

PUBLIC PERCEPTIONS OF UNIVERSAL HEALTHCARE ON SOCIAL MEDIA: A MACHINE LEARNING APPROACH WITH SENTIMENT ANALYSIS AND TOPIC MODELING

Michelle Chen, Cranbrook Kingswood Upper School
Chen Ye, Purdue University Northwest

ABSTRACT

In the United States, universal healthcare has been a long-term, open-to-debate topic, because the US is the only developed country without it, and health insurance imposes significant costs on employers and employees. Public opinions on universal healthcare are diverse and influenced by multiple factors, such as political campaigns, public health crises, social events, and technological advancement. The public perceptions on universal healthcare need to be precisely captured and understood for stakeholders to make the right decisions to cover the majority people's interests. Social media's rise and popularity offer an innovative method to track public sentiment over such societal topics. Prior research has shown that the rich source of information on social media can be used to understand the public's sentiment and opinions on major events. Therefore, in this study, we collected and analyzed 371 posts and 22,756 comments on Reddit related to universal healthcare, using Reddit Application Programming Interface (API) and Python programming. With Linguistics Inquiry and Word Count (LIWC), we discovered public sentiment shifted gradually in a positive direction on universal healthcare topics in the past 15 years. In addition, we used machine learning Latent Dirichlet Allocation (LDA) topic modeling to identify five key themes in the public discourse related to universal healthcare: costs and taxes, voting to change, comparison with other countries' healthcare systems, healthcare cost, and government authority expansion and socialism. Implications of the findings will be discussed.

Keywords: *universal healthcare, sentiment analysis, topic modeling, machine learning, social media analytics*

INTRODUCTION AND PREVIOUS RESEARCH

Universal healthcare (UHC) has been an open-to-debate topic in America for many years (Zieff et al., 2020). While proponents of UHC advocate its benefits to public health, the hefty financial cost to implement UHC is recognized as a formidable barrier (Bloom et al., 2018). Many countries and territories in the world have universal healthcare systems, such as Canada, the United Kingdom, Saudi Arabia, Japan, Brazil, and Australia. However, the United States is not one of them, and health insurance impose significant costs on both workers and businesses that hire them in the U.S. Several efforts have been made in American history towards universal healthcare, such as Medicare and Medicaid from the 1960s, Bill and Hillary Clinton's advocacy for universal health coverage, and the Affordable Healthcare Act in Barack Obama's presidential

election. However, a comprehensive universal healthcare system has yet to be implemented in the United States. Public opinions on universal healthcare in the United States are diverse and have been swinging over time alongside historical events and technological revolutions, such as presidential elections, the COVID-19 pandemic, and the new advancement of artificial intelligence (Bloom et al., 2018; Brennan, 2023).

Previously, public opinion on such public policy was evaluated primarily through traditional survey firms, such as Gallup and Pew Research, to interview certain groups of people with questionnaires (Brennan, 2023). Since the emergence of the internet and Web 2.0 technologies, the number of social media users has surged dramatically. As of 2022, there were 4.62 billion social media users worldwide, which represents sixty-two percent of the world population (Pabian & Pabian, 2023). Social media offers a new option for end users to express their opinions and emotions about a wide spectrum of subjects with minimal constraints. This genuine user-generated content on social media provides a rich source of information for businesses, governments, and policymakers. Social media content mining, sentiment analysis, and topic modeling based on computational analysis provide a novel approach to tracking public responses to universal healthcare topics.

Content mining refers to the process of extracting pertinent information from a natural language text. Within this domain, sentiment analysis is the field of study that specifically identifies people's subjective opinions, evaluations, and attitudes or sentiments towards products, services, political issues, education, and so on from the written language. The sentiment analysis methodologies can be classified into three main categories in general, i.e., lexicon-based techniques, machine learning techniques, and hybrids of both approaches. The lexicon-based technique is also known as a dictionary-based approach. It relies on a collection of known and predetermined sentimental terms, counting positive and negative words in the related data set, and assumes that the overall sentiment expressed in a given text document can be calculated as the average of the sentiment scores of its words and phrases. Linguistic Inquiry and Word Count (LIWC) is the major lexicon-based sentiment analysis tool. Machine learning approaches are based on various algorithms for training data sets to detect sentiments from test data sets; hybrid approaches are based on a combination of these two (Drus & Khalid, 2019).

Topic modeling is a machine learning technique used to identify latent topics within a corpus of documents or texts. Latent Dirichlet Allocation (LDA) is a commonly used unsupervised learning algorithm for topic modeling in Natural Language Processing (NLP) and acts as an intermediate layer between words and documents. LDA is based on the Bayesian probability model, assuming that each topic is a mixture over a set of words and each document is a mixture over sets of topic probabilities (Blei, 2012). In iterative learning processes, the LDA model assigns topics to a given corpus based on the words present.

The above-mentioned natural language processing techniques have been used in numerous studies (Sanders et al., 2021; Ye & Zhao, 2021; Melton et al., 2022; Ruan & Lv, 2022) to surveil social media posts. X (Twitter) is the leading social media platform that is used to collect user opinion for social media content analysis (Drus & Khalid, 2019). However, our study utilizes Reddit as the content resource for the following reasons: unlike X, which limits tweets up to 280 characters, Reddit posts can reach 40,000 characters, while comments can extend to 10,000 characters. The expanded character space allows for more in-depth and nuanced discussions on Reddit, which is an invaluable source of data for Natural Language Processing as it provides a deeper insight into user opinions and attitudes. In addition, the majority of user accounts on Reddit are anonymous, which encourages unrestricted sentiment expression

compared to X, which are mostly identifiable profiles. Moreover, Reddit represents the platforms that have not been studied as extensively as X, and it is worth further exploration. The key comparisons between Reddit and X are listed in Table 1.

Table 1 Comparison of Reddit and X (Twitter)		
	Reddit	X (Twitter)
Social Media Type	Blogs	Micro-blogs
Length of Post	40,000 characters on posts and 10,000 on comments	280 characters
Content focus	Wide range of topics, large and diverse user base, detailed discussion	Short updates of current events, politics, breaking news and cultural topics
Privacy	Authors are anonymous and user driven	More popular with celebrities and influencers, most public profiles
Interaction Style	Upvotes, downvotes, and threaded comments	Direct interaction through replies, retweets, and likes

To our best knowledge, no social media content sentiment analysis or latent topic modeling research has been done on universal healthcare topics on Reddit. In this study, the public's opinion data from Reddit was retrieved and analyzed with NLP techniques, LIWC, and LDA, aiming to answer the following questions:

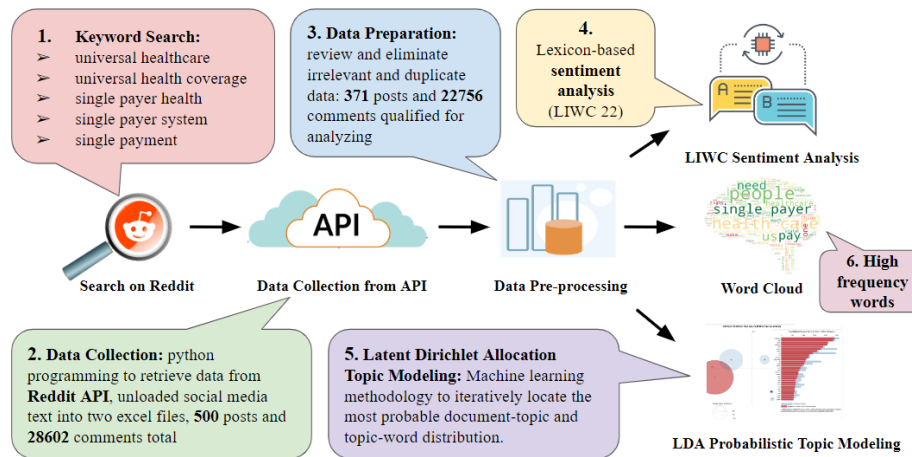
RQ1: What are the public perceptions of universal healthcare expressed in related social media discussion?

RQ2: What are the main topics in the public discourse concerning universal healthcare?

METHODS

The overall data collection, data preparation, and data analysis procedures of this study are indicated in Figure 1. We started with a Reddit platform search, and data was retrieved from the Reddit API. The returned data was pre-processed before further analysis using Natural Language Processing tools.

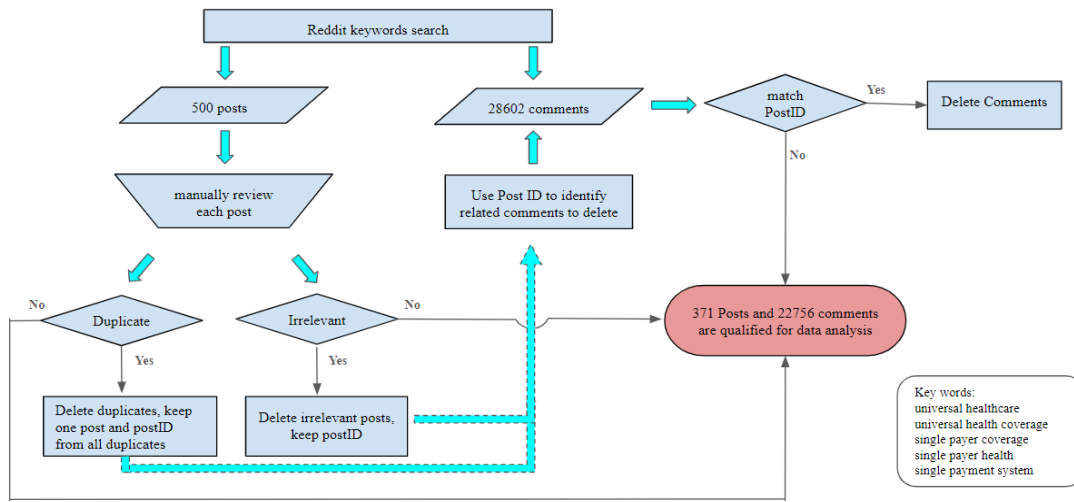
Figure 1
Data Collection, Data Preparation, and Data Analysis Process with NLP Tools



Data Collection and Data Preparation

The Reddit posts and comments written in English were searched and retrieved via the Reddit Application Programming Interface (API), which is a program interfacing between applications and web services to perform specific functions. Our data retrieval application is developed in the Python programming language. The interchangeable phrases “universal healthcare,” “universal health coverage,” “single payer coverage,” and “single payer health,” and “single payment system” were used as search keywords. The search through the Reddit API returned the top 100 most relevant posts for each keyword, with a time range from 2009 to 2023. Each post also contains a cluster of responding comments, and the comments are linked to the post by a unique postID. In total, 500 posts with 28,602 comments were retrieved initially. Each post was reviewed individually to eliminate duplicates and posts unrelated to universal healthcare in the United States. Follow-up comments associated with the eliminated posts were also removed. After the review and elimination process, 371 posts and 22756 comments were retained for data analysis. The data collection and data preparation workflow are illustrated in Figure 2.

Figure 2
Data Collection and Preparation Workflow



This study utilized Linguistic Inquiry and Word Count (LIWC), a widely adopted text language analysis tool, to analyze the various emotional, cognitive, and structural elements within people’s spoken and written language samples (Boyd et al., 2022). The LIWC text analysis operates by calculating the percentage of words in the target text matching its dictionary. LIWC-22, the most recently released version, has expanded dictionaries and improved user-friendly software design. The LIWC-22 analyzes over 100 dimensions of text and is capable of capturing 80% - 90% of all words people use in oral and written English, including words frequently used on social media (Boyd et al., 2022). The psychological constructs and processes in LIWC-22 include affect, cognition, and perception. Positive emotion and negative positive emotion are both measured constructs as part of the affective process. For example, LIWC would score the sentence “Pizza is my favorite food” as 20% for positive emotion because “favorite” is the only word of positive emotion out of the total of five words in the sentence. Prior research has proved the reliability of LIWC in identifying emotional expression (Tausczik & Pennebaker 2010). The direction and strength of the sentiment can be calculated by the sentiment polarity of each post using the following formula: $\text{polarity} = \text{positive score} - \text{negative score}$ (Salehan & Kim 2016), where both scores range from 0 to 100. Therefore, the sentiment polarity score falls within the range [-100, 100], with 0 representing a neutral sentiment.

This study utilized Latent Dirichlet Allocation (Blei et al., 2003), a probabilistic topic modeling algorithm, to discover latent discussion topics related to universal healthcare. The application was developed in the Python programming language with the SciKitLearn package, and LDAVis, a web-based interactive visualization tool, is used to display the frequency and relativity of words and topics. LDA uses unsupervised machine learning to assign words to the topics and topics to the documents in a corpus. An iterative process helps to discover optimal assignments, and the coherence score measures how semantically similar the top words of each topic are, with a scale of [0-1] (Syed and Spruit, 2017). A higher coherence score indicates that the topics are more meaningful and interpretable.

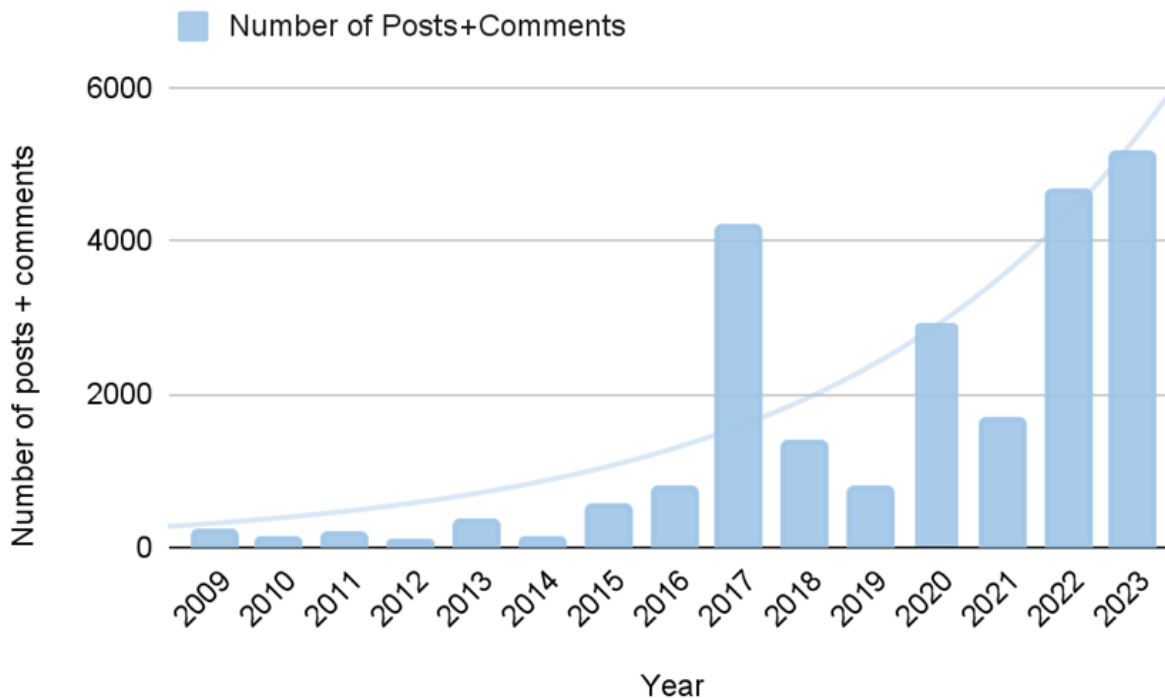
RESULTS

Dataset

After manual review and elimination, 371 posts and 22,756 comments related to universal health care were retained from the keyword search on Reddit. The number of posts and comments for each year is listed in Table 2, and the graphic view is shown in Figure 3.

Table 2 Summary of Number of Posts and Comments Each Year			
Year	Number of Posts	Number of Comments	Number of Post + Comments
2009	3	224	227
2010	2	128	130
2011	3	169	172
2012	1	80	81
2013	5	356	361
2014	2	107	109
2015	12	542	554
2016	15	768	783
2017	88	4094	4182
2018	19	1343	1362
2019	11	760	771
2020	34	2903	2937
2021	30	1637	1667
2022	64	4587	4651
2023	82	5058	5140
Total	371	22756	23127

Figure 3
Annual Total Number of Posts and Comments on Universal Healthcare Related Topics



The total number of relevant posts and comments on Reddit increased exponentially over the past fifteen years from 2009 to 2023. There was a sharp spike in the number in 2017, which indicates public attention was attracted to universal healthcare during 2017. Close examination of the public discourse reveals that the large spike occurred when former president Donald Trump started to repeal the Affordable Care Act and Senator Bernie Sanders advocated for single payer health, fueling substantial public discussions. Additionally, universal healthcare topics regained public interest in 2020 and have received continuously heightened ongoing attention and discourse. The second surge of public interest occurred amid the COVID-19 pandemic when millions of people lost their jobs and their employer-sponsored healthcare insurance, bringing public attention back to universal healthcare.

To address the potential impacts of the same users posting multiple opinions on the results, we also tabulated the posting frequency distribution in Table 3. A total of 20137 unique user IDs are identified from the posts and comments. Only less than ten users out of 20137 posted more than 8 posts and comments, which account for less than 0.05% of total user IDs. The majority of the users post less than or equal to three times, which account for 99.5% of the total user IDs. This result suggests that the data set represents a wide range of users' opinions, instead of a few users with a high volume of biased opinions.

Number of Posts + Comments by the Same User ID	Number of User ID	Percent in Total Accounts
20	1	0.0050%
12	1	0.0050%
9	3	0.015%
8	4	0.020%
7	2	0.0099%
6	6	0.030%
5	27	0.13%
4	49	0.24%
3	490	2.4%
2	1631	8.1%
1	17923	89%
Total	20137	100%

Sentiment Analysis

The retained posts and comments were analyzed by the LIWC-22 application, with each item receiving a positive tone score and a negative tone score. The polarity of each item is calculated as:

$$\text{polarity} = \text{positive tone score} - \text{negative tone score}$$

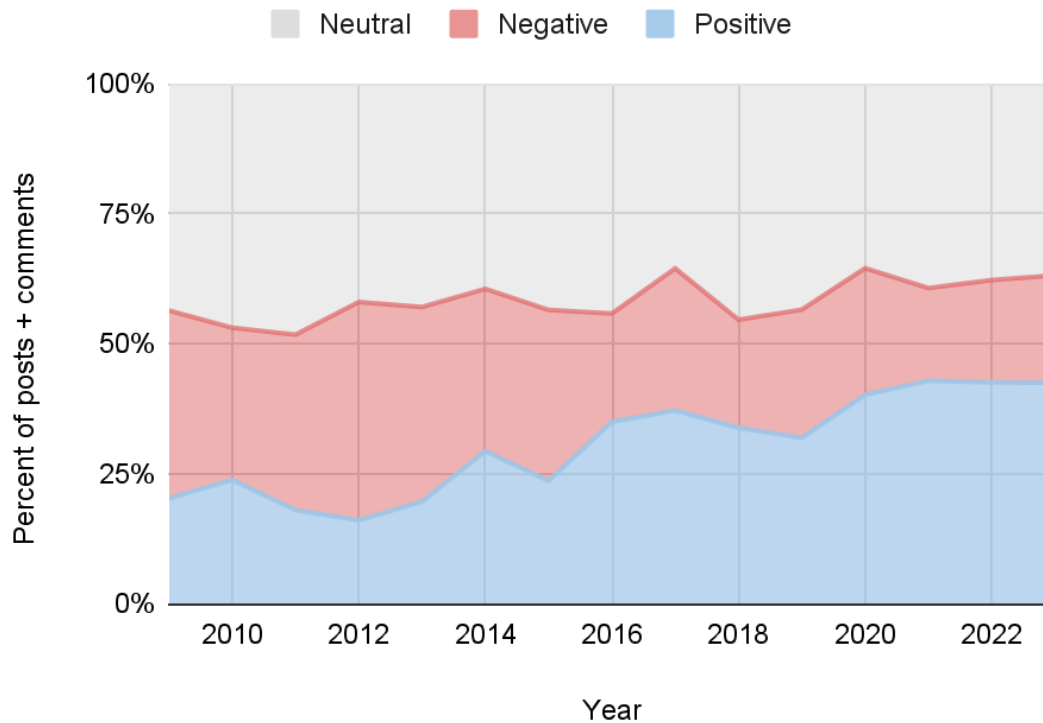
Positive polarity scores mean more favorable opinions on the topic; zero means neutral; and a negative score means unfavorable opinions. Sample posts and comments with positive, neutral, and negative polarity scores are provided in Table 4.

Table 4		
Sample Reddit Posts and Comments with Polarity Scores		
Sentiment	Sample Reddit Posts	Polarity
Positive	<i>100% yes. Health care is a basic human right and should be free to everyone</i>	13.33
	<i>It must be a lot of work for doctors to keep all the current insurance plans in line. A single payer plan would help simplify that job.</i>	7.41
	<i>Many of us already pay insurance, deductible, and co-pays that are nearly what we would pay under universal health care. The main thing we lose is insurance companies paying bonuses to the CEO and going bankrupt ourselves. We gain everybody having coverage.</i>	2.38
Neutral	<i>Just FYI, my company pays part of my premium, my share was 1,200 for a family.</i>	0
	<i>And somehow, dental and optometry don't make the cut?</i>	0
	<i>The answer is unpredictable. It's the same as betting at what age you will develop cancer.</i>	0
Negative	<i>Apparently universal healthcare implies slavery of doctors...</i>	-14.29
	<i>No. Quality and service goes down. Korea did universal HC but it was so bad many people opted for private insurance.</i>	-4.76
	<i>No. I don't want my taxes going to pay medical bills for people living unhealthy lifestyles. If there were some kind of stipulations about things like smoking, alcohol use, drug use, diet induced obesity I'd reconsider.</i>	-2.78

In terms of sentiment polarity, the percentage of positive, negative, and neutral posts and comments from each year is presented in Table 5 and Figure 4.

Year	Number of posts + comments	Number of positive posts + comments	Number of negative posts + comments	Number of neutral posts + comments	Percent of positive posts + comments	Percent of negative posts + comments	Percent of neutral posts + comments
2009	227	46	82	99	20.3%	36.1%	43.6%
2010	130	31	38	61	23.9%	29.2%	46.9%
2011	172	31	58	83	18.0%	33.7%	48.3%
2012	81	13	34	34	16.1%	42.0%	42.0%
2013	361	71	135	155	19.7%	37.4%	43.0%
2014	109	32	34	43	29.4%	31.2%	39.5%
2015	554	131	182	241	23.7%	32.9%	43.5%
2016	783	274	163	346	35.0%	20.8%	44.2%
2017	4182	1555	1142	1485	37.2%	27.3%	35.5%
2018	1362	461	283	618	33.9%	20.8%	45.4%
2019	771	246	190	335	31.9%	24.6%	43.5%
2020	2937	1180	715	1042	40.2%	24.3%	35.5%
2021	1667	715	297	655	42.9%	17.8%	39.3%
2022	4651	1980	915	1756	42.6%	19.7%	37.8%
2023	5140	2184	1062	1894	42.5%	20.7%	36.9%

Figure 4
Sentiment Distribution Over Time



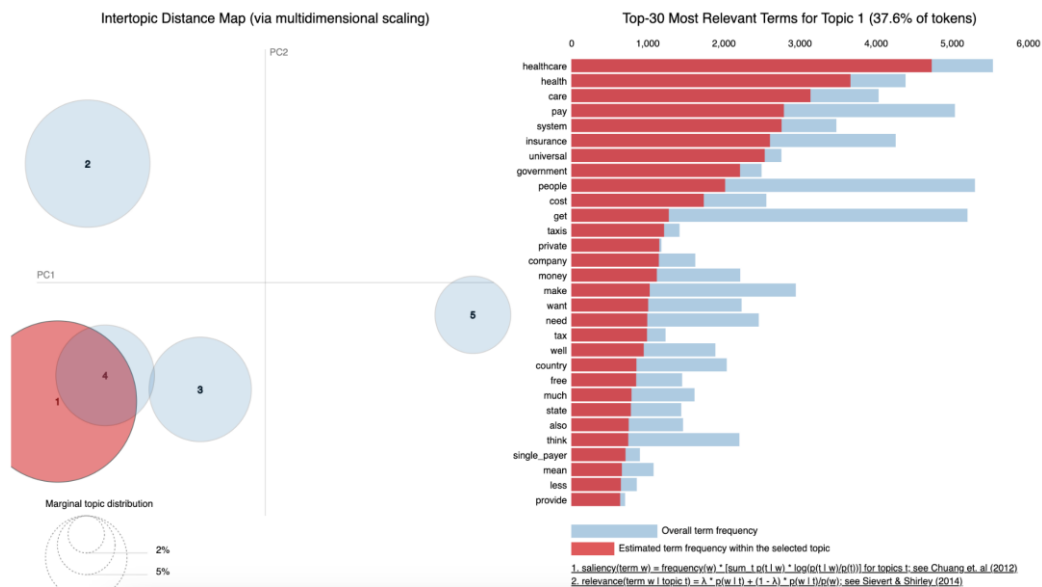
From 2009 to 2015, the percentage of negative posts and comments outweighed the percentage of positive posts and comments on universal healthcare discussion, while the majority remained neutral. After 2015, the percentage of negative posts and comments gradually decreased, while the percentage of positive posts and comments rose, indicating more favorable opinions toward universal healthcare, even though a large number of people still remain neutral.

Topical Modeling with Latent Dirichlet Allocation

We first use word cloud to highlight the popular terms. The font size and color of each word is proportional to its frequency in the posts and comments. Most prominent words can be identified in a quick glance in Figure 5, such as health care, people, single payer, universal healthcare, need, and pay.

words in the topic. The location of circles represents how the topics are related. The closer the circles, the topics are more related, and vice versa. Topic 1 and Topic 4 are more related than other topics in Figure 6. The coherence score is 0.613 for the LDA analysis in this study, which means a good semantic similarity between topical words and each topic (Syed and Spruit, 2017).

Figure 6
Example of LDA Visualization on Topic 1



Next, we will provide a brief overview of these LDA identified latent topics, with sample Reddit posts.

Costs and taxes (topic 1): The first latent topic generally addresses the question of “how to pay for a universal healthcare system”? The main reason that people have concerns about universal healthcare is possible tax increases to cover the costs of a universal healthcare system. Sample comments include:

- *“I believe these polls have generally been true, but the numbers drop significantly when you add the term, “Though it will raise taxes.””*
- *“This will never pass. Tax increases never pass in CO and a fucking 10% increase will be laughed right off the ballot. Ppl advocating for single payer just screwed the idea if this is the best they can do. It kills the middle class, not the rich.”*
- *“Would you accept higher taxes to enable universal mental health care?”*

Voting to change (topic 2): many posters recognize the critical role of voting for individuals to exert their power in shaping the healthcare system. Some comments read:

- *“Don’t forget people, get yourself a union that fights for free insurance. I pay \$0 for mine because the Teamsters did well. And voting helps too. To get us Medicare for all.”*
- *“As long as Americans vote for Republicans we will NEVER have universal health care. The ONLY ones I see advocating for it are the Democrats.”*
- *“I prefer the public option, but if we had the votes to get single payer I’d be happy for that as well.”*
- *“Vote for better people. Let them know that it's important to you.”*

Comparison with other countries’ healthcare systems (topic 3): people compare the American healthcare system with those of other countries, as the United States is the only developed country without universal healthcare, while even some developing countries have universal healthcare. Some sample posts state that:

- *“The United States is the only major country on Earth without universal healthcare, that's an absolute disgrace...”*
- *“I have said for a long time that the reason so many Americans hate the idea of universal health care is that they really believe we have the greatest system in the world, and that's because they have no experience with systems in other countries. I broke my wrist while on vacation in Greece. I had multiple sets of x-rays, an ER visit, a consult with an orthopedist. Cost me nothing. Absolutely nothing. Insurance is basically a for-profit way to keep people from getting healthcare. Right now, I am paying almost \$300 a month for insurance (with an \$8,000 deductible) and all of the providers I've used locally have stopped taking "marketplace" plans. So I have insurance, sure, but right now I have no access to health care.”*
- *“Yes. A former conservative, I traveled around the world and found that universal health care can work beautifully. And we would save all the costs like insurance and out-of-pocket costs. We could create the best universal health care system in the world. This issue led me to leave my former party and become an Independent who leans Progressive.”*

Healthcare cost (topic 4): the cost of healthcare at the individual and organizational levels emerges as another cluster of discussion. Whether people can afford the cost of healthcare under the current healthcare system is the issue that bothers many Redditors. The affordability of employee health plans to employers also attracts substantial discussion. Showing some overlapping with topic 1 (see Figure 6), the issues of costs and taxes are also brought up in some comments under this topic. Sample comments are:

- *“Just started a new job. For me and my spouse, I pay ~\$269.00 per pay period (every two weeks). My employer's contribution is ~\$530.00. Oh and the majority of it still only pays 80% after the deductible, not 100%. Fucking Bullshit...”*
- *“I am a contract employee and have really shitty healthcare. Would love another option. I doubt it would cost more than I already pay in premiums and co-pay.”*
- *“If every company in America continued to pay what they’re currently paying for health care into a singular fund we likely wouldn’t even need to increase personal taxes much, if at all. The divided market only creates inefficiencies and increases costs.”*

Government authority expansion and socialism (topic 5): many people consider the government provided healthcare a form of socialism and are against the expanded control from government. Some of those users commented that:

- *“When is Sanders going to start teaching college courses on socialism for free, full time? It's our human right after all. Then maybe when he's done taking 12 years of free medical courses, he can operate on me for free too. I really need that surgery Bernie, hop to it!”*
- *“100+ years of socialism has been systematically killing America. The solution is obviously more socialism. Plus, I see nothing wrong with stealing other people's money and redistributing it. As long as I *FEEL* it's the right thing to do - fuck our Constitution, freedom, and liberty.”*
- *“Keep the government out of the bedroom, bathroom and women's bodies” but now we want the government to have full control of our healthcare. Socializing the healthcare system is a horrible idea... like socialism.”*

DISCUSSION AND CONCLUSION

The objectives of this study were to apply computational social media content analysis to explore public sentiments, examine public perception trends, and identify the main topics in public discourse concerning universal healthcare. Posts and comments related to universal healthcare over the past fifteen years were collected from the social media blog site Reddit and analyzed. Results indicate an upward trend in public interest in universal healthcare topics over time.

Public Sentiment Towards Universal Healthcare

In response to the first research question, 371 Reddit posts and 22,756 comments related to universal healthcare and single payer systems topics from 2009 to 2023 were analyzed. The annual total number of posts and comments increased exponentially, indicating rising public interest in universal healthcare topics on Reddit. We also observed substantially heightened public discussions on universal healthcare in 2017 and 2020, triggered by significant political and public health events. The elevated public attention to universal healthcare continues to present, likely due to multiple factors including post-pandemic lifestyle change, such as retiring earlier and reducing working hours, younger generations desiring flexible employment opportunities, and the concerns of Artificial Intelligence replacing human jobs. These trends signal a more permanent shift towards valuing healthcare portability.

As the United States is the only developed country without universal healthcare coverage, people have mixed feelings towards this topic. Many people have the concern that government-provided healthcare is a form of socialism and will cause increased taxes and costs while reducing the quality of healthcare services. There was a higher percentage of people who showed unfavorable opinions on universal healthcare before 2015. However, public opinions changed gradually after 2015. This turning point is possibly due to the post-implementation effect of the Affordable Care Act, which launched open enrollment for the public in 2014. After a year of enrollment, more people accepted the government subsidized healthcare coverage. From 2015 to present, a slowly increasing percent of public opinions showed positive sentiment towards

universal healthcare. After years of implementation of the Affordable Care Act, more people are covered under this system. In addition, societal and cultural changes in lifestyle, generational changes, and technological advancements all impact the public's perception of universal healthcare, despite the large amount of negative sentiment that still exists. In summary, the slow but steadily increasing positive sentiment reflects the public's gradual favorable swing towards universal healthcare.

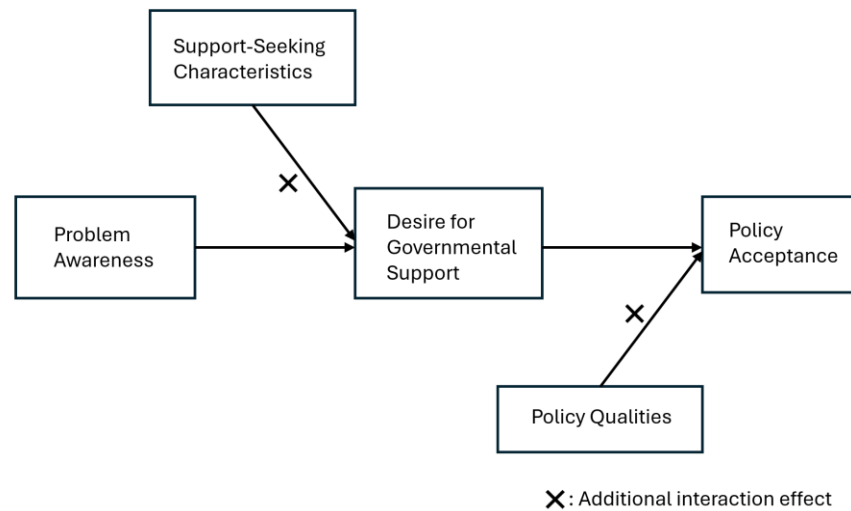
Topics of the Public Discourse on Universal Healthcare

The second research question of our study concerns the latent topics prevailing in the public discourse related to universal healthcare. The LDA topic modeling analysis provided insights into the key themes, such as the struggle with current healthcare insurance costs and desires to reduce payment. However, worries over potential tax increases caused by funding universal healthcare and government authority expansion pose barriers to expanding favorable support for universal healthcare. The fragmented American healthcare system was compared with the unified healthcare systems in other countries around the world, and questions were brought up about whether universal healthcare can be implemented in the United States. Voting was commonly cited as the optimal democratic mechanism to transform healthcare and address these challenges. The LDA analysis results revealed the multifaceted issues, opinions, and potential solutions comprising public discourse on this complex policy topic.

Theoretical and Practical Implications

To our knowledge, there has not been significant theoretical work that specifically addresses the citizens' attitudes toward universal healthcare as a public policy. However, researchers have recognized the importance of public acceptance to the success of public policies and investigated the psychological determinants of people's attitudes toward government policies and interventions in other domains of decision-making. Recently, Grelle and Hofmann (2024) conducted a review of over one hundred relevant empirical studies in fields ranging from psychology to political science and summarized the findings into the Integrative Public-Policy Acceptance (IPAC) framework, as shown in Figure 7. For scholars interested in understanding the formation of policy acceptance as a personal attitude in any given domain of policymaking, the IPAC framework serves as a concise and generalized framework that could be readily contextualized for a specific domain (Grelle and Hofmann, 2024). While it is outside the scope of this brief discussion to fully elaborate on the IPAC framework, it provides a suitable theoretical lens to interpret the findings and contributions of the present study.

Figure 7
The Integrative Public-Policy Acceptance (IPAC) framework (Grelle and Hofmann, 2024)



The first fundamental premise of the IPAC framework is that a citizen's awareness of a specific problem is the primary driving force behind their acceptance of a public policy aiming at addressing the issue. For example, those aware of the negative health effects of both firsthand and secondhand smoking will be more receptive to government interventions aiming at reducing public consumption of tobacco products and restricting smoking in public spaces. In alignment with this principle, the present study reveals awareness of the burden of healthcare costs on employers and individuals in the US as a prominent factor in the public mindset related to universal healthcare. In addition, knowledge of universal healthcare as a functioning policy in other countries may also be a contributing factor to citizens' awareness of the issue.

The IPAC framework also incorporates people's general desire for support from the government in a particular domain as a crucial mediator between problem awareness and public policy acceptance. Moreover, a citizen's perceptions of the characteristics of a certain public policy also contributes to policy acceptance both as a direct antecedent and as a moderator of the effect of desire for government support. Grelle and Hofmann (2024) describe perceived policy characteristics such as effectiveness, intrusiveness, transparency, fairness, and cost that could be salient in different problem domains as "policy qualities". For example, if a citizen perceives an outright ban by the government of all alcohol consumption to be an intrusion on private freedom, they may be less supportive of such policy even with higher degrees of awareness of the alcohol problem and desire for government support in tackling such problems in their country. The current study suggests that the cost of implementing a universal healthcare system to the taxpayers in the US to be one critical consideration concerning policy quality. Furthermore, some Redditors expressed their view of universal healthcare as a symbol of government overreach and socialism, suggesting that perceived intrusiveness on the private sector may also play a similar role in deciding acceptance of a national healthcare system operated by the government.

Additionally, the IPAC framework includes individual characteristics such as personal attitudes and beliefs that may affect a citizen's inclination to accept government support in a

specific domain. One such characteristic is a person's belief of control attribution, or the degree to which they attribute the outcomes of policymaking to the influence imposed through their own actions. The construct of perceived behavioral control has long been understood as a predictor of behavioral intention and actual behavior in Ajzen's theory of planned behavior (1991). Our finding points to a citizen's belief in their ability to control the direction of policymaking concerning universal healthcare by exercising their voting rights as one relevant individual characteristic.

In summary, the IPAC framework provides a valid theoretical underpinning for further understanding of public acceptance of universal healthcare. Successful application of a general framework such as IPAC in a policy domain requires researchers to contextualize key constructs by answering these questions: *what problem, which individual characteristics, and which policy characteristics?* By helping to answer these questions in the domain of universal healthcare, the present study lays the groundwork for contextualization and operationalization of the IPAC framework and contributes to further theoretical development in this area of research.

A deeper understanding of the factors that influence individual motivation to accept a universal healthcare system is also valuable to the practitioners in developing strategies to promote universal healthcare to the populace. Experts in health-related fields have weighed the pros and cons of universal healthcare. They advocate that universal healthcare will help reduce the economic burden of healthcare at the societal level in the long run, therefore the advantages of promoting a healthier population outweigh the initial implementation costs (Zieff et al. 2020). However, at the individual level, each citizen's interest in and disposition toward universal healthcare is understandably shaped by their socioeconomic and demographic characteristics, personal situations, and local conditions. Our findings highlight the importance of raising public awareness of the unsustainable costs of healthcare as a problem for the nation as a whole. Policymakers should also ensure transparency in the rulemaking process of the universal healthcare system to alleviate the concern of government overreach. To achieve broad support for universal healthcare, it is helpful to encourage each citizen to actively participate in the discussion and establish the belief that the future of healthcare is in their hands through the democratic process, i.e., casting votes in political elections at different levels.

LIMITATIONS AND FUTURE RESEARCH

One of the limitations of this research is that the Reddit API returns the top 100 most relevant posts per keyword search based on the Reddit matching algorithm. There is a possibility that posts that are relevant to universal healthcare but not returned by Reddit. We did multiple keyword searches and manually reviewed each search result. In each keyword search result, there are irrelevant posts being returned and there are duplicate posts returned across different keyword searches that are manually excluded from the analysis, which means our search method captures as many relevant posts as possible, but we cannot control the Reddit returning algorithm. Also, because Reddit accounts are anonymous, the analysis cannot be further categorized by users' age, gender, location, career or income levels, but this information will have impacts on the users' position on universal healthcare topics.

This study is also limited to the examination of public opinion expressed on one social media website during a particular period. To gain broader and deeper insights into the American public's attitude toward and potential acceptance of universal healthcare, we suggest several directions for future research such as expanding to other social media platforms and including a

more diverse user population. The research design can also be refined to unveil new findings and reveal the nuances that were not the focus of this study, such as sentiment analysis of each latent topic, or sentiment and topical analysis during major political and social events and movements to better understand the underlying dynamics of public sentiment shifts. Furthermore, the computational approach can be compared and combined with traditional survey methods to achieve a comprehensive understanding of public perceptions and validate the accuracy of computational findings against survey data. Lastly, there is fertile ground for scholars to apply the same analytic approach to address research questions related to other public policies with nationwide impacts on businesses and employees.

CONCLUSION

The findings of this study suggest that universal healthcare has attracted increasing public attention and will continue to be a spotlight on social media. The public's sentiment toward universal healthcare is moving in a favorable direction over time. The prevalent themes about which the public is most concerned about universal healthcare include healthcare affordability, potential tax increases, international healthcare comparison, government authority expansion, and voting being the optimal method of change. Also, this study demonstrates that social media content mining, sentiment analysis, and topic modeling have proved to be efficient and productive computational approaches for public opinion tracking and complex policy assessment. The data-driven governance and evidence-based policymaking approach provides leaders with an additional feedback channel beyond traditional polling.

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77- 84.
- Blei, D.M, Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning research*, 3, 993-1022.
- Bloom, D. E., Khoury, A., & Subbaraman, R. (2018). The promise and peril of universal health care. *Science*, 361(6404), p.eaat9644.
- Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin.
- Brenan, M. (2023, January 23), Majority in U.S. still say Gov't should ensure healthcare, Gallup, <https://news.gallup.com/poll/468401/majority-say-gov-ensure-healthcare.aspx>
- Drus, Z., & Khalid, H. (2019). Sentiment analysis in social Media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707-714.
- Grelle, S., & Hofmann, W. (2024). When and why do people accept public-policy interventions? An integrative public-policy-acceptance framework. *Perspectives on Psychological Science*, 19(1), 258-279.
- Kwok, S. W. H., Vadde, S. K. and Wang, G. (2021), Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis, *Journal of Medical Internet research*, 23(5).
- Melton, C. A., Olusanya, O.A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, 14(10), 1505-1512.
- Pabian, A., & Pabian, B. (2023). Role of social media in managing knowledge of the young generation in the sustainability area. *Sustainability*, 15(7), 6008.
- Ruan, T., & Lv, Q. (2022). Public perception of electric vehicles on Reddit over the past decade. *Communications in Transportation Research*, 2, 100070.

- Salehan, M. & Kim, D.J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30-40.
- Sanders, A.C., White, R.C., Severson, L. S., Ma, R., McQueen, R. Paulo, Haniel C. A., Zhang, Y., Erikson, J.S., & Bennet, K.P. (2021). Unmasking the conversation on masks: natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. *AMIA Jt Summits Transl Sci Proc*, 555-564.
<https://pubmed.ncbi.nlm.nih.gov/34457171/>
- Syed, S., & Spruit, M. (2017, October). Full-text or abstract? examining topic coherence scores using Latent Dirichlet Allocation. In *2017 IEEE International conference on data science and advanced analytics (DSAA)* (pp. 165-174). Ieee.
- Tausczik, Y., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54.
<https://journals.sagepub.com/doi/abs/10.1177/0261927x09351676>
- Ye, C., & Zhao, L.(2021). Public perceptions of Facebook's Libra digital currency initiative: Text mining on Twitter. *Proceedings of the 54th Hawaii International Conference on System Sciences*.
- Zieff, G., Kerr, Z. Y., Moore, J. B., & Stoner, L. (2020). Universal healthcare in the United States of America: A healthy debate. *Medicina*, 56(11), 580.