Are Prompts All You Need?: Chatting with ChatGPT on Disinformation Policy Understanding

Chen, Haihua University of North Texas, USA | haihua.chen@unt.edu
Cherukuri, Komala Subramanyam University of North Texas, USA | KomalaSubramanyamCherukuri@my.unt.edu
Zhu, Xiaohua (Awa) University of Tennessee-Knoxville, USA | xzhu12@utk.edu
Yang, Shengnan Western University, Canada | syang859@uwo.ca

ABSTRACT

ChatGPT has shown promise in assisting qualitative researchers with coding. Previous efforts have primarily focused on datasets derived from interviews and observations, leaving document analysis, another crucial data source, relatively unexplored. In this project, we address the rapidly emerging topic of disinformation regulatory policy as a pilot to investigate ChatGPT's potential for document analysis. We adapt our existing qualitative research framework, which identifies five key components of disinformation policy: context, actors, issue, instrument, and channel, to sketch out policy documents. We then designed a two-stage experiment employing a multi-layer workflow using a dataset with highly relevant policy documents from US federal government departments. Through iteratively developing and refining six different prompt strategies, we identified an effective few-shot learning strategy that achieved 72.0% accuracy and a 70.8% F-score with the optimal prompt. Our experimental process and outcomes explore the feasibility of using ChatGPT to support manual coding for policy documents and suggest a coding approach for conducting explicit document analysis through an interactive process between researchers and ChatGPT. Furthermore, our results initiate a wider debate on how to integrate human logic with ChatGPT logic, along with the evolving relationship between researchers and AI tools.

KEYWORDS

Disinformation regulation, Information policy, Automatic content analysis, Generative AI, ChatGPT, Prompt engineering.

INTRODUCTION

The rise of disinformation has become a prominent concern and a major topic of public discussion; the negative impact of misinformation or disinformation on society and individuals has become evident (Aïmeur, Amri, & Brassard, 2023). Therefore, intensive efforts among scholars have been dedicated to battling disinformation (Jiang et al., 2023; Nguyen et al., 2022; Santos, 2023; Shahbazi & Bunker, 2024). With the rise in public discussion on how to combat disinformation, government engagement and the production of related public documents have significantly increased (Rice et al., 2021; Tenove, 2020). Traditional analyses of these documents typically rely on limited datasets for detailed examination and comparison, often employing labor-intensive manual coding (Mende et al., 2024). However, the escalating volume of documents necessitates an urgent shift towards systematic analysis. Crucially, the advanced abilities of Generative AI, such as ChatGTP, in analyzing and creating text, images, code, and more have facilitated its broad application in areas, such as text classification, sentiment analysis, and data augmentation, among others (Gozalo-Brizuela & Garrido-Merchán, 2023; Henrickson & Meroño-Peñuela, 2023). A few initial studies have started exploring how ChatGTP can assist in qualitative coding (e.g., Katz et al., 2023; Zhang et al., 2023). Most of these studies have examined the utilities of ChatGPT by using interview or observation data in educational settings (Katz et al., 2023; Morgan, 2023). This leaves a wide range of areas unexplored. Existing experiments have highlighted the capability of GPT to free researchers from labor-intensive and repetitive tasks, thereby enhancing the overall efficiency of the coding process. This is achieved through a multi-step workflow, particularly beneficial for topics and themes demanding less interpretive analysis (Katz et al., 2023).

In light of data availability, technological capabilities, and the gaps in existing AI-assisted qualitative research, our study delves into the application of GPT-facilitated coding in analyzing public policy documents, particularly within the disinformation sphere. Given that public documents tend to feature more explicit and defined discourse patterns, this corpus may be particularly amenable to GPT analysis. Specifically, we aim to address the following questions:

RQ1: Can ChatGPT be used to analyze policy documents automatically and effectively?

RQ2: How to best design and optimize prompts for automatic policy understanding?

RQ3: What is the most useful information and best practice for performance improvement of automatic policy understanding using ChatGPT?

To answer the above research questions, we adapted a multidimensional framework from our previous research for disinformation policy analysis, which contains five critical components: context, actors, issue, instrument, and channel (Zhu & Yang, 2023). Specifically, *context* involves the situations of participants and action, creating a social space where socio-political entities interact. *Actors* encompasses both the regulators (government agencies and other entities taking specific actions against disinformation) and the targets of disinformation stakeholders (e.g., consumers, creators, and distributors) that the policy seeks to influence. *Channel* focuses on how disinformation is generated and propagated through ICT-based carriers of disinformation creation, distribution, and consumption. *Issue* covers the specific types, topics, and manifestations of disinformation, such as political disinformation and conspiracy theories. *Instrument* represents the array of tools, techniques, strategies, and other actions that regulators (plan to) use to achieve policy objectives. We then used ChatGPT (ChatGPT-3.5) to distinguish the five components of 37 disinformation policy documents on the sentence level, based on prompt engineering. We designed two zero-shot prompts incorporating instructions and definitions of each component, and four enhanced few-shot prompts using labeled data, reasoning, and unique aspects learned by ChatGPT (ChatGPT-4), and modified definitions with the few-shot learning strategy. Empirical evaluations demonstrate the feasibility of automatic policy analysis and the importance of prompt optimization. The data, code, more examples about the prompts and discussions of the results are available on GitHub at: https://github.com/kc6699c/ASIST2024-ChatGPT-Disinformation-Policy.

RESEARCH DESIGN

To test prompts facilitating a Generative AI model like ChatGPT to accurately label policy content, we conducted iterative experiments across phases P1 through P6, outlined in Figure 1. Each phase progressed with increasingly detailed and intricate prompts. Initially, basic prompts featuring simple definitions and examples were used, evolving towards more sophisticated techniques utilizing annotated data, leveraging ChatGPT's learning capabilities, and incorporating elaborate definitions to refine the Generative AI model's comprehension and categorization. Specific prompts were designed not only to illustrate the learning of clearer definitions and examples but also to include labeled data and more accurate terminology to increase the prompts' effectiveness. Using this trial-and-error approach, with gradual improvement of prompts, we expected to see an improvement in the output quality; that is, more accurate policy content labeling.

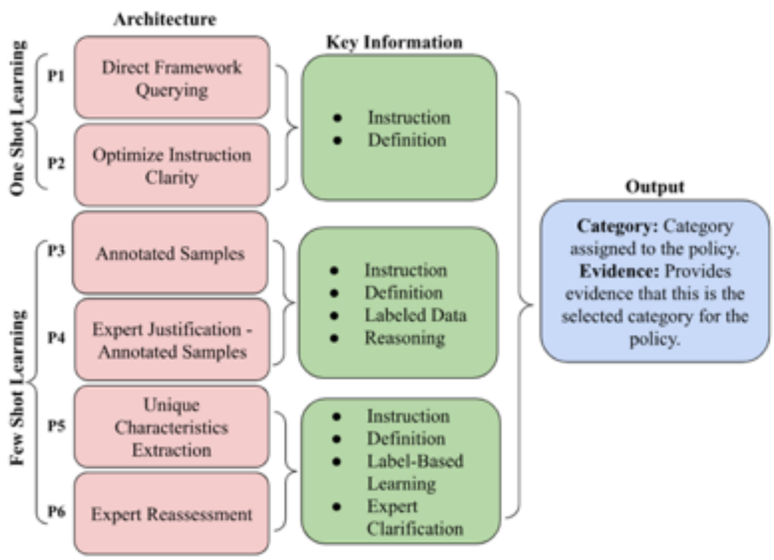

Figure 1. A Framework for Prompt Engineering and Optimization for Automatic Text Analysis

DATA COLLECTION AND PREPROCESSING

We searched exhaustively for disinformation policy documents issued by the US federal government, including congressional bills, regulations, presidential documents, congressional reports and hearings, and government announcements, between January 1, 2020, and December 31, 2021, using multiple government information sources, such as govinfo.gov and the websites of relevant government agencies. The keywords we used included "fake news," "misinformation," "disinformation," "malign information," and "conspiracy theories." In total, 92 highly relevant documents were selected, and manual coding was conducted for the qualitative analysis component of the larger policy study. The overall unit of the manual coding was on the document level—finding, labeling, and interpreting the relevant content within each document, rather than sentence-level or paragraph-level labeling.

In the automatic coding tests, we used a set of these highly relevant policies, a total of 37 bills, hearings, and presidential documents. We extracted every sentence from these documents, resulting in 337 excerpts for GPT labeling. Each excerpt was labeled using the aforementioned five components defined in the framework; namely, context, actors, issue, instrument, and channel, with up to four labels for each excerpt. At the end of each experiment, the labels assigned by ChatGPT were compared against the manual labeling for each excerpt.

EXPERIMENTAL RESULTS AND EVALUATION

Prompt Engineering

Prompt Engineering plays a crucial role as a preliminary stage that prepares ChatGPT for the advanced exploration of Zero-Shot and Few-Shot Learning research. By selectively adjusting the prompts for the model based on the experts' input, we developed a solid foundation necessary for the model to categorize policy content correctly. These preparations are crucial components in helping ChatGPT tackle the complexities of policy categorization successfully during the detailed Zero-Shot Learning ChatGPT (ZSLC) and Few-Shot Learning ChatGPT (FSLC). The specific prompt components are presented in Table 1. In this study, for any excerpt, a match between ChatGPT-generated label(s) with any of the manual labels associated with that excerpt is considered correct/accurate. The evaluation results are presented in Table 2.

	Prompt	Prompt Components	Objective
ZSLC	P1	Instruction + Expert Definition	Introduce basic categorization framework.
	P2	Instruction (modification) + Expert Definition	Evaluate the effect of instructional changes.
FSLC	P3	Instruction (P1) + Definition + Annotated Data	Enhance prompt understanding through examples.
	P4	Instruction (P1) + Definition + Annotated Data + Annotated Data Justification by expert	Expert clarification on complex categorization scenarios.
	P5	Instruction (P1) + Definition + ChatGPT unique characteristics extraction from labeled data	ChatGPT extracts category features from expert labeled data.
	P6	Instruction (P1) + Expert re-assessment on definition + unique characteristics by ChatGPT	Expert re-assessment on definitions and unique aspects by ChatGPT.

Table 1. Prompts for Automatic Disinformation Policy Understanding

	Prompt	F-1 Score						Accuracy
		Actors	Channel	Issue	Context	Instrument	Overall	Overall
ZSLC	P1	0.04	0.11	0.13	0.08	0.15	0.102	0.32
	P2	0.04	0.09	0.11	0.05	0.16	0.09	0.29
FSLC	P3	0.05	0.03	0.10	0.08	0.17	0.086	0.31
	P4	0.05	0.07	0.11	0.06	0.16	0.09	0.30
	P5	0.66	0.5	0.48	0.58	0.62	0.568	0.59
	P6	0.74	0.68	0.65	0.72	0.75	0.708	0.72

Table 2. Evaluation Results of Different Prompts

Zero-Shot Learning ChatGPT (ZSLC)

The use of ChatGPT in ZSLC seeks to leverage the advantages of both computational efficiency and human expertise. The study is initiated by taking a zero-shot learning perspective, which is an approach that uses instruction to guide ChatGPT, along with the definition in the task of classifying policy excerpts into one or more categories. Instruction in the prompt, a very key component, determines the model's success and sets the path. Two different experiments were conducted, both involving direct instruction within the prompt. The first experiment successfully demonstrated that the model received sufficient guidance from the experiment to adjust and analyze policies with higher accuracy. Subsequently, the first instruction method was adopted for further trials.

The distinction between the performance on Trials 1 and 2 (especially the lower scores for Actors and Channel with P2) suggests how a structured instruction can impact the model's accuracy. A well-engineered prompt instruction contains the right balance of guidance and flexibility, allowing ChatGPT to apply its capabilities effectively without being misled by overly narrow or overly broad instructions. This process of prompt engineering is critical, particularly in tasks that require a high level of conceptual understanding. By embedding Analysis Instructions within prompts, researchers can direct ChatGPT to not just process information but to analyze and reason through it, approaching the task with a nuanced understanding as required by the specific classification task.

Few-Shot Learning ChatGPT (FSLC)

The first few-shot learning prompt (P3) is structured with the prompt from Trial 1 (Instruction and the expert-provided category definitions) and the addition of expert-annotated data samples to the prompt. Nonetheless, this method did not result in a major improvement in prompt learning, showing that the mere presence of the annotated samples is not a very accurate predictor of the level of gained knowledge.

The next trial (P4) on the detailed justification of why particular samples were assigned to the policy categories, was an attempt to help ChatGPT better comprehend the nature of policy components. Apparently, presenting samples with justifications does not adequately support ChatGPT's concept understanding. Therefore, we then tested whether "distinguishing aspects" of each category could be a more efficient strategy for increasing understanding.

We used manually labeled policy documents as training data (all 92 documents, as mentioned in the Data Collection section) in the final two trials (P5 and P6). ChatGPT was given the label data to identify and summarize the unique characteristics of each category. The prompts in P5 incorporated instructions, experts' definitions, and characteristics that were generated by ChatGPT, which led to a good performance (0.59 in overall accuracy). Expert re-evaluation and definition clarification played a vital role in the continual development of the prompt in P6, where two experts re-evaluated, refined, and revised both the ChatGPT-generated characteristics and the definitions. The detailed process of fine-tuning eventually led to a significant increase in the precision of policy classification—the accuracy score jumped to 0.72.

Commitments to this sequence of techniques from zero-shot to few-shot learning, which have been enhanced by the help of expert assessment and iterative refinement, show the complexity of ChatGPT's categorization tasks and the necessity for increased adaptability. The significant improvements of unique characteristics that ChatGPT was exposed to underline the importance of learning from labeled data. By labeling these unique aspects within the training data, ChatGPT is trained to recognize and understand these subtle cues, which results in much more effective categorization.

DISCUSSION

In this exploratory study, we discovered that through specific optimization strategies, the accuracy of ChatGPT-driven automatic coding significantly improves. Starting with basic zero-shot results, we refined our methodology using a series of follow-up adjustments, including human explanations (prompts 3 and 4), unique characteristics summarized by ChatGPT itself, and a combination of human insights with GPT's capabilities. The integration of expert knowledge and GPT's output emerged as the most effective strategy, highlighting the potential of this collaborative approach.

Unexpectedly, we observed a notable difference in outcomes between prompts 3 and 4, and prompt 5. The initial prompts aimed to enhance human initiative with detailed examples and explanations, yet GPT's performance did not meet expectations. In contrast, prompt 5, which tapped into GPT's own logic, showed a marked improvement, indicating GPT's ability to achieve goals without fully adopting human reasoning patterns. This raises questions about understanding GPT's logic and its application in complex interpretative tasks, given the ongoing discussions about AI's "black box" nature.

GPT demonstrates the ability to learn yet does not mimic the intricacies of human reasoning. This fact persuades us to consider how we might grasp GPT's underlying logic and identify situations where the logic of machines becomes comprehensible. In this study, GPT was assigned to descriptive coding, a task for which it's theoretically well-suited. This scenario underscores the importance of delving deeper into GPT's potential and its constraints in contexts requiring more nuanced interpretation. Additionally, bridging the gap between the logic understood by machines and human reasoning emerges as an essential area for continued inquiry.

Moreover, the dynamics between humans and AI warrant careful consideration. In this context, researchers assume a role more akin to overseers than active participants, marking a departure from methods endorsed by Christou (2023) and traditional qualitative research practices. Surprisingly, the direct cognitive input from researchers was found to be less effective. Contrary to what might be expected, it became apparent that experts needed to dial back their direct contributions to facilitate superior coding outcomes. This shift challenges preconceived notions about the integration of AI into research and underscores the evolving nature of human-AI collaboration.

GENERATIVE AI USE

We confirm that we did not use generative AI tools/services to author this submission.

AUTHOR ATTRIBUTION

Haihua Chen (First Author): conceptualization, methodology, formal analysis, project administration, supervision, writing – original draft; Komala Subramanyam Cherukuri (Second Author): methodology, formal analysis, validation, visualization, writing – original draft; Xiaohua (Awa) Zhu (Third Author): conceptualization, data curation, investigation, writing – review and editing; Shengnan Yang (Fourth Author): conceptualization, data curation, writing – review and editing.

ACKNOWLEDGMENTS

The authors would like to thank Marie Bloechle at the University of North Texas for the copy-editing.

REFERENCES

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30.

Christou, P. A. (2023). How to use Artificial Intelligence (AI) as a resource, methodological and analysis tool in qualitative research?. *The Qualitative Report*, 28(7), 1968-1980.

Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). A survey of generative AI applications. *arXiv preprint arXiv*:2306.02781.

Henrickson, L., & Meroño-Peñuela, A. (2023). Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. *AI & Society*, 1-16.

Jiang, B., Tan, Z., Nirmal, A., & Liu, H. (2023). Disinformation detection: An evolving challenge in the age of LLMs. *arXiv preprint arXiv*:2309.15847.

Katz, A., Wei, S., Nanda, G., Brinton, C., & Ohland, M. (2023). Exploring the efficacy of ChatGPT in analyzing student teamwork feedback with an existing taxonomy. *arXiv preprint arXiv*:2305.11882.

Mende, M., Ubal, V. O., Cozac, M., Vallen, B., & Berry, C. (2024). Fighting infodemics: labels as antidotes to mis-and Disinformation?!. *Journal of Public Policy & Marketing*, 43(1), 31-52.

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT. *International Journal of Qualitative Methods*, 22, 16094069231211248.

Nguyen, H., Ogbadu-Oladapo, L., Ali, I., Chen, H., & Chen, J. (2023). Fighting misinformation: where are we and where to go?. *International Conference on Information* (pp. 371-394). Springer, Cham.

Rice, D., Siddiki, S., Frey, S., Kwon, J. H., & Sawyer, A. (2021). Machine coding of policy texts with the institutional grammar. *Public Administration*, 99(2), 248-262.

Santos, F. C. C. (2023). Artificial intelligence in automated detection of disinformation: a thematic analysis. *Journalism and Media*, 4(2), 679-687.

Shahbazi, M., & Bunker, D. (2024). Social media trust: Fighting misinformation in the time of crisis. *International Journal of Information Management*, 77, 102780.

Tenove, C. (2020). Protecting democracy from disinformation: Normative threats and policy responses. *The International Journal of Press/Politics*, 25(3), 517-537.

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2023). Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *arXiv preprint arXiv*:2309.10771.

Zhu, X., & Yang, S. (2023). Toward a sociotechnical framework for misinformation policy analysis. *The Usage and Impact of ICTs during the COVID-19 Pandemic* (pp. 11-45). Routledge.