台，文本平台选取了微博，音频平台选取了喜马拉雅，视频平台选取了抖音。而目前的内容和情绪的差异到底是模态的差异还是平台的差异，这是一个复杂的问题，未来可以在文本平台、音频平台、视频平台选取两个或者两个以上的社交媒体平台进行数据分析，比如可以再次选取文本平台知乎、音频平台蜻蜓 FM、视频平台哔哩哔哩与现有的结果进行对比分析，从而探究内容差异和情绪差异的深层次的原因。可以将研究方法应用于更广泛的社交媒体平台，并使用健康信念模型的四个维度，从而排除单个社交媒体平台的属性，探究社交媒体平台模态层面的差异。

第二，未来的研究可以进一步探索文本、音频、视频模态之间的差异。当前的研究主要聚焦于文本、音频、视频三类不同社交媒体平台的内容差异。然而，由于模态差异研究的局限性，目前尚未找到恰当的数据处理方法，以在内容层面上将文本、声音特征、视频特征逐一对应起来。因此，本研究不得不采取一种折衷的方法，即将音频和视频模态中讨论 HPV 疫苗的内容转换到文本层面，进而在文本层面进行对照比较分析。尽管如此，如果把文本、音频、视频都通过向量化的方式导入预训练的多模态大模型中，虽然这一过程可能缺乏足够的可解释性，但这也为未来的研究提供了一个值得探索的方向。

第三，在探讨社交媒体差异形成的根源时，一个值得深入研究的领域是社交媒体用户是否构成了这些差异的关键因素。鉴于前文所述的局限性部分归因于不同平台模态所导致的差异，我们同样可以考察社交媒体用户本身的差异性是否也是造成内容差异的一个重要原因。本文目前仅对用户组成进行了初步分析，但未来研究可以进一步拓展，深入探讨三种不同社交媒体平台上内容生产者用户属性的差异。由于不同的社交媒体平台往往吸引着具有不同属性的用户群体，因此，下一步的研究可以通过提取这些平台的用户属性，并在经过严谨的人工编码后，于平台用户属性类别一致的基础上，细致探究用户之间的差异性。

第四，未来的研究可以超越“HPV 疫苗”的主题，涵盖多个主题，可以应用到如健康、政治等领域，可以推广应用到社交媒体平台上的任何领域的话题，可以推进社交媒体平台的治理和公众政策宣传。

参考文献

- [1] KIM J N. Digital networked information society and public health: problems and promises of networked health communication of lay publics[J/OL]. Health Communication, 2018, 33(1): 1-4[2024-01-03]. <https://doi.org/10.1080/10410236.2016.1242039>.
- [2] AHMED R, MAO Y. Communication research on health disparities and coping strategies in COVID-19 related crises[J/OL]. Health Communication, 2022, 37(12): 1455-1456[2024-01-05]. <https://doi.org/10.1080/10410236.2022.2111633>.
- [3] STRØMME H, MOSDØL A, NORDHEIM L W, et al. NIPH systematic reviews: executive summaries: effects of using social media in health awareness campaigns to promote healthy lifestyle habits[M/OL]. Oslo, Norway: Knowledge Centre for the Health Services at The Norwegian Institute of Public Health (NIPH), 2014[2023-10-29]. <http://www.ncbi.nlm.nih.gov/books/NBK464906/>.
- [4] ALLINGTON D, DUFFY B, WESSELY S, et al. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency[J/OL]. Psychological Medicine, 2021, 51(10): 1763-1769[2024-01-05]. <https://doi.org/10.1017/S003329172000224X>.
- [5] 王宁刘硕. 2018 全球癌症统计报告解读[J/OL]. 肿瘤综合治疗电子杂志, 2019, 5(1): 87-97[2024-10-14]. <https://doi.org/10.12151/JMCM.2019.01-10>.
- [6] ICO/IARC HPV Information Centre. China: human papillomavirus and related cancers, 2023 [R/OL]. 2023. https://hpvcentre.net/statistics/reports/CHN_FS.pdf.
- [7] FRAZER I H. Prevention of cervical cancer through papillomavirus vaccination[J/OL]. Nature Reviews Immunology, 2004, 4(1): 46-55[2024-10-14]. <https://doi.org/10.1038/nri1260>.
- [8] ZHAO X L, HU S Y, HU J W, et al. Tackling barriers to scale up human papillomavirus vaccination in China: progress and the way forward[J/OL]. Infectious Diseases of Poverty, 2023, 12(1): 86[2024-01-05]. <https://doi.org/10.1186/s40249-023-01136-6>.
- [9] ZHAO Y. Nation plans to launch free HPV vaccinations[J/OL]. China Daily, 2022[2023-10-29]. <https://www.chinadaily.com.cn/a/202201/14/WS61e0aac6a310cdd39bc80e9a.html>.
- [10] FOWLER E F, GOLLUST S E, DEMPSEY A F, et al. Issue emergence, evolution of controversy, and implications for competitive framing: the case of the HPV vaccine[J/OL]. The International Journal of Press/Politics, 2012, 17(2): 169-189[2024-01-22]. <https://doi.org/10.1177/1940161211425687>.
- [11] LEE T K, SU L Y F. When a personal HPV story on a blog influences perceived social norms: the roles of personal experience, framing, perceived similarity, and social me-

- dia metrics[J/OL]. Health Communication, 2020, 35(4): 438-446[2024-01-22]. <https://doi.org/10.1080/10410236.2019.1567440>.
- [12] JIANG S, NG A Y K, NGIEN A. The effects of social media information discussion, perceived information overload and patient empowerment in influencing HPV knowledge[J/OL]. Journal of Health Communication, 2022, 27(6): 407-415[2024-01-24]. <https://doi.org/10.1080/10810730.2022.2115591>.
- [13] CHEN L, LING Q, CAO T, et al. Mislabeled, fragmented, and conspiracy-driven: a content analysis of the social media discourse about the HPV vaccine in China[J/OL]. Asian Journal of Communication, 2020, 30(6): 450-469[2023-08-18]. <https://doi.org/10.1080/01292986.2020.1817113>.
- [14] HIMELBOIM I, XIAO X, LEE D K L, et al. A social networks approach to understanding vaccine conversations on twitter: network clusters, sentiment, and certainty in HPV social networks[J/OL]. Health Communication, 2020, 35(5): 607-615[2024-01-25]. <https://doi.org/10.1080/10410236.2019.1573446>.
- [15] JIANG S, WANG P, LIU P L, et al. Social media communication about HPV vaccine in China: a study using topic modeling and survey[J/OL]. Health Communication, 2023, 38(5): 935-946[2023-10-24]. <https://doi.org/10.1080/10410236.2021.1983338>.
- [16] AGERGAARD T E, SMITH M I E, NIELSEN K H. Vaccine assemblages on three HPV vaccine-critical Facebook pages in Denmark from 2012 to 2019[J/OL]. Media and Communication, 2020, 8(2): 339-352[2024-01-22]. <https://doi.org/10.17645/mac.v8i2.2858>.
- [17] LI Y, GUO Y, LIN H. Feminized virus: a content analysis of social media's representation of HPV vaccine[J/OL]. Social Media + Society, 2022, 8(3): 20563051221104232[2023-07-16]. <https://doi.org/10.1177/20563051221104232>.
- [18] CAO W, CHEN H D, YU Y W, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020[J/OL]. Chinese Medical Journal, 2021, 134(7): 783[2024-01-05]. <https://doi.org/10.1097/CM9.0000000000001474>.
- [19] BRIANTI P, DE FLAMMINEIS E, MERCURI S R. Review of HPV-related diseases and cancers[J]. The New Microbiologica, 2017, 40(2): 80-85.
- [20] DUNNE E F, UNGER E R, STERNBERG M, et al. Prevalence of HPV infection among females in the United States[J/OL]. JAMA, 2007, 297(8): 813-819. <https://doi.org/10.1001/jama.297.8.813>.
- [21] CDC. Human papillomavirus (HPV) infection[J/OL]. Centers for Disease Control and Prevention, 2024[2024-05-31]. <https://www.cdc.gov/hpv/parents/about-hpv.html>.
- [22] Cervical cancer[EB/OL]. [2024-05-31]. <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>.

- [23] ORGANIZATION W H. Global strategy to accelerate the elimination of cervical cancer as a public health problem.[EB/OL]. 2020[2024-01-06]. <https://www.who.int/initiatives/cervical-cancer-elimination-initiative>.
- [24] ORGANIZATION W H. HPV vaccine included in national immunization programme [EB/OL]. 2023[2024-01-06]. <https://app.powerbi.com/view?r=eyJrIjojNDIxZTFkZGUtMDQ1Ny00MDZkLThiZDktYWFIYTdkOGU2NDcwIiwidCI6ImY2MTBjMGI3LWJkMjQ0NGIzOS04MTBiLTNkYzI4MGFmYjU5MCIslmMiOjh9>.
- [25] ORGANIZATION W H. Human papillomavirus (HPV) vaccination coverage[EB/OL]. 2023 [2024-01-06]. https://immunizationdata.who.int/pages/coverage/hpv.html?CODE=Global&ANTIGEN=15HPV1_F&YEAR=.
- [26] FERLAY J L F C M M L P M E A, Ervik M. Global cancer observatory: cancer today [EB/OL]. [2024-01-08]. <http://gco.iarc.fr/today>.
- [27] 药物临床试验登记与信息公示平台[EB/OL]. [2024-06-01]. <http://www.chinadrugtrials.org.cn/index.html>.
- [28] 普查数据 - 国家统计局[EB/OL]. [2024-06-01]. <https://www.stats.gov.cn/sj/pcsj/>.
- [29] COLOMBARA D V, WANG S M. The impact of HPV vaccination delays in China: lessons from HBV control programs[J/OL]. Vaccine, 2013, 31(38): 4057-4059[2024-06-01]. <https://doi.org/10.1016/j.vaccine.2013.06.031>.
- [30] GAO M, HU S, ZHAO X, et al. Health and economic impact of delaying large-scale HPV vaccination and screening implementation on cervical cancer in China: a modelling study [J/OL]. The Lancet Regional Health - Western Pacific, 2023, 36: 100768[2024-06-01]. <https://doi.org/10.1016/j.lanwpc.2023.100768>.
- [31] Nation plans to launch free HPV vaccinations - Chinadaily.com.cn[EB/OL]. [2024-01-05]. <https://www.chinadaily.com.cn/a/202201/14/WS61e0aac6a310cdd39bc80e9a.html>.
- [32] New government plan to speed up elimination of cervical cancer[EB/OL]. [2024-06-01]. https://english.www.gov.cn/policies/policywatch/202302/08/content_WS63e30acdc6d0a757729e681d.html.
- [33] OF THE PEOPLE'S REPUBLIC OF CHINA T S C. Circular of the state council on printing and issuing the program for the development of chinese women and the program for the development of chinese children[EB/OL]. [2024-06-01]. https://www.gov.cn/zhengce/content/2021-09/27/content_5639412.htm.
- [34] XIAO X, SU Y. Integrating reasoned action approach and message sidedness in the era of misinformation: the case of HPV vaccination promotion[J/OL]. Journal of Health Communication, 2021, 26(6): 371-380[2024-01-25]. <https://doi.org/10.1080/10810730.2021.1950873>.
- [35] KIM J, LEE J, HEO J, et al. Message strategies and viewer responses: content analysis of

- HPV vaccination videos on YouTube[J/OL]. *Journal of Health Communication*, 2021, 26(12): 818-827[2023-10-29]. <https://doi.org/10.1080/10810730.2021.2015644>.
- [36] KIM S C, VRAGA E K, COOK J. An eye tracking approach to understanding misinformation and correction strategies on social media: the mediating role of attention and credibility to reduce HPV vaccine misperceptions[J/OL]. *Health Communication*, 2021, 36(13): 1687-1696[2024-01-25]. <https://doi.org/10.1080/10410236.2020.1787933>.
- [37] COMAN I A, BAZAA U, GUIDRY J P D, et al. #EndtheSTigma: an exploratory analysis of the 2019 HPV awareness day conversation on Twitter & Instagram[J/OL]. *Health Communication*, 2023, 0(0): 1-10[2024-01-25]. <https://doi.org/10.1080/10410236.2023.2196466>.
- [38] 向菲; 冯思佳. 社会公众对 HPV 疫苗的关注热点分析-基于新浪微博数据[J]. *中国公共卫生*, 2021(04 vo 37): 738-740.
- [39] 郭小安; 王天翊. 新媒体接触、健康信念与 HPV 疫苗接种意向[J]. *新闻与传播研究*, 2020(06 vo 27): 58-74+127.
- [40] 姚志臻; 张斌; 司湘云. 基于网络问答社区的疫苗情绪语义网络分析[J/OL]. *信息资源管理学报*, 2021(01 vo 11): 70-79. DOI: 10.13365/j.jirm.2021.01.070.
- [41] 孙鹿童. HPV 疫苗如何走入大众视野?——基于微博相关视频的内容分析[J]. *医学与哲学*, 2022(09 vo 43): 18-21+26.
- [42] 商建辉; 赵昱. HPV 疫苗报道的框架效应探究——以微博情绪表达为例[J/OL]. *青年记者*, 2022(22 vo): 50-52. DOI: 10.15997/j.cnki.qnjz.2022.22.029.
- [43] MARÍN-GONZÁLEZ E, MALMUSI D, CAMPRUBÍ L, et al. The role of dissemination as a fundamental part of a research project: lessons learned from sophie[J/OL]. *International Journal of Health Services*, 2017, 47(2): 258-276[2024-02-03]. <https://doi.org/10.1177/0020731416676227>.
- [44] POTTER R F. The effects of voice changes on orienting and immediate cognitive overload in radio listeners[J/OL]. *Media Psychology*, 2000, 2(2): 147-177[2024-06-02]. https://doi.org/10.1207/S1532785XMEP0202_3.
- [45] POTTER R F, LANG A, BOLLS P D. Identifying structural features of audio[J/OL]. *Journal of Media Psychology*, 2008, 20(4): 168-177[2024-06-02]. <https://doi.org/10.1027/1864-1105.20.4.168>.
- [46] LANG A, GAO Y, POTTER R F, et al. Conceptualizing audio message complexity as available processing resources[J/OL]. *Communication Research*, 2015, 42(6): 759-778. <https://doi.org/10.1177/0093650213490722>.
- [47] SUNDAR S S. The main model: a heuristic approach to understanding technology effects on credibility[J/OL]. *Digital Media*. <https://doi.org/10.1162/dmal.9780262562324.073>.
- [48] HUANG E. Six cases of e-health videos on hospital web sites[J/OL]. *e-Service Journal*,

- 2009, 6: 56-71. <https://doi.org/10.2979/esj.2009.6.3.56>.
- [49] FERGUSON L A. Implementing a video education program to improve health literacy[J/OL]. *The Journal for Nurse Practitioners*, 2012, 8(8): e17-e22[2024-02-03]. <https://doi.org/10.1016/j.nurpra.2012.07.025>.
- [50] ARMSTRONG A W, IDRIS N Z, KIM R H. Effects of video-based, online education on behavioral and knowledge outcomes in sunscreen use: a randomized controlled trial[J/OL]. *Patient Education and Counseling*, 2011, 83(2): 273-277[2024-02-03]. <https://doi.org/10.1016/j.pec.2010.04.033>.
- [51] BRACKEN C C, DALESSANDRO A. Modality and medium characteristics[M/OL]//*The International Encyclopedia of Media Effects*. John Wiley & Sons, Ltd, 2017: 1-12[2024-01-19]. <https://doi.org/10.1002/9781118783764.wbieme0205>.
- [52] StatCounter. Desktop, mobile & tablet social media stats United States of America[J/OL]. StatCounter Global Stats, 2023[2023-10-29]. <https://gs.statcounter.com/social-media-stats/desktop-mobile-tablet/united-states-of-america>.
- [53] DOCKTER C E, LEE S, BOMAN C D, et al. The impact of retransmission and modality on communicating health research findings via social media[J/OL]. *Health communication*, 2021, 36(10): 1231-1241[2024-02-03]. <https://doi.org/10.1080/10410236.2020.1749354>.
- [54] DUNN A G, SURIAN D, LEASK J, et al. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States[J/OL]. *Vaccine*, 2017, 35(23): 3033-3040[2023-07-03]. <https://doi.org/10.1016/j.vaccine.2017.04.060>.
- [55] LOFT L, PEDERSEN E, JACOBSEN S, et al. Using Facebook to increase coverage of HPV vaccination among Danish girls: an assessment of a Danish social media campaign[J/OL]. *Vaccine*, 2020, 38(31): 4901-4908. <https://doi.org/10.1016/j.vaccine.2020.04.032>.
- [56] LI J, ZHENG H. Coverage of HPV-related information on Chinese social media: a content analysis of articles in Zhihu[J/OL]. *Human Vaccines & Immunotherapeutics*, 2020, 16(10): 2548-2554[2023-07-16]. <https://doi.org/10.1080/21645515.2020.1729028>.
- [57] ZHANG C, GOTSIS M, JORDAN-MARSH M. Social media microblogs as an HPV vaccination forum[J/OL]. *Human Vaccines & Immunotherapeutics*, 2013, 9(11): 2483-2489[2024-06-08]. <https://doi.org/10.4161/hv.25599>.
- [58] LUISI M. From bad to worse II: risk amplification of the HPV vaccine on Facebook[J/OL]. *vaccine*, 2021, 39(2): 303-308. <https://doi.org/10.1016/j.vaccine.2020.11.065>.
- [59] ZHANG J, XUE H, CALABRESE C, et al. Understanding human papillomavirus vaccine promotions and hesitancy in Northern California through examining public Facebook pages and groups[J/OL]. *Frontiers in Digital Health*, 2021, 3[2023-10-29]. <https://doi.org/10.3389/fdgth.2021.683090>.

- [60] KEARNEY M, SELVAN P, HAUER M, et al. Characterizing HPV vaccine sentiments and content on Instagram[J/OL]. *Health Education & Behavior*, 2019, 46(2_SUPPL): 37-48. <https://doi.org/10.1177/1090198119859412>.
- [61] SULLIVAN J L. The platforms of podcasting: past and present[J/OL]. *Social media+ society*, 2019, 5(4): 2056305119880002. <https://doi.org/10.1177/2056305119880002>.
- [62] BERRY R. Podcasting: considering the evolution of the medium and its association with the word ‘radio’ [J/OL]. *Radio Journal: International Studies in Broadcast & Audio Media*, 2016, 14(1): 7-22[2023-10-13]. https://doi.org/10.1386/rjao.14.1.7_1.
- [63] SPINELLI M, DANN L. Podcasting: the audio media revolution[M]. Bloomsbury Publishing USA, 2019.
- [64] MAÑAS-PELLEJERO M, PAZ E. Podcast fandom in Europe: audio audiences’ participation in a digital context[J/OL]. *Profesional de la información*, 2022, 31(5)[2023-10-12]. <https://doi.org/10.3145/epi.2022.sep.25>.
- [65] PERKS L G, TURNER J S. Podcasts and productivity: a qualitative uses and gratifications study[J/OL]. *Mass Communication and Society*, 2019, 22(1): 96-116[2023-10-12]. <https://doi.org/10.1080/15205436.2018.1490434>.
- [66] DMYTRYSHYN J, SELK A. Learning on the go: assessing knowledge gained from medical podcasts created for vulvovaginal disease education[J/OL]. *Journal of Lower Genital Tract Disease*, 2022, 26(2): 164-168[2023-10-12]. <https://doi.org/10.1097/LGT.0000000000000654>.
- [67] DUGAN M, MCALLISTER K D, BRENNAN E K W. Persuasion via performance: toward a handbook for youth to develop content that promotes HPV vaccination[J/OL]. *Critical Reviews in Eukaryotic Gene Expression*, 2022, 32(7)[2023-10-12]. <https://doi.org/10.1615/CritRevEukaryotGeneExpr.2022041298>.
- [68] LINDGREN M. Intimacy and emotions in podcast journalism: a study of award-winning australian and british podcasts[J/OL]. *Journalism Practice*, 2023, 17(4): 704-719[2023-10-30]. <https://doi.org/10.1080/17512786.2021.1943497>.
- [69] KIM Y, KIM Y, WANG Y. Selective exposure to podcast and political participation: the mediating role of emotions[J/OL]. *International Journal of Mobile Communications*, 2016, 14(2): 133-148[2023-10-30]. <https://doi.org/10.1504/IJMC.2016.075021>.
- [70] KAYE D B V, CHEN X, ZENG J. The co-evolution of two chinese mobile short video apps: parallel platformization of Douyin and TikTok[J/OL]. *Mobile Media & Communication*, 2021, 9(2): 229-253[2023-10-12]. <https://doi.org/10.1177/2050157920952120>.
- [71] BOATMAN D D, EASON S, CONN M E, et al. Human papillomavirus vaccine messaging on TikTok: social media content analysis[J/OL]. *Health Promotion Practice*, 2022, 23(3):

- 382-387[2023-10-12]. <https://doi.org/10.1177/15248399211013002>.
- [72] KIRKPATRICK C E, LAWRIE L L. Can videos on TikTok improve pap smear attitudes and intentions? effects of source and autonomy support in short-form health videos[J/OL]. *Health Communication*, 2023, 0(0): 1-13[2023-10-12]. <https://doi.org/10.1080/10410236.2023.2254962>.
- [73] BASCH C H, DONELLE L, FERA J, et al. Deconstructing TikTok videos on mental health: cross-sectional, descriptive content analysis[J/OL]. *JMIR Formative Research*, 2022, 6(5): e38340[2023-10-12]. <https://doi.org/10.2196/38340>.
- [74] HOPFER S, PHILLIPS K K D, WEINZIERL M, et al. Adaptation and dissemination of a national cancer institute HPV vaccine evidence-based cancer control program to the social media messaging environment[J/OL]. *Frontiers in Digital Health*, 2022, 4[2023-10-12]. <https://doi.org/10.3389/fdgth.2022.819228>.
- [75] LI Y, GUAN M, HAMMOND P, et al. Communicating COVID-19 information on TikTok: a content analysis of TikTok videos from official accounts featured in the COVID-19 information hub[J/OL]. *Health Education Research*, 2021, 36(3): 261-271[2023-10-29]. <https://doi.org/10.1093/her/cyab010>.
- [76] WANG X, HE R. Supporting vaccination on TikTok during the COVID-19 pandemic: vaccine beliefs, emotions, and comments[J/OL]. *Frontiers in Psychology*, 2022, 13[2023-10-29]. <https://doi.org/10.3389/fpsyg.2022.938377>.
- [77] LEWIS M, GRANTHAM S. From fork hands to microchips: an analysis of trending #CovidVaccine content on TikTok[J/OL]. *International Journal of Communication*, 2022, 16(0): 24[2023-10-29]. <https://ijoc.org/index.php/ijoc/article/view/18972>.
- [78] BRIONES R, NAN X, MADDEN K, et al. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube[J/OL]. *Health Communication*, 2012, 27(5): 478-485[2023-10-29]. <https://doi.org/10.1080/10410236.2011.610258>.
- [79] KAPLAN A M, HAENLEIN M. Users of the world, unite! the challenges and opportunities of social media[J/OL]. *Business Horizons*, 2010, 53(1): 59-68[2023-10-28]. <https://doi.org/10.1016/j.bushor.2009.09.003>.
- [80] THORSON K, DRISCOLL K, EKDALE B, et al. Youtube, Twitter and the occupy movement: connecting content and circulation practices[J/OL]. *Information, Communication & Society*, 2013, 16(3): 421-451[2023-10-28]. <https://doi.org/10.1080/1369118X.2012.756051>.
- [81] SMITH A N, FISCHER E, YONGJIAN C. How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?[J/OL]. *Journal of Interactive Marketing*, 2012, 26(2): 102-113[2023-10-28]. <https://doi.org/10.1016/j.intmar.2012.01.002>.

- [82] ROSENSTOCK I M. Historical origins of the health belief model[J/OL]. Health Education Monographs, 1974, 2(4): 328-335[2023-10-24]. <https://doi.org/10.1177/109019817400200403>.
- [83] JANZ N K, BECKER M H. The health belief model: a decade later[J/OL]. Health Education Quarterly, 1984, 11(1): 1-47[2023-10-24]. <https://doi.org/10.1177/109019818401100101>.
- [84] ROSENSTOCK I M. Why people use health services[J/OL]. The Milbank Quarterly, 2005, 83(4): 10.1111/j.1468-0009.2005.00425.x[2023-10-24]. <https://doi.org/10.1111/j.1468-0009.2005.00425.x>.
- [85] QUICK B L. Applying the health belief model to examine news coverage regarding steroids in sports by ABC, CBS, and NBC between march 1990 and may 2008[J/OL]. Health Communication, 2010, 25(3): 247-257[2023-10-28]. <https://doi.org/10.1080/10410231003698929>.
- [86] SEITER C R. Benefit, barrier, and self-efficacy messages in advance care planning education materials[J/OL]. Health Communication, 2021, 36(11): 1397-1404[2023-10-28]. <https://doi.org/10.1080/10410236.2020.1754601>.
- [87] GUIDRY J P, MEGANCK S L, PERRIN P B, et al. #Ebola: tweeting and pinning an epidemic[J/OL]. Atlantic Journal of Communication, 2021, 29(2): 79-92[2023-07-18]. <https://doi.org/10.1080/15456870.2019.1707202>.
- [88] YANG B, LI Y, TERÁNC L, et al. COVID-19 vaccines #ForYou: analyzing COVID-19 vaccine videos on TikTok during the early phase of the vaccine rollout in the U.S.[J/OL]. Health Communication, 2023[2023-10-12]. <https://doi.org/10.1080/10410236.2023.2227431>.
- [89] LI W, NOWAK G, JIN Y, et al. Inadequate and incomplete : chinese newspapers' coverage of the first licensed human papillomavirus (HPV) vaccine in china[J/OL]. Journal of Health Communication, 2018, 23(6): 581-590[2023-10-12]. <https://doi.org/10.1080/10810730.2018.1493060>.
- [90] DU J, LUO C, SHEGOG R, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine[J/OL]. JAMA Network Open, 2020, 3(11): e2022025[2023-02-26]. <https://doi.org/10.1001/jamanetworkopen.2020.22025>.
- [91] GUIDRY J P, VRAGA E K, LAESTADIUS L I, et al. HPV vaccine searches on Pinterest: before and after Pinterest's actions to moderate content[J/OL]. American Journal of Public Health, 2020, 110(S3): S305-S311[2023-10-12]. <https://doi.org/10.2105/AJPH.2020.305827>.
- [92] LI J, LI H. Research on product feature extraction and sentiment classification of short online review based on deep learning[J]. Chinese Journal of Information Studies: Theory & Application, 2018, 41(2): 143-148.

- [93] TURIAN J, RATINOV L A, BENGIO Y. Word representations: a simple and general method for semi-supervised learning[C/OL]//HAJIČ J, CARBERRY S, CLARK S, et al. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden: Association for Computational Linguistics, 2010: 384-394[2024-02-24]. <https://dl.acm.org/doi/pdf/10.5555/1858681.1858721>.
- [94] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[M/OL]. arXiv, 2013[2024-02-24]. <https://doi.org/10.48550/arXiv.1310.4546>.
- [95] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation [C/OL]//MOSCHITTI A, PANG B, DAELEMANS W. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014: 1532-1543[2024-02-24]. <https://doi.org/10.3115/v1/D14-1162>.
- [96] BOJANOWSKI P, GRAVE E, JOULIN A, et al. Enriching word vectors with subword information[M/OL]. arXiv, 2017[2024-02-24]. <http://dio.org/10.48550/arXiv.1607.04606>.
- [97] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[M/OL]. arXiv, 2019[2024-02-24]. <http://dio.org/10.48550/arXiv.1810.04805>.
- [98] LEWIS M, LIU Y, GOYAL N, et al. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[C/OL]//JURAFSKY D, CHAI J, SCHLUTER N, et al. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 7871-7880[2024-02-24]. <https://doi.org/10.18653/v1/2020.acl-main.703>.
- [99] DONG L, YANG N, WANG W, et al. Unified language model pre-training for natural language understanding and generation[M/OL]. arXiv, 2019[2024-02-24]. <http://dio.org/10.48550/arXiv.1905.03197>.
- [100] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[M/OL]. arXiv, 2023[2024-02-24]. <http://dio.org/10.48550/arXiv.1706.03762>.
- [101] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[M/OL]. arXiv, 2018[2024-02-24]. <http://dio.org/10.48550/arXiv.1801.06146>.
- [102] BENAÏM S, WOLF L. One-shot unsupervised cross domain translation[M/OL]. arXiv, 2018 [2024-02-24]. <http://dio.org/10.48550/arXiv.1806.06029>.
- [103] MOTIIAN S, JONES Q, IRANMANESH S M, et al. Few-Shot adversarial domain adaptation [M/OL]. arXiv, 2017[2024-02-24]. <https://doi.org/10.48550/arXiv.1711.02536>.
- [104] BERTINETTO L, HENRIQUES J F, VALMADRE J, et al. Learning feed-forward one-shot

- learners[M/OL]. arXiv, 2016[2024-02-24]. <https://doi.org/10.48550/arXiv.1606.05233>.
- [105] FINN C, XU K, LEVINE S. Probabilistic model-agnostic meta-learning[A/OL]. 2018[2024-02-24]. <https://doi.org/10.48550/arXiv.1806.02817>.
- [106] HENNEQUIN R, KHLIF A, VOITURET F, et al. Spleeter: a fast and efficient music source separation tool with pre-trained models[J/OL]. Journal of Open Source Software, 2020, 5 (50): 2154[2024-02-07]. <https://doi.org/10.21105/joss.02154>.
- [107] iFLYTEK. Speech recognition API—convert audio to text online[EB/OL]. 2023[2023-10-22]. <https://global.xfyun.cn/products/lfasr>.
- [108] FFmpeg Developers. FFmpeg[EB/OL]. 2023[2023-10-22]. <http://ffmpeg.org/>.
- [109] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach [M/OL]. arXiv, 2019[2023-10-12]. <http://dio.org/10.48550/arXiv.1907.11692>.
- [110] brightmart/roberta_zh: RoBERTa 中文预训练模型: RoBERTa for Chinese[EB/OL]. 2022 [2024-03-23]. https://github.com/brightmart/roberta_zh.
- [111] GROOTENDORST M. BERTopic: neural topic modeling with a class-based TF-IDF procedure[M/OL]. arXiv, 2022[2023-10-12]. <http://dio.org/10.48550/arXiv.2203.05794>.
- [112] REIMERS N, GUREVYCH I. Sentence-BERT: sentence embeddings using siamese BERT-networks[M/OL]. arXiv, 2019[2024-02-20]. <http://dio.org/10.48550/arXiv.1908.10084>.
- [113] THAKUR N, REIMERS N, DAXENBERGER J, et al. Augmented SBERT: data augmentation method for improving bi-encoders for pairwise sentence scoring tasks[M/OL]. arXiv, 2021[2024-02-20]. <http://dio.org/10.48550/arXiv.2010.08240>.
- [114] MCINNES L, HEALY J, MELVILLE J. UMAP: uniform manifold approximation and projection for dimension reduction[M/OL]. arXiv, 2020[2024-02-20]. <https://doi.org/10.48550/arXiv.1802.03426>.
- [115] MCINNES L, HEALY J. Accelerated hierarchical density based clustering[C/OL]//2017 IEEE International Conference on Data Mining Workshops (ICDMW). 2017: 33-42[2024-02-20]. <https://doi.org/10.1109/ICDMW.2017.12>.
- [116] SUN A. jieba[EB/OL]. 2023[2023-10-21]. <https://github.com/fxsjy/jieba>.
- [117] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for Chinese BERT[J/OL]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3504-3514 [2023-10-12]. <https://doi.org/10.1109/TASLP.2021.3124365>.
- [118] jinhuakst. weibo_senti_100k[J/OL]. GitHub, 2023[2023-10-22]. https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/weibo_senti_100k.
- [119] LIN, GAO C, PIAO J, et al. An exploratory study of information cocoon on short-form video platform[C/OL]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta GA USA: ACM, 2022: 4178-4182[2024-03-26]. <https://doi.org/10.1145/354182>.

[//doi.org/10.1145/3511808.3557548](https://doi.org/10.1145/3511808.3557548).

附录 A 补充材料

A.1 文本预处理无关词表

A.1.1 微博无关词表

免费咨询、预约找我客户、打扰一下深度烧伤的孤儿、代理招商加盟、超声刀、申请创建超话、免费体检、抢先排队、券后价、咨询预约顾问、鞠婧祎、光子嫩肤、私人订制、开团、超低价、水光针、优惠有限、美白丸、皮秒激光、美白抗糖、特惠、香港房产、皮肤管理、VIP 房、微博小店、鞠婧祎超话、马水振宝宝超话、香港医美超话、香港仁和 Android、关爱小俊逸超话、和田玉超话、高以翔超话、关爱鹏鹏超话、马水振宝宝超话、手机淘宝、优惠券、淘口令、明格帆、港兆健康、韩国整形

A.1.2 微博无关用户

孤独的暴发户、北京三甲医院挂号效率高、全仁医务中心 TrinityMedical、秀目庄园官博、听说取个长名字的人有个性、北京嘉禾妇儿医院、在首尔旅游网、DA 整形外科院长团队、康福國際健康管理、甜奈小姐姐-Annie、日本 tna 美白抗糖饮总部、香港仁和体检 HPV9 价基因检测医美、TNA 甜奈小姐姐-JENNIE、美安预约网、香港仁和預約助理 Hazel、丹东汇强私人健身、深圳市利宝亨投资顾问有限公司、段芝堂、抗衰老专家-严文、慈铭体检深圳分公司、司徒敬豪醫生、高先生还是个宝宝、松赢在线、香港现代医学专科、香港美亚预约网、拜恩健康、美德美医团购、2827403370_225、商业时报、人生如此无极限、靠谱的 Cindy-香港保诚、香港康心 CAYA、菡菡小花花、香港医美 or 体检-Tina、淘券优惠券分享、香港新生植发官博、香港 BMU 境外医疗医美顾问 kayla

致 谢

从第一次踏入科大，到现在我在这座安静的校园中经历了很多。收到了很多人善意的帮助和温暖，也有很多成长的阵痛，回想这段岁月，非常让我骄傲的是，我一直保持着自由而开放的自我探索和好奇心。也非常感谢科大拥有开放、自由、包容的环境让我非常自由的成长和发展。

首先感谢的是我的父母和亲人，为我提供了选择背后的支撑。虽然在科大以来，我自诩要做一个独立的人，在情感和经济上都要自立，但是在面临艰难时刻时，感谢父母无私质朴的伸出援手。时至今日，父母或许并不能理解我个人想要和追求的是什么，但是还是很感谢他们对我无私质朴的爱，给予我支撑和力量，能让我追寻我想要的东西。

感谢我的导师朱孟潇教授，朱老师是一位优秀的学者，她深厚的学术功底、敏锐的思维和严谨的治学态度使我受益匪浅，也让我领略到了计算社会科学这个交叉领域的魅力。在生活上，朱老师是一位慈爱的女性，给了我很多生活中的关照和温暖。本文从选题到研究设计到撰写，离不开朱老师的悉心指导。本文的完成离不开朱老师对我悉心的指导，记得 HPV 项目刚开始启动的时候，她有些时候凌晨都在跟我线上讨论课题和项目，那个时候在半夜走出西区科技楼实验室的时候，是充满希冀和对未来的向往的，也非常感谢朱老师能够跟我一起半夜奋战修改投稿的论文。转专业到科技传播系我感受到了非常自由和温暖的关怀，给予了我在研究之余心灵上的慰藉和鼓舞，这是一段非常奇妙的体验和非常美好的经历。感谢纪娇娇老师对我的照顾和关怀，纪老师不仅仅在计算传播学上给过我指导，同时也在我经历低谷的时候作为班主任老师对我有非常大的帮助，我对她充满感激与感谢，同时也要感谢在科技传播系学习期间董军峰老师、马子颂老师等老师对我的照拂和关怀。

同时需要特别感谢德州大学奥斯汀分校的栗若潇博士，若潇师姐在研究上给了我很多的启发和指点，作为材料转传播的同路人，跟若潇师姐的多次讨论使我入门了社会科学这个领域的研究，并且收获了很多。在若潇师姐读博士期间，依然给了我的研究很多的帮助和启发，在此我深表感谢。感谢信息学院的张立成博士，立成师兄给了我很多的帮助，常常半夜都在找他讨论问题。同时感谢赵晗一直以来非常给力和靠谱的合作。感谢王玉蕾在本文人工编码部分对待科研数据有着哈工大人“规格严格，功夫到家”一丝不苟的严谨精神。感谢王欣在我回复 ICWSM 审稿意见的时候给予的帮助和反馈。也一并感谢项目合作者和实验室的各位同学，他们是柳林、张帆、王晓雅、张林昕、汪先涛、王禧玥、史晗、焦凯杰等人。

在我硕士学习期间，我也有幸在语言科学交叉研究中心参与了科研项目，我要特别感谢袁家宏教授对我的指导。袁老师在语言学、自然语言处理、语音分析和处理方面给了我非常细致的指导。同时需要感谢欧洲科学院的院士、荷兰莱顿大学的 Yiya Chen 教授对我的指导，她让我觉得实验语音学是一个非常有趣有意思的领域，同时非常感谢她跟我讨论科研，那些突然的灵光乍现的激动和欣喜让我非常有收获。感谢语言科学交叉研究中心的王怡哲老师对我在计算语言学和大语言模型方面的指导。同时非常感谢爱丁堡大学的李远超博士对我的指导和合作，远超师兄对于我在语音处理和研究的指导非常细致，跟他合作让我觉得受益匪浅。同时也要感谢冯锐，韩欣然，胡子谦，王一鸣，杨映川等人在研究和生活中的帮助。

生活中，许多人给了我宝贵的帮助。感谢我在科大遇到的许多朋友。管理学院 SA19 同学、地空学院 BA21 同学，在我初到科大时给予了许多关怀，三年中也在无数重要时刻给予了支持，见证了我的成长。感谢计算机学院 SA21 同学和物理学院的 PB17 同学，伴随我一起成长，让我获得了许多科研、职业与人生的见解，也一同经历了低谷。感谢大数据学院的 SA20 同学，在许多思想和见解上对我有很大的启发。感谢 PB2100 同学在我经历低谷的时候能够陪伴我，让在低谷期的我有朋友的支撑和寄托。同时，感谢来自科大三百六十行、南七茶馆、校网球协会、校游泳协会及科技传播系 20 级、21 级、22 级的同学，在一些时刻给予我温暖和帮助。

同时非常感谢科大三百六十行，作为最开始来科大的时候和 PB17 同学聊起的一个 idea，我们就去做了，建立了这个社群，在能够帮助科大同学扩展眼界的同时，极大的丰富了我对各个行业在各个层面的认知，这是一笔非常宝贵的投资和收获。

特别感谢科大网球场，尤其是少院、生院、计算机学院、物院的 PB20、PB21、PB22、PB23 的朋友们，在我经历一些低谷时，他们陪伴我一起打球，给予了我内心的慰藉。同时感谢科大的游泳馆，在那些独处的时刻让我感到治愈与安宁。

感谢科大的物理空间和其中的个体以及他们的思想。感谢东区图书馆，东区郭沫若广场、东区操场，东区网球场，东区 1958，中区游泳馆，中区游泳馆的管理姐姐，中区游泳馆天台，中区网球场，中区健身房，西区科技楼，西区也西湖及周边，西区操场，食堂，宿舍，Fabio，东区人文楼。这些地方和地方的个体在某个时间和空间里面满足了我心血来潮和一时兴起的午夜听歌乱逛，日常的学习、交流和生活，给予了我很难忘而又珍贵的回忆。感谢科大人所有有趣的一切，包括各种讲座、有趣的人和故事、有趣的校友和科大所承载的一切，让我在这几年里面有趣而有会心一笑的东西。

同时感谢我经历的低谷，一路走来，因为自己的任性和自由的天性并不容

易，感谢在我遭遇挫折和困境的时候鼓励我的人，帮助我的人，这些人中有些甚至还是素昧平生的网友，感谢他们的善良，感谢他们的坚韧，感谢他们对生存的斗志，在夜晚的某个时刻温暖和激励了我，能够一直走下去。

何浏

于金寨路 96 号科大东区

二零二五年五月

在读期间发表的学术论文与取得的研究成果

研究报告

1. 主流价值观传播中微博用户行为画像. 全媒体信息传播理论与基础服务技术研究: 面向场景化内容的用户画像和影响力建模. 科技部重点研发项目 (SQ2020YFB40025). 2022.

会议论文

* equal contribution

1. Zhu, M.*, **He, L.***, Zhao, H., Su, R., Zhang, L., Hu, B. Same Vaccine, Different Voices: A Cross-Modality Analysis of HPV Vaccine Discourse on Social Media, Proceedings of the International AAAI Conference on Web and Social Media Regular Track (ICWSM 2025), June 2025.

会议海报

1. ICA Regional Hub 2023 Hefei, 海报展示

参与的科研项目

1. 全媒体信息传播理论与基础服务技术研究: 面向场景化内容的用户画像和影响力建模, 科技部重点研发项目 (SQ2020YFB40025). 2022.
2. 社会语言学、病理语言学和人工智能三重视角下基于语言变化的 AD 语言生物标志物研究, 国家社会科学基金项目 (23YYA01185). 2023

获奖的比赛

1. 第二届中国科学技术大学量化交易研究大赛, 第一名, 2023
2. 第五届传播数据挖掘竞赛, 三等奖, 2023