# Cover Letter

Huanrui CHEN
University of Chicago

May 26, 2024

## 1 Instructor's review

### 1.1 Reduction of repetitive/redundant paragraphs

I revised the opening sentences of most paragraphs in the literature review section to ensure they summarize the content of each paragraph effectively while avoiding repetitive phrasing and content. I also eliminated redundant information, including overly detailed descriptions of data processing procedures and variables irrelevant to the research question. This includes removing references to specific data processing packages like pandas, variables such as the sentiment and phrasing of posts, and unnecessary tokenization and characterization of the data.

### 1.2 Provide a more clear explanation of the experimental process of Kobayashi, Song, and Chan (2021).

I rewrote the description of the experiment.

**Original text:**

Previous studies have primarily relied on surveys or experiments to collect data and have contributed to discussions in this field. For example, Kobayashi, Song, and Chan (2021) observed that although public street protests decreased after the enactment of the law, covert support for opposition demands persisted and, in some cases, even strengthened. They utilized a randomized survey experiment design, collecting data on participants' attitudes and behaviors before and after the law's enactment, thereby comparing

the changes in support of these two-time points. Kobayashi, Song, and Chan (2021) noted that under strict laws, public actions became riskier, leading people to move their activities to less conspicuous but potentially safer platforms and forms of communication. The persistence and intensification of this support occur because repression while dampening visible dissent, does not necessarily diminish deeply held grievances and political convictions. Instead, repression can reinforce the resolve of those who perceive the government's actions as unjust, prompting them to continue supporting opposition demands through more covert or less overtly confrontational means.

**Revised text:**

The study by Kobayashi, Song, and Chan (2021) reveals changes in protest dynamics, showing that although visible street protests have declined following the enactment of stringent repressive laws, covert support for opposition movements has not only persisted but sometimes has even intensified. The researchers used conjoint experiments to assess public sentiment and strategic preferences under stringent laws. In these experiments, conducted before and after implementing the Hong Kong National Security Law, a repressive policy, participants were presented with various hypothetical scenarios featuring diverse policy attributes related to governmental responses to protests. Each policy profile varied systematically in attributes such as the severity of law enforcement actions and the rights to public assembly, allowing researchers to isolate which attributes influenced participants' preferences for supporting or opposing policy measures. By analyzing the choices made by participants, Kobayashi, Song, and Chan (2021) could infer a shift in protest dynamics. They noted that, after the enactment of the Hong Kong National Security Law, experiment participants showed a decreased preference for overt protest actions, which may reflect the increased risks associated with visible dissent under stricter laws. However, simultaneously, the experiments also indicated that when participants faced the condition of a stringent repressive policy, those participating in the experiments after the enactment of the Hong Kong National Security Law emotionally displayed more support and identification with the opposition. The persistence and intensification of this support occur because repression while dampening visible dissent, does not necessarily diminish deeply held grievances and political convictions. Instead, repression can reinforce the resolve of those who perceive the government's actions as unjust, prompting them to continue supporting opposition demands through more covert or less overtly confrontational means.

However, one question left unanswered by this study concerns whether public discontent translates into behaviors beyond street protests when repression does not diminish citizens' support for the opposition.

## 1.3 Revision the part of data encoding and classification

I removed all coding content that is not directly related to the research or difficult to annotate experimentally, including the sentiment of posts, and the rhetoric and tone of posts. I retained only the annotations related to whether the posts were political discussions and the sensitivity levels of the posts. The sensitivity levels of the posts are based on King, Pan, and Roberts' (2013) study of China's internet censorship system and are categorized into the following five levels:

- **Posts Mobilizing Collective Action:** Posts that directly call for or could lead to collective action, such as protests, demonstrations, or other public gatherings. Since these posts can immediately trigger public collective actions, they are usually the main target of censorship.
- **Pornography and Criticism of Censors:** These posts are consistently censored but not as aggressively as those with collective action potential. Criticism of censorship mechanisms or entities and posts containing pornography fall into this category.
- **Direct Criticism of government:** Posts that discuss government policies and political reforms or criticize the government. These posts may not directly mobilize action, and although they are considered relatively sensitive, they will not necessarily be censored and deleted.
- **General Political Discussion:** Less sensitive general political discussions that do not address criticism of government or advocate collective action. These might include discussions on political theory, non-contentious political commentary, or historical political analysis.
- **Non-Political Content:** Posts with no political content or relevance to sensitive social issues are minimally sensitive. These would include everyday discussions, entertainment news, or trivial content.

Based on these standards and encoding requirements, I reevaluated the data annotation capabilities of large language models. I sampled 25 posts each from the entertainment and current affairs sections of LIHKG posted on May 21, 2024. I manually annotated the sensitivity of the posts and whether they belonged to political discussions. Subsequently, I used ChatGPT-4.0 to classify the sensitivity and determine whether the 50 posts were political discussions. ChatGPT-4.0 achieved a 100% accuracy rate in identifying whether posts were political discussions and a 96% accuracy rate in assessing the sensitivity of the posts.

**I also added a paragraph explaining why it is important to annotate the sensitivity of the posts and consider it as a variable:**

Although the website categorizes posts and designates a specific section for political discussions, this study aims not only to capture the differences between political and non-political discussions but also to identify political discussion posts that may mobilize collective movements, as well as other types of everyday political discussions or criticisms because this study would like to explore the chilling effect, which implies that due to vague and harsh laws, citizens might choose to avoid participating in and discussing topics they consider sensitive out of fear of potential punishment. The more extensive the topics avoided and the lower the perceived sensitivity of these topics, the more pronounced the chilling effect becomes.

## 1.4 Give clearer explanations of some parts of the proposal:

**I added a paragraph to explain Hong Kong's turn towards authoritarianism:**

The law mandates an educational curriculum that aligns with the historical and ideological viewpoints supported by the Chinese Communist Party (CCP). These changes include revisions to textbooks and new guidelines for teachers, aiming to standardize the teaching of history and civics in a way that emphasizes aspects of governance and policy as seen from the perspective of the CCP. The selection process for judges handling national security cases has been adjusted in the judicial arena. Authority has been granted to the Chief Executive to appoint judges for these cases. This development re-

flects a significant shift in the administrative process, bringing about changes in the judicial selection that have prompted discussions on judicial independence. The modifications are seen as a move to ensure that the judiciary can effectively manage national security cases, thus shifting Hong Kong's politics and society towards authoritarianism.

**I added a paragraph to explain the research hypothesis:**

This is because the ambiguous yet stringent law may lead ordinary citizens to struggle to discern whether their discussion topics and statements may violate the law while fearing potentially severe punishments. This concern prompts excessive self-censorship when they engage with topics they perceive as sensitive, thereby decreasing participation in public discussions and collective actions, ultimately suppressing social dynamics and critical voices (Penney, 2017). Furthermore, since the law currently targets participants and activists from the 2019 Anti-Extradition Law Amendment Bill Movement, and labels popular slogans from the movement, like "Liberate Hong Kong, the revolution of our times," as terrorist statements under the National Security Law, the public may consider discussions related to social movements and collective actions as particularly sensitive and taboo. This fear leads to heightened caution and even fear regarding discussions on related topics.

**I added a paragraph to explain the role and process of manual coding:**

To facilitate the use of ChatGPT for annotating the sensitivity and categorization of posts, researchers and research assistants will manually encode the posts using the sensitivity framework developed by King, Pan, and Roberts (2013), distinguishing whether the posts are related to political discussions. Subsequently, these manually annotated posts will serve as a training dataset for fine-tuning ChatGPT. With this annotated dataset, researchers will utilize ChatGPT's contextual capabilities by providing the manually annotated posts to the ChatGPT API and setting them as the context in a conversational exchange, enabling the manually annotated posts to provide context clues for ChatGPT's annotations to assist it in annotating unlabeled posts. This process leverages ChatGPT's context embedding feature, restricting ChatGPT's output style and content by including targeted content in the input prompts, ensuring its outputs conform to the sensitivity definitions in the training dataset data. Additionally, during the initial

text annotation process, researchers can monitor ChatGPT's annotation outputs, reannotate and provide targeted explanations for any potential errors, thus creating a new contextual background to optimize ChatGPT's annotation capabilities further. After obtaining the annotated data, to ensure the dataset's accuracy and representativity are sufficient to support the complexity and demands of the research, researchers will undertake a series of detailed statistical methods for data validation and analysis. Initially, samples will be randomly selected from the entire annotated dataset, ensuring the randomness and comprehensiveness of the samples. Subsequently, researchers will re-annotate the sampled data and calculate the true positive and true negative rates to determine the consistency between machine annotations and manual reviews, thus establishing the annotation accuracy on the sample data. Finally, researchers will use degrees of freedom and p-values to test the sample accuracy to determine whether the sample accuracy can be generalized to the entire large language model annotated dataset, thereby estimating the accuracy of the large language model annotated data.

# 2 Peer's review

## 2.1 Expand the literature review on suppression:

**I added a paragraph to supplement how the internet is utilized to aid in suppression:**

For example, according to Bernot & Davies (2023), the Chinese government employs a method known as 'social sorting' to monitor LGBTQ+ activists. This process involves analyzing digital footprints and physical behaviors to categorize individuals based on their perceived threat or conformity to social norms. Social organizations or activists perceived as threatening are subjected to continuous surveillance through both overt measures, such as police monitoring, and covert ones, like the tracking of online activities and communications. Despite these pressures, LGBTQ+ communities adapt by using encrypted messaging apps and anonymous social media platforms, creatively circumventing the pervasive surveillance to continue their advocacy efforts.

## 2.2 Creating a flowchart for data collection and processing procedures:

**I created a flowchart illustrating the data collection and processing procedures:**

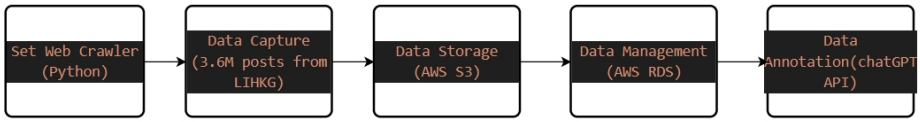


Figure 1: Flowchart of the Data Handling Process

## 2.3 Include ethical statement to ensure the safety and privacy of users on the platform:

**I added a section for the ethical statement:**

The data used in this study is sourced from publicly accessible posts on LIHKG, involving no processes that require login or following accounts, nor does it involve the disclosure of user privacy. Additionally, users on this platform are anonymous, the data could not be used to trace the real identity of the speakers through usernames and user IDs. The researchers will encrypt the data after collection to prevent the sensitivity labels assigned to posts in this research data from being used for purposes that could potentially threaten the posters. This encryption ensures that even if the data is obtained externally, the sensitivity level of each post cannot be directly determined. Researchers will only share limited data with academic collaborators

in Chicago or Hong Kong for scholarly purposes, not with third parties without prior approval. Furthermore, the researchers will store the data outside China to avoid potential risks.

## 2.4 Provide a more detailed explanation of the models used in data analysis and include a equation:

**I revised the models used for data analysis and incorporated a simulation equation:**

$$N_{t,s} = (1 - \sum_{i=1}^{p} \phi_{i,s} L^i)(1 - L)^d N_{t,s} = \left(1 + \sum_{j=1}^{q} \theta_{j,s} L^j\right) \varepsilon_{t,s} + \delta_s I_t$$

Figure 2: Simulated Autoregressive Integrated Moving Average Model

In this model, $N_{t,s}$ represents the number of posts at time $t$ with a specific sensitivity level $s$. $\phi_{i,s}$ and $\theta_{j,s}$ are the autoregressive and moving average parameters for sensitivity level $s$, reflecting the data's autocorrelation and moving average behavior. $\delta_s$ is the parameter for the intervention effect, quantifying the immediate impact of implementing the National Security Law on the number of posts of corresponding sensitivity. The intervention variable $I_t$ indicates the implementation status of the National Security Law, with a value of 0 before implementation and 1 thereafter. The stationarity of the time series is ensured through the differencing operation $(1 - L)^d$.

## 2.5 Improve transitions in the introduction:

See question 1 in the Instructor's review section.

## 2.6 Provide more detailed rules for annotating posts that may inspire collective action:

See question 3 in the Instructor's review section.

# How does digital surveillance affect public political discussions on the Internet?

Huanrui CHEN
University of Chicago

May 26, 2024

**Abstract**

Previous studies suggest that when individuals perceive their expressions as being monitored, they may engage in self-censorship to avoid legal repercussions or moral criticism, especially under ambiguous laws that carry severe penalties. However, some argue that regulatory control over speech and suppression of dissent may provoke public discontent, imparting a sense of urgency and legitimacy to resistance, thereby intensifying conflict. To empirically explore this topic, this study examines the impact of the enactment of the National Security Law by the Hong Kong government in 2020. This law intended to suppress acts such as secession, subversion, terrorism, and collusion with foreign forces, the law has been critiqued for its vague and expansive definitions. This research utilizes a large-scale data analysis approach, focusing on Hong Kong's largest online forum, LIHKG. Employing a large language model-assisted annotation design, the study analyzes three million posts using sentiment analysis and time-series methods to assess the shifts in political discussions following the law's implementation. The findings indicate that the law not only affected the topics defined as illegal activities but also significantly reduced overall political discourse on the forum. Discussions with the potential to mobilize collective movements showed the most significant decline, highlighting the potential for suppressing political mobilization through digital monitoring and vaguely defined crimes.

# 1 Introduction

Prior research suggests that when individuals perceive their expressions as being monitored, they may self-censor to avoid legal consequences or moral condemnation, especially when ambiguous laws carry severe penalties (B¨uchi et al., 2022; Penney, 2017). This phenomenon occurs because when the legal boundaries are unclear, individuals cannot confidently predict what behaviors may lead to punishment. This uncertainty, compounded by the fear of severe repercussions, drives them to limit their actions and expressions, even in cases where the behavior is legal (Stevens et al., 2023; Penney, 2017). Such monitoring can arise from various sources, whether it is the explicit knowledge that one's actions are being recorded by the government (Kappeler et al., 2023) or awareness that service providers might access one's digital traces (Strycharz & Segijn, 2024). These forms of surveillance often lead citizens to perceive an infringement of their freedoms, causing anxiety and a defensive reduction in their use of certain software or platforms (Rubel, 2007; Strycharz & Segijn, 2024). Furthermore, individuals may use less contentious expressions in potentially monitored situations to avoid the risk of incriminating themselves (Stoycheff et al., 2019).

Studies show that the chilling effects of surveillance and harsh penalties could act as a social control mechanism, suppressing potential dissent and reducing collective action (Stevens et al., 2023; Bernot & Davies, 2023). This control mechanism creates an atmosphere where individuals and organizations preemptively conform to government expectations to avoid potential repercussions. Ambiguous laws create widespread uncertainty, thereby heightening the risk associated with engaging in activities that, although potentially lawful, may be subject to legal interpretation as unlawful. This uncertainty inhibits actions that critique government policies or could potentially mobilize public sentiment against state actions" (Penney, 2017). The chilling effect is sustained by two primary factors: the laws' ambiguity and the penalties' severity. Ambiguous laws cause continuous uncertainty about the legality of actions, leading to widespread self-censorship as individuals opt for safer, less controversial expressions to avoid the harsh penalties that might follow even minor transgressions (Schauer, 1978; Solove, 2007). Additionally, the widespread nature of surveillance further amplifies this effect, as the fear of being monitored makes individuals less likely to participate in or initiate collective actions (Lyon, 2006; Marthews & Tucker, 2014). In this way, the chilling effect strategically leverages legal uncertainty and fear

of punishment to maintain social and political control, thereby preserving existing power structures. It silences potential opposition not through overt suppression but through internalizing fear and caution within the populace.

Although some research indicates that social control and repression can provoke a backlash, the outcomes are not uniformly predictable (Sullivan & Davenport, 2017). Severe forms of state repression can significantly diminish political opposition, yet, paradoxically, they can also trigger anger and resistance, fueling further mobilization (Gupta et al., 1993; Meyer & Minkoff, 2004). Franceschini and Nesossi (2018) provide a view of this phenomenon in the context of Chinese labor NGOs. Despite escalating repression under the Chinese government, which includes sophisticated legal strategies aimed at curtailing the activities and funding of NGOs, labor activists have sometimes responded with increased solidarity and tactical innovation. Instead of succumbing to fear, some labor NGOs have adapted by intensifying their advocacy for collective bargaining and labor rights, turning state repression into a catalyst for further collective action (Franceschini & Nesossi, 2018). This suggests that governments' unreasonable actions, while intended to suppress, may inadvertently create new mobilization opportunities for social movements, thus enhancing the rationality and urgency of collective action (Meyer & Minkoff, 2004). New political opportunities can emerge from increased risks as state repression changes the political landscape and alters the calculation of costs and benefits for activists. When traditional channels of political participation are closed, repression can increase the perceived legitimacy of resistance among the wider public and activists, thereby encouraging more visible and confrontational forms of collective action (Tarrow, 1998).

In the digital age, state surveillance and citizen resistance have transformed as governments leverage advanced monitoring technologies, and citizens can adopt new methods to evade control and express dissent. Governments possess more means to monitor citizens, such as analyzing internet traffic to monitor and intervene in citizens' online activities (Marczak et al., 2015) or utilizing social media surveillance to gather information (Tai & Fu, 2020). Concurrently, citizens have more flexible ways to express resistance against the government, such as posting critical content on anonymous online platforms (Tufekci, 2017) and using encrypted communication tools to evade surveillance (McCoy et al., 2008). For example, according to Bernot & Davies (2023), the Chinese government employs a method known as 'social sorting' to monitor LGBTQ+ activists. This process involves analyzing digital footprints and physical behaviors to categorize individuals based on

their perceived threat or conformity to social norms. Social organizations or activists perceived as threatening are subjected to continuous surveillance through both overt measures, such as police monitoring, and covert ones, like the tracking of online activities and communications. Despite these pressures, LGBTQ+ communities adapt by using encrypted messaging apps and anonymous social media platforms, creatively circumventing the pervasive surveillance to continue their advocacy efforts. This discrepancy suggests that governments may implement control and deterrence through more indirect and non-violent means, while citizens can protest using more adaptable methods. This dynamic interaction between state surveillance and citizen resistance underscores the complexity of modern protests, suggesting that traditional models of understanding political activism may need to be updated to capture the nuances of digital resistance effectively.

The study by Kobayashi, Song, and Chan (2021) reveals changes in protest dynamics, showing that although visible street protests have declined following the enactment of stringent repressive laws, covert support for opposition movements has not only persisted but sometimes has even intensified. The researchers used conjoint experiments to assess public sentiment and strategic preferences under stringent laws. In these experiments, conducted before and after implementing the Hong Kong National Security Law, a repressive policy, participants were presented with various hypothetical scenarios featuring diverse policy attributes related to governmental responses to protests. Each policy profile varied systematically in attributes such as the severity of law enforcement actions and the rights to public assembly, allowing researchers to isolate which attributes influenced participants' preferences for supporting or opposing policy measures. By analyzing the choices made by participants, Kobayashi, Song, and Chan (2021) could infer a shift in protest dynamics. They noted that, after the enactment of the Hong Kong National Security Law, experiment participants showed a decreased preference for overt protest actions, which may reflect the increased risks associated with visible dissent under stricter laws. However, simultaneously, the experiments also indicated that when participants faced the condition of a stringent repressive policy, those participating in the experiments after the enactment of the Hong Kong National Security Law emotionally displayed more support and identification with the opposition. The persistence and intensification of this support occur because repression while dampening visible dissent, does not necessarily diminish deeply held grievances and political convictions. Instead, repression can reinforce the resolve of those who perceive the

government's actions as unjust, prompting them to continue supporting opposition demands through more covert or less overtly confrontational means. However, one question left unanswered by this study concerns whether public discontent translates into behaviors beyond street protests when repression does not diminish citizens' support for the opposition. This is because the experiments in this study only measured changes in public opinion, but did not collect empirical data to assess changes in behavior. This leaves the potential for public support of the opposition to translate into specific actions unexplored. A potential behavior is that adaptive strategies come into play when individuals seek alternative ways to express dissent. Encrypted communications and anonymous online platforms become tools for covert resistance, allowing individuals to support the cause of the opposition through online discussions without exposing themselves to the direct risks associated with actual protests (Tufekci, 2017).

However, some studies give the opposite opinion, suggesting that censorship and the enactment of repressive policies can similarly limit protest behavior online. A study based on a large dataset addresses this issue, suggesting that national-level online censorship and surveillance are strongly negatively correlated with political participation. Although there is a robust positive relationship between online news and political participation, online censorship and surveillance significantly inhibit this positive effect by increasing the costs of political engagement. In countries with strict censorship, it becomes more difficult for people to access information that motivates collective action, thereby suppressing public political activity (Chan et al., 2022). Although some cases indicate that governmental repression may trigger a backlash effect, increasing political participation, in most instances, the constant threat posed by surveillance effectively curtails public political engagement (Chan et al., 2022). However, the data used in this article are self-reported survey data, which may be biased due to the sensitive nature of the issue. Moreover, the data in the article measure citizens' online political engagement through three dimensions: signing electronic petitions, encouraging others to take political action, and organizing political activities, but do not include a sufficiently broad range of political participation forms, such as discussions of daily political news or issues and criticism that does not aim to initiate collective movements. This implies that current research lacks effective primary data to assess the impact of repressive policies and speech control on online political discussions. It raises the question of whether online censorship and surveillance broadly suppress all forms of online political

participation, including everyday political discussions and grievances, or only affect discussions that might mobilize collective movements. Alternatively, it might be that such repression incites citizens' dissatisfaction, prompting them to use the Internet as a medium to express their protest.

To investigate the impact of suppression and surveillance in the digital age on dissent, this proposal examines the changes in online discourse in Hong Kong before and after the implementation of the National Security Law, utilizing all post data from LIHKG, the largest online forum in Hong Kong, to comprehensively measure citizens' real behaviors in the digital space. The 2019 Anti-Extradition Law Amendment Bill Movement in Hong Kong is considered a paradigm of digital activism, with protestors leveraging social media to organize rallies and disseminate information (Zhong & Zhou, 2022). Throughout this process, social media and online platforms served as tools for mobilization and communication and as arenas for an information war, presenting this decentralized social movement in a process that attracted international attention, leaving a wealth of data on platforms such as LIHKG (Ku, 2020). However, the movement concluded in 2020 following the Chinese government's enactment of the Hong Kong National Security Law on June 30, 2020. The law claims to safeguard national security within the Hong Kong Special Administrative Region by targeting acts of secession, subversion of state power, terrorist activities, and collusion with foreign forces. Under this legislation, a significant number of dissenters and political activists who participated in the movement were prosecuted by the Hong Kong government based on their speech. Research indicates that enacting the National Security Law curtailed street protests (Kobayashi et al., 2021). In addition, the National Security Law has modified Hong Kong's educational and judicial systems. The law mandates an educational curriculum that aligns with the historical and ideological viewpoints supported by the Chinese Communist Party (CCP). These changes include revisions to textbooks and new guidelines for teachers, aiming to standardize the teaching of history and civics in a way that emphasizes aspects of governance and policy as seen from the perspective of the CCP (U.S.-China Economic and Security Review Commission, 2021). The selection process for judges handling national security cases has been adjusted in the judicial arena. Authority has been granted to the Chief Executive to appoint judges for these cases. This development reflects a significant shift in the administrative process, bringing about changes in the judicial selection that have prompted discussions on judicial independence. The modifications are seen as a move to ensure that the judiciary can effec-

tively manage national security cases, thus shifting Hong Kong's politics and society towards authoritarianism (U.S.-China Economic and Security Review Commission, 2021).

The unique scenario of Hong Kong's transition from a democratic to an authoritarian society, along with the control and surveillance imposed on its citizens following the law's implementation, makes Hong Kong a pertinent case study for examining the effects of digital surveillance on citizen dissent and resistance in the digital era. This discussion explores whether the National Security Law, as a legal mechanism imposing surveillance pressure on citizens, can suppress online dissent as successfully as it did street protests or whether it will further ignite public anger, prompting them to express their grievances in the anonymous spaces of the internet. This study hypothesizes that due to the vague yet severe characteristics of the National Security Law, it is likely to trigger a widespread chilling effect, thus suppressing all forms of citizens' engagement with political topics, particularly reducing the volume of content that could trigger collective action. This is because the ambiguous yet stringent law may lead ordinary citizens to struggle to discern whether their discussion topics and statements may violate the law while fearing potentially severe punishments. This concern prompts excessive self-censorship when they engage with topics they perceive as sensitive, thereby decreasing participation in public discussions and collective actions, ultimately suppressing social dynamics and critical voices (Penney, 2017). Furthermore, since the law currently targets participants and activists from the 2019 Anti-Extradition Law Amendment Bill Movement, and labels popular slogans from the movement, like "Liberate Hong Kong, the revolution of our times," as terrorist statements under the National Security Law, the public may consider discussions related to social movements and collective actions as particularly sensitive and taboo. This fear leads to heightened caution and even fear regarding discussions on related topics.

# 2 Methods and Data

This study plans to use the posting on LIHKG, the most famous online forum in Hong Kong, focusing on the period before and after the implementation of the Hong Kong national security law as the research object. During the Anti-Extradition Law Amendment Bill Movement in 2019, LIHKG was a pivotal platform for participant interaction and information dissemination.

The survey data indicated that 33% of respondents actively participated in discussions on the forum by replying to posts related to the movement, while 10.8% contributed by creating new posts (Ku, 2020). Using this most active online forum with representative user groups and characteristics as samples aims to minimize sample selection bias and comprehensively understand the dynamics of online discourse in Hong Kong. "The study employs LIHKG, a widely-used online forum in Hong Kong, as its primary data source. This forum was chosen for its active user engagement and broad demographic reach, which could reduce sample selection bias. The aim is to capture a representative snapshot of online discourse and to understand communication patterns in Hong Kong's socio-political environment. To address potential variability in political discussions that could be influenced by external events or seasonal trends, the study encompasses all LIHKG posts as the dataset, consisting of approximately 3,600,000 posts. This approach is designed to provide a robust dataset that supports the creation of a control group, which is fundamental for conducting a rigorous and structured analysis. Based on Amazon's cloud platform's computing resources and storage capabilities, an automated parallel web crawler program has been designed using Python and Scrapy frameworks. This program can efficiently capture a large amount of data quickly, ensuring data integrity and update speed, thus making data analysis more accurate and timely. To manage this large amount of data, Amazon S3 is used as a data storage solution to facilitate the processing and access of massive forum posts. Additionally, this study utilizes Amazon RDS for data storage and management to ensure structured data and query efficiency.

Data from the end of April 2024 to the end of January 2022 was crawled within two weeks, totaling approximately 800,000 posts. Each post includes the poster's ID, posting time, posting section, post content, post title, likes, and dislikes. Although the website categorizes posts and designates a specific section for political discussions, this study aims not only to capture the differences between political and non-political discussions but also to identify political discussion posts that may mobilize collective movements, as well as other types of everyday political discussions or criticisms because this study would like to explore the chilling effect, which implies that due to vague and harsh laws, citizens might choose to avoid participating in and discussing topics they consider sensitive out of fear of potential punishment. The more extensive the topics avoided and the lower the perceived sensitivity of these topics, the more pronounced the chilling effect becomes. This research will

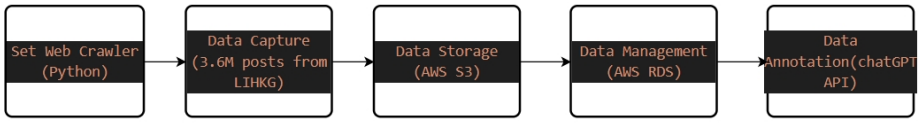


Figure 1: Flowchart of the Data Handling Process

classify the potential sensitivity of posts according to the coding method adopted by King, Pan, and Roberts (2013). Based on their analysis of the frequency and context of deleted posts, King, Pan, and Roberts (2013) summarized how the Chinese government censors online posts and categorized the sensitivity of online posts into five levels:

- **Posts Mobilizing Collective Action:** Posts that directly call for or could lead to collective action, such as protests, demonstrations, or other public gatherings. Since these posts can immediately trigger public collective actions, they are usually the main target of censorship.
- **Pornography and Criticism of Censors:** These posts are consistently censored but not as aggressively as those with collective action potential. Criticism of censorship mechanisms or entities and posts containing pornography fall into this category.
- **Direct Criticism of government:** Posts that discuss government policies and political reforms or criticize the government. These posts may not directly mobilize action, and although they are considered relatively sensitive, they will not necessarily be censored and deleted.
- **General Political Discussion:** Less sensitive general political discussions that do not address criticism of government or advocate collective action. These might include discussions on political theory, non-contentious political commentary, or historical political analysis.

- **Non-Political Content:** Posts with no political content or relevance to sensitive social issues are minimally sensitive. These would include everyday discussions, entertainment news, or trivial content.

Given the need to annotate large volumes of data and Cantonese's relatively niche linguistic type, this study employs a large language model to annotate the posts. Specific large language models have demonstrated excellent performance in recognizing Cantonese data; research shows that ChatGPT-4.0 achieves a sentiment annotation accuracy rate of 95.3% on Cantonese data (Fu et al., 2023). To facilitate the use of ChatGPT for annotating the sensitivity and categorization of posts, researchers and research assistants will manually encode the posts using the sensitivity framework developed by King, Pan, and Roberts (2013), distinguishing whether the posts are related to political discussions. Subsequently, these manually annotated posts will serve as a training dataset for fine-tuning ChatGPT. With this annotated dataset, researchers will utilize ChatGPT's contextual capabilities by providing the manually annotated posts to the ChatGPT API and setting them as the context in a conversational exchange, enabling the manually annotated posts to provide context clues for ChatGPT's annotations to assist it in annotating unlabeled posts. This process leverages ChatGPT's context embedding feature, restricting ChatGPT's output style and content by including targeted content in the input prompts, ensuring its outputs conform to the sensitivity definitions in the training dataset data. Additionally, during the initial text annotation process, researchers can monitor ChatGPT's annotation outputs, reannotate and provide targeted explanations for any potential errors, thus creating a new contextual background to optimize ChatGPT's annotation capabilities further. After obtaining the annotated data, to ensure the dataset's accuracy and representativity are sufficient to support the complexity and demands of the research, researchers will undertake a series of detailed statistical methods for data validation and analysis. Initially, samples will be randomly selected from the entire annotated dataset, ensuring the randomness and comprehensiveness of the samples. Subsequently, researchers will re-annotate the sampled data and calculate the true positive and true negative rates to determine the consistency between machine annotations and manual reviews, thus establishing the annotation accuracy on the sample data. Finally, researchers will use degrees of freedom and p-values to test the sample accuracy to determine whether the sample accuracy can be generalized to the entire large language model annotated dataset, thereby

estimating the accuracy of the large language model annotated data.

After collecting the data, this study plans to use Python's Dask package for batch preprocessing to ensure that the posts are converted into a uniform format. The research dataset will include whether each post is a political discussion, its sensitivity level, and the posting date. Subsequently, the data will be reorganized based on the date, providing the frequency of posts on political/non-political topics and different sensitivity levels for each day from the website's inception in December 2016 until April 2024. This study uses the Autoregressive Integrated Moving Average model (ARIMA) to analyze the time series data of the number and sentiment of posts before and after implementing the National Security Law. The ARIMA is a statistical analysis model that predicts and describes points within time series data. It combines autoregressive (AR) and moving average (MA) models and includes differencing (I) to achieve data stationarity, allowing the model to handle potential non-stationary trends effectively. By differencing, we eliminate the seasonal and trend components in the data, thus focusing on short-term fluctuations. During model establishment, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) will be used to determine the parameters (p, d, q) of the ARIMA model, thereby precisely capturing the intrinsic dynamics of the data. Next, the data will be divided into two phases: before and after the implementation of the National Security Law, based on the specific date of its enactment. This division is fundamental for conducting an intervention analysis, as it allows us to accurately assess the impact of this law on the dynamics of online discussions. In implementing the intervention analysis, this research will construct a time series model that includes an intervention variable to analyze the immediate and long-term effects of the law's implementation. The intervention variable (typically represented as a binary variable of 0 and 1) will be inserted at the date of the National Security Law's implementation, where before the law, it is 0. After the law, it is 1. Specifically, this study will use the Transfer Function Model in the intervention time series analysis to assess the impact of the National Security Law on social media discussions. This model considers the impact of the intervention event and analyzes the behavioral changes in the data before and after the intervention. The model will include estimates of the intervention effect, evaluating the specific impact of the National Security Law on the frequency of different topics and different sensitivity levels through response functions. Additionally, impulse response analysis will be conducted to observe changes in post characteristics following the law's im-

plementation, including shifts in discussion topics and sensitivity levels.

$$N_{t,s} = (1 - \sum_{i=1}^{p} \phi_{i,s} L^i)(1 - L)^d N_{t,s} = \left(1 + \sum_{j=1}^{q} \theta_{j,s} L^j\right) \varepsilon_{t,s} + \delta_s I_t$$

Figure 2: Simulated Autoregressive Integrated Moving Average Model

In this model, $N_{t,s}$ represents the number of posts at time $t$ with a specific sensitivity level $s$. $\phi_{i,s}$ and $\theta_{j,s}$ are the autoregressive and moving average parameters for sensitivity level $s$, reflecting the data's autocorrelation and moving average behavior. $\delta_s$ is the parameter for the intervention effect, quantifying the immediate impact of implementing the National Security Law on the number of posts of corresponding sensitivity. The intervention variable $I_t$ indicates the implementation status of the National Security Law, with a value of 0 before implementation and 1 thereafter. The stationarity of the time series is ensured through the differencing operation $(1 - L)^d$.

# 3 Preliminary Work

This study successfully designed a large-scale parallel web crawling pipeline based on Amazon Cloud, which collected approximately 800,000 rows of data within two weeks, accounting for about a quarter of the total data volume (3,695,252). The researchers tested the data annotation capabilities of ChatGPT-4.0 and found that it performed well on Cantonese data even without fine-tuning and contextual prompts. Researchers sampled 25 posts each from the entertainment and current affairs sections of LIHKG posted on May 21, 2024, and manually annotated the sensitivity of the posts and whether they belonged to political discussions. Subsequently, the researchers used ChatGPT-4.0 to classify the sensitivity and annotate whether the 50 posts were political discussions. They then compared these annotations with the manually annotated results. The findings revealed that ChatGPT-4.0 had a 100% accuracy rate in identifying whether posts were political discussions and a 96% accuracy rate in assessing the sensitivity of the posts. Additionally, the researchers tested the cost of using ChatGPT-4o turbo for annotation and found that the feasible cost to complete the entire data annotation was about $600, a manageable expense as the researchers plan to

fund the study with a $30,000 scholarship provided in the second year of the University of Chicago Computational Social Science program.

Ideally, this study aims to demonstrate the following hypothesis/concept: If data analysis concludes that the frequency of online political discussions significantly decreases, and discussions on sensitive topics are reduced, then the hypothesis is confirmed. This result would support previous research that the sensation of being monitored, especially under ambiguous laws with harsh penalties, compels individuals to self-censor. Furthermore, if the research findings show that discussions potentially mobilizing collective action, specifically those of the highest sensitivity, have significantly decreased, this would further confirm the hypothesis that the National Security Law serves as a deterrent to expressing opinions that might be seen as threatening public order or national security. This aligns with the theoretical framework that ambiguous laws disproportionately affect communication about public concerns. This could lead to broader suppression of free expression, weakening democratic participation. If data shows that overall participation in political discussions remains stable, the study plans to analyze the sensitivity of all posts further. If posts with the potential to trigger collective action significantly decrease, but posts criticizing the government and the National Security Law increase, this might reflect how legal restrictions reshape communication strategies. Individuals and groups might choose less direct forms of resistance or expression, adjusting their behavior to reduce risk while attempting to voice dissent more covertly or subtly.

# 4 Proposed Timeline and Feasibility Assessment

Regarding data collection, given the progress of collecting a quarter of the data in two weeks, it appears feasible to complete all data collection by the end of June to early July. From a budgetary perspective, the cost of utilizing ChatGPT-4.0 turbo for annotation is approximately $600, manageable within allocated funds. For the part of data analysis, having taken a course in causal inference this quarter, I plan to further my studies in quantitative analysis and causal inference next quarter, focusing on time series and text analysis methods. I also intend to self-study these topics during the summer to enhance my analytical skills. For the project's theoretical

framework, I will continuously revise my research proposal before completing data collection. In addition to receiving feedback and making revisions in the MACS30200 course, I plan to send the first version of my research proposal (the version submitted on May 24) to scholars who specialize in Hong Kong social movements, given that both my undergraduate institution and the case study of the proposal are based in Hong Kong. This step could give me valuable advice on refining the theory and approach. Furthermore, during spring break, a scholar from the Sociology Department at my undergraduate university expressed significant interest in my research, offering his guidance and potentially connecting me with other academics who could assist further. These wonderful potential mentors, including Dr. Shihlin Jia, Dr. Molly Over-West, and Dr. Nick Feemster, are familiar with my research program and interests, which could help to modify my research approach.

The timeline for this project is structured to complete data collection by early July, finalize the research proposal revision by the same time, complete data analysis by mid-August, and finish the project's first draft by the end of September.

# 5 Ethical Statement

The data used in this study is sourced from publicly accessible posts on LIHKG, involving no processes that require login or following accounts, nor does it involve the disclosure of user privacy. Additionally, users on this platform are anonymous, the data could not be used to trace the real identity of the speakers through usernames and user IDs. The researchers will encrypt the data after collection to prevent the sensitivity labels assigned to posts in this research data from being used for purposes that could potentially threaten the posters. This encryption ensures that even if the data is obtained externally, the sensitivity level of each post cannot be directly determined. Researchers will only share limited data with collaborators in Chicago or Hong Kong for academic purposes, not with third parties without prior approval. Furthermore, the researchers will store the data outside China to avoid potential risks.

# 6 Project Repository

Visit the project repository on GitHub at https://github.com/hchen0628/MACS30200.

Please see this link for the currently available data: https://docs.google.com/spreadsheets/d/1HV38-rp1jabZFsbEgiiZgAtbJaHLWcHw/edit?usp=sharing&ouid=100247803780513432805&rtpof=true&sd=true.

# References


Bernot, A., & Davies, S. E. (2024). The "fish tank": Social sorting of lgbtq+ activists in china. *INTERNATIONAL FEMINIST JOURNAL OF POLITICS*, *26*(2), 351–374. https://doi.org/10.1080/14616742.2023.2261948

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Büchi, M., Festic, N., & Latzer, M. (2022). The chilling effects of digital dataveillance: A theoretical model and an empirical research agenda. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517211065368

Chan, M., Yi, J., & Kuznetsov, D. (2022). Government digital repression and political engagement: A cross-national multilevel analysis examining the roles of online surveillance and censorship. *The International Journal of Press/Politics*, *29*(2), 371–393. https://doi.org/10.1177/19401612221117106

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Franceschini, I., & Nesossi, E. (2018). State repression of chinese labor ngos: A chilling effect? *The China Journal*, *80*, 111–131. https://doi.org/10.1086/698275

Fu, Z., Hsu, Y., Chan, C., Lau, C., Liu, J., & Yip, P. (2023). Efficacy of chatgpt in cantonese sentiment analysis: A comparative study [Preprint]. *Journal of Medical Internet Research*, *26*. https://doi.org/10.2196/51069

Gupta, D. K., Singh, H., & Sprague, T. (1993). Government coercion of dissidents: Deterrence or provocation? *Journal of Conflict Resolution*, *37*(2), 301–339. https://doi.org/10.1177/0022002793037002004

Kappeler, K., Festic, N., & Latzer, M. (2023). Dataveillance imaginaries and their role in chilling effects online. *International Journal of Human-Computer Studies*, *179*, 103120. https://doi.org/10.1016/j.ijhcs.2023.103120

Kobayashi, T., Song, J., & Chan, P. (2021). Does repression undermine opposition demands? the case of the hong kong national security law. *Japanese Journal of Political Science*, *22*(4), 268–286. https://doi.org/10.1017/S1468109921000256

Ku, A. S. (2020). New forms of youth activism—hong kong's anti-extradition bill movement in the local-national-global nexus. *Space and Polity*, *24*(1), 111–117.

Marczak, B., Scott-Railton, J., Marquis-Boire, M., & Paxson, V. (2015). When governments hack opponents: A look at actors and technology. *USENIX Security Symposium*.

McCoy, D., Bauer, K., Grunwald, D., Kohno, T., & Sicker, D. (2008). Shining light in dark places: Understanding the tor network. *Privacy Enhancing Technologies: 8th International Symposium, PETS 2008 Leuven, Belgium, July 23-25, 2008 Proceedings*, 63–76.

Meyer, D. S., & Minkoff, D. C. (2004). Conceptualizing political opportunity. *Social Forces*, *82*(4), 1457–1492. https://doi.org/10.1353/sof.2004.0082

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Penney, J. W. (2017). Internet surveillance, regulation, and chilling effects online: A comparative case study. *Internet Policy Review*, *6*(2). https://doi.org/10.14763/2017.2.692

Prinz, J. (2010). The moral emotions. In P. Goldie (Ed.), *The oxford handbook of philosophy of emotion* (Online ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0024

Rubel, A. (2007). Privacy and the usa patriot act: Rights, the value of rights, and autonomy. *Law & Philosophy*, *26*, 119.

Schauer, F. (1978). *Fear, risk, and the first amendment: Unraveling the chilling effect*. Beacon Press.

Stevens, A., Fussey, P., Murray, D., Hove, K., & Saki, O. (2023). 'i started seeing shadows everywhere': The diverse chilling effects of surveillance

in zimbabwe. *Big Data & Society*, *10*(1), 205395172311586. https://doi.org/10.1177/20539517231158631

Stoycheff, E., Liu, J., Xu, K., & Wibowo, K. (2019). Privacy and the panopticon: Online mass surveillance's deterrence and chilling effects. *New Media & Society*, *21*(3), 602–619. https://doi.org/10.1177/1461444818801317

Strycharz, J., & Segijn, C. (2024). Chilling effects as a result of corporate surveillance in digital communication: A comparison between american and dutch media users. *International Journal of Communication*, *18*, 320–343.

Sullivan, C. M., & Davenport, C. (2017). The rebel alliance strikes back: Understanding the politics of backlash mobilization. *Mobilization: An International Quarterly*, *22*(1), 39–56. https://doi.org/10.17813/1086-671X-22-1-39

Tai, Y., & Fu, K. W. (2020). Specificity, conflict, and focal point: A systematic investigation into social media censorship in china. *Journal of Communication*, *70*(6), 842–867.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

U.S.-China Economic and Security Review Commission. (2021). Hong kong's government embraces authoritarianism. https://www.uscc.gov/annual-report/2021-annual-report-congress

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert. *arXiv preprint*, *arXiv:2302.10198*. https://doi.org/10.48550/arXiv.2302.10198

Zhong, Z.-J., & Zhou, J.-L. (2022). Trends creators and discoverers in online political discussion: A study based on the golden and lihkg forums in hong kong. *Communication and Society*, *59*, 47–79.