Leveraging the Power of ChatGPT: Evaluating Its Effectiveness for Content Analysis and Framing Research in Mass Communication

Hyerin Kwon
Univeristy of Wisconsin-Madison
hyerin.kwon@wisc.edu

Linqi Lu
University of Wisconsin-Madison
llu84@wisc.edu

Jiwon Kang
University of Wisconsin-Madison
jkang76@wisc.edu

Douglas M. McLeod
University of Wisconsin-Madison
dmmcleod@wisc.edu

Abstract

This study explores the application of ChatGPT (GPT) to content analysis within the context of framing research, specifically examining its effectiveness in identifying public health, economic stability, and civic vitality frames in COVID-19 press releases. Our methodology is grounded in the Semantic Architecture Model (SAM), which conceptualizes framing as a process by which meaning is embedded in content units at various textual levels (i.e., concepts, assertions, arguments and narratives). In addition, this study underlines the necessity of AI prompt engineering to improve GPT's coding performance in identifying frames at the concept, assertion, and thematic argument levels. The findings indicate the transformative potential of AI in communication research, highlighting its ability to analyze complex message framing across diverse contexts.

Keywords: GPT, prompt engineering, content analysis, framing, Semantic Architecture Model

1. Introduction

As information technologies have rapidly expanded over recent decades, new frontiers have opened for researchers studying media content and its effects. Emerging technologies have catapulted our capacity to produce, disseminate, collect, store and analyze massive amounts of digital trace data that provide insights into human knowledge, attitudes and behaviors. While efforts to capture and analyze such data (including mediated communication) are driven by financial interests to capitalize these data for commercial gain, this information and the tools to analyze it also offer great promise for social scientists.

Communication researchers are increasingly at the forefront of social science researchers interested in analyzing digital trace data as it provides tremendous potential for unobtrusively revealing the nature of human communication behaviors, answering questions about what kinds of people are consuming and disseminating what kinds of information and how people are using and responding to the information that they receive. The data that can be collected and analyzed are generally more easily accessible and more voluminous than that afforded by traditional social science methods. Moreover, researchers are rapidly developing new techniques to analyze and display data to understand important aspects of human behavior as they relate to important processes such as monitoring social change and relational dynamics.

This rapid growth in analytical tools has been driven by huge investments in the development of artificial intelligence (AI), which have burst into prominence as these technologies assume greater roles in human activities. AI is changing the nature of human existence, just as it is providing new techniques for data analysis.

Most importantly, the expanding capabilities to capture data, the rapid development of tools to efficiently analyze data, and the increasing presence of AI in daily life are all opening up vast unexplored territories for consequential research questions that can be addressed by mass communication researchers. While it is true that there are a multitude of academic disciplines that have been blessed by the promise of access to large datasets, innovations in analytical tools and the ascension of AI, communication researchers may be particularly drawn to the new avenues for inquiry created by these developments. Much of the content in digital trace data is the result of human communication, not only our attempts to engage with other humans, but also our attempts to engage with media organizations and other social institutions. As communication researchers, we have an important role to play in studying the content of communication.

The purpose of this paper is to document how AI (in the form of GPT) can be adapted to provide an analytical tool to analyze communication content. We describe our efforts to train and validate GPT as a form

URI: https://hdl.handle.net/10125/109119
978-0-9981331-8-8

of content analysis that offers advantages over both human coding and other forms of computer-assisted content analysis. In the process, we link the use of GPT for content analysis to the Semantic Architecture Model (SAM) of media framing (McLeod et al., 2022) to unitize content and assess how meaning is embedded in media messages. Additionally, we employ prompt engineering to facilitate the generative AI model in providing an optimized prompt. The importance of this work is underscored by changes in the way that citizens communicate and interact with online media in the 21st Century Digital Age, as well as by the tremendous growth in the storage of massive amounts of digital trace data that are available to researchers, organizations and institutions. As more and more resources are being dedicated to developing AI technologies, it is important to conduct and share research on how such technologies can be used effectively to analyze the burgeoning array of digital trace data, particularly as it is applied to areas of communication research.

2. Research Background

2.1. Analyzing communication content

Analyzing communication content has always been at the heart of mass communication research. Many researchers have focused on describing the nature of mediated content using quantitative content analysis following social science principles (Holsti, 1969; Krippendorff, 1980). Others have employed more qualitative techniques such as critical discourse analysis (Van Dijk, 1985) and participant observation (Gans, 1979; Gitlin, 1980; Tuchman, 1978) to reveal the impact of content antecedents such as underlying power relations and message production processes to understand the nature of mediated content.

Furthermore, researchers interested in a variety of different types of media effects have examined media content to isolate significant content features that can be isolated in experimental stimulus variations that constitute potentially influential factors that can explain message effects on audiences (Grabe & Westley, 2003; Thorson et al., 2012).

2.2. The framing perspective

One theoretical framework that has emerged to understand the nature of message content (and ultimately its potential effects) is “frame analysis.” Content researchers from a variety of different theoretical perspectives have contributed to the development and use of frame analysis to understand how meaning is constructed within communication messages (Entman, 1993; Pan & Kosicki, 1993). Frame analysis has often focused on the narrative frame of the message as a whole to identify a central narrative framework that emphasizes a particular meaning or way of understanding the information and events presented in a message. From this perspective, researchers often seek to identify common narrative patterns in message construction such as the “the protest paradigm” that often characterizes news coverage of social protests (Chan & Lee, 1984; McLeod & Hertog, 1999).

McLeod et al. (2022) articulate a Semantic Architecture Model (SAM) that extends frame analysis beyond narrative frames to examine how smaller units of text can be framed to convey meaning. Following an analogy to house construction, message construction often begins with an abstract blueprint for what the narrative of a message will ultimately look like, but the construction process begins with smaller units. Words with particular suggested meanings are chosen to represent a “concept” frame, the equivalent of bricks in the house analogy. Concepts are used to construct “assertion” frames (the walls), which are then put together into “thematic” frames (the rooms), which are compiled to build the “narrative” frames (the house). Thus meaning is embedded with the message by the construction decisions made at each of these levels, and as such can work together to amplify the power of messages.

The SAM’s architectural conception of framing can be illustrated with a message drawn from the COVID-19 press releases that were analyzed in this study. This particular press release was issued by the office of Michigan Governor Gretchen Whitmer. Though this entire message (i.e., the house) was framed around the public health implications of the COVID-19 pandemic, our procedures (as detailed below) selected direct quotations from governors within each press release for further analysis. These direct quotations (i.e., the room) are built from health related concept frames (i.e., the bricks), which are combined into health-related assertions (i.e., the walls).

In her statement, Governor Whitmer uses public health concepts as the “omicron variant,” “kn95 masks,” and the “MDHHS office” (Michigan Department of Health and Human Services) as she builds public health assertions frames (e.g., “By distributing 10 million highly-effective kn95 masks, we can keep families and communities safe.”). Such assertions are combined to form an argument that frames meaning (as conceptualized by Entman, 1993) to *define*, *diagnose*, *evaluate* and *prescribe* solutions to the COVID-19 problem. Applying the SAM to communication research has a number of benefits.

First, it is helpful in identifying the message components that carry the frames as well as the meanings that are conveyed by those components. Second, once categories are established to capture alternative meanings, the various frames can be used as outcome variables to examine their relationship to antecedent variables such as the characteristics of the creator and the creator's organizational affiliation. Third, it provides a lens through which one can observe content that helps in making indirect inferences about the intentions of the message creator. And finally, it can be used to isolate important message characteristics that can be manipulated in experimental research examining the influence of message frames on various perceptual, affective, emotional and behavior outcomes.

2.3. Analyzing message content

The application of the SAM has parallels to quantitative content analysis methodology in terms of "*unitization,*" one of the basic elements of quantitative content analysis. One of the first choices that must be made in designing content analysis centers on the "unit of analysis," the basic textual unit that first must be isolated and then coded for the content variables of interest. Each of these variables must be linked to a list of categories that represent an exhaustive and mutually exclusive set of content categories that reflect the important potential content differences between different content units (e.g., the presence or absence of a frame within that unit). Coders could also assess the intensity (or *saturation*) of the frame within that unit (e.g., a major or minor frame).

Along these lines, the SAM suggests that a given message (such as a newspaper article or a press release) could be unitized and coded in terms of concepts, assertions, arguments and narrative that work together to convey meaning. Manual coders could be trained to isolate a particular unit for coding according to the categories for each variable that describes that textual unit. Coding content at the concept-level is a relatively simple task, but is very time consuming. As content unitization moves to higher levels (i.e., assertions to thematic arguments to the narrative of the entire message), coding becomes more efficient in terms of coding time, but the judgment becomes more complex as multiple frames may come into play and the saturation level of a particular frame (i.e., how salient the frame is within the unit) varies from one unit to another (often reducing the intercoder agreement of human coding).

Thus, human content coding involves a priori methodological decision regarding unitization that considers implications for validity, reliability, coding time and expense, task complexity and coder fatigue, while trying to capture textual meaning sufficient to answering research questions and testing hypotheses. All of these considerations impose potential limitations.

2.4. Computer-assisted content analysis

With the advent of digital media and the explosion of online content, traditional manual content analysis began to face new challenges. The vast amounts of data generated in the digital age necessitated the integration of computational techniques that could efficiently manage and analyze these large datasets (Dakhel et al., 2024). As such, computational methods allow for more efficient sampling and coding of extensive text corpora, enabling the analysis of patterns across large bodies of text (Lewis et al., 2013).

Computers assist in analyzing content through various Natural Language Processing (NLP) tasks, leveraging computational power to manage large volumes of text data. They offer several methods for content analysis, allowing communication researchers to systematically process and interpret the large-scale text data. (Chowdhary & Chowdhary, 2020).

Word Frequency Analysis: This fundamental method involves counting the occurrences of specific words or phrases within a text corpus. It provides insights into the prominence of certain topics or themes in the dataset, helping to identify areas of focus or concern (Manning & Schütze, 1999).

Dictionary-Based Sentiment Analysis: This approach utilizes predefined lists of words associated with positive, negative, or neutral sentiments to analyze the overall emotional tone of a text. It aids in understanding the sentiment expressed in communication, providing a systematic assessment of public opinion and emotional responses (Liu, 2012).

Topic Modeling: Techniques such as Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (STM) are employed to discover hidden themes in large text corpora (Blei et al., 2003). Topic modeling identifies groups of words that frequently appear together, thereby uncovering the underlying topics within the text and offering a nuanced understanding of its thematic structure.

Network Analysis: Network analysis examines the relationships between different entities (e.g., words, individuals, organizations) within a text (Abraham et al., 2009; Tabassum et al., 2018). It helps visualize and understand the connections and structures within communication content, shedding light on the interactions and relationships present in the data.

Traditional Machine Learning Classification: Machine learning algorithms can be trained to classify text into predefined categories based on patterns learned from annotated training data (Thai, 2022; Parmar et al., 2023). This method is useful for spam detection and sentiment classification tasks, allowing for efficient categorization and analysis of large text datasets.

By employing these computational methods, researchers can uncover patterns, sentiments, and structures within vast amounts of digital data that might not be easily discernible through manual analysis.

2.5. The use of AI in content analysis

Recently, computer-assisted content analysis has been advanced by developments in artificial intelligence (AI). Unlike simple keyword matching and topic modeling, AI models employ sophisticated algorithms to delve into the deeper meanings and contexts embedded within a text (Orrù et al., 2023). By analyzing syntax, semantics, and context, these models can accurately identify message frames and underlying themes, even in complex and nuanced language. Moreover, AI enables automated and scalable analysis, processing vast amounts of data at speeds far beyond human capabilities (Zhang & Lu, 2021). Utilizing parallel computing and distributed systems, AI algorithms can swiftly analyze massive text corpora, making large-scale content analysis feasible and efficient. This scalability is instrumental in managing the deluge of digital information generated daily across various platforms and sources. Additionally, AI models engage in dynamic learning, continuously improving their accuracy and adaptability through exposure to new data (Chang et el., 2023). Using reinforcement learning and online learning techniques, these models iteratively refine their understanding and predictive capabilities based on real-time feedback. This ongoing learning process ensures that the analysis remains abreast of evolving communication trends, adapting to shifts in language usage, cultural context, and societal dynamic.

Despite these advancements, the use of AI in content analysis also faces several challenges. First, compared to the growing demand for AI in research, there is a scarcity of studies focusing on applying Large Language Models (LLMs) in communication research. This gap results in a lack of clear roadmaps and guides for using state-of-the-art models effectively. Second, while AI models have advanced significantly, they are not infallible. Issues with accuracy and reliability persist, especially when dealing with nuanced or context-specific content. Third, few studies have explored sophisticated framing and content analysis using AI. The typical computer-assisted approaches often simplify the task by identifying whether a certain frame, stance, or sentiment appears (i.e., binary classification) without delving into the more intricate aspects of framing.

2.6. Research questions

In order to advance the use of AI in content analysis, this paper explores the application of GPT to the analysis of content. We approach this content analysis from the perspective of framing analysis, using GPT to reveal meaning is embedded in message content. We illustrate this process by using a case study of how we trained GPT to identify message frames at various levels. To guide this analysis presented below, we propose the following research questions:

RQ1): How can GPT be used as an effective tool for content analysis in the context of framing research?

RQ2): How do distinct prompts impact the performance of GPT in analyzing message frames?

RQ3): How consistent is GPT with human coders?

3. Methodology

3.1. Data collection & preprocessing

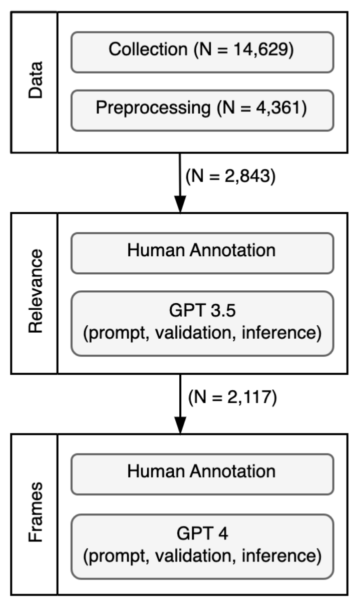

Figure 1. The GPT frame analysis process

To demonstrate our application of GPT to frame analysis (see Figure 1), we analyzed the statements of 12 U.S. governors that appeared in press releases regarding the COVID-19 pandemic. We began by using ScrapeStorm, a web scraping tool, to gather press releases from state government archives spanning January 2020 to May 2023 (N = 14,629). We then selected only those press releases that contained the words "covid," "coronavirus," and "pandemic" yielding a total of 4,361 COVID-19 press releases.

From these press releases, we unitized content for coding by isolating governors' direct quotes to represent how these governors were framing COVID-19 concerns. This was accomplished by identifying three common quotation formats used in press releases: a) *"[Statement]" said Governor [Name]*; b) *"[Statement]," Governor/Gov. [Name] said*; and c) *"[Statement]" [Name] said/added*. Using these patterns, we employed regular expressions in R to systematically isolate these quotes. This process yielded 2,843 press releases with their governor quotation passages isolated for coding.

3.2. Human annotation for data cleaning

The above procedures yielded a small number of governor quotation passages that were irrelevant to COVID-19. As the first step in the process to train GPT to recognize and remove these irrelevant quotations (i.e., build a relevance classifier), we coded a random sample of 200 press releases drawn from the preprocessed dataset to identify relevant quotations in which the governor's statement was related to the pandemic. Press releases were coded as '1' if deemed relevant to COVID-19 and '0' if not (Krippendorff's alpha = .87). We subsequently used this dataset of 200 human-annotated documents to conduct a validation test of GPT's identification of COVID-relevant quotations as described below.

3.3. Using GPT for data cleaning

We developed a specific GPT prompt to filter out COVID-irrelevant quotations. The prompt used was: "{Governor's quotation passage} If the given text is related to the context of COVID-19, answer '1' otherwise '0.' Your answer MUST be either '1' or '0'. Do not include your explanation." Incorporating the phrases, "Your task is" and "You MUST," resulted in improved model performance (Bsharat et al., 2023), prompting us to include 'MUST' in the query. Additionally, the prompt's last sentence suppresses the often lengthy explanations GPT provides with its coding decisions.

After preparing this prompt and our sample of 200 quotations, we submitted them to GPT for label generation. Initially, we defined hyperparameters such as 'model,' 'messages,' and 'temperature.' As the relevance assessment task is relatively simple, we chose to use the cost-effective 'GPT 3.5' model (i.e., 'gpt-3.5-turbo-0125').

Within GPT's 'messages' parameter, we specified two 'role' configurations: 'system' with the message 'You are the expert of content analysis of COVID-19.' and 'user' with our prompt, allowing GPT to generate responses based on the provided role and instructions. Finally, we set the 'temperature' of the model to 0, which regulates randomness. Temperature closer to 1 yields more creative but less factual outputs (Gray et al., 2023).

To assess the performance of the GPT on this relevance assessment task, we used common metrics: precision, recall, and F1 score (Zhang et al., 2019). We achieved high precision (.95), recall (.95) , and F1 scores (.95), so no further updates were necessary for the existing prompt.

In machine learning, inference refers to the process of applying a trained model to predict outcomes based on previously unseen data (Web3.com Ventures, 2023). For the relevance task, out of 2,843 documents, the model categorized 2,117 as COVID-relevant governor quotes (typically multiple related assertions) and 726 as irrelevant. These quotes (essentially *argument* frames from the SAM typology) became the unit that was coded in our frame analysis.

3.4. COVID-19 frame categories

Once the quotation passages were isolated for coding, we developed categories to represent potential frames that align with major public concerns stemming from the COVID-19 pandemic. Past research has focused on public health and economic threats as major concerns (Kassas & Nayga, 2021; Knapp et al., 2023; Zhong & Broniatowski, 2023). Also apparent in public discourse, were concerns about threats to civic life, as participation in schools, churches, and other community events were widely curtailed. To capture these concerns, we created three frame categories: *public health*, *economic stability*, and *civic vitality.*

These three categories not only represent central themes of the COVID-19 discourse, but they also fit Entman's (1993) observation that frames convey *problem definition, causal diagnosis, moral judgments,* and *prescriptive solutions.* We illustrate these categories and demonstrate their fit with Entman's observation using the following examples:

Public Health. Michigan Governor Gretchen Whitmer's statement that, "COVID-19 has had an immense impact on our state's healthcare system and its ability to provide quality care in critical conditions," clearly contains elements of problem definition and causal diagnosis. Her statement that, "as we continue facing COVID, the best thing you can do to protect yourself and your loved ones is to get vaccinated, and if you're eligible, get your booster shot," reflects both a moral judgment and a prescriptive solution.

Economic Stability. Missouri Governor Michael Parson's quote, "COVID-19 has had severe impacts on our anticipated economic growth. This is truly unlike anything we have ever experienced before, and we are now expecting significant revenue declines," illustrates both problem definition and causal diagnosis. Wisconsin Governor Tony Evers's stated that, "we want small businesses to know that help is on the way. and once we receive federal funds, we aren't going to wait to get these funds out quickly to help small businesses restock shelves, catch up on bills, rehire and retain workers, and continue to help keep their customers, employees, and our communities safe as we work to bounce back together," includes elements of problem definition, causal diagnosis, moral judgment and prescriptive solution.

Civic Vitality. Kansas Governor Laura Kelly's quote illustrates problem definition and causal diagnosis: "Opportunities for in-person voting registration are among the many normal routines that have become more difficult as a result of covid-19." Governor Whitmer statement, "My number one priority right now is protecting Michigan families from the spread of covid-19. For the sake of our students, their families, and the more than 100,000 teachers and staff in our state, I have made the difficult decision to close our school facilities for the remainder of the school year," reflects problem definition, causal diagnosis, moral judgment and prescriptive solution.

3.5. Human annotation for frame analysis

Following the identification of COVID-19 relevant governor quotation passages, the second stage of annotation involved both human and computer annotation of the presence of our frames of interest. We began with human coding of the presence and intensity of these frames within the quotation passages for later assessment of the GPT analysis. This stage involves more complex analysis as this larger textual unit that potentially constitutes an *argument frame* provides the capacity for the speaker to co-mingle multiple sub-frames through the incorporation of different assertion and concept frames. That is, the quotation passage might contain elements of all three of COVID-19 frames (i.e., *public health*, *economic stability,* and *civic vitality*). By focusing on indicative concept frames within the quotation passage, we assessed the presence or absence of each of these three frames, as well as their intensity (saturation) within the quotation passage. As such, we coded whether each of these frames had a 'major,' 'minor,' or 'no' presence.

For example, we analyzed the following quotation passage: "It's great to see Wisconsinites rolling up their sleeves and doing their part to make sure our state and our economy continue to recover," said Gov. Evers. "The vaccine is safe, effective, and is the best way to keep yourself and your loved ones healthy. I encourage Wisconsinites to drop by our vaccine clinic at the state fair to get your shot–and a free cream puff, too!" Here we coded *public health* as major, *economic stability* as minor, and *civic vitality* as none. In this passage, the public health frame was central to the message, encouraging vaccinations to keep people healthy. Economic stability was identified as a side benefit of people rolling up their sleeves to get vaccinated, while no mention of the civic vitality frame.

After several rounds of training, human coders analyzed a random sample of 200 quotation passages for the presence and intensity of the three frames (Krippendorff's alpha = .82). The results of this human coding were subsequently used to assess the accuracy of GPT coding.

3.6. Using GPT for frame analysis

One of the main objectives of our study is to compare GPT's performance in content analysis to that of human coding. Enhancing the accuracy and performance of the LLM in generating outputs requires designing various prompts (instructions) through trial-and-error experimentation until an optimized prompt is developed.

Our prompt development consisted of three rounds of refinement. The initial baseline instruction to analyze the presence and intensity of the three frames in the quotation passage is illustrated in Figure 2. After giving GPT examples of the three types of *concept frames*, we instructed GPT to identify the *thematic argument frames* in each passage along with a confidence rating based on the presence of concept indicators. As a cut-off criterion, frames with a confidence rating greater than 5 were labeled as 'major,' those scoring between 1 and 5 were classified as 'minor,' and frames that did not appear were categorized as 'none.'

The cost-effective *GPT 3.5* model used in the relevance task was not sufficient to conduct the more complex frame analysis, so we opted for the more expensive and more powerful 'gpt-4-turbo-preview' model (*GPT 4*) to conduct the framing analysis. In terms of the parameters and temperature, we retained the same values applied in the relevance assessment task.

Using the initial prompt, the comparison to human coding yielded less than satisfactory results. The precision, recall and F1 scores were: .46, .50, and .47

for the public health frame; .36, .35, and .35 for the economic stability frame; and .59, .46, and .50 for the civic vitality frame. To enhance the model's efficacy, prompt engineering strategies were considered in subsequent rounds of frame analysis.

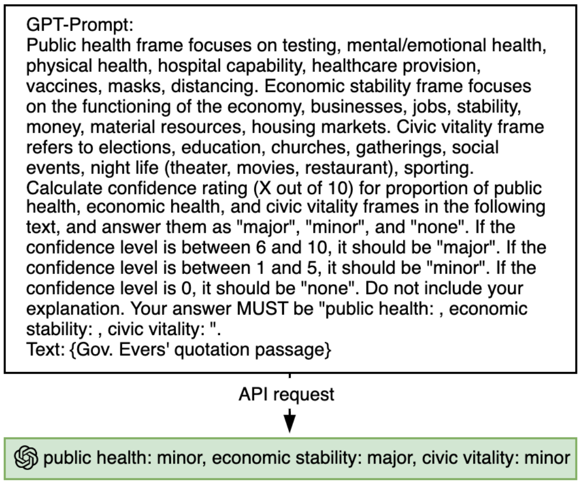

Figure 2. Example of the initial GPT prompt

3.7. Prompt engineering

Prompt engineering involves deliberate procedures to design and optimize prompts (instructions) to enhance the accuracy of LLMs (Zhang et al., 2023). Numerous studies in Computer Science and Engineering have explored and evaluated prompts using diverse strategies and principles (Bsharat et al., 2024; Hatakeyama-Sato et al., 2023; Velásquez-Henao et al., 2023).

Past research (Liyanage et al., 2023; Wang et al., 2022) has supported the efficacy of the "Chain-of-Thought (CoT)," in which prompts are delivered step-by-step. Instead of expecting the LLMs to generate an answer in one step, the task is broken down into smaller, logical steps that guide the model toward the solution. This sequential approach reduces task ambiguity and enhances comprehension, and ultimately improves the model's performance.

Figure 3 illustrates our implementation of three separate prompts designed to guide GPT through a sequential reasoning process. Upon reviewing the initial GPT annotation results, we observed challenges in the model's ability to predict major frames despite their explicit presence in the text. This difficulty may stem from the lengthiness of press releases. To address this, our revised approach first condensed the entire document into a brief summary (of 30 words). Then, GPT was asked to identify major frames in the summary, applying the same confidence rating procedure that was used in our initial analysis. Once major frames had been identified, GPT was then asked to examine the original quotation passage to identify minor frames, excluding those already categorized as major. The updated prompt exhibited improved performance in precision, recall and F1 scores: public health (.76, .78, and .75 respectively); economic stability (.71, .74, and .69); and civic vitality (.62, .54, and .57).

To further increase the effectiveness of the GPT's performance, we revised the second prompt to clarify the quantification of frame intensity by specifying more concrete procedures for classifying 'major,' 'minor' frames.

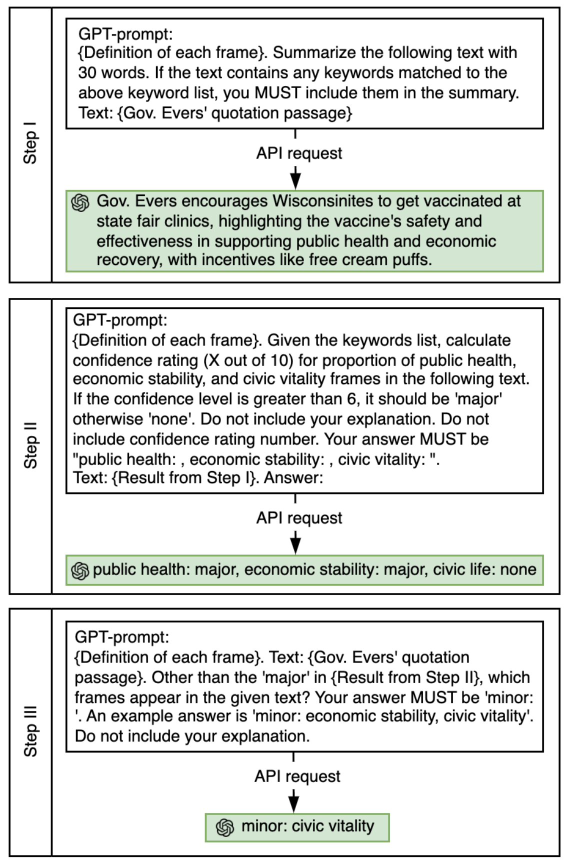

Figure 3. Example of the second GPT prompt

Here, we followed the SAM's principle that meaning can be embedded in different textual levels to develop a four-step process. First, each document was separated (tokenized) into discrete sentences (at the assertion level). Second, within each sentence, keywords (concept frames) indicative of particular frames (i.e., public health, economic stability, and civic vitality) were isolated.

Third, the proportion of sentences containing each frame-relevant keyword was calculated and divided by the total number of sentences. This information was then used to create a code of 'major', 'minor' or 'none' to indicate the saturation of particular frames within the message. In addition, we followed examples from Bsharat et all. (2024) to further enhance our CoT approach by combining our step-by-step instructions into a single prompt (i.e., instead of making three separate API requests, we consolidated them into a single API request). In the process, we enhanced the clarity of our GPT instructions by employing '#', '\n', and alphabetical ordering to distinctly delineate each step: '###Frames and Keywords###', '###Instruction###', and '###Question###' (see Figure 4).

Our latest prompt yielded satisfactory results in terms of public health (.95 for all three metrics), economic stability (.90, .89, and .89) and civic vitality (.89, .88, and .88). Using this satisfactory prompt, we are able to use GPT to conduct frame analysis to provide descriptive results of the frequency and intensity of frames found in the 2,117 governor quotation passages in our dataset (see Table 1).

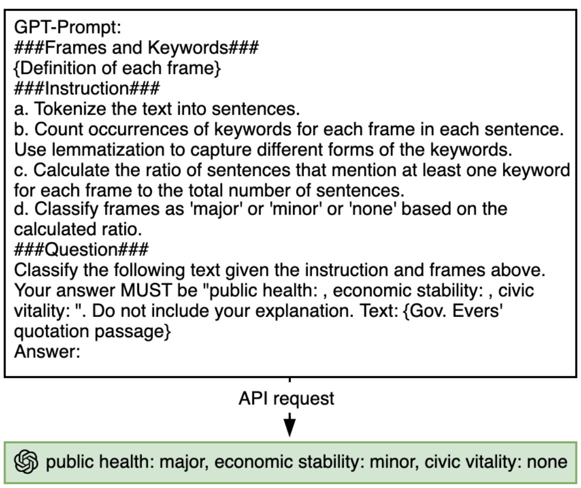

Figure 4. Example of the final GPT prompt

This GPT frame analysis could then be used to test theoretically-derived hypotheses such as whether there were significant differences between Democratic and Republican, or male and female governors.

Table 1. Inference results of framing analysis

	Public Health	Economic Stability	Civic Vitality
Major	1359 (64.2%)	639 (30.2%)	108 (5.1%)
Minor	184 (8.7%)	463 (21.9%)	638 (30.1%)
None	574 (27.1%)	1015 (47.9%)	1371 (64.8%)

4. Discussion and Conclusion

This study adapting GPT for framing analysis offers a roadmap to guide communication researchers in leveraging cutting-edge computational techniques for research. This approach is particularly valuable to framing and other research that analyzes the content of large social media datasets.

We illustrated the utility of GPT as an effective content analysis tool by examining how statements by U.S. governors reported in official press releases framed arguments related to the COVID-19 pandemic. GPT efficiently performed a series of tasks, including relevance assessment, unitization and the identification of the presence and intensity of frames, as validated by comparisons to human coding.

Our framing analysis was guided by the Semantic Architecture Model (McLeod et al., 2022), which posits that framed meaning can be embedded in textual units including the author's selection of concepts, assertions and arguments that are used to build a message. As the goal of this analysis was to examine how governors framed COVID-19 to address this issue, we isolated the direct quotations within each press release (quotation passages that potentially carry what SAM refers to as *argument frames*). Our initial manual assessment of these passages indicated the presence of multiple COVID-19 frames, suggesting that meaning assessment should include the examination of smaller SAM units *assertion* and *concept frames*. Our procedures to guide the GPT application to framing analysis reflect this orientation.

After isolating the quotation passages, we inductively developed three categories to represent inherent argument frames (*public health, economic stability,* and *civic vitality*) that present different definitional, diagnostic, evaluative and prescriptive understandings of COVID-19 concerns.

In training GPT to assess these frames, we designed our initial prompt by providing example *concept frames* for each category and instructing the AI model to identify *thematic argument frames* and calculate a confidence rating for each identified frame. To improve this initial coding, we then used two prompt engineering techniques: isolating the prompt instructions one step at a time and adding a preliminary step of asking GPT to write a 30-word summary of each quotation passage before proceeding to analyze the entire passage. For our final prompt, we added the sentence tokenization instruction (mirroring the SAM's *assertion* level framing) and provided specific instructions for calculating the confidence rating. After this final round of prompt revision, we were able to produce satisfactory GPT coding.

The validity of GPT coding was demonstrated by comparing it to reliable human coding using quantitative evaluation metrics (i.e., precision, recall, and F1 scores). For the relevance assessment task, we found that GPT's prediction was highly consistent with human coding (.95). For the frame analysis task, after prompt engineering, we achieved highly satisfactory results, with all three metrics exceeding .8 for *public health, economic health, and civic vitality frames.*

Building on these findings, we anticipate that future research leveraging this dataset will explore more comprehensive aspects of framing analysis. Specifically, subsequent studies could test hypotheses regarding the influence of antecedent variables (e.g., political affiliation and gender of the governors) on the deployment of the COVID-19 public health, economic stability and civic vitality frames. Future research could also employ experiments to test the outcomes of frames and frame combinations within these messages to assess their influence on audience perceptions, attitudes and behavioral intentions.

Our application of GPT to identify and evaluate message frames has some limitations. Knowledge and techniques based on LLMs are evolving rapidly, which may necessitate further development of our GPT procedures. However, the broader insights identified here such as the importance of prompt engineering informed by the semantic architecture approach to framing analysis may be important to future iterations of ChatGPT and other language modeling applications used to analyze communication messages.

Additionally, it is important to consider that an over-reliance on algorithmic approaches can make it challenging to draw meaningful inferences about social phenomena by oversimplifying complex human communication. This may exacerbate the gap between theoretical frameworks and the computational methods used for text analysis (Baden et al., 2022; Zamith & Lewis, 2015).

To address these limitations, communication researchers should adopt a balanced, hybrid approach that combines the strengths of traditional qualitative and quantitative content analysis with the power of computational methods. In fact, this study provides an example of such a hybrid approach as we used a simple, inductive qualitative approach to identify COVID-19 frames, and developed a theory-driven GPT-based computational technique that was evaluated by comparisons to traditional quantitative content analyses.

Furthermore, the processes of fostering collaboration between computational and social science researchers, developing methods aligning with theoretical needs, and improving validation techniques are essential for more comprehensive future research.

5. References

Abraham, A., Hassanien, A. E., & Snášel, V. (Eds.). (2009). *Computational social network analysis: Trends, tools and research advances*. Springer.

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*, 993–1022.

Bsharat, S. M., Myrzakhan, A., & Shen, Z. (2023). Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4. *arXiv* preprint arXiv:2312.16171.

Chan, J. M., & Lee, C. C. (1984). Journalistic paradigms on civil protests: A case study of Hong Kong. In A. Arno & W. Dissanayake (Eds.), *The news media in national and international conflict* (pp. 183–202). Westview Press.

Chang, J. D., Brantley, K., Ramamurthy, R., Misra, D., & Sun, W. (2023). Learning to generate better than your LLM. *arXiv* preprint arXiv:2306.11816.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45.

Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603–649.

Dakhel, A. M., Nikanjam, A., Khomh, F., Desmarais, M. C., & Washizaki, H. (2024). An Overview on Large

Language Models. *Generative AI for Effective Software Development*, 3–21.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58.

Gans, H. J. (1979). *Deciding what's news: A study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. Random House.

Gitlin, T. (1980). *The whole world is watching: Mass media in the making and unmaking of the New Left.* University of California Press.

Grabe, M. E., & B. H. Westley (2003). The controlled experiment. In G. H. Stempel, D. H. Weaver, & G. C. Wilhoit (Eds.), *Mass communication cesearch and theory* (pp. 267–298). Allyn & Bacon.

Gray, M., Savelka, J., Oliver, W., & Ashley, K. (2023). Can GPT alleviate the burden of annotation?. In *Legal Knowledge and Information Systems* (pp. 157–166). IOS Press.

Hatakeyama-Sato, K., Yamane, N., Igarashi, Y., Nabae, Y., & Hayakawa, T. (2023). Prompt engineering of GPT-4 for chemical research: What can/cannot be done? *Science and Technology of Advanced Materials: Methods*, *3*(1), 2260300.

Holsti, O. (1969). *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley Publishing.

Kassas, B., & Nayga Jr, R. M. (2021). Promoting higher social distancing and stay-at-home decisions during COVID-19: The underlying conflict between public health and the economy. *Safety Science*, *140*, 105300.

Knapp, E. R., Smith, B. A., & Motta, M. P. (2023). Complementary or competing frames? The Impact of economic and public health messages on COVID-19 attitudes. *Journal of Experimental Political Science*, *10*(1), 21–33.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. SAGE Publications.

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of broadcasting & electronic media*, *57*(1), 34–52.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Liyanage, C., Gokani, R., & Mago, V. (2023). GPT-4 as a Twitter data annotator: Unraveling its performance on a stance classification Task. *Authorea Preprints*.

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

McLeod, D. M., Choung, H., Su, M. H., Kim, S., Tao R., Liu, J., & Lee, B. (2022). Navigating a diverse paradigm: A conceptual framework for experimental framing effects research. *Review of Communication Research*, *10*, 1–58.

McLeod, D. M., & J.K. Hertog. (1999). Social control and the mass media's role in the regulation of protest groups: The communicative acts perspective. In D. Demers and K. Viswanath (Eds.), *Mass media, social control and social change* (pp. 305–330). Iowa State University Press.

Orrù, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in artificial intelligence*, *6*, Article 1199350.

Pan, Z., & G. M. Kosicki (1993). Framing analysis: An approach to news discourse. *Political Communication*, *10*(1), 55–76.

Parmar, J., Chouhan, S., Raychoudhury, V., & Rathore, S. (2023). Open-world machine learning: applications, challenges, and opportunities. *ACM Computing Surveys*, *55*(10), 1–37.

Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1256.

Thai, H. T. (2022, April). Machine learning for structural engineering: A state-of-the-art review. In *Structures* (Vol. 38, pp. 448–491). Elsevier.

Thorson, E., Wicks, R., & Leshner, G. (2012). Experimental methodology in journalism and mass communication research. *Journalism and Mass Communication Quarterly*, *89*(1), 112–124.

Tuchman, G. (1978). *Making news: A study in the construction of reality.* Free Press.

Van Dijk, T. A. (1985). *Discourse and communication*. de Gruyter.

Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *Dyna*, *90*(230), 9–17.

Wang, B., Min, S., Deng, X., Shen, J., Wu, Y., Zettlemoyer, L., & Sun, H. (2022). Towards understanding chain-of-thought prompting: An empirical study of what matters. *arXiv* preprint arXiv:2212.10001.

Web3.com Ventures. (2023, November 29). Inference, bridging AI model and reality. *Medium*. https://medium.com/@Web3comVC/inference-bridging-ai-model-and-reality-cf2261746840

Zamith, R., & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 307–318.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, *23*, Article 100224.

Zhang, Y., Shah, D., Foley, J., Abhishek, A., Lukito, J., Suk, J., Kim, S.J., Sun, Z., Pevehouse, J., & Garlough, C. (2019). Whose lives matter? Mass shootings and social media discourses of sympathy and policy, 2012–2014. *Journal of Computer-Mediated Communication*, *24*(4), 182–202.

Zhang, H., Wu, C., Xie, J., Kim, C., & Carroll, J. M. (2023). QualiGPT: GPT as an easy-to-use tool for qualitative coding. *arXiv* preprint arXiv:2310.07061.

Zhong, W., & Broniatowski, D. A. (2023). Economic risk framing increases intention to vaccinate among Republican COVID-19 vaccine refusers. *Social Science & Medicine*, *317*, Article 115594.