

Leveraging Prompt-Based Large Language Models: Predicting Pandemic Health Decisions and Outcomes Through Social Media Language

Xiaohan Ding
Department of Computer Science,
Virginia Tech
Blacksburg, USA
xiaohan@vt.edu

Buse Carik
Department of Computer Science,
Virginia Tech
Blacksburg, USA
buse@vt.edu

Uma Sushmitha Gunturi
Department of Computer Science,
Virginia Tech
Blacksburg, USA
umasushmitha@vt.edu

Valerie Reyna
Human Neuroscience Institute,
Cornell University
New York, USA
vr53@cornell.edu

Eugenia H. Rho
Department of Computer Science,
Virginia Tech
Blacksburg, USA
eugenia@vt.edu

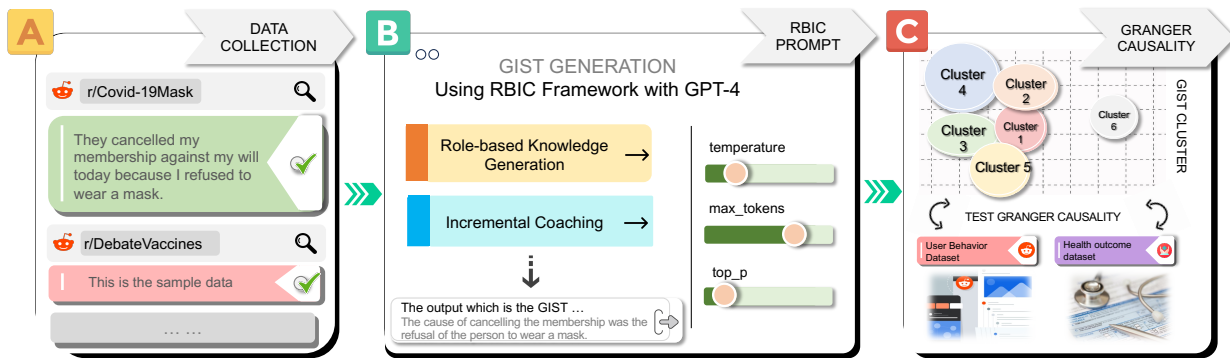


Figure 1: Our research empirically links social media conversations to national health outcomes related to COVID-19. We introduce Role-Based Incremental Coaching (RBIC), a Large Language Model (LLM) prompting framework. (A) Collecting Reddit datasets focused on communities known for opposing COVID-19 health practices. (B) Guided by Fuzzy-Trace Theory, we introduce a novel LLM framework called Role-Based Incremental Coaching (RBIC) to extract cause-effect pairs and formulate coherent gists capturing causal relations in texts. (C) Granger-causality tests and data analytics reveal the impact of these gists on community engagement and national health outcomes such as vaccination uptake and hospitalization rates.

ABSTRACT

We introduce a multi-step reasoning framework using prompt-based LLMs to examine the relationship between social media language patterns and trends in national health outcomes. Grounded in fuzzy-trace theory, which emphasizes the importance of “gists” of causal coherence in effective health communication, we introduce Role-Based Incremental Coaching (RBIC), a prompt-based LLM framework, to identify gists at-scale. Using RBIC, we systematically extract gists from subreddit discussions opposing COVID-19 health

measures (Study 1). We then track how these gists evolve across key events (Study 2) and assess their influence on online engagement (Study 3). Finally, we investigate how the volume of gists is associated with national health trends like vaccine uptake and hospitalizations (Study 4). Our work is the first to empirically link social media linguistic patterns to real-world public health trends, highlighting the potential of prompt-based LLMs in identifying critical online discussion patterns that can form the basis of public health communication strategies.



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642117>

CCS CONCEPTS

• Human-centered computing → Empirical studies in HCI; • Social Media/Online Communities; • Natural language processing → Large Language Models; • Quantitative Methods;

KEYWORDS

large language models, prompting, social media, online communities, health decisions, health outcomes, covid-19, pandemic

ACM Reference Format:

Xiaohan Ding, Buse Carik, Uma Sushmitha Gunturi, Valerie Reyna, and Eugenia H. Rho. 2024. Leveraging Prompt-Based Large Language Models: Predicting Pandemic Health Decisions and Outcomes Through Social Media Language. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3613904.3642117>

1 INTRODUCTION

During the COVID-19 pandemic, social media was at the center of proliferating mass antipathy and distrust towards government health policies and recommendations [26, 55]. Millions took to social media to oppose federal and state health practices, criticize medical professionals, or organize anti-vaccine and mask-wearing rallies [62]. The viral growth of such online conversations fueled animosity and extremist views that encouraged people to resist public health guidelines [2, 26]. Disregarding public health practices, such as wearing masks, maintaining social distance, and getting vaccinated resulted in significant societal costs. Between November and December of 2021 alone, over 692,000 preventable hospitalizations were reported among unvaccinated patients, leading to a staggering \$13.8 billion [31]. Soaring COVID-19 infection cases put a massive burden on healthcare systems, depleting medical resources and contributing to severe employee burn-outs and shortages of healthcare workers [53]. Meanwhile, COVID-19 conspiracies and hyper-partisan news on social media led to nationwide protests, obstruction of medical facilities [6], and even fatal assaults of employees requesting customers to wear masks [8].

According to fuzzy-trace theory (FTT), texts that clearly establish cause-and-effect relationships facilitate humans extraction of *gist* mental representations, helping people understand and remember information better than texts without any causal coherence [83, 87]. This aligns with previous studies in decision sciences, which have shown that causal coherence of *gists* in texts plays a crucial role in how individuals perceive risks and make health-related decisions [24, 88]. Throughout the pandemic, social media conversations refuting COVID-19 public health practices based on mis-/disinformation and identity politics continued to obscure people's knowledge of safe health practices, making well-informed health decisions extremely difficult [113]. Using evidence-based theories like FTT allows us to create psychologically descriptive models that transform language into analyzable units shown to predict human behavior [24, 85].

In this paper, we leverage the capabilities of prompt-based Large Language Models (LLMs) to delve into the nuanced language patterns in social media discussions opposing COVID-19 public health practices through a theory-driven approach using fuzzy-trace theory. By leveraging prompt-based LLMs, we dissect the language around resistance to COVID-19 health practices through the lens of FTT and its central concept of *gist*. Specifically, we examine how causal language patterns or *gists* manifest across social media communities that denounce pandemic health practices, contribute

to trends in people's health decisions, and by extension, impact national health outcomes. We divide our work into four main studies to address the following research questions:

- RQ1. How can we efficiently predict gists across social media discourse at-scale? (Study 1)
- RQ2. What kind of gists characterize how and why people oppose COVID-19 public health practices? How do these gists evolve over time across key events? (Study 2)
- RQ3. Do gist patterns significantly predict patterns in online engagement across users in banned subreddits that oppose COVID-19 health practices? (Study 3)
- RQ4. Do gist patterns significantly predict trends in national health outcomes? (Study 4)

We answer RQ1 by leveraging LLMs and their prompt-based capabilities to identify gists in social media conversations at-scale (Study 1). We do so by developing a novel prompting framework that detects and extracts cause-effect pairs in sentences from a corpus of online discussions collected from banned Reddit communities known for opposing public COVID-19 health practices. Study 1 allows us to identify the causal language (cause-effect pairs that form *gists*) that underlie how people argue against COVID-19 health practices on social media. We answer RQ2 by clustering sentence embeddings of gists (sentences with causal relations identified from Study 1) to identify the most salient gist clusters, and demonstrate how they evolve across key events (Study 2). Finally, we answer RQs 3 and 4 by using Granger-causality to test whether causal discourse (*gists*) on social media can significantly predict online engagement patterns (Study 3), and trends in national health decisions and outcomes in the U.S. (Study 4).

Contributions: This work's intellectual merits are methodological and theoretical. The computational techniques introduced in this work enable efficient and scaled prediction of gists on social media, and thus can be used to better identify and understand underlying mental representations that motivate health decisions and attitudes towards public health practices (Study 1). The clustering and evolution of gists in Study 2 identify the most salient themes associated with how people causally argue against pandemic health practices online. Patterns in gist volumes across cluster topics fluctuate closely with topically-related high-profile events, including federal health announcements, congressional policies, and remarks by a country's leader. Study 3 empirically confirms how gist volumes significantly drive subreddit engagement patterns (upvotes and comments), providing implications for how causal language may play a role in monitoring conversations in content-moderation practices of controversial online health communities. Finally, gist patterns within subreddits that support anti-pandemic health practices were significantly interrelated with nationwide trends in important health decisions and outcomes (Study 4). To the best of our knowledge, our research is the first to empirically establish Granger causality between linguistic patterns in social media discussions about COVID-19 health measures and real-world trends in public health outcomes. Our work entails the following contributions:

- The task of accurately predicting causal language patterns and generating coherent gists (causal statements) is a complex challenge [65, 86]. We overcome this by introducing

a multi-step prompting framework: Role-Based Incremental Coaching (RBIC). RBIC is a prompting mechanism that allows efficient prediction of gists across social media conversations at-scale. RBIC integrates role-based cognition with effective learning in sub-tasks to enhance the model’s overall understanding of a given task prior to generating a final output. We overcome prior challenges in detecting subtle and complex expressions of semantic causality in noisy text by leveraging RBIC. By doing so, this work advances state-of-the-art approaches in detecting gists at-scale, yielding a novel, psychologically relevant, and efficient technique for identifying and examining bottom-line meanings in massive amounts of textual data.

- We demonstrate the novel application of prompt-based LLMs in advancing computational social science (CSS) methods in Human-Computer Interaction (HCI) research. Generic Natural Language Processing (NLP) models and LLMs typically lack multi-step reasoning capabilities [116]. This limitation makes it difficult to apply such models in performing nuanced and complex text analyses in CSS research [123]. By applying RBIC, we overcome this limitation and demonstrate the versatility and effectiveness of prompt-based LLMs in identifying and synthesizing nuanced linguistic patterns. In so doing, we contribute to broadening the potential application of prompt-based LLMs for theory-driven textual analysis in CSS research in the HCI domain.
- Our research enhances the analytical depth and scope of insights into the causal discourse surrounding people’s opposition to public health practices on social media. We identify the most salient gist clusters that embody the core topics at the center of how and why people oppose public health practices throughout COVID-19, from May 2020 to October 2021. We use sentence embeddings and clustering to provide a characterization of how the volume of gists across each topic fluctuates in relation to key events associated with the core topics embodied by the gist clusters. By doing so, we capture how causal online discourse surrounding anti-COVID-19 health practices evolves over time across real-world events. Such insights can, in turn, inform timely public health communication strategies and interventions that account for ongoing current events [85].
- Finally, we address the question of whether and how social media language patterns in the form of gists influence nationwide trends in vaccinations, COVID-19 cases, and hospitalization in the U.S., providing new evidence around how important health decisions and national health outcomes are impacted by causal linguistic signatures across social media health discussions—an important link that has not been empirically established at-scale in prior research.

2 RELATED WORK

2.1 Understanding the Impact of Social Media Language Patterns on Health Decisions and Outcomes

The COVID-19 pandemic has ignited an unprecedented increase in social media discourse on health decisions and practices [113, 122],

spurring a wave of computational social science research [39, 104] aimed at understanding this phenomenon in the field of HCI [63, 77] and CSCW [16]. Using text mining and computational linguistics, researchers have analyzed pandemic-related social media discourse through the lens of mental health [75], political views [17, 90], attitudes towards vaccines [79, 119], misinformation [49, 73, 99], and perceptions of health policies and government institutions [41]. Such studies have uncovered key insights on how language patterns reflect people’s beliefs [109], sentiments [54], and emotional well-being [10, 118] during Covid-19. For example, researchers have examined collective shifts in the public mood in response to the evolving pandemic news cycles by analyzing the daily sentiment of tweets [105]. Similarly, others have analyzed social media posts containing a subset of depression-indicative n-grams to track the fluctuation in mental health of social media users over the course of the pandemic [39].

While such studies have made valuable contributions to understanding the role of language patterns in health-related discourse on social media [9, 30], there remains an opportunity to explore their impact on real-world health decisions and outcomes. To the best of our knowledge, there has been a lack of research that examines how social media discussion patterns surrounding health practices can predict patterns in health decisions and outcomes in the real world. Our research aims to fill this gap. Some emerging research, such as the study by Nyawa et al. (2022), has started to explore this link by applying computational linguistics to categorize individuals as either vaccine-accepting or vaccine-hesitant based on their online language patterns [71]. Yet, the majority of empirical studies examining the impact of social media discourse on real-world behavior thus far have leaned heavily on survey-based methods [78, 118]. These surveys often depend on self-reported metrics about social media use and health behaviors, thereby offering only a limited perspective on the complex relationship between social media discourse patterns and actual health decisions. This limitation underscores the existing challenges in understanding how health-related discussions on the internet translate into or shape real-world outcomes and decisions [7]. Our research aims to address this challenge by investigating how language patterns in social media conversations can serve as predictive markers for understanding real-world trends in people’s health decisions and outcomes during the pandemic.

2.2 Understanding Health Discourse Through the Lens of Fuzzy-Trace Theory and Its Core Concept of Gist

Scholars have used fuzzy-trace theory (FTT) as a theoretical lens to explore risk perceptions and decisions underlying health practices and discussions in various contexts, including vaccines [113], cancer [115], HIV-AIDS [114] and the prescription of antibiotics [52]. These studies support FTT’s core tenet that gists are stronger and more effective forms of communication than verbatim representations in the sense that they are (a) better remembered and (b) more likely to influence decisions [83, 87]. For example, a study comparing articles on vaccines posted on Facebook showed that those containing gists (e.g., bottom-line meaning) are shared 2.4 times more often on average than articles with verbatim details

(e.g., statistics) [15]. Having a story or images did not add unique variance to predictions once gist was accounted for. The study's results show that communications about vaccines are more widespread when they express a clear gist explaining the bottom-line meaning of the statistics rather than just the data themselves. Likewise, scholars have also used FTT as a theoretical framework to examine people's behavior across diverse contexts, such as law, medicine, public health, systems engineering, and HCI [61, 88, 124]. For example, in HCI, researchers have used FTT to examine people's behavior in online social tagging [93] and to improve speech-to-text interface design through gist-based communication [65]. Others have used FTT in designing a web-based intelligent tutoring system for communicating the genetic risk of breast cancer through gists [115]. Overall, FTT's theoretical breadth and empirical support as a cognitive explanation of how people process and communicate information related to health decisions makes FTT a well-suited theoretical lens to examine resistance towards public health practices in our research.

Further, gists that causally link some event, actor, or outcome tend to facilitate more effective uptake of information than those that are less causally coherent [57, 85]. In fact, causal coherence is one of the most important semantic aspects of gists that make gist-based communications effective [40]. For example, in a study analyzing 9,845 vaccine-related tweets, researchers discovered that tweets containing explicitly causal gists (e.g., "vaccines cause autism") were far more likely to be retweeted and to go viral. This was in contrast to tweets that suggested a link between vaccines and autism but emphasized details and lacked a meaningful causal connection [15]. Simply, information with stronger causal structure produces more meaningful gists in people, who then are more likely to remember, apply, and share that information [86]. Fuzzy-trace theory draws on psycholinguistic research on mental representations of narratives that underlies both human memory models and computational models in which causal connections are a central feature of common gists [89, 103]. Hence, we focus on causal gists, or gists that contain a cause-effect relation. From hereon, we refer to causal gists as gists.

2.3 Challenges in Predicting Semantic Causality in Online Health Discourse

Extracting cause-effect relations in text is one of the many open challenges in NLP research that has seen significant breakthroughs in recent years through the development of generative Large Language Models [120]. However, computational social science research has yet to take advantage of these advancements [123], particularly in examining gists related to health practices. For example, scholars have used topic modeling, such as Latent Dirichlet Allocation (LDA) [11] to identify gists in vaccine hesitancy [40]. While useful, these methods do not enable granular detection of gists at the sentence or phrase level. For instance, LDA only allows the detection of gists at the corpus level, where each identified topic across the entire dataset is treated as a proxy identification of one gist. Recent scholarship in medical informatics has examined health-related attitudes in social media by extracting causality through machine learning approaches with rule-based dependency parsing and named entity

recognition [19, 29, 67, 68, 80]. While such approaches are an improvement, they can only detect intra-sentential (within a single sentence) and not inter-sentential causality where cause and effect lie in different sentences (e.g., God made us to breathe naturally. I won't be forced to wear masks.). More recently, transformer models such as InferBERT and CausalBERT, specifically designed for extracting causal relationships, have yielded more promising results [50, 111]. However, the token limit of these models significantly reduces performance when dealing with longer texts [4]. Additionally, like humans, these models struggle to discern subtle forms of semantic causality in noisy or incoherent data. Our research aims to not only identify causality in text, but also generate coherent gists based on the identified cause-effect pairs. To achieve this, we address prior limitations by leveraging recent advancements in pretrained LLMs and their prompt-based approaches to develop a novel prompting framework to systematically predict gists [112].

3 STUDY 1: PREDICTING GISTS IN SOCIAL MEDIA CONVERSATIONS AT-SCALE

As a first step to analyzing how causal language patterns on social media impact health decisions and outcomes, we leverage the power of prompt-based LLMs in Study 1. Specifically, we develop and apply a multi-step prompting framework called *Role-Based Incremental Coaching (RBIC)* to **efficiently predict gists across social media discourse at-scale**. Role-Based Incremental Coaching is a prompting framework (Fig. 2) built with few-shot demonstrations using GPT-4, which consists of two primary prompting techniques: *Role-Based Knowledge Generation* and *Incremental Coaching*. Combined together, RBIC allows the model to (1) learn its role for a given task by generating role-specific knowledge as a task-performing agent and (2) perform a series of small sub-tasks to refine its understanding and quality of the final output by incrementally building upon the sub-task responses. RBIC allows us to systematically identify the presence of semantic causality in a given post, and generate causally coherent gists across large volumes of textual corpora at-scale.

3.1 Data

We collected all publicly available posts from 20 anti-COVID-19 subreddits that were banned for denouncing COVID-19 public health practices. These subreddits were chosen based on their community size, as well as the significant media attention they received from major news outlets [44], and their virality among American social media users [43]. We obtained all posts and corresponding metadata (comments, post id, timestamp, up/down-vote ratio, etc.) for each of these subreddits using Pushift API. This resulted in a total of 79,680 posts spanning from May 2020 to October 2021 from the following subreddits: *conspiracy_commons*, *CoronavirusCirclejerk*, *CoronavirusFOS*, *Coronavirus_Rights*, *COVID19*, *covid19_testimonials*, *covidvaccinateduncut*, *VaxKampf*, *DebateVaccines*, *FauciForPrison*, *ivermectin*, *lockdownskepticism*, *NoNewNormal*, *trueantivaccination*, *vaccinelonghaulers*, *VAERSreports*, *Wuhan_Flu*, *CovidIsAFraud*, *COVID19Origin*, *churchofcovid*.

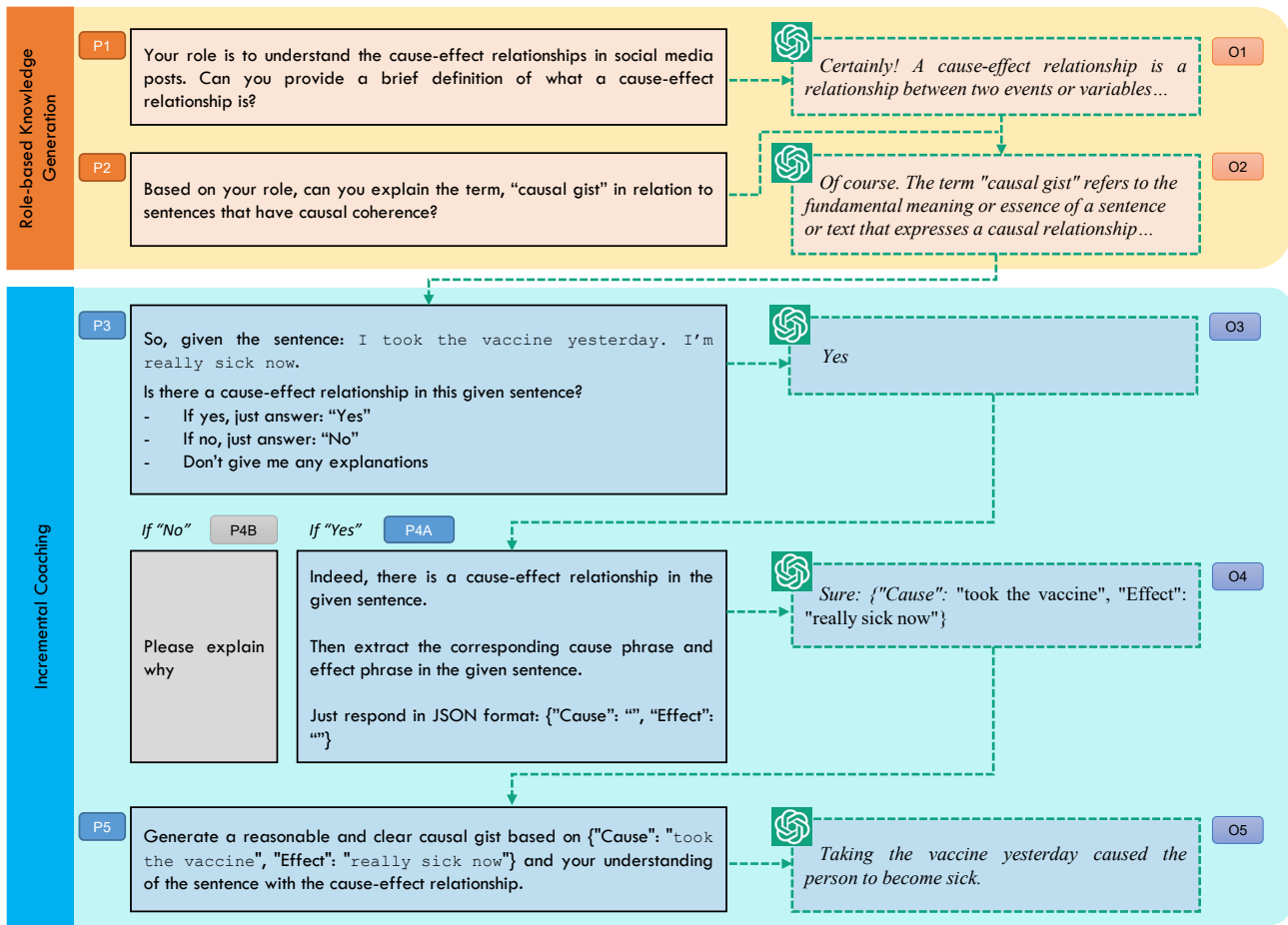


Figure 2: Illustration of the Role-Based Incremental Coaching (RBIC) prompting framework: RBIC incorporates role-based cognition and sub-task training to improve the model’s comprehension of a specific task before generating the final output.

3.2 Method: Role-Based Incremental Coaching (RBIC)

Role-Based Knowledge Generation. Drawing inspiration from prior NLP research that leverages multi-step reasoning capabilities in LLMs [58], we developed Role-Based Knowledge Generation as the initial grounding component of our prompting framework. Before producing a final response from LLMs, asking LLMs to generate potentially useful information about a given task improves the final response [58]. For example, as shown in an open online course "Learn Prompting" ¹, when prompted with "Which country is larger, Congo or South Africa?", GPT-3 answers incorrectly. However, when the model is prompted to "Generate some knowledge about the sizes of South Africa and Congo", before answering the final question, the model uses the output to the intermediate prompt ("South Africa [has] an area of...") to generate the correct answer: Congo is larger than South Africa. We leverage this prompting intuition in Role-Based Knowledge Generation to enhance the model’s understanding of its role as a task-performing agent. By

¹<https://learnprompting.org/>

doing so, the model can achieve better performance by accessing potentially relevant contextual information, as shown in prompts, P1 and P2 (Fig. 2). The corresponding outputs to P1 and P2 - O1 and O2, respectively - are then integrated with a task-specific prompt (P3) in the following step. The role-based knowledge outputs (O1 and O2) allow the model to perform tasks more accurately given its enhanced understanding of its specific role for achieving the task.

Incremental Coaching. Inspired by Chain of Thought (CoT) [112], Incremental Coaching is a technique within the Role-Based Incremental Coaching (RBIC) framework that involves breaking down a complex task into smaller, manageable sub-tasks as shown in P3-P5 in Fig. 2. The role-based agent is coached through a series of sub-tasks in a step-by-step manner, with each sub-task building upon the previous one. To implement Incremental Coaching effectively within RBIC, it is necessary to follow a logical sequence of sub-task prompts that allows the model to build understanding and confidence in performing the final task by generating incremental outputs (O3-O4). By breaking down the final task into a series of incremental sub-tasks, the role-based agent can gradually improve

its comprehension of the final task to deliver a more accurate final response.

Algorithm 1 Role-Based Incremental Coaching (RBIC)

Require: User Input (P), Role-Based Agent (RBA)
Ensure: Knowledge Base (KB) Final Task Output (F)

- 1: **Comment: Role-Based Knowledge Generation**
- 2: **Input:** User Input P
- 3: **Output:** Intermediate knowledge K
- 4: $K \leftarrow \text{RBA.GenerateKnowledge}(P)$
- 5: **Update** Knowledge Base (KB) with K
- 6: **Comment: Incremental Coaching**
- 7: **Input:** Final Complex Task T , Sequence of Sub-Tasks: $T = \{S_1, S_2, \dots, S_n\}$
- 8: **Output:** Incremental outputs $O_i = \{O_1, O_2, \dots, O_n\}$ and Final Task Output F
- 9: **for** $i = 1$ **to** n **do**
- 10: $O_i \leftarrow \text{RBA.Coach}(S_i, \text{KB})$
- 11: **Update** KB with O_i
- 12: **end for**
- 13: $F \leftarrow \text{RBA.FinalOutput}(\text{KB}, T)$
- 14: **return** F

Application of RBIC. Here, we demonstrate the algorithmic conceptualization of the RBIC prompting framework in the context of generating gists. The essence of the Role-Based Incremental Coaching (RBIC) framework lies in its two core algorithmic components: Role-Based Knowledge Generation and Incremental Coaching, as shown in Algorithm 1. The RBIC algorithm requires the following inputs:

- User Input P : The RBIC is initialized by the user input. For example, in our study, we operationalized user input as $P = (P1, P2, P3, P4A \text{ or } P4B, P5)$, as shown in Fig.2.
- Role-Based Agent: Essentially, this can be any prompt-based LLM. For our study, we used GPT-4 as our Role-Based Agent.

Next, the RBIC algorithm will generate the following output:

- Knowledge Base (KB): The first phase of the RBIC algorithm, denoted as Role-Based Knowledge Generation, is symbolized by the function $\text{RBA.GenerateKnowledge}(P)$. In this step, the Role-Based Agent (in our case, we use GPT-4, but this can be substituted with any prompt-based LLMs) is prompted with a user input P to elicit relevant background knowledge K . This knowledge forms the basis for task execution and is stored in an initial Knowledge Base (KB).

$$K \leftarrow \text{RBA.GenerateKnowledge}(P) \quad (1)$$

Here, K represents the knowledge generated, and P represents the user input posed by the user. \leftarrow signifies the assignment of generated knowledge K to the Knowledge Base (KB), thus creating a dynamic knowledge architecture that adapts over time. For instance, in our study, K comprised of $O1$ and $O2$ (as shown in the upper right of Fig. 2), which collectively formed our Knowledge Base (KB).

- Final Task Output (F): The subsequent phase, known as Incremental Coaching, is predicated on a sequence of sub-tasks $\{S_1, S_2, \dots, S_n\}$ and their corresponding outputs:

$$\{O_1, O_2, \dots, O_n\}.$$

$$O_i \leftarrow \text{RBA.Coach}(S_i, \text{KB}) \quad (2)$$

In this phase, each sub-task S_i leverages the updated Knowledge Base (KB) to produce an output O_i . O_i is then used to update the KB, thus iteratively coaching the model through a series of sub-tasks in a step-by-step manner. Breaking down the final Complex Task T into simpler sub-tasks S_i allows the model to incrementally build up the necessary knowledge and skills to tackle the final task. Therefore, this incremental knowledge building across sub-tasks enables the model to better understand and perform the final Complex Task T . In our case, our Complex Task (T) generates a "gist" based on the cause-effect pairs. The individual sub-tasks that contribute to this complex task are labeled as P3, P4A, P4B and P5 (Fig. 2). The algorithm proceeds sequentially, producing intermediate outputs O3, O4, and ultimately culminating in O5, which is the gist generated from the cause and effect pairs identified in sub-task P4A.

When applied to predicting gists in social media conversations, RBIC instructs the model to understand the concept of cause-effect relations as a task-performing agent. The model then incrementally performs sub-tasks to recognize and extract cause-effect pairs, and finally generates a concise gist that captures the essence of the identified causal relationship.

3.2.1 Human Evaluation. To assess the effectiveness of RBIC's application in predicting gists in our data, we conducted a human evaluation of the RBIC-generated outputs. We recruited 6 human evaluators to evaluate the presence of causal coherence (O3), cause-effect pairs (O4), and gists (O5) for each Reddit post based on the following criteria:

- **Accuracy (classification):** Is there a cause-effect relationship in the post (1/0; Yes/No)?
- **Relevance (extraction):** How well does the cause-effect pair capture the primary causal relationship in the post (1-5; not well at all, slightly well, moderately well, very well, extremely well)?
- **Conciseness (generation):** How well does the gist concisely summarize the cause-effect relationship in the post (1-5; not well at all, slightly well, moderately well, very well, extremely well)?

To mitigate error propagation, the evaluation was designed as a sequential process but with checks for accuracy. First, evaluators focused on 'Accuracy', verifying the presence of a cause-effect relationship. Second, 'Relevance' was examined to ensure the identified cause-effect pairs accurately reflected the post's main causal relationship. The final and third evaluation stage, 'Conciseness', was only evaluated in posts that had already met the 'Accuracy' and 'Relevance' criteria. This approach minimized propagation of errors from earlier stages.

The accuracy criteria assesses the model's performance in identifying the presence of a causal relationship in a post. Relevance evaluates the model's ability to correctly extract the cause and effect phrases that are most salient to the core message of the post's content. Conciseness assesses the model's generative performance in concisely synthesizing a coherent gist based on the identified

Table 1: Sample results from applying the RBIC method for extracting cause-effect relationships and generating gists from Reddit posts discussing health mandates between May 2020 and October 2021, including the post content

No.	Reddit Post	Label	Cause	Effect	Gist
1	99.995% of children survive cv infection, why are they pushing so hard to have kids take an experimental vaccine?	Yes	they are pushing so hard to have kids take an experimental vaccine	99.995% of children survive cv infection	Despite the high survival rate of children from cv infection, there is a push to have kids take an experimental vaccine.
2	Had my Pfizer jab last Wed and have felt like death since.	Yes	Had my Pfizer jab	have felt like death since	The cause of feeling like death is having the Pfizer jab last Wednesday.
3	Imagine pointing and laughing at a single father of 3 who's now jobless and has to take care of 3 kids with no income all because he didn't want to wear a face diaper or take the experimental gene modification.	Yes	he didn't want to wear a face diaper or take the experimental gene modification	he's now jobless and has to take care of 3 kids with no income	The man's refusal to wear a face mask or take the experimental gene modification led to him losing his job and being unable to provide for his three children, resulting in financial hardship and increased responsibility for him.
4	LA Fitness cancelled my membership against my will today because I refused to wear a mask.	Yes	I refused to wear a mask	LA Fitness cancelled my membership	The cause of LA Fitness cancelling the membership was the refusal of the person to wear a mask, which led to the effect of the membership being canceled against their will.
5	I've been thinking a lot about COVID data that's been circulating and want to share some thoughts. I think it's essential to remember that COVID data is not beyond skepticism, because what counts as a case varies.	Yes	what counts as a case varies	COVID data is not beyond skepticism	The variation in what is considered a COVID case has led to skepticism about the accuracy and reliability of COVID data.
6	I stumbled upon some news. Governor Wolf has a false positive, won't admit it because it would be admitting the tests are unreliable. And do you think it's possible that politicians might hide their own false positive results to maintain confidence in the testing system?	Yes	false positive, won't admit it	it would be admitting the tests are unreliable	Governor Wolf won't admit a false positive because it would undermine COVID-19 test reliability, potentially affecting public health and safety measures.

cause and effect phrases. In total, each of the 6 annotators evaluated 3,100 posts that were randomly selected from the entire dataset. For each criteria, each post received three evaluation scores from three annotators. The evaluators' assessment of the model's performance across the three criteria were generally high based on inter-rater agreement scores using Fleiss kappa (k) [32]: accuracy ($k = 0.892$); relevance (mean = 4.3, $k = 0.839$); conciseness (mean = 4.5, $k = 0.864$).

3.3 Result

Table 1 presents the results of RBIC's application, demonstrating the effectiveness of our prompting framework in predicting gists at-scale. We identified a total of 6,861 gists in our data. As shown, RBIC cannot only detect semantic causality (O3), but also extract verbatim phrases corresponding to the main cause-effect pairs (O4), and generate coherent gists (O5) based on the identified pairs. In the first example, RBIC detects sentences where causality is implied with nuance, as well as those that are more explicitly stated. Although most of the gists accurately capture the semantic essence of the causal relationship, some are more eloquent than others. For instance, the gists in examples 2 and 4 use sentence inversions, beginning with "the cause of", while others are more semantically fluid. We also performed a comparison using fine-tuned language models (BERT, RoBERTa and XLNet), as detailed in the appendix

(Table 6), which showed that RBIC outperformed the baseline models in extracting cause-effect pairs (O4) by 26.6% in F1-score when comparing RBIC to the best-performing baseline model (RoBERTa with 0.814 F1-score).

4 STUDY 2: HOW GISTS EVOLVE OVER TIME

Given the rapidly evolving public health discussions on social media, it is crucial to examine how they evolve over time [28, 38]. This enables a better understanding of shifts in public opinion and emerging concerns across contentious debates around health practices like vaccinations, mask-wearing, and social-distancing [33]. Hence in Study 2, we build upon our Study 1 findings to address: **What kind of gists characterize how and why people oppose COVID-19 public health practices? How do these gists evolve over time across key events?** To answer these questions, we extract sentence embeddings from each of the gists identified from Study 1, and cluster the embeddings to identify distinct gist clusters that characterize the core topics at the center of how people argue against COVID-19 health practices.

4.1 Method

4.1.1 Extracting Sentence Embeddings from Gists. To identify the most salient topics across the causal language (gists) surrounding

the social media discourse against public health practices, we use Sentence-BERT (S-BERT) to extract semantically rich representations of the gists identified in Study 1. S-BERT is a transformer-based model designed to produce contextualized sentence embeddings, which are particularly valuable in clustering texts [82, 101]. After preprocessing the gists with standard text cleaning operations (lowercasing, removal of special characters, tokenization), we implemented S-BERT using the SentenceTransformer to extract embeddings from our gists. The S-BERT model comprises 12 hidden layers, with each layer producing an output representation of $1(\mathcal{N}) \times 768(\mathcal{M})$ dimensions. To obtain high-quality embeddings, we extracted the output representations from each of the last three hidden layers of the model (layers 10-12), and computed their means. By doing so, we are able to capture and generate semantically rich representation of each gist as high dimensional vectors.

4.1.2 Clustering of Sentence Embeddings. After obtaining the sentence embeddings, we applied Principal Component Analysis (PCA) [13] to reduce the dimensionality of the embeddings prior to the clustering step. This was done to better visualize the language embeddings in a lower-dimensional space and to facilitate a more effective interpretation of the embedding results. We selected PCA as our method, given its frequent use and proven effectiveness in reducing dimensionality, especially for language embeddings [97]. We used k-means [69] for clustering, as it is especially reliable for clustering semantic word representations [121]. The k-means algorithm iteratively assigns each embedding to a cluster with the closest centroid, and updates the centroid by calculating the mean of the embeddings assigned to the cluster [121]. This process continues until the centroids stabilize. To enhance the reliability and robustness of our clustering approach, we incorporated sentence embeddings of posts that did not contain any gists, following Samosir’s study [92]. This step allows us to assess the quality of sentence embeddings by verifying that embeddings from sentences that do not contain gists cluster apart from embeddings derived from gists. Finally, we used the elbow method [70] to determine the optimal number of clusters by calculating the sum of squared errors (SSE) in ascending order of cluster numbers until additional clusters resulted in diminishing returns [64].

4.1.3 Verifying Gist Clusters. The first author initially identified the primary themes of each cluster through categorization, screening, and summarization of 200 randomly selected gists from each cluster. Next, we recruited 6 annotators to manually evaluate and verify five primary gist clusters, as shown in Table 3. Annotators manually evaluated the clustering results by iteratively examining and discussing the themes across 200 randomly selected gists belonging to each cluster (1/0). See annotation agreement in Table 2. The verification process also includes two additional steps: (1) refining cluster descriptions such that they were thematically salient and representative of the core ideas and topics embodied by the gists in each cluster and (2) examining the sentences in the non-gist cluster that did not include any gists (non-gist cluster is C6).

4.2 Result

A representative sample of gists from each cluster is presented in Table 3, illustrating the core topics that characterize the opposition

discourse surrounding pandemic health practices. In Fig. 3, (top panel), we visualize the evolution of our gist clusters across four time points (May 2020 - October 2021).

The visualization reveals interesting relations between the clusters. For instance, **cluster 4**, which embodies gists discussing the impact of COVID-19 on the economy and society at-large wraps around **cluster 3** gists related to the impact of lockdown policies. This spatial proximity suggests that causal discussions on the broader consequences of the pandemic are closely intertwined with gist-based conversations on the impact of lockdown measures, shedding light on the interconnectedness of how people talk about these two topics in a causal manner. Similarly, **clusters 1** and **2** are not only close in proximity, but also similar, in terms of position and shape: both clusters are diagonally positioned from top left to the bottom right, and run parallel to each other. Given that both clusters represent gists concerning specific health practices (vaccinations, mask-wearing), it is likely that these topics may share similarities in the causal manner in which people talk about the effectiveness of such health practices. **Cluster 5**, which represents gists related to conspiracy theories, domestic politics, and foreign countries appears to lack a clear boundary and is spatially dispersed compared to other clusters. This could be due to the fact that **cluster 5** encompasses multiple topics, as indicated by its description, in contrast to other clusters that are more uniformly focused on specific health practices, government measures, or particular aspects of the pandemic’s impact. The lower inter-rater reliability agreement for **cluster 5** (Fleiss $k = 0.821$) further supports the notion that it is a heterogeneous cluster consisting of various topics compared to other clusters.

The bottom half of Fig. 3 demonstrates that peak volumes of gists within each cluster align closely with key events related to the respective topics embodied by those clusters. To identify these key events, we relied on reports from major health organizations including the Centers for Disease Control and Prevention (CDC), World Health Organization (WHO), and United Nations (UN) for announcements related to public health interventions like lockdowns and vaccine rollouts [48, 107]. News reports from these organizations were widely recognized as authoritative information sources across the global community. Hence, we used such announcements and reports from these sources that highlighted key pandemic events, public health announcements, and significant milestones across the timeline of COVID-19.

In addition to these organization reports, we also analyzed articles related to COVID-19 published by major news outlets, such as AP News, Reuters, CNN, Fox News, Wall Street Journal, New York Times, and NPR. We then identified highly mediatized events by using the number of shares and article comments. This process also entailed iterative discussions among all the authors to ensure a comprehensive and balanced selection of events. Our approach aimed to minimize biases by incorporating a diverse range of sources and validating the significance of events through multiple indicators such as media coverage intensity and public engagement.

For example, **cluster 1** peak occurs in November 2020, coinciding with the country’s initial phase of vaccine distributions to healthcare workers and high-risk groups [48]. Similarly, **cluster 2** gists (mask-wearing) peaks in June 2021, the same month in which the federal mask mandate is lifted [107]. **Cluster 3** gists, which relate

Table 2: Inter-rater reliability scores for human evaluation of topic clustering results. Fleiss’ kappa coefficient was calculated to assess agreement between six annotators judging whether each gist was correctly assigned to one of the five clusters listed.

Cluster	IRR value (Fleiss <i>k</i>)
C1. Implications of Vaccine Policies, Efficacy, Side-Effects	0.926
C2. Controversies Related to Masks-Wearing Practices	0.894
C3. Impact of Lockdown	0.884
C4. Societal and Economic (Macro) Impact of COVID-19	0.902
C5. Conspiracy Theories, Domestic Politics, Foreign Countries	0.821
C6. Lack of distinct causal relationship or coherent gists	0.980

Table 3: Representative examples are shown for each cluster from C1 to C5, highlighting the main ideas identified in the health mandate debate on Reddit. Cluster 6 is not included, as it lacks a distinct causal relationship or coherent gist.

Cluster	Gist
C1. Implications of Vaccine Policies, Efficiency, Side-Effects	<ul style="list-style-type: none"> The implementation of a vaccine mandate has resulted in people losing their jobs. The use of experimental COVID vaccines is causing an increase in COVID deaths. The vaccine was ineffective against new variants, which led to the death of 7,000 people who received the spike protein mRNA jab, including little kids. This suggests that the vaccine was administered for no reason, as it failed to provide protection against the new variants.
C2. Controversies Related to Masks-Wearing Practices	<ul style="list-style-type: none"> If a person refuses to wear a mask at a business for medical reasons, the business may deny them services. The lifting of mask mandates for vaccinated individuals has caused the proliferation of a deadly biohazard, which could lead to the CDC and other agencies being charged with involuntary manslaughter. Wearing masks prevents people from seeing each other’s faces, which leads to difficulties in understanding and building trust with others.
C3. Impact of Lockdown	<ul style="list-style-type: none"> The lockdowns have caused tourism-dependent islands in Thailand to suffer from a lack of income, leading to a situation where they have been on food aid for over a year. The lockdowns caused a loved one to almost commit suicide, highlighting the negative impact of lockdowns on mental health. The prolonged lockdown imposed by Cuomo for six months has resulted in the inability of the speaker to pay their bills.
C4. Societal and Economic (Macro) Impact of COVID-19	<ul style="list-style-type: none"> The outbreak of COVID-19 has caused people to struggle with their livelihood, leading to financial difficulties and economic instability. The COVID-19 pandemic has caused the biggest drop in US life expectancy since the second world war. The COVID-19 shutdowns have resulted in 1 in 5 churches facing permanent closure within 18 months due to the financial strain caused by the pandemic.
C5. Conspiracy Theories, Domestic Politics, Foreign Countries	<ul style="list-style-type: none"> People refuse to share a table or work with certain people because they see "certain people" as sub-human because of their vaccination status. The sentence suggests that if COVID-19 was intentionally released, it would lead to a major benefit for China and billionaires. The implication is that the cause of COVID-19’s intentional release would be to bring about this benefit for these parties. The lack of information on the epidemic from people on whether they think something is safe or not is preventing the speaker from being able to debate with their conspiracy theory friends.

to the impact of lockdowns, peak in May 2020, by which approximately 4.2 billion or 54% of the world’s population was under lockdown [42]. In December 2020, the U.S. Congress passed a bill to distribute \$90 billion in stimulus checks to households, as nearly 30 million American adults reported food and income insecurity in the same month [106]. These events temporally coincide with the peak in cluster 4 gists, which concerns the socioeconomic consequences of the pandemic. Finally, cluster 5 gists, which are related to conspiracy theories, politics, and foreign countries, reached their peak volume in August 2020, around the time when President Trump

retweeted a popular online conspiracy theory [100] and referred to the "China virus" in his White House briefing [12]. Our findings imply that trends in gist volumes are linked with real-world events.

5 STUDY 3: HOW SOCIAL MEDIA GIST PATTERNS INFLUENCE ONLINE ENGAGEMENT BEHAVIOR

Delineating key semantic patterns (e.g., gists) that drive online behavior can help gain insight into how social media language impacts the dissemination of health information online. This, in turn, can

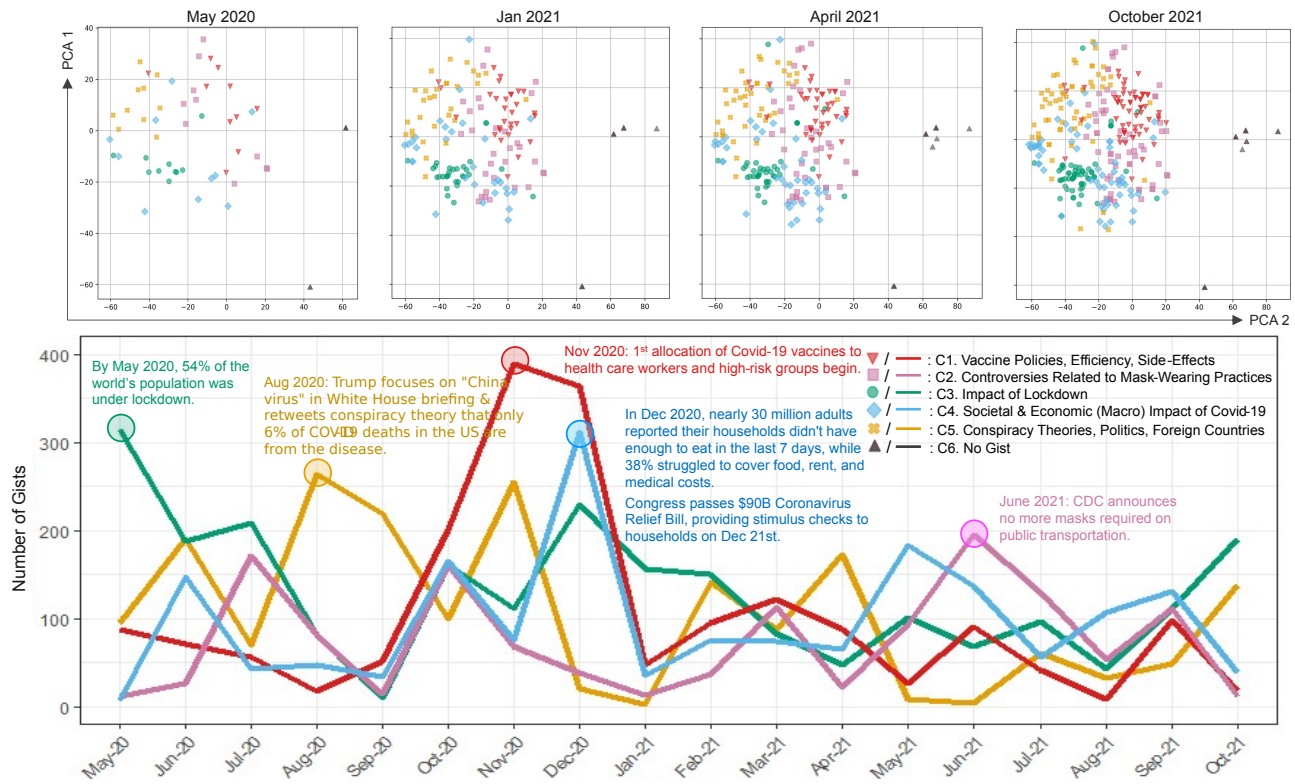


Figure 3: The upper portion of the illustration displays the progression of clusters across four-month periods. The line graph illustrates the month-by-month evolution of the number of posts containing gists, representing the central themes discussed on Reddit concerning health mandates. The graph highlights specific dates when each topic was most prominently discussed and presents relevant news events related to COVID-19 and health mandates during those periods.

better inform public communication strategies for time-sensitive health interventions. Hence, in Study 3, we use Granger-causality to **examine the extent to which gist patterns influence online engagement**, such as up-voting and commenting in subreddit communities that oppose COVID-19 health practices.

5.1 Hypothesis Testing with Granger Causality

Granger causality determines whether a time series X is meaningful in forecasting another time series Y [35]. For two aligned time series X and Y , it can be said that X Granger-causes Y if past values $X_{t-l} \in X$ lead to better predictions of the current $Y_t \in Y$ than do the past values $Y_{t-l} \in Y$ alone, where t is the time point and l is the lag time or the time interval unit in which changes in X are observed in Y . Lag time l in Granger causality refers to the delay between a change in one time series potentially causing a change in another, indicating the time it takes for the effect to be observed. We used Granger-causality to test hypotheses 1-2, as shown below, as well as the reversed variations of H1 and H2 (H1R and H2R) where i ranges from 1 to 5.

H1. The daily volume of gists in cluster i significantly Granger-causes the upvote ratio of Reddit posts containing gists in

cluster i .

H2. The daily volume of gists in cluster i significantly Granger-causes the number of comments associated with Reddit posts containing gists in cluster i .

5.2 Method and Analysis

First, we constructed the time series data, \mathcal{T}_G for each cluster, where \mathcal{T}_G represents the daily number of gists in **cluster i** spanning from May 2020 to October 2021. We then created two more temporally corresponding time series data, \mathcal{T}_U and \mathcal{T}_C , which represent the daily upvote ratio and the daily comment count for each Reddit post containing gists from cluster i , respectively. We conducted a total of 20 Granger causality tests (5 clusters \times 4 hypotheses - H1, H2, H1R, H2R), using time lags ranging from 1 to 14 days. To ensure that the value of the time series was not merely a function of time, we conducted the Augmented Dickey-Fuller (ADF) test [21] using the serial difference method to achieve stationarity with ADF test values exceeding the 5% threshold.

Table 4: Granger causality test results analyzing the relationships between the daily volume of gists in Clusters (C1-C5) and online engagement behavior - upvote ratios (UR), number of comments (NC) - across Reddit discussions.

Cluster	Behavior	Hypotheses	Lag (days)	F-value	p-value
C1. Vaccination Implications	upvote ratio (UR)	C1 \nrightarrow UR	4	4.228	0.022
	upvote ratio (UR)	UR \nrightarrow C1	3	1.739	0.277
	number of comments (NC)	C1 \nrightarrow NC	4	3.090	0.032
	number of comments (NC)	NC \nrightarrow C1	7	1.107	0.445
C2. Controversies and Policies Related to Masks	upvote ratio (UR)	C2 \nrightarrow UR	6	5.818	0.012
	number of comments (NC)	C2 \nrightarrow NC	7	6.007	0.014
	upvote ratio (UR)	UR \nrightarrow C2	6	4.738	0.022
	number of comments (NC)	NC \nrightarrow C2	9	0.771	0.463
C3. Impact of Lockdown	upvote ratio (UR)	C3 \nrightarrow UR	2	0.712	0.545
	upvote ratio (UR)	UR \nrightarrow C3	7	3.410	0.033
	number of comments (NC)	C3 \nrightarrow NC	3	0.715	0.496
	number of comments (NC)	NC \nrightarrow C3	8	6.121	0.016
C4. Societal and Economic (Macro) Impact of Covid-19	upvote ratio (UR)	C4 \nrightarrow UR	3	6.485	0.011
	upvote ratio (UR)	UR \nrightarrow C4	7	2.011	0.092
	number of comments (NC)	C4 \nrightarrow NC	2	7.912	0.002
	number of comments (NC)	NC \nrightarrow C4	4	0.815	0.413
C5. Social Issues, Health, and Personal Experiences	upvote ratio (UR)	C5 \nrightarrow UR	5	1.441	0.512
	upvote ratio (UR)	UR \nrightarrow C5	2	1.715	0.289
	number of comments (NC)	C5 \nrightarrow NC	4	2.412	0.089
	number of comments (NC)	NC \nrightarrow C5	4	1.135	0.530

5.3 Results

Table 4 shows significant Granger causal results ($p < 0.05$). Gists across certain topics are significantly predictive of up-voting and commenting patterns, and vice-versa, in banned subreddits that oppose pandemic health practices. Specifically, the daily volume of gists significantly forecasts up-voting and commenting behavior across the topic of vaccines (cluster 1), mask-wearing (cluster 2), and macro-impacts of the pandemic (cluster 4) with significant lag lengths ranging from 2-7 days. These results align with prior research highlighting the linguistic power of gists in spreading online information. The reverse (H1R and H2R) is true for gists discussing the impact of lockdowns (cluster 3): up-voting and commenting behavior both significantly forecast fluctuations in the volume of lockdown related gists.

5.3.1 Bidirectional Causality: Notably for cluster 2, which pertains to controversies and policies related to masks-wearing, we observe an interesting feedback loop between gist volumes and commenting behavior. As the volume of gists related to mask-wearing practices increases, corresponding online engagement around posts containing such gists, also increases in the form of up-votes. This behavior, in turn, further influences the volume of gists that are topically related to mask-wearing practices. In other words, there is a mutually reinforcing effect between causal language and online behavior in the context of mask-related discussions.

6 STUDY 4: HOW SOCIAL MEDIA GIST PATTERNS INFLUENCE NATIONWIDE TRENDS IN HEALTH OUTCOMES

In Study 4, we address the question of **whether and how social media language patterns in the form of gists influence health**

decisions and outcomes in the U.S. We follow Study 3’s application of Granger causality to examine the relationship between gists patterns and important health decisions and outcomes related to COVID-19 in America. Considering the extensive attention the subreddits we analyzed received from the American public and the media [43], we focus on U.S. health outcomes.

6.1 COVID-19 Data on Health Outcomes

We used the following data from Our World in Data ², a trusted source for COVID-19 health data for our analysis:

- **Number of Vaccinations (NV):** the total number of COVID-19 vaccine doses administered on a given day.
- **General Hospitalization (GH):** the number of individuals hospitalized due to COVID-19 on a given day.
- **ICU Hospitalization (ICU):** the number of patients with COVID-19 who are in the ICU on a given day.
- **Total Daily COVID-19 Cases (TC):** the total number of confirmed COVID-19 cases, including probable cases.
- **New Daily COVID-19 Cases (NC):** the number of newly confirmed COVID-19 cases, including probable cases.

6.2 Hypothesis Testing with Granger Causality

Following Study 3, we Granger-test the relationship between the daily volume of gists and patterns in people’s health decisions (vaccinations) and national health outcomes (General/ ICU Hospitalization, Total/ New Daily COVID-19 Cases) through H3 and its reversed variation (H3R):

²<https://github.com/owid/covid-19-data/tree/master/public/data>

Table 5: Result of Granger causality test for relationships between Reddit discussion clusters (C1-C4) and health outcomes dataset. Cluster 5 is not included due to the absence of significant Granger causality findings. Notes: See Appendix A for complete statistical results (Table 7 and 8).

Cluster	Health Outcomes (HO)	Hypotheses	Lag	F-val	p-val	Direction of Impact
C1. Vaccination Implications	Number of Vaccinations	C1 \nrightarrow NV	4	4.778	0.027	bidirectional
	Number of Vaccinations	NV \nrightarrow C1	14	7.771	0.007	
C2. Controversies and Policies Related to Masks	Total Daily Covid Cases	C2 \nrightarrow TC	5	9.395	0.005	gists impacts HO
	New Daily Covid Cases	C2 \nrightarrow NC	5	11.829	0.004	
C3. Impact of Lockdown	ICU Hospitalization	ICU \nrightarrow C3	14	4.000	0.039	HO impacts gists
	General Hospitalization	GH \nrightarrow C3	9	8.822	0.006	
C4. Societal and Economic (Macro) Impact of COVID-19	Total Daily Covid Cases	TC \nrightarrow C4	9	3.663	0.031	HO impacts gists
	New Daily Covid Cases	NC \nrightarrow C4	9	3.092	0.032	

H3. The daily frequency of gists (Cluster i) significantly Granger-causes people’s health decisions and/or national health outcomes, where i ranges from 1 to 5.

H3R. People’s health decisions and/or national health outcomes significantly Granger-causes the daily frequency of gists (Cluster i), where i ranges from 1 to 5.

We created five time series data, \mathcal{T}_{NV} , \mathcal{T}_{GH} , \mathcal{T}_{ICU} , \mathcal{T}_{TC} , \mathcal{T}_{NC} , corresponding to the five health outcome data described above. We temporally align our data with the time frame for Studies 1-3. We performed 25 Granger causality tests (5 clusters \times 5 health outcome data) with a range of lag times from 1 to 14 days. We conducted ADF tests using the serial difference method to ensure statistical robustness.

6.3 Results

Table 5 show shows significant Granger-causal results with corresponding lag lengths ($p < 0.05$). We summarize our findings below.

Causal Talk Around Vaccines and National Vaccination Trends are Bidirectional. Our results demonstrate bidirectional causality between causal discourse patterns related to vaccines and the number of vaccinations administered in the U.S. The daily volume of **cluster 1** gists, which consists of causal arguments related to vaccine regulations, efficacy, and side effects, is predictive of vaccination patterns across the U.S., and vice-versa. However, there is a difference in the lag lengths between H3 and H3R. It takes 4 days for gist patterns to influence vaccine adoptions (H3), while it takes two weeks for vaccination trends to shape how people talk about vaccine-related topics in a causal manner (H3R) across COVID-19 subreddits known for vaccine skepticism. In addition to a more significant Granger-causal relationship, we also observe a higher Pearson correlation for H3 ($r = 0.413$, $p = 0.005$) compared to H3R ($r = 0.105$, $p = 0.028$), indicating that national vaccination patterns have a greater impact on shaping vaccine-related causal language on social media than the other way around. There are two possible explanations: first, as more people get vaccinated, online discussions on the experiences and potential side effects of vaccines may become more prevalent - leading people to talk in a causal manner about the side effects of vaccines (e.g., "Had my Pfizer jab

last Wed and have felt like death since"). Another possible explanation is that the increasing vaccination requirements by corporations and governments as a condition for work or travel (and therefore, nationwide uptick in vaccinations) during the pandemic may have compelled vaccine-skeptics to argue more vehemently against vaccines [25]. Previous research has shown that vaccine skeptics are susceptible to confirmation bias, as are most individuals, such that initial beliefs lead to polarization [66]. That is, vaccine skeptics are likely to seek out and discuss information about vaccines that confirms pre-existing beliefs when presented with opposing information or situated in contexts that challenge their views. Our findings align with this research, suggesting that as national vaccination uptake increases, vaccine skeptics might increasingly argue against vaccines in a causal manner (e.g., "If you take the vaccine, it's probably because you're unhealthy."), as commonly expressed in posts that contain **cluster 1** gists.

Causal Talk Around Mask-Wearing Practices Significantly Predicts Trends in COVID-19 Cases. Our Granger-causal results show that national health outcomes, such as the total and new daily COVID-19 cases can be significantly predicted by the volume of mask-related gists (**cluster 2**) with a lag of 5 days. The mask mandate was one of the most controversial health practices that impacted people of all ages and occupations during the pandemic [62, 98]. Parents were polarized over school mask requirements to the extent of resorting to violence [94]. Employees who asked customers to wear masks were physically assaulted [8]. Although people initially adhered to wearing masks, more individuals started to protest mask mandates both on and offline, citing physical distress ("If having healthy lungs is important for COVID, why would we wear masks that reduce lung function?") or invasion of personal rights: "They will call you a 'coward' or 'scared' for not wanting an intrusive mask over your face (for no reason)", as exemplified by posts containing **cluster 2** gists in our data. Over time, the proliferation of anti-mask views, followed by extreme resistance as demonstrated by violent altercations and wide-scale protests across the nation, may have led people to abandon mask-wearing practices [36], which in turn may have led to an increase in COVID-19 cases within a relatively short time-frame of 5 days, as indicated in our results.

Rising Hospitalization Trends Prompt Causal Talk on Lockdown Impact. Our findings show that nationwide trends in the number of patients hospitalized in both general and intensive

care units significantly prompt more gists discussing the impact of lockdowns with a lag of 9 and 14 days, respectively (Table 5). Nationwide lockdowns were implemented to curb steep rises in COVID-19 cases and hospitalization rates. In fact, some posts containing **cluster 3** gists often explicitly link lockdowns with hospitalizations: *"The main reason for implementing restrictions or lockdowns was to prevent ICUs from overflowing."* Despite its necessity and intended benefit as a public health measure, studies have shown that lockdowns significantly contributed to social isolation, decrease in mental health, and rise in domestic violence across the U.S. [18]. As the lockdown continued to amplify challenges and problems in people's lives, rising hospitalization trends across the country may have heightened people's fear and distress, leading to more intensified and causal online discourse on the lockdown's impact on everyday life. Such sentiments are clearly expressed across posts containing **cluster 3** gists: *"People are literally starting to go hungry because of lockdown restrictions"; "The implementation of lockdowns has resulted in more harm than good."*

Rising Trends in COVID-19 Cases Prompt Causal Talk on the Pandemic's Macro-Level Impact. Nationwide trends in COVID-19 cases significantly Granger-causes the volume of gists discussing the pandemic's impact on society at large, with a lag of 9 days for both total and new cases. In other words, increasing trends in COVID-19 cases seem to nudge people to talk casually about the macro-level consequences of COVID-19. COVID-19 presented major economic and social setbacks that impacted all aspects of society. Some of these concerns were expressed across posts containing **cluster 4** gists that linked the pandemic with economic crises (*"The pandemic caused one of the largest economic crises, which in turn led to one of the largest poverty and hunger crises"*), decreased life expectancy (*"The COVID-19 pandemic has caused the biggest drop in US life expectancy since the second world war"*), potentially oppressive public health measures (*"The cause of the next deadly pandemic will lead to the implementation of authoritarian prevention measures"*), and even racism (*"The fact that Covid19 affects people of color more than whites is the cause of the conclusion that Covid19 is racist"*). With COVID-19 cases rising and situations continuing to remain unpredictable, people may have become more anxious and distressed about the long-term effects on society. Consequently, this may have led individuals to discuss the pandemic's impact in a causal manner on social media, as they try to make sense of its far-reaching consequences on society [84].

7 DISCUSSION

In summary, our findings underscore RBIC's effectiveness in efficiently predicting social media gists at scale (Study 1), thereby enriching our insight into the underlying mental constructs that shape people's health decisions and attitudes towards public health practices. In Study 2, we cluster and track the evolution of such gists, revealing key themes in online arguments against pandemic health practices. These gist volumes closely align with significant topical events, such as health announcements, policy changes, and leadership statements. In Study 3, we empirically demonstrate how gist volumes significantly drive subreddit engagement patterns (up-votes and comments). Finally, Study 4 reveals the interplay between

gist patterns in anti-Covid-19 subreddits and nationwide health trends. We discuss the implications of these findings below.

7.1 Harnessing Large Language Models in Computational Social Science (CSS) Research in HCI

Prompt-based LLMs are increasingly used in the CHI community [20, 56, 76, 110], primarily contributing to the development of applications like chatbots [46] and tools for co-writing [56], virtual simulations [108], story-telling [23], and visualization enhancement [96]. Such studies have primarily focused on using LLMs as *production* tools [56] rather than tools for analysis. More recently, computational social scientists in HCI have used prompt-based LLMs for text analyses [34, 102, 123]. However, there remain several challenges for using LLMs in nuanced examination of social media discourse. First, traditional NLP models and commonly used LLMs in CSS research often lack reasoning capabilities [116]. For instance, LLMs like BERT-based models, which are extensively used in HCI research that analyze large volumes of social media data [27, 60], are typically fine-tuned for specific discrete downstream tasks (e.g., classification). While these pretrained language models have shown promise in performing discrete analyses, some emerging HCI research [116, 117] demonstrate the additional value of prompting LLMs to perform multi-step reasoning for a more comprehensive analysis. Building on these prior insights, RBIC aims to enable a more nuanced analysis of social media discourse by leveraging the multi-reasoning capabilities of large language models. To this end, RBIC operates by performing multiple, step-by-step interrelated sub-tasks (*question-answering, classification, extraction, generation*) prior to generating its final output. This incremental coaching mechanism enhances the model's overall understanding and performance of the final task, allowing us to analyze social media discourse with a more comprehensive and nuanced approach.

Second, LLM development paradigms often incentivize researchers to optimize model performance using established evaluation datasets [20, 58]. While valuable for comparing an LLM's performance with other models, this approach may not result in high performance when applied to new, unseen, in-the-wild datasets [22, 123] or with tasks that are slightly different from those that the model was evaluated on [123]. As a result, this may limit the potential application of such LLMs for analyzing intricate, heterogeneous in-the-wild data, such as unstructured social media conversations. The role-based cognition component of RBIC addresses this limitation by allowing researchers to define and customize the role of any prompt-based LLM to perform a complex and nuanced language task. By introducing and applying RBIC in the analysis of social media conversations, we demonstrate the versatility and effectiveness of prompt-based LLMs in identifying and synthesizing nuanced linguistic patterns, thus broadening the potential application of prompt-based LLMs for theory-driven textual analysis in CSS research in the HCI domain.

7.2 Leveraging Causal Language Patterns in Online Content Moderation Practices

Our results show that the volume of gists across certain topics are significantly predictive of up-voting and commenting patterns, and

vice-versa, in banned subreddits that oppose pandemic health practices. For example, daily gist volumes significantly predict up-voting and commenting behavior across topics related to vaccines, masks, and the pandemic's impacts, highlighting the linguistic power of gists in spreading online information as demonstrated in prior literature [85, 86]. Similarly, our findings show that increasing trends in vaccine adoptions in the U.S. are strongly predictive of the growing volumes of vaccine-related gists in subreddits whose members are generally skeptical of vaccines. While a nationwide rise in vaccine uptake is certainly beneficial, such conditions may present challenging contexts that may reinforce vaccine skeptics to become further entrenched in their views. Vaccine opponents exposed to situations that contradict their perceptions are especially vulnerable to confirmation biases [5], which may lead to an increased tendency to express their anti-vaccine sentiments in online communities in a causal manner, as implied by our findings.

These insights underscore the critical role of understanding and monitoring causal language patterns in public health discourse, particularly within online spaces. Current content moderation practices that rely on language models traditionally focus on flagging hate speech or monitoring specific keywords [91]. However, our research suggests that monitoring causal language patterns can be a valuable addition to these content moderation practices, especially in controversial online communities where people exchange and learn health information. By leveraging nuanced insights from gists across various health topics, content moderation can become more effective in identifying and managing discussions that may contribute to the spread of online health misinformation or resistance to public health guidelines.

7.2.1 Design Implications for Moderation Dashboard: Prior studies have shown that the design of a social media platform plays an important role in promoting transparency in content moderation [47]. Moderators often fail to articulate what aspect of the content prompted moderation or why such moderation was necessary [47]. The approach taken in our study can be built on to effectively inform users about the consequences of their posting behavior, and which aspects of their posts can potentially lead to negative outcomes. The results can also inform design strategies that platforms can undertake to assist moderators in communicating such information to users.

Understanding and identifying causality can be difficult for humans as causality may be expressed implicitly and across sentences or intersententially [93]. Currently, there is no automated mechanism for moderators to systematically identify and understand the impact of causal language across online discussions. A design feature in the moderation dashboard, such as the one shown in Fig. 4 (Appendix D), serves as an illustrative example of how RBIC may address this gap. For example, when a moderator clicks on a button called 'Enable Gist Detection (RBIC)', an RBIC-powered extension can automatically scan posts, highlight the cause-and-effect pairs, and identify the overarching gists within the posts. This functionality may also allow moderators to see a list of top gists across community discussions in descending order of gist volumes, and an option to organize these gists based on engagement metrics, including the upvote ratio and comment volume. Additionally, the system may be designed such that the moderator may be able to

drill down into posts that pertain to each of these top gists, in which the system can highlight the relevant text spans that pertain to the cause and effect in each post.

7.2.2 Improving Moderation and Community Guidelines. Identifying posts that do not contain moderator-specified keywords (e.g., profanity) or those that exclude explicit causal language, can still violate community norms or include misleading information in subtle ways [72]. Traditional keyword-based filters fall short in identifying such content [45]. This can lead to difficulties in setting specific rules for moderation practices, explaining moderation-decisions, or adapting community guidelines during critical times, such as a global pandemic. With RBIC-powered gist detection, moderators can scale the searching of such posts to identify those that reflect common and theoretically predicted disconnects between the public and public health experts. This mechanism can potentially enable moderators to use concrete examples to better explain moderation decisions, as well as improve community guidelines to explain how posts that contain implicit causal narratives may impact people's knowledge and decisions around safe health practices, as shown in our work.

7.3 Broader Implications for Understanding Engagement Patterns Across Online Communities and Offline Health Outcomes During Public Health Crises

Our work shows that capturing psychologically important language patterns across social media, in the form of gists, can be useful in predicting human behavior and, consequently, health outcomes. In Study 2, we demonstrate that fluctuations in the volume of gists can significantly predict online engagement patterns, specifically in terms of up-vote ratio patterns (H1) and the volume of comments (H2). This has important implications for researchers studying user behaviors in online communities [37, 118]. Researchers have shown that the virality of online content is often influenced by a positivity bias in engagement metrics [51], such as up-votes and comments: posts receiving higher engagement are more visible and thus have a greater likelihood of going viral [3, 81]. This tendency can exacerbate the spread of misinformation, especially during public health crises [1, 95]. Posts challenging pandemic health practices are often laden with misleading information [55], and online posts embedded with gists are more likely to attract more user engagement compared to those without gists [14]. H1 and H2 results demonstrate that such user engagement patterns are predictable through gist volumes, thus highlighting the potential of using RBIC for gist analysis to track and understand the dynamics of how health-related content, especially during pandemics, resonates with and influences online user engagement. This insight is crucial for developing strategies to combat misinformation and guide public health communication effectively.

Furthermore, HCI research in crisis informatics has contributed to advancing public health monitoring systems by developing tools that track public health outcomes, online engagement patterns, or health-related topics on social media [59, 74]. Some of these tools that monitor online conversations extract various linguistic aspects from social media discourse, such as sentiment [59] and topical

keywords [104]. While these advancements have been valuable in providing descriptive insights, most do not go the full distance in linking such linguistic patterns to real-world health decisions and outcomes [55]. Our work addresses this gap by demonstrating how RBIC can be leveraged to better connect online conversation patterns to offline health outcome trends. Study 4 results show that online causal talk related to controversial health practices, such as face-masks, are significantly predictive of total and new daily COVID-19 cases across the U.S. Likewise, our findings show that the uncertainty arising from deteriorating trends in national health outcomes may prompt people to increasingly engage online in causal discussions on the pandemic's influence on their lives and society as a whole. For example, nationwide COVID-19 cases and hospitalization patterns significantly drive up the volume of gist-based conversations concerning the pandemic's impact on society, economy, and individuals under lockdown. These findings imply that integrating gist-based language patterns into public health monitoring systems can hold promise for gaining valuable insights into the cognition that underlies skepticism and resistance to public health practices and, by extension, their impact on real-world health outcomes. Integrating RBIC-powered gist detection and real-time analysis of national health indicators into tools can potentially enhance public health agencies' ability to understand and respond to critical health challenges in relation to people's online behavior.

8 CONCLUSION & LIMITATIONS

This research synthesizes LLM techniques with theoretical perspectives from cognitive and social psychology to advance the knowledge of health decisions and outcomes in the context of the most recent pandemic. Our work is the first to systematically identify and characterize how causal language patterns surrounding anti-pandemic health practices on social media are significantly predictive of national health outcomes. These findings carry crucial implications for public health communication and policy interventions. By recognizing the influential role of causal language patterns across social media in shaping national health outcomes, public health efforts and online moderation practices can be tailored to address and mitigate the impact of social media conversations that adversely affect public health consequences.

Our study has a limitation in our data source: it concentrates on Reddit posts and omits comments. This exclusion is primarily due to certain months of comment data being either restricted or deleted in compliance with Reddit's policies by Archive administrators. While this focus allows for an in-depth analysis of original posts, it may not capture the full discourse, including diverse viewpoints and nuanced discussions that often take place in the comments section. Consequently, our findings may offer a limited perspective on the topic under study. Future work might consider alternate ways to capture community discourse, such as through interviews or surveys, to complement the data from Reddit posts. Furthermore, as datasets from the future expand, integrating machine learning models that are capable of detecting subtle changes in discourse over time and adjust to extensive datasets may offer a dynamic view of how gists evolve. This method has the potential to uncover patterns and trends that may not be immediately obvious when using a traditional unsupervised clustering approach.

In summary, we built an LLM-based model to identify psychologically influential mental representations—gists—from social media posts, demonstrated the links between these gists and public health events, and verified associations with user engagement and national health trends, with implications for HCI design and the promotion of public health.

REFERENCES

- [1] Saifuddin Ahmed and Muhammad Ehab Rasul. 2022. Social Media News Use and COVID-19 Misinformation Engagement: Survey Study. *Journal of Medical Internet Research* 24 (2022). <https://api.semanticscholar.org/CorpusID:252109554>
- [2] M. Al-Ramahi, A. Elnoshokaty, O. El-Gayar, Tareq Nasrallah, and A. Wahbeh. 2020. Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data. *JMIR Public Health and Surveillance* 7 (2020). <https://doi.org/10.2196/26780>
- [3] Kholoud Khalil Aldous, Jisun An, and Bernard Jim Jansen. 2019. View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. In *International Conference on Web and Social Media*. <https://api.semanticscholar.org/CorpusID:189818831>
- [4] Wajid Ali, Wanli Zuo, Rahman Ali, Gohar Rahman, Xianglin Zuo, and Inam Ullah. 2022. Towards Improving Causality Mining using BERT with Multi-level Feature Networks. *KSI Transactions on Internet & Information Systems* 16, 10 (2022), 3230–3255. <https://doi.org/10.3837/tiis.2022.10.002>
- [5] Hossein Azarpanah, Mohsen Farhadloo, Rostam M. Vahidov, and Louise Pilot. 2021. Vaccine hesitancy: evidence from an adverse events following immunization database, and the role of cognitive biases. *BMC Public Health* 21 (2021). <https://api.semanticscholar.org/CorpusID:237522699>
- [6] David W. Baker. 2020. Trust in Health Care in the Time of COVID-19. *JAMA* 324, 23 (2020), 2373–2375. <https://doi.org/10.1001/jama.2020.23343>
- [7] Jay J. Van Bavel, Katherine Baicker, Paulo S. Boggio, and et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* 4, 5 (2020), 460–471. Number: 5 Publisher: Nature Publishing Group.
- [8] Abha Bhattarai. 2020. Retail workers are being pulled into the latest culture war: Getting customers to wear masks. *The Washington Post* (2020).
- [9] Laura Biester, Katie Matton, Janarthanan Rajendran, Emily Mower Provost, and Rada Mihalcea. 2021. Understanding the Impact of COVID-19 on Online Mental Health Forums. *ACM Trans. Manage. Inf. Syst.* 12, 4, Article 31 (sep 2021), 28 pages. <https://doi.org/10.1145/3458770>
- [10] Johnna Blair, Chi-Yang Hsu, and et al. 2021. Using Tweets to Assess Mental Well-Being of Essential Workers During the COVID-19 Pandemic. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI EA '21*). Association for Computing Machinery, New York, NY, USA, Article 236, 6 pages.
- [11] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 4 (2003), 993–1022.
- [12] James S. Brady. 2020. Remarks by President Trump in Press Briefing August 19, 2020. <https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-press-briefing-august-19-2020/> The White House, National Archives.
- [13] Rasmus Bro and Age K. Smilde. 2014. Principal component analysis. *Analytical Methods* 6, 9 (2014), 2812–2831. <https://doi.org/10.1039/C3AY41907J>
- [14] David Broniatowski and Valerie F. Reyna. 2018. Causal Gist Predicts Spread of Tweets about Vaccination. In *40th Annual Meeting of the Society for Medical Decision Making*. SMDM. <https://smdm.confex.com/smdm/2018/meetingapp.cgi/Paper/11774>
- [15] David Andre Broniatowski, Karen M. Hilyard, and Mark Dredze. 2016. Effective Vaccine Communication during the Disneyland Measles Outbreak. *Vaccine* 34, 28 (2016), 3225–3228. <https://doi.org/10.1016/j.vaccine.2016.04.044>
- [16] Clara Caldeira, Cleidson R.B. de Souza, Leticia Machado, Marcelo Perin, and Pernille Bjørn. 2023. Crisis Readiness: Revisiting the Distance Framework During the COVID-19 Pandemic. 32, 2 (2023), 237–273. <https://doi.org/10.1007/s10606-022-09427-6>
- [17] Dustin P. Calvillo, Bryan J. Ross, Ryan J. B. Garcia, Thomas J. Smelter, and Abraham M. Rutchick. 2020. Political Ideology Predicts Perceptions of the Threat of COVID-19 (and Susceptibility to Fake News About It). *Social Psychological and Personality Science* 11, 8 (2020), 1119–1128. <https://doi.org/10.1177/1948550620940539>
- [18] Clare E. B. Cannon, Regardt Ferreira, Frederick Buttell, and Jennifer First. 2021. COVID-19, Intimate Partner Violence, and Communication Ecologies. *American Behavioral Scientist* 65, 7 (2021), 992–1013. <https://doi.org/10.1177/0002764221992826>
- [19] Tommaso Caselli and Piek Vossen. 2017. The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction. In *Proceedings of the Events and Stories in the News Workshop*. Tommaso Caselli, Ben Miller, Marieke

- van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, Teruko Mitamura, and David Caswell (Eds.). Association for Computational Linguistics, Vancouver, Canada, 77–86. <https://doi.org/10.18653/v1/W17-2711>
- [20] Jiangjie Chen, Wei Shi, Ziquan Fu, Sijie Cheng, Lei Li, and Yanghua Xiao. 2023. Say What You Mean! Large Language Models Speak Too Positively about Negative Commonsense Knowledge. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Toronto, Canada, 9890–9908. <https://doi.org/10.18653/v1/2023.acl-long.550>
- [21] Yin-Wong Cheung and Kon S. Lai. 1995. Lag Order and Critical Values of the Augmented Dickey–Fuller Test. *Journal of Business & Economic Statistics* 13, 3 (July 1995), 277–280.
- [22] Hyung Won Chung, Le Hou, S. Longpre, and et al. 2022. Scaling Instruction-Finetuned Language Models. *ArXiv abs/2210.11416* (2022). <https://api.semanticscholar.org/CorpusID:253018554>
- [23] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. <https://doi.org/10.1145/3491102.3501819>
- [24] Jonathan C. Corbin, Valerie F. Reyna, Rebecca B. Weldon, and Charles J. Brainerd. 2015. How Reasoning, Judgment, and Decision Making are Colored by Gist-based Intuition: A Fuzzy-Trace Theory Approach. *Journal of Applied Research in Memory and Cognition* 4, 4 (2015), 344–355. <https://doi.org/10.1016/j.jarmac.2015.09.001>
- [25] Kali DeDominicis, Alison M. Buttenheim, Amanda C. Howa, Paul L. Delamater, Daniel A. Salmon, Saad B. Omer, and Nicola P. Klein. 2020. Shouting at each other into the void: A linguistic network analysis of vaccine hesitance and support in online discourse regarding California law SB277. *Social science & medicine* 266 (2020), 113216. <https://api.semanticscholar.org/CorpusID:225202630>
- [26] John Demuyakor, Isaac Newton Nyatuame, and S. Obiri. 2021. Unmasking COVID-19 Vaccine “Infodemic” in the Social Media. *Online Journal of Communication and Media Technologies* (2021). <https://doi.org/10.30935/ojcm/11200>
- [27] Keyan Ding, Ronggang Wang, and Shiqi Wang. 2019. Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) (MM '19). Association for Computing Machinery, New York, NY, USA, 2682–2686. <https://doi.org/10.1145/3343031.3356062>
- [28] Xiaohan Ding, Michael A. Horning, and Eugenia Ha Rim Rho. 2023. Same Words, Different Meanings: Semantic Polarization in Broadcast Media Language Forecasts Polarity in Online Public Discourse. *Proceedings of the International AAAI Conference on Web and Social Media* (2023). <https://api.semanticscholar.org/CorpusID:259430890>
- [29] Son Doan, Elly W. Yang, Sameer S. Tilak, Peter W. Li, Daniel S. Zisook, and Manabu Torii. 2019. Extracting health-related causality from twitter messages using natural language processing. *BMC Medical Informatics and Decision Making* 19, 3 (2019), 79.
- [30] Anna Fang and Haiyi Zhu. 2023. Measuring the Stigmatizing Effects of a Highly Publicized Event on Online Mental Health Discourse. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 482, 18 pages. <https://doi.org/10.1145/3544548.3581284>
- [31] Paige M. Farrenkopf. 2022. The Cost of Ignoring Vaccines. *The Yale Journal of Biology and Medicine* 95, 2 (2022), 265–269. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9235251/>
- [32] Joseph L. Fleiss. 1981. *Statistical Methods for Rates and Proportions* (2nd ed.). John Wiley & Sons, New York.
- [33] Amandine Gagneux-Brunon, Elisabeth Botelho-Nevers, Marion Bonneton, Patrick Peretti-Watel, Pierre Verger, Odile Launay, and Jeremy K. Ward. 2022. Public opinion on a mandatory COVID-19 vaccination policy in France: a cross-sectional survey. *Clinical Microbiology and Infection* 28, 3 (2022), 433–439. <https://doi.org/10.1016/j.cmi.2021.10.016>
- [34] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences of the United States of America* 120 (2023). <https://api.semanticscholar.org/CorpusID:257766307>
- [35] C. W. J. Granger. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2 (Jan. 1980), 329–352.
- [36] Jordan Grunawalt. 2021. The Villain Unmasked: COVID-19 and the Necropolitics of the Anti-Mask Movement. *Disability Studies Quarterly* 41, 3 (Sep. 2021), 8343. <https://doi.org/10.18061/dsq.v41i3.8343>
- [37] Cheng Guo. 2020. Identity and User Behavior in Online Communities. In *Companion Proceedings of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) (GROUP '20). Association for Computing Machinery, New York, NY, USA, 35–38. <https://doi.org/10.1145/3323994.3371018>
- [38] William L. Hamilton and et al. 2018. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. <https://doi.org/10.48550/arXiv.1605.09096>
- [39] Keith Harrigan and Mark Dredze. 2022. The Problem of Semantic Shift in Longitudinal Monitoring of Social Media: A Case Study on Mental Health During the COVID-19 Pandemic. In *Proceedings of the 14th ACM Web Science Conference 2022* (Barcelona, Spain) (WebSci '22). Association for Computing Machinery, New York, NY, USA, 208–218. <https://doi.org/10.1145/3501247.3531566>
- [40] Pedram Hosseini, Mona Diab, and David A. Broniatowski. 2019. Does Causal Coherence Predict Online Spread of Social Media?. In *Social, Cultural, and Behavioral Modeling*, Robert Thomson, Halil Bisgin, Christopher Dancy, and Ayaz Hyder (Eds.). Springer International Publishing, Cham, 184–193.
- [41] Amir Hussain and Aziz Sheikh. 2021. Opportunities for Artificial Intelligence–Enabled Social Media Analysis of Public Attitudes Toward Covid-19 Vaccines. *Catalyst non-issue content* 2, 1 (2021). <https://doi.org/10.1056/CAT.20.0649> arXiv:https://catalyst.nejm.org/doi/pdf/10.1056/CAT.20.0649
- [42] IEA. 2020. Global Energy Review 2020. International Energy Agency.
- [43] Mike Isaac. 2020. Reddit, acting against hate speech, bans “The_donald” subreddit. <https://www.nytimes.com/2020/06/29/technology/reddit-hate-speech.html>
- [44] Rishi Iyengar. 2021. Reddit takes action against groups spreading Covid misinformation | CNN business. <https://www.cnn.com/2021/09/01/tech/reddit-covid-misinformation-ban/index.html>
- [45] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26 (2019), 1–35. <https://api.semanticscholar.org/CorpusID:198180039>
- [46] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 18, 16 pages. <https://doi.org/10.1145/3544548.3581503>
- [47] Prerna Juneja, Deepika Rama Subramanian, and Tanushree Mitra. 2020. Through the Looking Glass: Study of Transparency in Reddit’s Moderation Practices. 4, GROUP, Article 17 (jan 2020), 35 pages. <https://doi.org/10.1145/3491102.3517622>
- [48] Sheila Kaplan, Katherine J. Wu, and Katie Thomas. 2020. C.D.C. Tells States How to Prepare for Covid-19 Vaccine by Early November. *The New York Times* (2020). <https://www.nytimes.com/2020/09/02/health/covid-19-vaccine-cdc-plans.html>
- [49] Naveena Karusala and Richard Anderson. 2022. Towards Conviviality in Navigating Health Information on Social Media. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 43, 14 pages. <https://doi.org/10.1145/3491102.3517622>
- [50] Vivek Khetan, Roshni Ramnani, Mayuresh Anand, Shubhashis Sengupta, and et al. 2021. Causal BERT : Language models for causality detection between events expressed in text. <https://doi.org/10.48550/arXiv.2012.05453>
- [51] Jiyoun Kim. 2020. The Meaning of Numbers: Effect of Social Media Engagement Metrics in Risk Communication. *Communication Studies* 72 (2020), 195–213. <https://api.semanticscholar.org/CorpusID:224960500>
- [52] Eili Y. Klein, Elena M. Martinez, Larissa May, and et al. 2017. Categorical Risk Perception Drives Variability in Antibiotic Prescribing in the Emergency Department: A Mixed Methods Observational Study. *Journal of General Internal Medicine* 32, 10 (2017), 1083–1089. <https://doi.org/10.1007/s11606-017-4099-6>
- [53] Irene A. Kretchy, Michelle Asiedu-Danso, and James-Paul Kretchy. 2021. Medication management and adherence during the COVID-19 pandemic: Perspectives and experiences from low-and middle-income countries. *Research in Social and Administrative Pharmacy* 17, 1 (2021), 2023–2026. <https://doi.org/10.1016/j.sapharm.2020.04.007>
- [54] D. Ashok Kumar and Anandan Chinnalagu. 2020. Sentiment and Emotion in Social Media COVID-19 Conversations: SAB-LSTM Approach. In *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*. IEEE, Moradabad, India, 463–467. <https://doi.org/10.1109/SMART50582.2020.9337098>
- [55] Jiyoun Lee, Jihyang Choi, and Rebecca K. Britt. 2023. Social Media as Risk-Attenuation and Misinformation-Amplification Station: How Social Media Interaction Affects Misperceptions about COVID-19. *Health Communication* 38, 6 (2023), 1232–1242. <https://doi.org/10.1080/10410236.2021.1996920> PMID: 34753361.
- [56] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. <https://doi.org/10.1145/3491102.3502030>
- [57] Tracy Liederholm, Michelle Gaddy Everson, Paul van den Broek, Maureen Mischinski, Alex Crittenden, and Jay Samuels. 2000. Effects of causal text revisions on more- and less-skilled readers’ comprehension of easy and difficult texts. *Cognition and Instruction* (2000), 525–556.
- [58] Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022. Generated Knowledge Prompting for Commonsense Reasoning. In *Proceedings of the 60th Annual Meeting of*

- the Association for Computational Linguistics* (Dublin, Ireland). Association for Computational Linguistics, Dublin, Ireland, 3154–3169. <https://doi.org/10.18653/v1/2022.acl-long.225>
- [59] Yafeng Lu, Xia Hu, Feng Wang, Shamanth Kumar, Huan Liu, and Ross Maciejewski. 2015. Visualizing Social Media Sentiment in Disaster Scenarios. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) (*WWW '15 Companion*). Association for Computing Machinery, New York, NY, USA, 1211–1215. <https://doi.org/10.1145/2740908.2741720>
- [60] Muhammad Shahid Iqbal Malik, Tahir Imran, and Jamjoom Mona Mamdouh. 2023. How to detect propaganda from social media? Exploitation of semantic and fine-tuned language models. *PeerJ Computer Science* 9 (2023). <https://api.semanticscholar.org/CorpusID:257060193>
- [61] Deniz Marti and David A. Broniatowski. 2020. Does gist drive NASA experts' design decisions? *Systems Engineering* 23, 4 (2020), 460–479. <https://doi.org/10.1002/sys.21538> [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sys.21538](https://onlinelibrary.wiley.com/doi/pdf/10.1002/sys.21538)
- [62] Sam Martin and Samantha Vanderslott. 2022. "Any idea how fast 'It's just a mask!' can turn into 'It's just a vaccine!'": From mask mandates to vaccine mandates during the COVID-19 pandemic. *Vaccine* 40, 51 (2022), 7488–7499. <https://doi.org/10.1016/j.vaccine.2021.10.031>
- [63] Suellen R. Martinelli and Luciana A. M. Zaina. 2021. Learning HCI from a Virtual Flipped Classroom: Improving the Students' Experience in Times of COVID-19. In *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems* (Virtual Event, Brazil) (*IHC '21*). Association for Computing Machinery, New York, NY, USA, Article 34, 11 pages. <https://doi.org/10.1145/3472301.3484326>
- [64] Dhendra Marutho, Sunarna Hendra Handaka, Ekaprana Wijaya, and Muljono. 2018. The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. In *2018 International Seminar on Application for Technology of Information and Communication*. IEEE, Semarang, Indonesia, 533–538. <https://doi.org/10.1109/ISEMANTIC.2018.8549751>
- [65] Brinda Mehra, Kejia Shen, Hen Chen Yen, and Can Liu. 2023. Gist and Verbatim: Understanding Speech to Inform New Interfaces for Verbal Text Composition. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) (*CUI '23*). Association for Computing Machinery, New York, NY, USA, Article 15, 11 pages. <https://doi.org/10.1145/3571884.3597134>
- [66] Corine S. Meppelink, Edith G. Smit, Marieke L. Fransen, and Nicola Diviani. 2019. "I was Right about Vaccination": Confirmation Bias and Health Literacy in Online Health Information Seeking. *Journal of Health Communication* 24, 2 (2019), 129–140.
- [67] Claudiu Mihăilă and Sophia Ananiadou. 2014. Semi-supervised learning of causal relations in biomedical scientific discourse. *BioMedical Engineering OnLine* 13, 2 (2014), S1.
- [68] Paramita Mirza and Sara Tonelli. 2016. CATENA: CAusal and TEmporal relation extraction from NATural language texts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Yuji Matsumoto and Rashmi Prasad (Eds.). The COLING 2016 Organizing Committee, Osaka, Japan, 64–75. <https://aclanthology.org/C16-1007>
- [69] Shi Na, Liu Xumin, and Guan Yong. 2010. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. In *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*. IEEE, Georgia, USA, 63–67. <https://doi.org/10.1109/IITSI.2010.74>
- [70] Rena Nainggolan, Resianta Perangin-angin, Emma Simarmata, and Astuti Feriani Tarigan. 2019. Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. , 012015 pages.
- [71] Serge Nyawa, Dieudonné Tchuente, and Samuel Fosso-Wamba. 2022. COVID-19 vaccine hesitancy: a social media analysis using deep learning. *Annals of Operations Research* (2022), 1–39. <https://doi.org/10.1007/s10479-022-04792-3>
- [72] Chan Young Park, Julia Mendelsohn, Karthik Radhakrishnan, Kinjal Jain, Tushar Kanakagiri, David Jurgens, and Yulia Tsvetkov. 2021. Detecting Community Sensitive Norm Violations in Online Conversations. In *Conference on Empirical Methods in Natural Language Processing*. <https://api.semanticscholar.org/CorpusID:238583283>
- [73] Jinkyung Park, Rahul Dev Ellezhuthil, Joseph Isaac, Christoph Mergerson, Lauren Feldman, and Vivek Singh. 2023. Misinformation Detection Algorithms and Fairness across Political Ideologies: The Impact of Article Level Labeling. In *Proceedings of the 15th ACM Web Science Conference 2023* (Austin, TX, USA) (*WebSci '23*). Association for Computing Machinery, New York, NY, USA, 107–116. <https://doi.org/10.1145/3578503.3583617>
- [74] Seungeun Park, Betty Bekemeier, Abraham Flaxman, and Melinda Schultz. 2022. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. *Informatics for Health and Social Care* 47, 2 (2022), 175–193. <https://doi.org/10.1080/17538157.2021.1982949> [arXiv:https://doi.org/10.1080/17538157.2021.1982949](https://doi.org/10.1080/17538157.2021.1982949) PMID: 34582297.
- [75] Umashanthi Pavalanathan and Munmun De Choudhury. 2015. Identity Management and Mental Health Discourse in Social Media. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) (*WWW '15 Companion*). Association for Computing Machinery, New York, NY, USA, 315–321. <https://doi.org/10.1145/2740908.2743049>
- [76] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 225, 16 pages. <https://doi.org/10.1145/3544548.3580907>
- [77] Manish Puri, Zachary Dau, and Aparna S. Varde. 2021. COVID and Social Media: Analysis of COVID-19 and Social Media Trends for Smart Living and Healthcare. *SIGWEB NewsL*. 2021, Autumn, Article 5 (dec 2021), 20 pages. <https://doi.org/10.1145/3494825.3494830>
- [78] Neha Puri, Eric A. Coomes, Hourmazd Haghbayan, and Keith Gunaratne. 2020. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics* 16, 11 (2020), 2586–2593. <https://doi.org/10.1080/21645515.2020.1780846> PMID: 32693678.
- [79] Miftahul Qorib, Timothy Oladunni, Max Denis, Esther Ososanya, and Paul Cota. 2023. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications* 212 (2023), 118715. <https://doi.org/10.1016/j.eswa.2022.118715>
- [80] Alan Lee Rashmi Prasad, Bonnie Webber and Aravind Joshi. 2019. Penn Discourse Treebank Version 3.0. <https://doi.org/10.35111/qebf-gk47>
- [81] Agha Ali Raza, Bilal Saleem, Shan Randhawa, Zain Tariq, Awais Athar, Umar Saif, and Roni Rosenfeld. 2018. Baang: A Viral Speech-Based Social Platform for Under-Connected Populations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3174217>
- [82] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. <https://doi.org/10.48550/arXiv.1908.10084>
- [83] Valerie F. Reyna. 2012. A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making* 7, 3 (2012), 332–359. <https://doi.org/10.1017/S1930297500002291>
- [84] Valerie F. Reyna. 2012. Risk perception and communication in vaccination decisions: a fuzzy-trace theory approach. *Vaccine* 30 25 (2012), 3790–7. <https://api.semanticscholar.org/CorpusID:29880616>
- [85] Valerie F. Reyna. 2021. A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proceedings of the National Academy of Sciences* 118, 15 (2021), e1912441117. <https://doi.org/10.1073/pnas.1912441117> [arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1912441117](https://www.pnas.org/doi/pdf/10.1073/pnas.1912441117)
- [86] Valerie F Reyna and Charles J Brainerd. 2023. Numeracy, gist, literal thinking and the value of nothing in decision making. *Nature Reviews Psychology* (2023), 1–19.
- [87] Valerie F. Reyna, Jonathan C. Corbin, Rebecca B. Weldon, and Charles J. Brainerd. 2016. How Fuzzy-Trace Theory Predicts True and False Memories for Words, Sentences, and Narratives. *Journal of Applied Research in Memory and Cognition* 5, 1 (2016), 1–9. <https://doi.org/10.1016/j.jarmac.2015.12.003>
- [88] Valerie F. Reyna, Sarah Edelson, Bridget Hayes, and David Garavito. 2022. Supporting Health and Medical Decision Making: Findings and Insights from Fuzzy-Trace Theory. *Medical Decision Making* 42, 6 (2022), 741–754. <https://doi.org/10.1177/0272989X221105473> Publisher: SAGE Publications Inc STM.
- [89] Valerie F. Reyna and Barbara J Kiernan. 1994. Development of Gist versus Verbatim Memory in Sentence Recognition: Effects of Lexical Familiarity, Semantic Content, Encoding Instructions, and Retention Interval. *Developmental Psychology* 30 (1994), 178–191. <https://api.semanticscholar.org/CorpusID:204322634>
- [90] Ludovic Rheault and Christopher Cochrane. 2020. Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis* 28, 1 (2020), 112–133. <https://doi.org/10.1017/pan.2019.26>
- [91] Eugenia Ha Rim Rho and Melissa Mazmanian. 2020. Fostering Civil Discourse Online: Linguistic Behavior in Comments of #MeToo Articles Across Political Perspectives. <https://api.semanticscholar.org/CorpusID:243831339>
- [92] Feliks Victor Parningotan Samosir, Hapnes Toba, and Mewati Ayub. 2022. BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper. *Jurnal Teknik Informatika dan Sistem Informatika* 8, 1 (2022), 202–217.
- [93] Paul Seitlinger and Tobias Ley. 2012. Implicit Imitation in Social Tagging: Familiarity and Semantic Reconstruction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (*CHI '12*). Association for Computing Machinery, New York, NY, USA, 1631–1640. <https://doi.org/10.1145/2207676.2208287>
- [94] Deepa Shivaram. 2021. The Topic Of Masks In Schools Is Polarizing Some Parents To The Point Of Violence. (2021). <https://www.npr.org/sections/back-to-school-live-updates/2021/08/20/1028841279/mask-mandates-school-protests-teachers>
- [95] Mirela Silva, Fabricio Ceschin, Prakash Shrestha, Christopher Brant, Juliana Fernandes, Catia S. Silva, Andr'e Gr'egio, Daniela Oliveira, and Luiz H. F. Giovanini. 2020. Predicting Misinformation and Engagement in COVID-19 Twitter

- Discourse in the First Months of the Outbreak. *ArXiv abs/2012.02164* (2020). <https://api.semanticscholar.org/CorpusID:227254664>
- [96] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2022. ProgPrompt: Generating Situated Robot Task Plans using Large Language Models. *2023 IEEE International Conference on Robotics and Automation (ICRA)* (2022), 11523–11530. <https://api.semanticscholar.org/CorpusID:252519594>
- [97] Eun Young Song, Hoe-Ryeon Choi, and Hong Chul Lee. 2019. A Study on Efficient Training Method for Named Entity Recognition Model with Word Embedding Applied to PCA. *Journal of the Korean Institute of Industrial Engineers* (2019). <https://doi.org/10.7232/JKIE.2019.45.1.030>
- [98] Oona St-Amant, J. Anneke Rummens, Henry Parada, and Karline Wilson-Mitchell. 2022. The COVID-19 Mask. *Ans. Advances in Nursing Science* 45, 2 (2022). <https://doi.org/10.1097/ANS.0000000000000393>
- [99] Qi Su, Clara M. Wan, Xiaoqian Liu, and Chu-Ren Huang. 2020. Motivations, Methods and Metrics of Misinformation Detection: An NLP Perspective. (2020). <https://doi.org/10.2991/nlpr.d.200522.001>
- [100] Peter Sullivan. 2020. Trump retweets conspiracy theory questioning COVID-19 death toll. <https://thehill.com/policy/healthcare/514430-trump-retweets-conspiracy-theory-questioning-covid-19-death-toll/> The Hill.
- [101] Nandan Thakur, Nils Reimers, Johannes Daxenberger, and Iryna Gurevych. 2021. Augmented SBERT: Data Augmentation Method for Improving Bi-Encoders for Pairwise Sentence Scoring Tasks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technology*. 296–310. <https://www.aclweb.org/anthology/2021-naacl-main.28>
- [102] Petter Törnberg. 2023. ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning. *ArXiv abs/2304.06588* (2023). <https://api.semanticscholar.org/CorpusID:258108255>
- [103] Tom Trabasso and Paul van den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language* 24 (1985), 612–630. <https://api.semanticscholar.org/CorpusID:146399871>
- [104] Milka Trajkova, A'aeshah Alhakamy, Francesco Cafaro, Sanika Vedak, Rashmi Mallappa, and Sreekanth R. Kankara. 2020. Exploring Casual COVID-19 Data Visualizations on Twitter: Topics and Challenges. 7, 3 (2020), 35. <https://doi.org/10.3390/informatics7030035> Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- [105] Danny Valdez, Marijn ten Thij, Krishna Bathina, Lauren A Rutter, and Johan Bollen. 2020. Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data. *J Med Internet Res* 22, 12 (14 Dec 2020), e21418. <https://doi.org/10.2196/21418>
- [106] Deirdre Walsh. 2020. Congress Passes \$900 Billion Coronavirus Relief Bill, Ending Months-Long Stalemate. *NPR* (2020).
- [107] Mischa Wanek-Libman. 2021. CDC to revise mask requirements concerning outdoor areas of transportation systems.
- [108] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandekar, Chaowei Xiao, Yuke Zhu, Linxi (Jim) Fan, and Anima Anandkumar. 2023. Voyager: An Open-Ended Embodied Agent with Large Language Models. *ArXiv abs/2305.16291* (2023). <https://api.semanticscholar.org/CorpusID:258887849>
- [109] Hanyin Wang, Yikuan Li, Meghan Hutch, Andrew Naidech, and Yuan Luo. 2021. Using Tweets to Understand How COVID-19-Related Health Beliefs Are Affected in the Age of Social Media: Twitter Data Analysis Study. *J Med Internet Res* 23, 2 (22 Feb 2021), e26302.
- [110] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for Conceptual Blending with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 435, 19 pages. <https://doi.org/10.1145/3544548.3580948>
- [111] Xingqiao Wang, Xiaowei Xu, Weida Tong, Ruth Roberts, and et al. 2021. InferBERT: A Transformer-Based Causal Inference Framework for Enhancing Pharmacovigilance. *Frontiers in Artificial Intelligence* (2021).
- [112] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. <https://doi.org/10.48550/arXiv.2201.11903>
- [113] Maxwell Weinzierl and Sanda Harabagiu. 2022. Identifying the Adoption or Rejection of Misinformation Targeting COVID-19 Vaccines in Twitter Discourse. In *Proceedings of the ACM Web Conference 2022* (New York, NY, USA) (WWW '22). Association for Computing Machinery, 3196–3205. <https://doi.org/10.1145/3485447.3512039>
- [114] Evan A. Wilhelms, Valerie F. Reyna, Priscila Brust-Renck, and et al. 2015. Gist Representations and Communication of Risks about HIV-AIDS: A Fuzzy-Trace Theory Approach. *Current HIV research* 13, 5 (2015), 399–407.
- [115] Christopher R. Wolfe, Valerie F. Reyna, Colin L. Widmer, and et al. 2015. Efficacy of a web-based intelligent tutoring system for communicating genetic risk of breast cancer: a fuzzy-trace theory approach. *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 35, 1 (2015), 46–59. <https://doi.org/10.1177/0272989X14535983>
- [116] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. <https://doi.org/10.1145/3491102.3517582>
- [117] Tongshuang Sherry Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J. Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (2022). <https://api.semanticscholar.org/CorpusID:247447133>
- [118] Yi Wu and Fei Shen. 2022. Exploring the impacts of media use and media trust on health behaviors during the COVID-19 pandemic in China. *Journal of Health Psychology* 27, 6 (2022), 1445–1461. <https://doi.org/10.1177/1359105321995964> PMID: 33646827.
- [119] Zhan Xu and Hao Guo. 2018. Using Text Mining to Compare Online Pro- and Anti-Vaccine Headlines: Word Usage, Sentiments, and Online Popularity. *Communication Studies* 69, 1 (2018), 103–122. <https://doi.org/10.1080/10510974.2017.1414068>
- [120] Jie Yang, Soyeon Caren Han, and Josiah Poon. 2022. A survey on extraction of causal relations from natural language text. *Knowledge and Information Systems* 64, 5 (2022), 1161–1186.
- [121] Baosheng Yin, Meishu Zhao, Lu Guo, and Liqi Qiao. 2023. Sentence-BERT and k-means Based Clustering Technology for Scientific and Technical Literature. In *2023 15th International Conference on Computer Research and Development (ICCRD)*. 15–20. <https://doi.org/10.1109/ICCRD56364.2023.10080830>
- [122] Renwen Zhang, Natalya N. Bazarova, and Madhu Reddy. 2021. Distress Disclosure across Social Media Platforms during the COVID-19 Pandemic: Untangling the Effects of Platforms, Affordances, and Audiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) (CHI '21). Association for Computing Machinery, 1–15. <https://doi.org/10.1145/3411764.3445134>
- [123] Caleb Ziemis, William Held, Omar Shaikh, Jiaao Chen, Zehao Zhang, and Diyi Yang. 2023. Can Large Language Models Transform Computational Social Science? <https://doi.org/10.48550/arXiv.2305.03514> arXiv:2305.03514 [cs]
- [124] Tina M. Zottoli, Rebecca K. Helm, Vanessa A. Edkins, and Michael T. Bixter. 2023. Developing a model of guilty plea decision-making: Fuzzy-trace theory, gist, and categorical boundaries. *Law and Human Behavior* 47, 3 (2023), 403–421. <https://doi.org/10.1037/lhb0000532> Place: US Publisher: Educational Publishing Foundation.

A STUDY 1 MODEL PERFORMANCE COMPARISON RESULTS

In our study, we trained several baseline models on our human-annotated Reddit dataset to evaluate their performance in cause-effect pair extraction. Table 6 summarizes the performance metrics of these models.

Table 6: Model performance on cause-effect pair extraction from our manually labeled Reddit dataset.

Model	Precision	Recall	F1-score	Accuracy
BERT	0.709	0.708	0.707	0.711
RoBERTa	0.826	0.829	0.814	0.846
XLNet	0.804	0.819	0.811	0.821
RBIC (GPT-3.5)	0.934	0.949	0.941	0.940
RBIC (GPT-4)	0.972	0.951	0.961	0.977

B STUDY 4 STATISTICAL RESULT

Table 7: Result of Granger causality test for relationships between gist clusters (C1, C2 and C3) and health outcomes dataset.

Cluster	Health Outcomes	Hypotheses	Lag	F-val	p-val	Direction of Impact
C1. Vaccination Implications	NV	C1 ↗ NV	4	4.778	0.027	gists impacts NV
	NV	NV ↗ C1	14	7.771	0.007	NV impacts gists
	GH	C1 ↗ GH	11	1.765	0.271	-
	GH	GH ↗ C1	5	1.015	0.410	-
	ICU	C1 ↗ ICU	6	1.118	0.402	-
	ICU	ICU ↗ C1	3	1.818	0.242	-
	TC	C1 ↗ TC	11	2.495	0.089	-
	TC	TC ↗ C1	11	2.960	0.072	-
	NC	C1 ↗ NC	4	2.986	0.068	-
	NC	NC ↗ C1	9	0.749	0.488	-
C2. Controversies and Policies Related to Masks	NV	C2 ↗ NV	17	0.654	0.5124	-
	NV	NV ↗ C2	7	0.745	0.4888	-
	GH	C2 ↗ GH	3	1.110	0.411	-
	GH	GH ↗ C2	5	2.210	0.091	-
	ICU	C2 ↗ ICU	4	2.282	0.090	-
	ICU	ICU ↗ C2	14	1.416	0.343	-
	TC	C2 ↗ TC	5	9.395	0.005	gists impacts TC
	TC	TC ↗ C2	8	1.101	0.414	-
	NC	C2 ↗ NC	5	11.829	0.004	gists impacts NC
	NC	NC ↗ C2	7	1.181	0.409	-
C3. Impact of Lockdown	NV	C3 ↗ NV	4	2.437	0.084	-
	NV	NV ↗ C3	4	1.101	0.413	-
	GH	C3 ↗ GH	12	1.118	0.410	-
	GH	GH ↗ C3	9	8.822	0.006	GH impacts gists
	ICU	C3 ↗ ICU	7	0.715	0.494	-
	ICU	ICU ↗ C3	14	4.000	0.039	ICU impacts gists
	TC	C3 ↗ TC	4	1.135	0.409	-
	TC	TC ↗ C3	11	0.796	0.470	-
	NC	C3 ↗ NC	5	2.357	0.088	-
NC	NC ↗ C3	1	1.178	0.382	-	

C COST AND TIME OF RUNNING RBIC

Cost estimation is important for researchers considering the RBIC framework in their studies, especially with large datasets. Applying the GPT-4 pricing model of \$0.06 per 1,000 tokens (to the combined total of prompts, input posts, and outputs), the total cost for using RBIC on our dataset of 79,680 posts amounted to \$621.51, translating to roughly \$0.0078 per post. The total duration of running RBIC on GPT-4 was 199.2 hours or 8 seconds per post, as shown in Table 9. While we did not use GPT-3.5 in this study, we show the estimated cost and running time based on GPT-3.5's pricing at the time of this study. While we used GPT-4, which can be expensive for large datasets, the RBIC framework is adaptable to other open-source large language models that may be more cost-effective.

Table 8: Result of Granger causality test for relationships between gist clusters (C4 and C5) and health outcomes dataset

Cluster	Health Outcomes	Hypotheses	Lag	F-val	p-val	Direction of Impact
C4. Societal and Economic (Macro) Impact of COVID-19	NV	C4 ↗ NV	3	2.579	0.072	-
	NV	NV ↗ C4	3	2.014	0.091	-
	GH	C4 ↗ GH	9	1.787	0.261	-
	GH	GH ↗ C4	4	2.072	0.080	-
	ICU	C4 ↗ ICU	15	2.618	0.066	-
	ICU	ICU ↗ C4	1	2.071	0.081	-
	TC	C4 ↗ TC	1	1.780	0.263	-
	TC	TC ↗ C4	9	3.663	0.031	TC impacts gists
	NC	C4 ↗ NC	4	1.100	0.412	-
C5. Social Issues, Health, and Personal Experiences	NC	NC ↗ C4	9	3.092	0.032	NC impacts gists
	NV	C5 ↗ NV	7	2.385	0.085	-
	NV	NV ↗ C5	4	1.101	0.415	-
	GH	C5 ↗ GH	12	0.786	0.471	-
	GH	GH ↗ C5	14	0.799	0.464	-
	ICU	C5 ↗ ICU	1	1.119	0.399	-
	ICU	ICU ↗ C5	9	2.282	0.090	-
	TC	C5 ↗ TC	5	2.001	0.107	-
	TC	TC ↗ C5	2	1.137	0.389	-
	NC	C5 ↗ NC	2	2.427	0.079	-
	NC	NC ↗ C5	1	1.710	0.212	-

Table 9: Comparative Analysis of Cost and Time Efficiency for RBIC Framework Using GPT-4 and GPT-3.5 Models. The table provides a detailed breakdown of the estimated costs and processing time required for analyzing 79,680 posts using two different OpenAI LLMs.

Model	Cost per Token (Input)	Cost per Token (Output)	Cost per Post	Total Cost	Total Time(hrs)
GPT-4	\$0.03 / 1K tokens	\$0.06 / 1K tokens	\$0.0078	\$621.50	199.20
GPT-3.5	\$0.0010 / 1K tokens	\$0.0020 / 1K tokens	\$0.00039	\$31.07	154.93

D DESIGN FEATURE IN THE MODERATION DASHBOARD

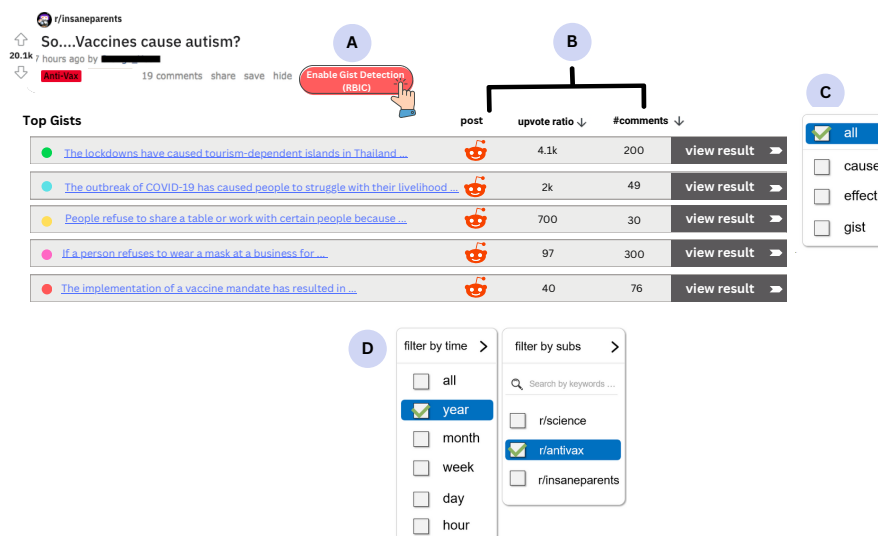


Figure 4: Design Implication: Illustration of moderator’s view of a submitted post or comment.