Investigating the Role of Silent Users on Social Media

by

Mansooreh Karami

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved August 2023 by the
Graduate Supervisory Committee:

Huan Liu, Chair
Arunabha Sen
Hasan Davulcu
Michelle V. Mancenido

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

Social media platforms provide a rich environment for analyzing user behavior. Recently, deep learning-based methods have been a mainstream approach for social media analysis models involving complex patterns. However, these methods are susceptible to biases in the training data, such as participation inequality. Basically, a mere 1% of users generate the majority of the content on social networking sites, while the remaining users, though engaged to varying degrees, tend to be less active in content creation and largely silent. These silent users consume and listen to information that is propagated on the platform. However, their voice, attitude, and interests are not reflected in the online content, making the decision of the current methods predisposed towards the opinion of the active users. So models can mistake the loudest users for the majority. To make the silent majority heard is to reveal the true landscape of the platform.

In this dissertation, to compensate for this bias in the data, which is related to user-level data scarcity, I introduce three pieces of research work. Two of these proposed solutions deal with the data on hand while the other tries to augment the current data. Specifically, the first proposed approach modifies the weight of users' activity/interaction in the input space, while the second approach involves reweighting the loss based on the users' activity levels during the downstream task training. Lastly, the third approach uses large language models (LLMs) and learns the user's writing behavior to expand the current data. In other words, by utilizing LLMs as a sophisticated knowledge base, this method aims to augment the silent user's data.

To my dear parents and my best- and first-ever teachers, ***Khodabakhsh Karami*** *and* ***Mahin Moorizadeh****, whose unwavering support has made it possible for me to chase my dreams. Their love and guidance have been my guiding stars, illuminating my path through this ethereal journey.*

I extend my heartfelt dedication to my beloved siblings, ***Nasibeh****,* ***Honeyeh****, and* ***Ahmadreza****, who have stood by my side through thick and thin. Their constant presence and encouragement have been the gentle breeze that carried me forward. Without their assistance and companionship, this wonderful odyssey would have remained but a distant reverie.*

And, to ***God****...*

ACKNOWLEDGEMENTS

With profound gratitude, I express my heartfelt appreciation for the invaluable support and guidance extended by my esteemed advisor, the eminent Dr. Huan Liu. Under his exceptional mentorship, I have been enriched with a myriad of skills and insights that have not only enhanced my academic pursuits but have also carved a rewarding career path for me. His kindness and generosity in offering suggestions and support for various aspects of my life have been immeasurable. I am still astounded by his remarkable capacity to swiftly comprehend intricate ideas and establish profound connections amid seemingly unrelated topics.

I extend my sincere appreciation to my committee members: Dr. Arunabha Sen, Dr. Hasan Davulcu, and Dr. Michelle Mancenido. Their invaluable suggestions have left an indelible impact on my academic journey. The thought-provoking inquiries they posed during and after my proposal defense have not only shaped the trajectory of my present research but also serve as a guiding light for my future endeavors.

I am grateful for the exceptional opportunities I had as an intern at Microsoft and Spotify USA, working alongside brilliant colleagues and mentors. To Yuting Jia, Kiyoung Yang, Anqi Bao, from Microsoft Dynamics 365, and Oscar Celma, Andrew Asman, Tymur Maryokhin, Tahora Nazer, from Spotify PZN, I extend my heartfelt thanks for making my experiences in the new environment enjoyable and rewarding. Your support played a pivotal role in contributing to exciting projects and broadening my knowledge.

I would link to thank my friends and colleagues at ASU and the Data Mining and Machine Learning Lab, whose constant support and encouragement were instrumental in this journey. In particular, my co-authors, Raha Moraffah, Paras Sheth, and David Mosallanezhad. Special thanks to Tahora Nazer, who taught me the ropes

of research. To my saviors from academic hardship, Dr. Douglas Sandy, Dr. Myra Schatzki, Dr. H. Russell Bernard, Sachin Grover, Sarath Sreedharan, and Tyler Black, I am forever grateful.

To the Shahlayan group, Tahora Nazer, Faezeh Kalantari, Zahra Zahedi, Fatemeh Zahedi, and Rana Pourmohamad, and to my housemate of 6-month but forever sister, Parisa Mahmoudidaryan, I will never forget our friendship. You all emerged like a celestial guardian, silently supporting me through my darkest hours when I least expected and yet desperately needed your presence.

At last, without my family, I would have never made it this far. They believed in me and nurtured my potential beyond the horizon of possibilities. My very essence is indebted to you − Dad, Mom, Nasibeh, Honeyeh, and Ahmadreza.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Why are you so afraid of silence,
Silence is the root of everything.
If you spiral into its void
a hundred voices will thunder
messages you long to hear.

– Jalāl al-Dīn Muḥammad Rūmī

Social media has become an integral part of our modern society, shaping the way we connect, communicate, and share information on a global scale. Platforms like Facebook, Twitter, and LinkedIn have revolutionized the dynamics of social interaction, enabling individuals to express themselves, form communities, and engage in discussions with unprecedented ease. The rise of these social media platforms has led to an explosion of user-generated content, creating an immense digital landscape filled with diverse opinions, experiences, and ideas. This vast volume of data serves as a treasure trove of information for researchers from a wide range of disciplines, including behavioral science, social science, and computer science. These researchers employ various techniques to extract patterns and trends from user-generated content. The knowledge gained from this analysis is frequently used to develop personalized recommendations, enhance customer experience, and predict user behavior.

Despite the ongoing efforts to analyze social media data, it is important to acknowledge that these models are susceptible to various biases, one of which is *participation inequality*. Participation inequality refers to the phenomenon where a small subset of users disproportionately contributes to the content creation activity within social networks. This pattern has been consistently observed among Online Social

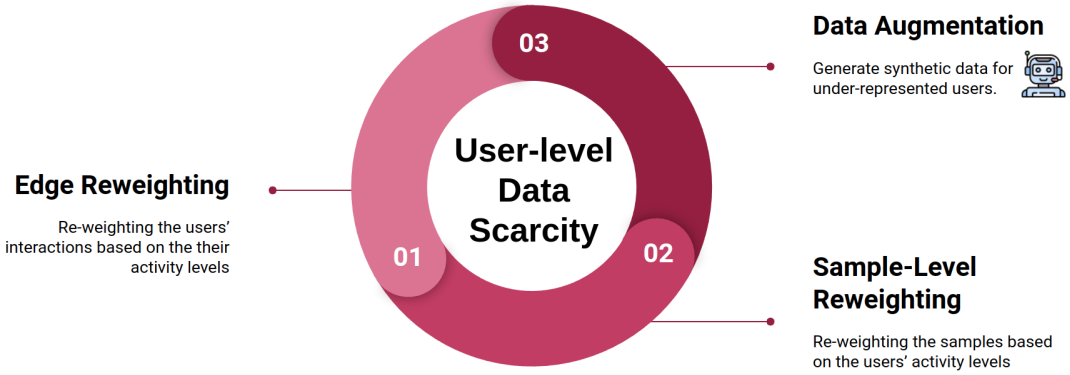
Network (OSN) users and can be further categorized into three distinct types. Firstly, we have the *lurkers* who make up approximately 90% of OSN users. Lurkers are individuals who rarely engage in content creation on social media platforms, accounting for only about 1% of the total postings. They primarily consume the content generated by others without actively contributing. Secondly, *engagers* comprise around 9% of social media users. Engagers are individuals who occasionally participate in content creation, contributing to approximately 9% of the total postings. They are more active than lurkers but still do not account for a substantial portion of the content created. Lastly, *contributors* constitute merely 1% of OSN users but are responsible for an overwhelming 90% of the content generated on social media platforms. These contributors are highly active and play a pivotal role in shaping the content landscape.

This pattern, often referred to as the *90-9-1 Rule* or the *1% Rule* by web usability experts [Nielsen, 2006], demonstrates the inherent biases present in the data used for current social media analysis applications. Relying solely on observable actions within these platforms might not necessarily reflect the true intentions or overall engagement levels of users within the community or platform as a whole [Kokkodis *et al.*, 2020].

Recognizing the limitations of existing approaches, the aim of this dissertation is to delve into the characteristics of silent users, specifically the lurkers, and explore methods for incorporating their perspectives into deep learning models that analyze social media data. By shedding light on the silent majority, this research seeks to bridge the gap and ensure a more comprehensive understanding of user behavior and engagement within online communities. By integrating these perspectives into existing deep learning models, we aim to develop a more nuanced and accurate understanding of social media dynamics, thus mitigating the biases introduced by participation inequality.

In this dissertation, we present three research endeavors. Among these, two of the

suggested methodologies deal with the data on hand while the other tries to augment the current data. Specifically, the first proposed strategy entails the modification of user activity/interaction weights within the input domain. The second approach adjusting the loss based on the users' activity levels during the downstream task training. Finally, the third technique employs large language models (LLMs) to comprehend user writing patterns and extend the current dataset. In other words, by harnessing LLMs as an advanced knowledge base, this approach seeks to augment the data of silent users (refer to Figure **??** for an overview of the three methods).



An Overview of Three Research Solutions Introduced in This Dissertation That Compensate for the Biases Related to User-Level Data Scarcity on Social Media.

1.1 Research Challenges

To study the role of silent users on social media, we are faced with several challenges:

- **Data scarcity and lack of identifiable information**. The lack of identifiable information is a significant challenge when investigating the role of silent users on social media. Silent users, by definition, refrain from actively participating or engaging in discussions on social media platforms. As a result, they leave behind minimal digital footprints, making it difficult to gather relevant information

about them. Unlike contributors who leave comments, post content, or interact with others, silent users may have limited or, in extreme cases, no public activity that can be used to identify them.

The absence of identifiable information hampers researchers' ability to understand the motives, affiliations, or intentions of silent users. Without such information, it becomes challenging to piece together a comprehensive picture of their behaviors, preferences, or patterns of engagement. Researchers may struggle to determine if silent users are intentionally remaining silent, or if they are simply disinterested or inactive on the platform.

Moreover, the absence of identifiable information also limits the potential for conducting targeted investigations. Researchers often rely on user profiles, connections, or public interactions to identify relevant individuals for further study. However, silent users typically provide little or no personal information on their profiles, making it challenging to differentiate them from other users or identify potential patterns of behavior.

Addressing this issue requires alternative approaches and researchers may need to rely on indirect indicators to gain insights into silent users.

- **Churners versus Lurkers**. The challenge of distinguishing lurkers from churners (i.e., inactive accounts) on social media platforms is a significant hurdle when investigating the role of these users. While lurkers intentionally choose not to engage or participate in discussions, inactive accounts may simply be the result of users abandoning the platform or losing interest over time. This distinction is crucial because lurkers may have specific motivations or behaviors that are different from those who are inactive.

  Identifying lurkers requires a deeper analysis of their patterns of behavior. Un-

like active users who may have a visible digital footprint, lurkers typically leave minimal traces of their presence. In contrast, inactive accounts may have a history of past activity, such as previous posts or interactions, before becoming dormant.

To overcome this challenge, investigators need to consider various factors to differentiate lurkers from inactive accounts. One approach is to examine the duration of inactivity. Lurkers are characterized by their prolonged absence from active engagement, often spanning weeks, months, or even years. Inactive accounts, on the other hand, may have a shorter period of inactivity, indicating a more recent disengagement from the platform.

Another consideration is the context in which the account was created. Lurkers may intentionally create accounts to observe or monitor social media activity without actively participating. They may follow specific individuals or groups, consume content, and gather information while avoiding direct engagement. In contrast, inactive accounts might have been created with the intent to participate initially but became dormant over time due to personal reasons or disinterest.

Moreover, investigating the account's connections and interactions can provide insights into its status. Silent users may follow or be followed by a limited number of accounts, indicating a more selective engagement strategy. Inactive accounts, on the other hand, may have a broader network of connections but lack recent activity. Analyzing the nature of these connections, such as the quality of relationships or shared interests, can also help in distinguishing lurkers from inactive accounts.

- **Difficulty in determining the user's intent or misinterpreting the si-**

lence. Intent refers to the underlying motivations, goals, or purposes that drive a person's actions or behaviors. Without explicit communication or active participation, researchers or investigators must rely on indirect cues or assumptions to infer intent.

In the absence of explicit communication, it becomes easy to misinterpret or speculate about the intentions of silent users. For example, a silent user who does not engage in discussions or share content may be seen as disinterested, uninvolved, or even suspicious. However, it is essential to consider other possibilities such as personal preference, privacy concerns, or a desire to observe rather than actively participate. Misinterpreting intent can lead to inaccurate conclusions and misrepresentation of silent users.

It is crucial to approach the issue of silence with caution, avoiding assumptions or stereotypes. Misinterpreting silence can lead to unfair judgments, false accusations, or misrepresentation of individuals, undermining the integrity of any investigation.

To address these challenges, investigators need to consider alternative methods of gathering information or understanding intent. This may involve analyzing patterns of behavior, exploring non-public data if available with appropriate permissions, or conducting interviews or surveys with a sample of silent users to gain insights into their motivations and preferences.

1.2 Contributions

The main contributions of this dissertation can be summarized as follows:

- Investigating previously unexplored issues related to comprehending the actions of silent users on social media.

- Developing systematic methodologies for leveraging signals provided by silent users based on established social theories, thereby enhancing the accuracy of fake news detection.

- Expanding upon currently available datasets to account for the type of users based on their activity and performing comprehensive experiments to validate the efficacy of the proposed frameworks.

- Designing and Evaluating different prompts for large language models specific to silent user data generation for the task of ideology detection.

1.3 Organization of the Dissertation

In this dissertation, we provide an introduction to the factors that drive online participation, categorize online users based on their activity, and analyze how to incorporate these categories into machine learning models. In detail, in chapter 2, we will look into the factors that motivate individuals to participate in online activities including social media, and explore the impact of these factors on the patterns of user activity observed in online networks. In chapter 3, we will categorize online users based on their activity, drawing on the participation inequality phenomenon. In particular, we will focus on describing in detail the silent users. Moreover, we will investigate how to incorporate these categories into machine learning models to mitigate the biases that arise from the participation inequality phenomenon. This analysis will explore different techniques to balance the representation of each category in the training data. In chapter 4, our investigation revolves around determining the data volume necessary for silent users in order to enhance the performance of machine learning models. The objective is to classify issue-specific silent users, a particular group that engages in discussions on certain topics while remaining silent on others.

Finally, in chapter 5, we will discuss how we can employ powerful tools such as large language models to generate synthetic data that closely mimics the characteristics of existing datasets and expand the limited data availability among silent users[1] .

[1]ChatGPT [OpenAI, 2023] was utilized to modify some sections of this dissertation including the readability and grammar checking.

Chapter 2

ONLINE PARTICIPATION

To address the biases caused by the activity of the users, we first need to delve deeper into understanding the underlying factors that drive user participation in online social communities. By examining these factors in detail, we can gain insights into how and why certain users are more active in content creation while others are not. We identify several motivational, behavioral, and psychological factors, including a range of intrinsic and extrinsic factors, that drive individuals to participate in online social networks. Our next step will be to categorize users according to the volume of content they generate on online social networks. This categorization allows us to identify different user segments and tailor interventions or strategies accordingly.

2.1 Motivating Factors for Online Participation

There are three primary types of behavioral factors that prompt online participation. These factors include a wide range of elements that influence individuals' engagement in online activities and interactions: (1) *individual-level factors*, which take into account the inherent and individual traits of users. Each person brings their unique characteristics, preferences, motivations, and goals to the online environment. Personal traits, such as personality, cognitive abilities, and demographic factors, can play a significant role in shaping an individual's online participation; (2) *community-level factors*, which is related to communities fostered by online platforms where like-minded individuals gather to share information or discuss topics of interest. The nature and dynamics of these communities can strongly influence individuals' participation; and (3) *environmental-level factors*, which examine external

environmental influences that can shape online participation. They include aspects such as the characteristics of the online system and the development process. Moreover, The interplay among these three levels gives rise to a multitude of additional factors that influence online participation.

In Figure 2.1, a summary of these factors has been provided, illustrating the intricate relationship between individual-level, community-level, and environmental-level factors and their impact on online participation. This summary serves as a useful reference to understand the various dimensions and influences that shape individuals' engagement in online activities. In what follows we will look into each level and discuss each factor in detail.

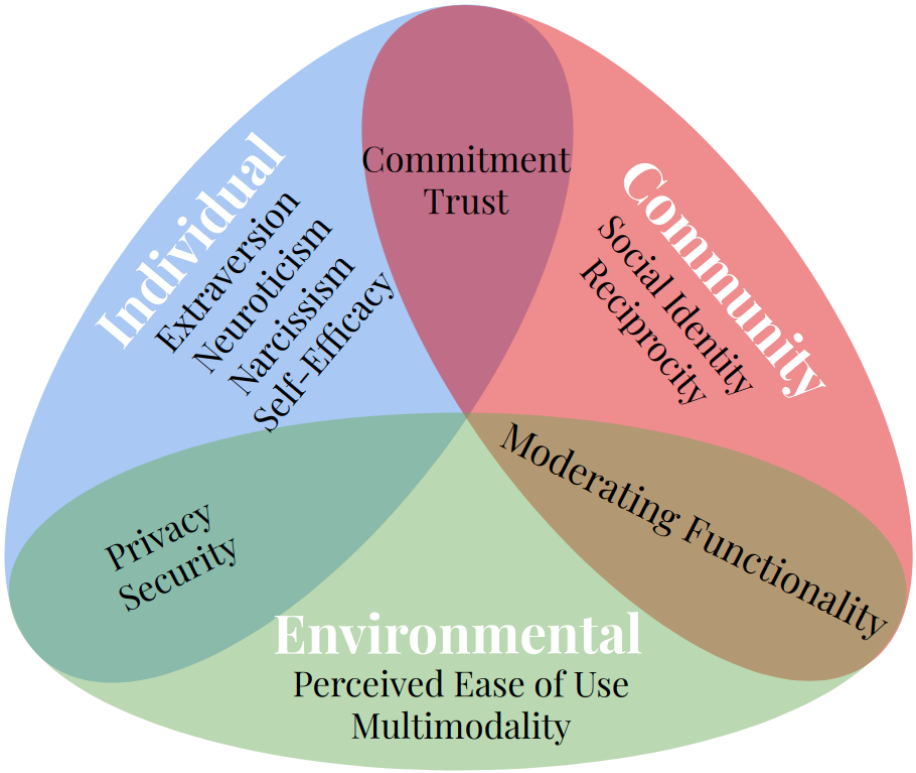


Figure 2.1: Three Categories of Behavioral Influence That Encourage Online Participation and User Engagement: Individual-, Community-, and Environmental-Level Factors.

2.1.1 Individual-Level Factors

Studies suggest that demographic features such as gender and age as well as personality traits play a significant role in providing distinctive patterns and motives for social media use [Liu and Campbell, 2017; Nonnecke and Preece, 2001]. Among the various personalities identified, there are four prevalent types associated with higher levels of online posting activity. Social media provides a platform for these users to meet their desire for self-presentation and empowerment. These four personality traits are as follows:

- **Extraversion**: which relates to the quantity and intensity of interpersonal interactions. Individuals with higher levels of extraversion tend to be more sociable and outgoing, leading them to engage in more frequent and expressive interactions on social media. As a result, they are likely to have a higher number of posts, receive more likes and comments, and build larger online networks [Blackwell *et al.*, 2017; Marengo *et al.*, 2020; Moore and Craciun, 2021].

- **Neuroticism**: which reflects an individual's susceptibility to emotional instability. Neurotic individuals tend to experience frequent mood swings and high-stress levels. In the context of social media, they express their emotions through more extensive use of words in their posts compared to emotionally stable users. This inclination towards using a greater number of words helps them convey their feelings and experiences more thoroughly [Blackwell *et al.*, 2017; Marengo *et al.*, 2020; Moore and Craciun, 2021].

- **Narcissism**: which indicates excessive self-promotional behavior. Grandiose narcissists[1] utilize social media as a means to maintain their high self-esteem.

[1]Not to be mistaken with pathological narcissistic personality disorder [Edition *et al.*, 2013].

They devote more time to these platforms, engage in self-centered activities such as posting numerous selfies and status updates, and actively seek to amass a large number of followers and friends. Their aim is to garner attention, admiration, and validation from others [McCain and Campbell, 2018].

- **Self-efficacy**: that captures self-confidence in one's own ability to achieve high-performance results. In other words, individuals with high levels of self-efficacy possess the necessary knowledge, technical skills, and confidence to contribute valuable content to the online community. They believe that their posts will be useful and well-received by many people. As a result, they actively participate in sharing their expertise, insights, and experiences through various forms of online content creation [Liu and Campbell, 2017].

2.1.2 Community-Level Factors

Social media platforms have become an integral part of people's lives, offering a wide range of features and functionalities that cater to users' social interaction needs. These platforms serve as virtual spaces where individuals can enhance their social skills, forge connections with like-minded individuals, and seek support or recognition within their chosen communities. By facilitating these interactions, social media plays a crucial role in fulfilling the innate human desire for social belonging and affiliation. Following are the key factors that significantly impact the level of communication and engagement on social media platforms.

- **Social Identity**: which refers to how individuals perceive themselves as part of a specific community, which often gives rise to *us vs. them* mentality, as described by Tajfel's seminal work [Tajfel, 1974]. People actively contribute to online communities by sharing content related to their shared interests, hob-

bies, or beliefs, thereby reinforcing their sense of social identity. Consequently, this heightened identification with a particular community strengthens their attachment to the virtual environment and motivates them to participate more actively in community-related activities [Yen, 2016; Mousavi *et al.*, 2017].

- **Reciprocity**: that examines the extent to which a community can provide benefits and resources to its members, and in turn, how much members are willing to reciprocate these actions [Sun *et al.*, 2014]. When individuals receive support, information, or assistance from their online communities, they often feel a sense of obligation to contribute back and reduce any perceived indebtedness. This reciprocity can manifest in various forms, such as sharing valuable information, actively participating in discussions, or providing support to other community members [Hsu *et al.*, 2018].

By fostering social identity and reciprocity, social media platforms create an ecosystem where users feel a sense of belonging and interconnectedness. They derive emotional and social satisfaction from being part of a community that shares similar interests and values. This virtual sense of belonging can have a profound impact on users' behavior, as they seek validation and recognition from their peers within these online spaces. Consequently, the more individuals identify with a particular community and experience reciprocal interactions, the more likely they are to actively engage and contribute to the collective activities within the social media platform.

2.1.3 Environmental-Level Factors

Users are more likely to engage with platforms that are intuitive, offer a variety of content creation options, and cater to their desire for real-time experiences. The following are the prominent environmental-level factors:

- **Perceived Ease of Use**: The primary driver of social media user engagement is *perceived ease of use*, which has been extensively studied and validated by the Technology Acceptance Model (TAM) [Davis, 1989]. Perceived ease of use refers to how easily a user can comprehend and navigate the features and functionalities of a social media platform. When the user interface of a social media platform is overly complex and bewildering, it can result in frustration and hesitation among users when it comes to posting content. Research has even identified this factor as one of the key reasons for lurking behavior, where users prefer to observe rather than actively participate [Nonnecke *et al.*, 2006; Preece *et al.*, 2004].

  To encourage active engagement, it is crucial for social media platforms to prioritize simplicity and user-friendly interfaces. However, while ease of use is important, it should not come at the expense of limited platform functionality. If a social media platform lacks essential features or fails to provide a diverse range of options for content creation and interaction, users may lose interest and engagement can decline.

- **Multimodality:** Social media platforms should strive to offer a rich variety of components and modalities to cater to the diverse demands and needs of their users [Nguyen, 2020]. For example, other than text that allows the users to express their thoughts and opinions effectively, modalities such as images and videos can enhance the engagement and interactive potential of a platform.

Users appreciate the ability to share visual content, as it enables them to convey emotions, experiences, and information in a more engaging and impactful manner.

2.1.4 Linked Factors

There are also several factors that could fall under two or more of the categories mentioned above. In the following section, we will introduce some of these factors.

- **Privacy and Security:** There is an increasingly pressing concern within society regarding the issue of privacy and security, specifically concerning the collection, storage, and analysis of users' personal information and social media accounts. As technology continues to advance and permeate various aspects of our lives, the potential risks and vulnerabilities associated with online platforms have become more prominent.

  In the current digital landscape, social media platforms play a pivotal role, as they encourage users to willingly disclose their personal information. The aim behind this is to establish diverse social networks and facilitate the maintenance of existing offline relationships through these online channels. However, this practice has led to a dilemma for many users who value their privacy and wish to protect their personal information from unauthorized access. These cautious behaviors may be associated with personality traits and characteristics, such as neuroticism and self-efficacy [Osatuyi, 2015; Popovac and Fullwood, 2019]. In other words, these traits may contribute to an individual's inclination to adopt a more reserved approach when it comes to online interactions and information sharing.

  Therefore, an individual's willingness to engage in online activities can be in-

fluenced by the security measures implemented by social media platforms. If platforms prioritize cyber-security and employ privacy-preserving algorithms, users are more likely to feel a sense of reassurance and be more comfortable engaging in various online activities [Osatuyi, 2015; Sun *et al.*, 2014]. For instance, platforms can employ robust encryption techniques to protect users' personal data from unauthorized access, ensuring that only authorized individuals have the necessary keys to decrypt and access the information. Additionally, incorporating strict access controls and authentication mechanisms can further bolster the security of users' personal information. Furthermore, social media platforms can educate users about the importance of privacy and security, providing them with the necessary knowledge and tools to protect themselves online. This can include guidelines on setting strong passwords, avoiding suspicious links or phishing attempts, and understanding the implications of sharing sensitive information.

By promoting digital literacy and empowering users with the skills to navigate the online world securely, platforms can foster a culture of responsible information sharing.

- **Trust and Relationship Commitment:** Commitment in a relationship is built on trust and can be described as the utmost exertion of effort that both parties invest in maintaining their mutual bond that is believed to be of paramount importance [Wu *et al.*, 2010]. Trust between an individual and the community is established through a combination of personal qualities, shared values, common experiences, and structural assurance [Cheng *et al.*, 2017]. Trust holds significant implications for users' perception of their community and their level of involvement and dedication to it.

When users perceive a community as highly trustworthy, their confidence in the community's intentions and actions increases. This heightened trust acts as a catalyst, motivating individuals to actively engage in community activities and contribute their time, knowledge, and resources [Sun *et al.*, 2014].

- **Moderating Functionality.** Studies have shown community managers play a crucial role in fostering engagement among members by creating content and initiating discussions [Lev-On, 2017]. In order to facilitate the norm of reciprocity and build trust among members, social media platforms must offer functionality that enables community managers to manage both content and members [Chen and Hung, 2010]. These functionalities can include content creation and curation, member management, and analytics and insights.

  By providing these functionalities, social media platforms empower community managers to cultivate trust, encourage engagement, and foster a sense of belonging among community members. This, in turn, leads to a more vibrant and active community.

2.2 Participation Inequality

Early mass media, in its nascent stages, primarily relied on one-way communication channels, offering limited opportunities for direct person-to-person interaction. This traditional media landscape fostered an environment where readers were largely passive participants, often limited to lurking and consuming content without actively engaging with it. The readers' role was predominantly that of an observer, with only a select few comments handpicked to be featured in specific sections or columns of the news media.

However, the advent of online social networks revolutionized the dynamics of me-

dia consumption and user engagement. Networking websites encouraged individuals who desired a voice to become active participants by creating and sharing their own content, as well as building interpersonal relationships. The newfound emphasis on user-generated content transformed the traditional media landscape into a more inclusive and interactive platform. Despite this encouragement for user engagement, not all social media users exhibit the same level of involvement in online activities such as posting, reposting, following, and favoring content. As a result, a phenomenon known as *participation inequality* emerged. User behavior in the realm of online activity follows a power-law distribution, where a small fraction of individuals is responsible for generating the majority of content or posts on social media platforms.

To address and account for this participation inequality, web usability experts introduced the concept known as the *1% Rule* or the *90-9-1 Rule*, as elucidated by Nielsen's work [Nielsen, 2006]. According to this rule, online social network users can be categorized into three distinct groups based on their level of engagement. The largest group, comprising approximately 90% of users, are the *lurkers* who primarily consume content without actively participating in online activities. They prefer to observe rather than contribute. The next group, constituting around 9% of users, are the *engagers*. These users exhibit a moderate level of involvement and contribute sporadically to content creation and interactions. They may occasionally post or engage in discussions but not at the same level of commitment as the contributors. The smallest group, comprising only 1% of users, are the *contributors*. These individuals actively generate and share the majority of content on social media platforms. They are enthusiastic and dedicated participants who contribute frequently and are instrumental in shaping the online discourse.

In the subsequent discussion, our focus will be on exploring the definition and characteristics of the lurker's group as documented in the existing literature. By

delving deeper into their behaviors and motivations, we can gain a comprehensive understanding of the diverse user dynamics within online social networks.

2.2.1 Silent Users or Lurkers

Nonnecke and Preece [Nonnecke and Preece, 2003] defined lurkers as "anyone who reads but seldom if ever publicly contributes to an online group". This behavior has been extensively studied in the literature, often accompanied by negative labels such as *passive actors* [Hemmings-Jarrett *et al.*, 2017], *abusers of common good* [Amichai-Hamburger *et al.*, 2016], and *free-riders* [Nonnecke and Preece, 2003], suggesting that they only take resources without contributing. However, a reevaluation of this behavior is necessary as it has been recognized as a normal, positive, active, and valuable form of online participation [Edelmann, 2013]. Lurking behavior is an inherent aspect of online communities and cannot be completely eliminated [Nielsen, 2006]. In fact, Kokkodis *et al.* [Kokkodis *et al.*, 2020] reported that approximately 79% of the individuals who engage in lurking never contribute any content. Despite this, gaining a better understanding of these users would greatly benefit online research, e-business, and e-government since lurkers constitute the largest group in terms of numbers within the online environment [Edelmann, 2013].

It is important to recognize that lurkers actively consume and listen to relevant information within the online community. They play a crucial role in creating connections and being receptive to the content shared by active contributors [Edelmann, 2013; Gong *et al.*, 2015]. Lurking can be seen as a valuable way for individuals to gather knowledge, learn from others, and stay informed about current discussions and trends.

Moreover, lurkers contribute to the overall ecosystem of online communities by providing an audience for active participants. Their presence and attention motivate

content creators to continue sharing valuable insights, spark discussions, and generate new ideas. Lurkers serve as an essential component that sustains the vitality of online platforms, even if they do not actively engage in content creation themselves.

2.2.2 Lurkers Categorization

Various types of silent users can be identified on social media. Silent users should not be confused with Churners. To this means, we introduce churners first and then describe different types of silent users.

- **Churners**. Users who are registered to an online platform or service but no longer actively use it, without having deleted their account. On the server side, these users can be identified by their last day of login, which provides a starting point for determining their inactive status. However, on the client side, it is essential to track their activity over a period of time to gain a comprehensive understanding of their engagement patterns and to differentiate them from others. To accurately identify churners, various user activities can be considered, extending beyond just the last login date. These activities encompass a range of interactions that indicate a user's level of engagement with the platform. For example, posting a new status update is a significant indicator of user activity. This can include creating original posts, reposting content, quoting others, liking posts, or engaging in reply/comment threads. Tracking these actions allows for a more nuanced evaluation of a user's involvement. In addition to these interactive activities, changes made to profile information can also serve as indicators of differentiating churners from others. Alterations in screen name, description, location, profile image, or banner image may suggest that a user is still using the platform. Churners at the time of their active period could have fallen into one of the lurkers, engages, or contributors categories.

Following lists different types of silent users or lurkers:

- **Content Consumers**. Lurkers who would never contribute to the platform, including looky-loos who are simply curious about what is happening. These lurkers are individuals who are intrigued by the platform's content and visit it solely to satisfy their curiosity. They often browse through various sections, read posts, view media, and explore the platform's features. However, despite being active in terms of consuming content, they refrain from engaging with the community by posting comments, creating content, or initiating discussions. These types of lurkers may spend a considerable amount of time on the platform, frequently revisiting it to stay updated on the latest happenings or to delve deeper into specific topics of interest. They may find pleasure in observing the interactions and discussions of others, but they choose not to actively participate themselves.

  These individuals would easily be identified by their hidden activities such as modifying their profile information. While these actions can be considered as activity from a technical standpoint, they remain hidden from the broader community, and their overall contribution to the platform remains negligible or nonexistent. However, the recently added *impression counts* feature on Twitter would record these hidden audiences' engagements and encourage content creators. It is important to note that the presence of content consumers is a common phenomenon in many online platforms. While they may not actively contribute, their presence contributes to the overall user base and can still generate valuable metrics in terms of traffic and user engagement.

- **Reaction Lurkers**: Refers to those lurkers who are often characterized by their tendency to only *like* or *repost* existing content rather than creating something new. They may show their appreciation for a post by clicking the like/heart button or sharing it with others, but they refrain from expressing their thoughts or opinions directly.

- **Highly Influential Lurkers**: These lurkers are individuals who hold a significant level of respect and admiration within a particular online community. While they may not actively engage in frequent content creation or participation, when they do decide to post something, their contributions tend to generate a substantial number of reactions from other community members. These reactions can manifest in the form of likes, shares, comments, or reposts, indicating that their words and ideas carry considerable weight and influence.

  Identifying the highly influential lurkers can be invaluable for community moderators as it allows them to understand and leverage the power of these individuals to create a more vibrant and engaging community. By encouraging these respected lurkers to become more active participants, moderators can tap into their ability to drive interactions and foster a sense of community engagement. When these influential lurkers share a post, it often has a ripple effect within the community. Their words are highly regarded and can inspire others to join the conversation, share their opinions, or contribute additional valuable content. This influx of participation breathes life into the community, creating a dynamic and thriving environment.

  Moreover, the impact of highly influential lurkers extends beyond their individual posts. Their reputation and standing within the community can positively influence the overall perception of the community itself. When other users

observe the active involvement of these respected individuals, it enhances the community's credibility and attractiveness, potentially drawing in more participants and widening the reach of the community as a whole.

- **Community-Specific Lurkers**: These lurkers actively engage and participate within certain online communities, while adopting a more passive role as lurkers in other communities. These individuals might be drawn to specific communities due to shared interests, hobbies, or professional affiliations. In these communities, they enthusiastically contribute to discussions, ask questions, provide insights, and offer support to fellow members. Their active participation demonstrates a desire to be an integral part of these communities and actively contribute to their growth. On the other hand, community-specific lurkers choose to maintain a passive role in other online communities. While they might be aware of the discussions and activities occurring within these communities, they prefer to observe silently without actively engaging.

  This behavior can be influenced by several factors, such as a lack of personal interest in the subject matter, a perceived lack of expertise in the community's domain, or simply a preference for consuming content without feeling the need to actively participate. It's important to note that community-specific lurkers are not necessarily disengaged or disinterested individuals. Rather, they strategically allocate their time and energy based on their level of connection and relevance to each community. They prioritize their active involvement in communities where they feel a stronger sense of belonging and derive personal value from their contributions. In doing so, they might gain a deeper understanding of the community's dynamics, build relationships with other members, and establish themselves as valuable contributors.

The phenomenon of community-specific lurking also highlights the nuanced nature of online identities. Individuals might have multiple online personas, each tailored to specific communities and reflective of their various interests and expertise. These lurkers possess the ability to adapt their online behavior to fit the context of each community, effectively navigating the intricacies of online social dynamics.

- **Issue-Specific Lurkers**: Introduced by Gong *et al.* [Gong *et al.*, 2016], these individuals exhibit a pattern of being highly involved in certain subjects while remaining relatively inactive in other subjects. From the perspective of server-side analysis, it is possible to identify the topics that these users are interested in by analyzing their behavior. One way to accomplish this is by observing the amount of time these users spend reading or interacting with posts related to specific topics before moving on to other content. Some platforms also provide users with a list of topics from which they can select their areas of interest. By examining the topics selected by the users, we can gain insights into the issues they are more likely to lurk on.

  On the client side, the identification of issue-specific lurkers requires at least one activity related to each topic of interest. By employing clustering models like topic modeling, it becomes feasible to discern the various subjects the user has discussed or interacted with in the past. This process involves analyzing the content of the user's posts or comments and determining the frequency and extent to which they have engaged with each topic. Through this approach, it becomes possible to categorize the different topics the user has participated in and ascertain the number of posts they have dedicated to each specific subject.

It's important to note that these categories are not mutually exclusive, and individuals can exhibit lurking behavior across multiple types depending on their interests, motivations, and social media habits.

Chapter 3

SILENT SPEAKS VOLUMES

Silence is an ocean.
Speech is a river.
When the ocean is searching for you,
don't walk into the river.
Listen to the ocean.

– Jalāl al-Dīn Muḥammad Rūmī

Deep learning methods have recently become prevalent in the field of social media analysis due to their ability to model complex and non-linear relations between input data. However, despite efforts to analyze social media data, these models often fail to capture diverse opinions. This is because they rely solely on observed data, which tends to be provided by engagers and contributors. As a result, the inferred user behavior is biased and does not accurately represent the true landscape of the platform. Figure 3.1 illustrates the percentage of the interactions from each group of users - lurkers, engagers, and contributors - for two different datasets. For example, a data point located in the lower right corner represents a news article where 100% of the interactions were from contributors and 0% from lurkers and engagers.

Early studies in behavioral and social science literature frequently defined lurkers as users who only consume resources without contributing to the community. This also influences machine learning researchers to overlook the contributions of the lurkers. However, we argue that lurkers' behavior can provide additional cues for social media analysis methods as these users actively consume and listen to the relevant information, create connections, and are receptive [Edelmann, 2013; Gong *et al.*, 2015]. This can be corroborated by recent efforts to drive user participation in online social

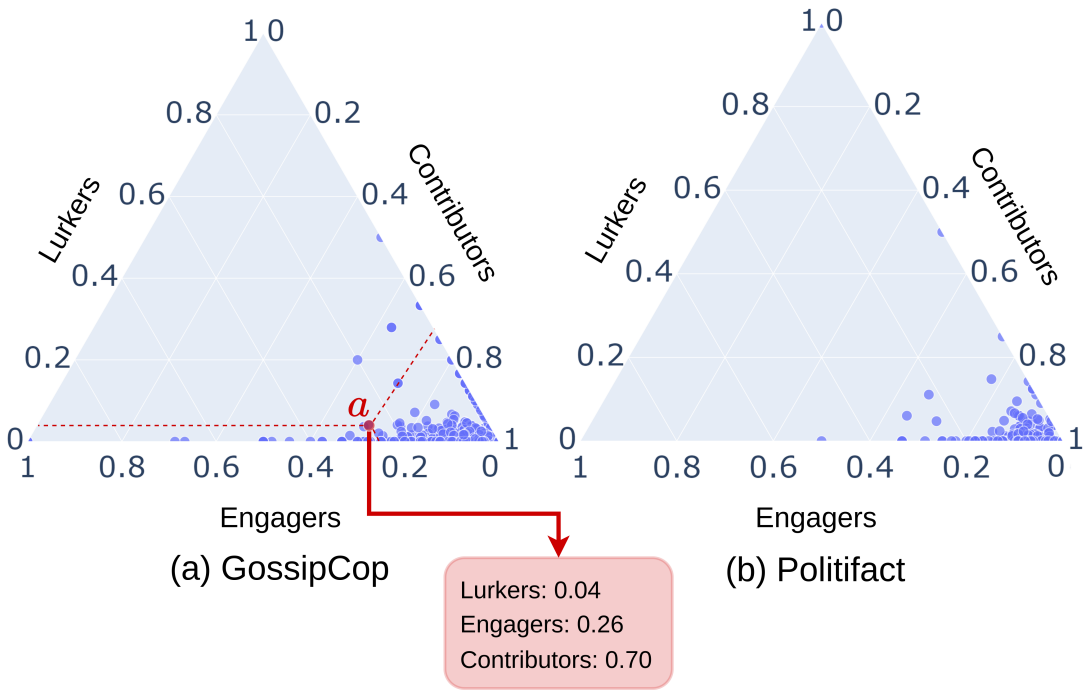


Figure 3.1: Ternary Plots of the Percentage of the Interactions on Social Media Created by Each of the Lurker, Engager, and Contributor Groups in (a) GossipCop and (b) Politifact Datasets. In General, As Expected, the Percentage of the Interactions Recorded by the Contributors Is More Than the Other Two Groups. Out of the Users Who Reacted to the News a, 4% Are Lurkers, 26% Are Engagers, and 70% Are Contributors.

communities. For example, among reasons listed in [Nonnecke and Preece, 2001] for the lurking behavior, a user's motivation to post diminishes if they cannot provide any vital or novel information. Furthermore, the authors in [Nguyen *et al.*, 2022] mention that one of the reasons a lurker becomes active on social networking sites is when they can gain knowledge as well as share it outside the community.

Given these reasons, a lurker might engage with a post when they have valuable information to add related to the topic. Thus, we hypothesize that giving importance to such interactions between the posts and lurkers may improve the performance of the different social media analysis applications. For instance, consider the task of fake news detection. This task entails classifying a news article as real or fake by benefiting from the user-news interactions obtained from social media data. However, directly

Figure 3.2: Example of a Piece of News Content From the Politifact Dataset, the Tweets of Users From Each Group. We Hypothesize That if a Piece of News Provokes a Lurker To Create Content on Social Media, Giving Importance to Such Interaction Might Improve the Performance of Fake News Detection Models.

utilizing this network may not be fruitful due to two reasons. First, as mentioned, this interaction may be biased toward the views of the contributors as they are the ones creating about 90% of the interactions. Second, unobserved interactions (i.e., unshared news) do not guarantee that the user was not exposed to the news. A user might be exposed to the article but may choose to refrain from expressing their opinions due to one or more reasons. For example, a user might doubt the post's veracity, or a user may feel like they might not add value to the already propagated content. In compliance with the earlier stated hypothesis, if a lurker engages with a news article, they might have more information about the news article. Thus, by

up-weighting the limited lurkers' interaction, although marginal, one may improve the detection capabilities of the fake news detection model. Furthermore, it can also aid the model in fairly representing the voice of the silent users. Figure 3.2 shows a motivational example from the Politifact dataset. The example includes the content of the news and different tweets that mention the news from three different types of users. In this work, we only utilized retweet interactions.

We propose to leverage re-weighting techniques to verify whether silence speaks volumes. We use the task of fake news detection and track their performance by differentiating between the interactions based on the aforementioned user categories. Our approach learns a fair representation based on the true landscape of the platform and up-weights those news articles that triggered the silent users more as they might provide additional information for detection.

3.1 Background: Fake News Detection

Due to the increasing amount of time spent on social media platforms, it is no surprise that people tend to receive their news content through social media more than before. One in five U.S. adults used social media as their main source of political and election news for the US presidential election in 2020 [Mitchell and Jurkowitz, 2020].

The high rate of engagement with online news can be mainly attributed to the nature of the social media platforms themselves. Social Media is typically inexpensive, provides easy access to users, and supports fast dissemination of information that is not possible through traditional media outlets. However, despite these advantages, the quality of news on social media is considered lower than that of traditional news outlets. A factor contributing to this low quality is the widespread nature of fake news articles online. Fake news is a piece of false information published by news outlets to mislead consumers [Zhou and Zafarani, 2020; Shu *et al.*, 2020a].

Fake news has several significant negative effects on civil society. First, people may accept deliberate lies as truths. The likelihood of accepting fake news as true increases after repeated exposure [Hasher *et al.*, 1977], especially when the content aligns with the user's beliefs [Weir, 2017; Jiang *et al.*, 2021a]. Second, fake news may change the way people respond to legitimate news. When people are inundated with fake news, the line between fake news and real news becomes more uncertain. Fake news spreaders make users doubt the nature of real news and create the idea that everything is biased and conflicted, and it is impossible to distinguish fake from real news [Lynch, 2016]. Finally, the prevalence of fake news has the potential to break the trustworthiness of the entire news ecosystem. For instance, despite traditional domains such as the New York Times, the Washington Post, and CNN being among the most shared COVID-19-related stories on Twitter, a fake news domain, Gateway Pundit, was ranked 4th in August and 6th in September of 2020 among the most shared domains for URLs about COVID-19 [Lazer *et al.*, 2020]. Therefore, it is critical to develop methods that detect and mitigate fake news, with the purpose of benefiting the general public and the entire news ecosystem.

Detecting fake news is a challenging task because it is designed to be indistinguishable from real news and intentionally misleading. As a result, the features extracted from the content are not enough to build an accurate detection method. For example, in the field of user-based fake news detection and fake news spreader profiling, researchers have utilized different conjunctions of user's profile information, user's activity, user's network connectivity, and user's generated content [Antelmi *et al.*, 2019; Karami *et al.*, 2021; Cheng *et al.*, 2021] to detect fake news. Cheng *et al.* proposed a model to identify the causal relationships between users' profiles and their susceptibility to sharing fake news articles [Cheng *et al.*, 2021]. The authors modeled the dissemination of fake news by creating implicit feedback based on the user's exposure

and interest in specific fake news. The learned fake news sharing behavior is then used in improving the detection of fake news. Karami *et al.* [Karami *et al.*, 2021] extracted some features from the user's profile information, generated content, and activity that represents their motivational behavior in spreading fake news. They showed the effectiveness of their model in determining which users are more likely to spread fake news. Cardaioli *et al.* [Cardaioli *et al.*, 2020] investigated how behavioral-based features such as Big Five personality and stylometric features extracted from the content of a user's timeline can be used to profile fake news spreaders. Shu *et al.* [Shu *et al.*, 2019c] investigated the importance of explicit features such as register time, follower and following count as well as implicit user meta information such as location and political bias inferred from their online behaviors and historical tweets in fake news detection.

Nevertheless, all the aforementioned methods do not distinguish between lurkers, engagers, and contributors, hence, generalizing the dissemination behavior for all types of users. Thus, they are biased toward the majority content created by the minority class of users.

3.2 Problem Statement

Let $\mathcal{X} = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ denote a set of n news articles with labels $y = 0$ for true and $y = 1$ for fake news. Each news article x_i consists of three components: (1) the news content, $a_i \in \mathcal{A}$, which is a sequence of k words $\{w_1, w_2, ..., w_k\}$, (2) a set of m comments containing different views of the users' opinion related to the corresponding news article, $c_i = \{c_{1i}, c_{2i}, ..., c_{mi}\} \in \mathcal{C}$, and (3) a user-news interaction $u_{ji} \in \mathcal{U}$ with p number of users.

Typically, $\mathcal{U}$ is a *binary* matrix representing interaction between user j and news i: if j interacts with i then $u_{ji} = 1$, otherwise $u_{ji} = 0$. Note that $u_{ji} = 0$ can be

interpreted as either the user j was not exposed to the news article i or was exposed to but due to some reasons (e.g., not sure of the veracity of the news [**?**]) chose not to propagate it. Based on our hypothesis, to investigate the impact of interactions with under-represented users, we aim to design a fake news detection function that considers the type of users in terms of their activity, $\mathcal{G} = \{L, E, C\}$.

Formally, we can represent the model as follows:

> Given news articles $\mathcal{A}$, users' comments $\mathcal{C}$, and a user-news interaction $\mathcal{U}$, learn a fake news detection function $f(\mathcal{A}, \mathcal{C}, \mathcal{U}, \mathcal{G}) \rightarrow \hat{y}$ with respect to the users belonging to one of the lurkers (L), engagers (E), and contributors (C) groups $\mathcal{G}$.

3.3 Designing Fake News Detection Model

Previous methods in fake news detection either do not consider user-news interaction in their model, or it is appended as a binary matrix with 1 showing the user tweeted or retweeted about specific news. Similar to other social media analysis studies, this news dissemination data in online environments is also biased toward the users who create the majority of the social media content. In other words, the user-news interaction matrix is biased towards the views of the users that are more eager on asserting their opinion about the news but belong to only 1% of the social media population - i.e. the contributors. The focus of this paper is to provide a fair representation by giving more value to the interactions created by lurkers.

We designed two approaches as illustrated in Figure 3.3. The first method balances the user-news interaction matrix which later will be added to the baseline models as a weighted matrix. The second method will apply sample re-weighting based on the activity of the users to see whether this would improve the performance of the downstream task.

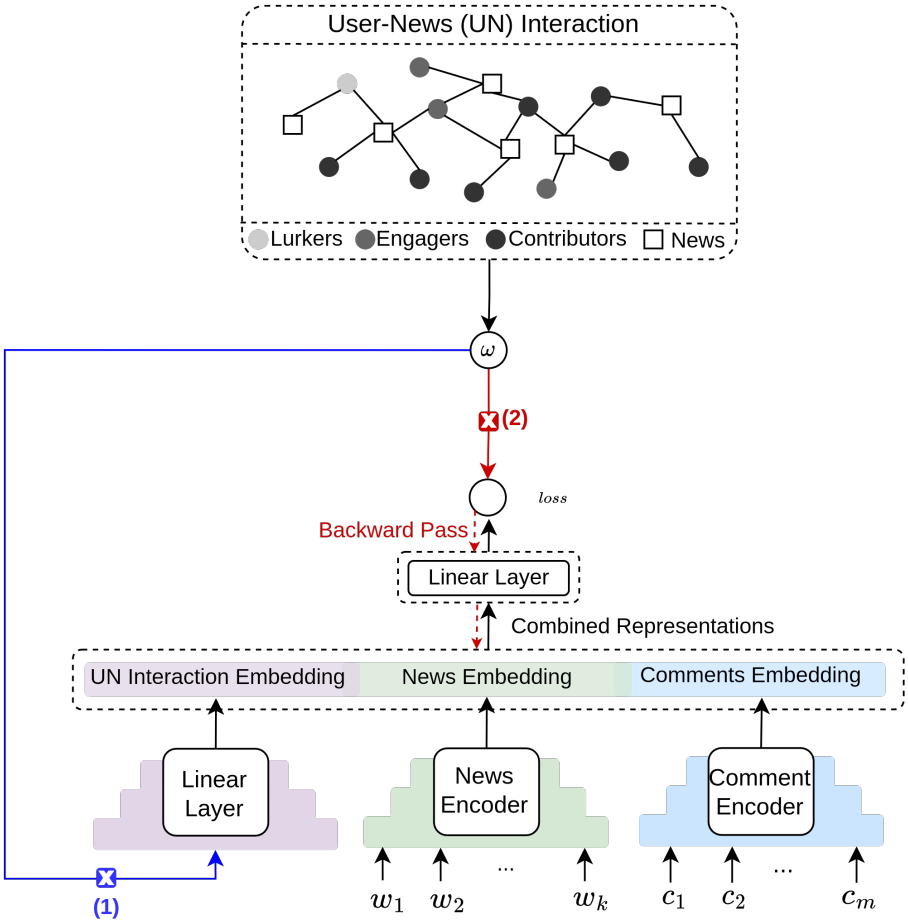


Figure 3.3: Two Re-Weighting Strategies Were Used to Learn a Balanced Representation for the Task of Fake News Detection: (1) Edge Re-weighting (Section 3.3.2) and (2) Sample-Level Re-weighting (Section 3.3.3).

In this section, we will briefly talk about the text representation learning for news articles as well as the news comments and then introduce our weighting mechanisms.

3.3.1 News Articles and Users' Comments Representations

To generate a vector representation of the news content as well as the users' comments, different models apply different text representations. In the task of fake news detection, earlier methods use word-level and sentence-level features such as bag-of-words and n-grams. Recent models use deep learning-based methods such as

Recurrent neural networks (RNN), Long Short Term Memory (LSTM), and Transformers to model sequential data. Transformers use a self-attention mechanism to extract vital information from the input data. Both the news and the comment encoder inputs are text sequences, and they output the vector representation of text. Formally, if we show the article's content and the comment encoder as $g_a(\cdot)$ and $g_c(\cdot)$ functions, respectively, then for each news i,

$$z_{ia} = g_a(w_1, w_2, ..., w_k) \quad \text{and} \quad z_{ic} = g_c(c_1, c_3, ..., c_m) \tag{3.1}$$

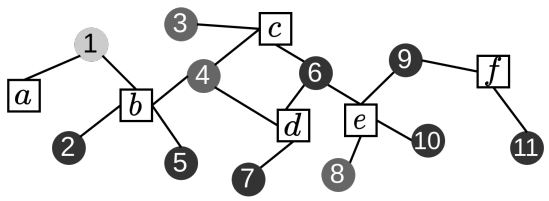
where z_{ia} and z_{ic} are the embedding vectors for the news content and the user comments, respectively, $w_1, w_2, ..., w_k$ is the sequence of the words in the news articles and $c_1, c_2, ..., c_m$ are its corresponding comments.

3.3.2 Edge Re-weighting Mechanism for News Dissemination Network

The news dissemination network consists of two different types of nodes: users and news. In Figure 3.4, users are denoted by circles while the news pieces are illustrated by squares. Each user node can belong to one category of lurkers, engagers, or contributors.

To handle the imbalancedness of the user types on social media, we propose a weighting mechanism based on the 90-9-1 Rule. The calculated weight would be applied to all the edges connected to a square-shaped node based on the type of all its connected circle-shaped nodes. Formally, we substitute the binary user-news interaction matrix ($\mathcal{U}$) in our formulation of the fake news detection function with a normalized weighted version ($\overline{\mathcal{U}}$). We propose the following weighting mechanism:

$$\overline{u}_i = u_i \cdot \left(1 + \frac{\omega_i}{\parallel \omega \parallel}\right)^{\alpha} \quad \forall i \in \{1, ..., n\} \tag{3.2}$$



	1	2	3	4	5	6	7	8	9	10	11	ω
a	1	0	0	0	0	0	0	0	0	0	0	$0.9 \cdot 1$
b	1	1	0	1	1	0	0	0	0	0	0	$0.9 \cdot 1 + 0.09 \cdot 1 + 0.01 \cdot 2$
c	0	0	1	1	0	1	0	0	0	0	0	$0.09 \cdot 2 + 0.01 \cdot 1$
d	0	0	0	1	0	1	1	0	0	0	0	$0.01 \cdot 1 + 0.01 \cdot 2$
e	0	0	0	0	0	1	0	1	1	1	0	$0.09 \cdot 1 + 0.01 \cdot 3$
f	0	0	0	0	0	0	0	0	1	0	1	$0.01 \cdot 2$

Figure 3.4: An Example of a Network With 11 Users (1 Lurker, 3 Engagers, and 7 Contributors) Interacting With 6 Pieces of News. This Interaction Vector Is a Binary Vector With 1 Indicating the Existence of an Interaction. The Weights Are Calculated Based on Equation 3.3.

where ω_i is calculated as follows:

$$\begin{aligned} \omega_i = 0.9 \cdot \sum_{j=1}^{p} \mathbb{1}_L(j) \cdot u_{ji} + 0.09 \cdot \sum_{j=1}^{p} \mathbb{1}_E(j) \cdot u_{ji} \\ + 0.01 \cdot \sum_{j=1}^{p} \mathbb{1}_C(j) \cdot u_{ji} \end{aligned} \tag{3.3}$$

In the above equations, u_i is a vector showing the user's interaction activity (i.e., 0 or 1) with all the news. n and p are the number of news articles and users, respectively. L, E, and C are the list of lurkers, engagers, and contributors. The $\alpha \geq 0$ is a hyperparameter that controls the intensity of the weighting mechanism. For example, $\alpha = 1$ will apply a weighting based on the 90-9-1 Rule on each user type while $\alpha = \frac{1}{2}$

is the smoother version of it. Moreover, $\mathbb{1}_S(j)$ is an indicator function and is 1 if $j \in S$, otherwise, it is 0, where S is one of the user types. The indicator functions defines which type a specific user belongs to. An example is given in Figure 3.4. In this figure, for instance, four users interacted with news b, out of which one is a lurker, one is an engager, and two are contributors. The weight is calculated as:

$$\begin{aligned} \omega_b &= 0.9 \cdot (\#\text{ of lurkers}) + 0.09 \cdot (\#\text{ of engagers}) \\ &\quad + 0.01 \cdot (\#\text{ of contributors}) = 0.9 \cdot 1 + 0.09 \cdot 1 \\ &\quad + 0.01 \cdot 2 = 1.01 \end{aligned} \tag{3.4}$$

3.3.3 Sample-level Re-weighting Mechanism for News Representation

Sample re-weighting has been a mainstream approach in creating a robust model when dealing with imbalanced training data [Cao *et al.*, 2019; Cui *et al.*, 2019]. Inspired by this, we trained the models by applying a sample-level re-weighting method based on the users belonging to lurker, engager, or contributor groups. In other words, for the news article i and M number of samples in a batch, the normalized weight is integrated into the loss function to model a balanced fake news detection. Formally,

$$\mathcal{L}_{balanced} = -\frac{1}{M} \sum_{i=1}^{M} \left(1 + \frac{\omega_i}{\parallel \omega \parallel}\right)^{\alpha} \cdot \mathcal{L}_{CE}(y_i, \hat{y}_i) \tag{3.5}$$

where y_i and $\hat{y}_i$ is the true and the predicted labels, respectively. The weights are calculated as a batch-wise version of equation 3.3. Moreover, $\mathcal{L}_{CE}$ is the cross-entropy loss, formulated as:

$$\mathcal{L}_{CE}(y_i, \hat{y}_i) = y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{3.6}$$

The batch-wise learning process of the balanced fake news detection and the weighting procedure is provided in Algorithm 1.

Algorithm 1: Learning to Re-weight News Representations Based on the User Types.

Input : $\mathcal{X}_{train}$; θ^0; epochs; UN Matrix $[u_{ji}]$; α; Lurkers (L), Engagers (E), and Contributors (C) sets.

Output: θ^T

for $e = 0, ..., \text{epochs}$ **do**:
  for $t = 0, ..., T-1$ **do**:
    $\mathcal{X}^t_{train} \leftarrow \text{SampleMiniBatch}(\mathcal{X}_{train}, t)$
    $\hat{y}^t_{train} \leftarrow \text{Forward}(\mathcal{X}^t_{train}, \theta^t)$
    $\omega^t_{train} \leftarrow \sum_{S \in \{L,E,C\}} w_S \cdot \sum_{j=1}^{p} \mathbb{1}_S(j) \cdot u_{ji}$
    $loss = mean\left[\left(1 + \frac{\omega^t_{train}}{\|\omega\|}\right) \mathcal{L}_{CE}(y^t_{train}, \hat{y}^t_{train})\right]$
    $\nabla\theta^t \leftarrow \text{Backward}(loss, \theta^t)$
    $\theta^{t+1} \leftarrow \text{OptimizerStep}(\theta^t, \nabla\theta^t)$
  end for
end for

3.4 Experimental Setting

In this section, we describe the details of the experimental setup including the benchmark datasets, dataset preparation, baseline methods, and implementation details.

3.4.1 Datasets and Dataset Preparation

We used two datasets from the FakeNewsNet repository as the seed datasets for the evaluation: Politifact and GossipCop [Shu *et al.*, 2020b].

- Politifact[1] : a fact-checking website where reporters and editors from the media fact-check political news articles. The URLs of news articles are available on the Politifact website and are used to collect tweets related to them.

[1]https://www.politifact.com/

- GossipCop[2] : a website for fact-checking entertainment stories aggregated from various media outlets. On the GossipCop website, articles get a score between 0 and 10 as the degree from fake to real.

In these datasets, along with the content of the news, the news comments and IDs of the Twitter users who reposted these fake and real stories are also included. The textual data (i.e., news content and news comments) were pre-processed to remove punctuation, out-of-vocabulary words, URLs, hashtags, and mentions. We utilized the Twitter user ids to create the user-news interaction matrix.

We also collected the history of the activities of each of the Twitter users identified in the Politifact and GossipCop datasets. Some of these users were deleted or suspended accounts and we were not able to access their activity and profile information anymore (9,537 of the GossipCop users and 13,181 of the Politifact users). We ignored these users in our matrix creation. For the rest, to categorize them into three groups of lurkers, engagers, and contributors, we calculate the average number of activities per day. We set the thresholds for the average number of activities per day in creating the lurkers and engagers to 0.025 and 0.15, respectively, such that it approximately follows the 90-9-1 Rule [Nielsen, 2006] as well as the definition provided in social science behavioral papers [Sun *et al.*, 2014]. Statistics of the created datasets are summarized in Table 3.1.

3.4.2 Baselines

In this section, for evaluation, we consider state-of-the-art baselines that use both news content and users' comments. To also include the BERT [Kenton and Toutanova, 2019] model to the group of baselines, we integrate BERT with a comment encoder for a fair comparison.

[2]https://www.gossipcop.com/

		Politifact	GossipCop
Number of News	***Real***	132	3,588
	Fake	319	2,230
	Total	451	5,818
Number of Interactions	***Lurkers***	482	382
	Engagers	4,295	3,945
	Contributors	41,738	30,054
	Total	46,515	34,381
Number of Comments		89,999	231,269

Table 3.1: Statistics of the Datasets.

The followings are the details regarding each baseline:

- CSI [Ruchansky *et al.*, 2017]. This method applies a hybrid deep model to capture the characteristics of fake news such as the text of the article, the set of tweets in which users commented about the fake news, and the source of the article such as the structure of the URL or the credibility of the media source. For a fair comparison, we disregarded the news source feature.

- dEFEND [Shu *et al.*, 2019a]. This model applies deep hierarchical sentence-comment co-attention network. dEFEND learns feature representations of the content and the comments for fake news detection and jointly discovers explainable sentences from these two sources.

- TCNN-URG [Qian *et al.*, 2018]. Based on convolutional neural network idea for text classification [Kim, 2014], this model tries to capture semantic information from the article's text using Two-level Convolutional Neural Network (TCNN). Moreover, it incorporates a User Response Generator (URG) module to learn a generative model (Variational Autoencoder) of user responses to the article and utilizes the learned model in generating responses for unseen news articles.

- BERT+HAN. We created a variant of the BERT model that includes the comments to match the other baseline models. We added the Hierarchical Attention Network for training the news comment section following Mosallanezhad *et al.* [Mosallanezhad *et al.*, 2022] which models the importance of each comment along with the salient word features.

3.4.3 Implementation Details

Traditional fake news detection methods only utilize the text of the news for detecting the fake from the real. However, integrating auxiliary information would provide a comprehensive representation of the samples and help in improving the performance of the models. For example, news comments provide useful signals for fake news detection [Shu *et al.*, 2019a; Mosallanezhad *et al.*, 2022], since semantic cues such as signals supporting or doubting the veracity of the content can be extracted from the comments. On the other hand, user-news interactions can highlight the type of items a user interacts with and further improve the understanding of user behaviors [Mosallanezhad *et al.*, 2022; Shu *et al.*, 2019b, 2022]. Moreover, it has been well documented that, fake news tends to spread faster than true news articles on social media sites such as twitter. Thus, incorporating user-item interactions provides additional cues to enhance fake news detection.

To study the effectiveness of our weighting mechanism in the task of fake news detection, we integrated this user-news interaction component into each of the baseline models. In other words, the output of the news and comment encoders were concatenated to the user-news interaction encoder which is a feed-forward network, and was fed to a dense layer to be trained for the fake news detection task, similar to the illustration provided in the Figure **??**. Table 3.5 shows the performance (accuracy) of these models with the original architecture, when the binary user-news interaction

is added, and when we incorporate the two proposed weighting techniques.

To improve the training process time of the BERT+HAN models, we initialize the news and comments encoder by fine-tuning them with the news content and users' comments, respectively. Due to BERT's input size limitation, we truncate each news content and comment to include its first 512 words. The embedding dimension for the HAN architecture is set to 100. Both the news content and user comments networks were trained using a simple feed-forward fake news classifier on top of it which was removed in the final architecture of the model. Once pre-trained, we merged the news and comments encoders in the BERT+HAN model with the user-news interaction encoder. With passing the news elements (i.e., news content, user comments, and user-news interaction matrix) through this integrated network, we train the final fake news classifier.

We trained the models with early stopping for all the baselines. For the edge re-weighting mechanism, instead of the binary user-news interaction matrix, we fed the weighted version, while for the sample-level re-weighting, we changed the loss based on the equation 3.5. Moreover, we tracked all the experiments using the Weights & Biases tool [Biewald, 2023] where applicable. The hyperparameters tuned are the batch size, epochs, and learning rate.

3.5 Experimental Results

In this section, we review the designed experiments using the task of fake news detection. We specifically are looking to answer the following research questions:

Q1. How much effect do the designed weighting mechanisms have on the performance of the models?

Q2. Which weighting mechanism would capture the voice of the silence better?

Using the available data, one way to investigate whether the voices of the silent users make a difference is to up-weight the silent users' signals and compare the performance of the downstream task with the original case. To be able to apply the weighting procedures based on the designed architecture, at first, we need to integrate the user-news interaction module (i.e., the UN interaction Embedding in Figure **??**) to different baselines introduced in Section 3.4.2 and record their performance. Comparing the two accuracy columns in Table 3.2, we can see that user-news interaction conveys valuable information when added to the current fake news detection algorithms. The average improvement in the accuracy of the models for Politifact news is +4.63% while the average improvement of +8.14% has been observed in the GossipCop dataset.

Dataset	**Model**	Original	With Binary User-News Interaction Module (+UN)
		Accuracy	**Accuracy**
Politifact	CSI	81.10 ± 1.07	85.93 ± 2.63
	dEFEND	81.48 ± 1.50	84.36 ± 2.20
	TCNN-URG	80.32 ± 2.06	86.92 ± 1.24
	BERT+HAN	83.04 ± 1.35	87.25 ± 1.32
GossipCop	CSI	85.98 ± 0.29	88.77 ± 0.50
	dEFEND	78.34 ± 1.55	87.62 ± 0.84
	TCNN-URG	81.42 ± 2.62	85.66 ± 0.46
	BERT+HAN	71.86 ± 0.00	88.14 ± 0.41

Table 3.2: The Average Performance on the Original Architecture of the Baselines Along With a Variation That Includes the Binary User-News Interaction Component (+UN).

Moreover, the best accuracy for the original architecture and when binary user-news interaction has been added has been shown in Table 3.3 and Table 3.4.

Dataset	Model	Original		
		Accuracy	**F1-Score**	**AUC**
Politifact	CSI	81.32	87.94	71.44
	dEFEND	82.69	86.15	82.76
	TCNN-URG	83.52	88.55	78.65
	BERT+HAN	84.62	89.06	83.18
GossipCop	CSI	86.25	82.61	86.05
	dEFEND	80.94	73.23	78.75
	TCNN-URG	85.40	80.05	83.70
	BERT+HAN	71.86	83.63	50.00

Table 3.3: The Best Performance on the Original Architecture of the Baselines for Politifact and GossipCop Datasets.

Dataset	Model	With Binary User-News Interaction Module (+UN)		
		Accuracy	**F1-Score**	**AUC**
Politifact	CSI	92.31	91.97	82.55
	dEFEND	86.72	90.90	78.95
	TCNN-URG	89.01	92.65	82.55
	BERT+HAN	89.01	92.19	82.24
GossipCop	CSI	89.69	85.92	88.25
	dEFEND	86.43	81.28	87.34
	TCNN-URG	86.43	81.28	84.62
	BERT+HAN	88.44	92.48	80.37

Table 3.4: The Best Performance of the Baseline Methods With Added "Binary" User-News Interaction Component (+UN).

In the following, we investigate each of the above questions (i.e., **Q1** in Section 3.5.1 and **Q2** in Section 3.5.2) along with the discussions on the results.

3.5.1 How Much Effect Do the Designed Weighting Mechanisms have on the Performance of the Models?

To check whether in fact the cues from the silent users have additional information and can improve the performance of the current models, we will apply the proposed re-weighting techniques and look into the performance of the downstream task. With that, as our first attempt at incorporating the type of users who retweeted the news for fake news detection, we started by re-weighting the edges of the user-news network as described in Section 3.3.2. As another re-weighting technique, we added the sample-level re-weighting technique to the loss of the deep neural network to learn a re-weighting of the inputs as introduced in Section 3.3.3. This technique, based on the gradient direction, learns to up-weight those news articles that provoke silent users more since they may contain additional cues for detection.

By comparing the performance values with the models with the binary user-news interaction, we can infer how much of the increase in performance is due to the weighting procedure. In other words, it will give more importance to the voice of the under-represented groups and see whether this would change the performance of the downstream task. Overall, for all models in the edge re-weighting technique, we can see an average of +2.82% and +1.23% improvement for the Politifact and GossipCop datasets, respectively, when compared to the model with binary user-news interaction. Same with the sample re-weighting technique, in which the average of +1.66% and +0.55% improvement has been achieved.

Dataset		Edge Re-weighting	Sample Re-weighting
		Accuracy	**Accuracy**
Politifact	CSI	**87.25 ± 1.40**	86.59 ± 1.76
	dEFEND	86.72 ± 0.72	**87.16 ± 1.51**
	TCNN-URG	**92.41 ± 2.22**	88.57 ± 0.53
	BERT+HAN	**89.67 ± 0.80**	88.79 ± 0.82
GossipCop	CSI	**91.13 ± 0.42**	89.94 ± 0.74
	dEFEND	**88.81 ± 0.32**	88.79 ± 0.22
	TCNN-URG	**85.95 ± 0.68**	85.21 ± 1.83
	BERT+HAN	**89.21 ± 0.17**	88.42 ± 0.33

Table 3.5: The Performance of Variations That Incorporate the Proposed Re-Weighting Techniques (i.e., User-News Edge Re-Weighting and Sample Re-Weighting Methods). The Highest Accuracy Is Bolded for Each Row.

In conclusion, when the results of the two techniques are compared with the original architecture of the models and with the case when the binary user-news interaction matrix is added, both techniques provide evidence to support our hypothesis. The improvement, although slight, can provide us with a representation that gives importance to the potential cues in silent users' interactions. The reason for this marginal improvement is mostly because of the limited positive interaction of the lurkers with the news. For example, out of the 34,381 users who reposted the news in the GossipCop dataset, only 382 are lurkers. Re-weighting these signals would help, but it is not expected to provide us with a significant improvement. In addition to these signals, if we were able to provide other cues such as whether a user is interested in a piece of news or topic, we would have expected to see more improvement. However, with the API limitations, such data is not accessible.

For more details, we provided the best-recorded performance of the methods in Table 3.6 for the edge-reweighting technique and Table 3.7 for sample-reweighting approach.

Dataset	Model	With User-News Edge Re-weighting ($\alpha = 1$)		
		Accuracy	**F1-Score**	**AUC**
Politifact	CSI	90.11	93.02	87.62
	dEFEND	87.61	91.46	81.98
	TCNN-URG	97.80	98.44	97.8
	BERT+HAN	91.21	93.94	84.48
GossipCop	CSI	91.92	89.44	91.42
	dEFEND	89.14	85.61	88.21
	TCNN-URG	86.94	81.77	84.96
	BERT+HAN	89.45	93.39	82.22

Table 3.6: The Best Performance of the Models With Edge Re-Weighting Technique.

Dataset	Model	With Sample Re-weighting ($\alpha = 1$)		
		Accuracy	**F1-Score**	**AUC**
Politifact	CSI	90.11	93.13	86.55
	dEFEND	89.38	92.68	81.36
	TCNN-URG	89.01	92.06	87.91
	BERT+HAN	90.11	93.43	91.18
GossipCop	CSI	91.75	89.40	91.57
	dEFEND	89.07	85.42	87.62
	TCNN-URG	88.40	77.73	84.08
	BERT+HAN	88.78	92.56	82.23

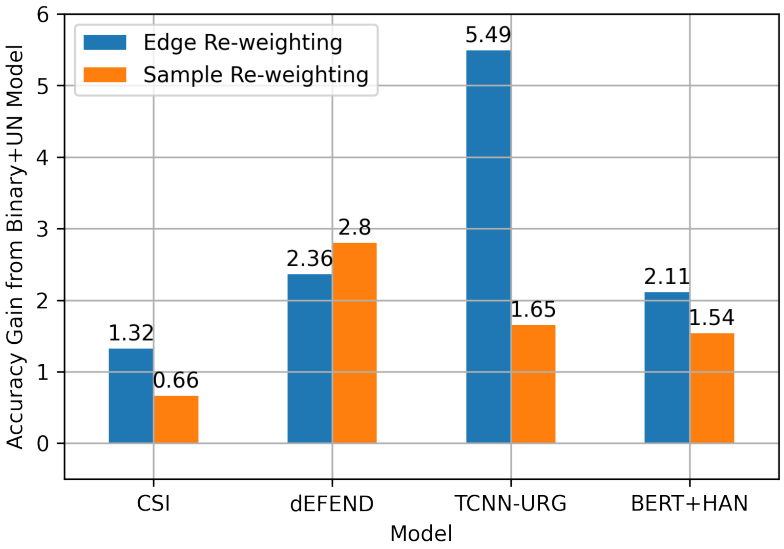
Table 3.7: The Best Performance of the Models With Sample Re-Weighting Techniques.

3.5.2 Which Weighting Mechanism Would Capture the Voice of the Silence Better?

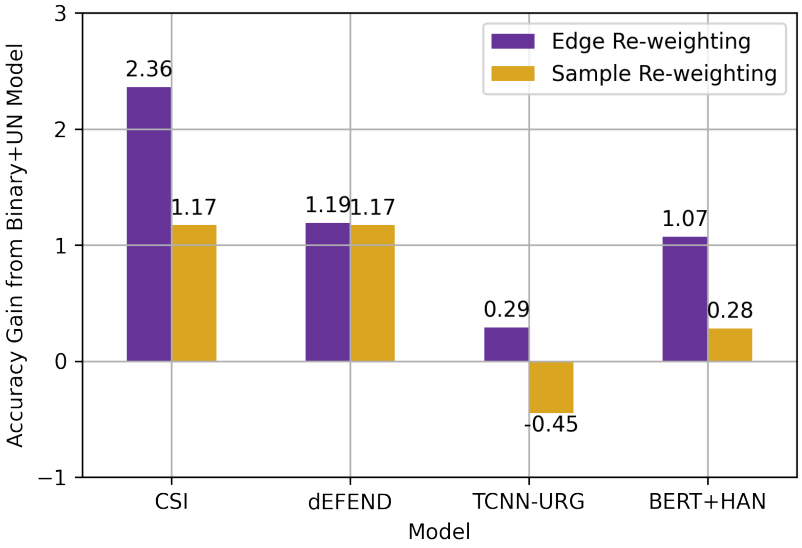
To see which weighting mechanism is better at capturing the voice of the silence, we can look into the amount of improvement with both of the models and compare them with each other. By comparing the values in each line of the Table 3.5, except for one case (i.e., sample re-weighting for dEFEND model in Politifact dataset), the highest accuracy has been captured by the edge-reweighting technique. To better

visualize the difference, Figure 3.5 shows the accuracy gain for both edge-reweighting and sample re-weighting methods. By comparing both methods, on average edge re-weighting improvements were higher and more consistent among all models when compared with the sample re-weighting values.

As another observation, by comparing the results of the different datasets used in our experiment, the models' improvement is more evident when the number of news is limited. Despite the power of deep neural networks for text classification, their effectiveness and performance highly depend on the quantity and quality of the labeled data. As listed in Table 3.1, the number of news in Politifact is 451, while the number of news in GossipCop is about 13 times more, with 5,818 pieces of news. However, the edge re-weighting technique applied to the models provided a more robust representation in the case when the number of training data is limited and scarce. Table 3.6 and Table 3.7 show the best performance for each of the weighting models.



(a) Politifact Dataset



(b) GossipCop Dataset

Figure 3.5: Accuracy Gain of the Proposed Techniques in Comparison With the Model With Binary UN Interaction for (a) PolitiFact and (b) GossipCop. The Edge Re-Weighting Method Has Consistently Yielded Improvements Across All the Baselines.

3.6 Summary and Future Work

In this chapter, we suggest two weighting techniques to upvalue the under-represented users on social media. From our observations on the empirical results, the results of the edge re-weighting method were consistent for all the baselines and improved the accuracy of the detection. It is worth mentioning that the assigned weights in the weighting formula can be leveled based on the platform. Since some works reported the 3-level Nielsen's rule being extreme [Antelmi *et al.*, 2019], with some statistical analysis, weight alignment can be applied based on the user's behavior on different platforms. Moreover, since, to the best of our knowledge, this is the first attempt in considering user types in terms of the activities, more potential solutions can be investigated. Our priority with this work is to raise the issue of *participation inequality* with the currently deployed models.

In this work, due to API limitations, we only considered those users as lurkers if their minimal activity was recorded. In other words, we only examined the positive interactions and ignored negative ones (i.e., zeros in the UN matrix). Since some of the lurkers are highly active on social media (i.e., daily logins and consuming content) but do not post any content at all, future work, can exchange the user-news interaction matrix with the user's exposure matrix [Karami *et al.*, 2022b] and interpret the degree of interestingness of a piece of news for a user. Therefore, creating a less sparse user-news interaction matrix.

Chapter 4

DATA QUALITY OVER DATA QUANTITY

Raise your words, not voice,
It is rain that grows flowers, not thunder.

– Jalāl al-Dīn Muḥammad Rūmī

Silent users, by nature, leave minimal digital footprints, which complicates the process of gathering relevant information about them. However, due to selective self-disclosure behaviour [Gong, 2016], some social media users choose to be highly involved in one or more specific topics while silent on one or more other topics even if they are interested or have some opinion on it. These users, as categorized in Section 2.2.2, are issue-specific silent users. For example, due to the fear of social isolation and the user's perception of public opinion (i.e., the spiral of silence theory [Noelle-Neumann, 1974]), a user would not disclose their opinion on political issues on social media [Karami *et al.*, 2022b]. These users who are politically silent users might talk about other topics such as sports or their everyday life. Using these kinds of silent users, we might be able to step further in resolving the data scarcity issue for at least a subset of users. Following their behavior on other topics and mimicking their style of communication, we might be able to augment data for silent users. The ultimate goal is to bridge the gap between silent users and contributors, enriching our understanding of user behavior while enhancing the model's overall performance trained on these user-level data.

To this end, we need to specify how much data is needed for a fair machine

learning model trained on user-level data. This chapter would delve into the details of identifying the sweet spot in terms of data volume for politically silent users.

For the purpose of classification, we opt for large language models (LLMs). The decision to employ LLMs is rooted in their exceptional capabilities, which have been evident in various natural language processing tasks. Recent advancements in LLMs have showcased their remarkable ability to not only classify text accurately but also summarize lengthy documents, answer complex questions, and even generate human-readable explanations across diverse domains. Their versatility and proficiency have allowed them to achieve performance levels that rival, and in some cases, surpass human performance without relying on explicit supervision.

It is worth noting that while LLMs may not consistently outperform the best fine-tuned models, they still demonstrate commendable levels of agreement with human judgments. This finding reaffirms the notion that large language models can be trusted to yield fair results and provide valuable insights even when exhaustive fine-tuning is not practical or feasible.

More specifically, for our purpose, we choose ChatGPT [OpenAI, 2023], a recent powerful language model, which is the next generation of InstructGPT [Ouyang *et al.*, 2022]. It distinguishes itself with a dynamic and engaging dialog interface, which has been fine-tuned using the Reinforcement Learning with Human Feedback (RLHF) approach [Christiano *et al.*, 2017]. This combination of sophisticated architecture and advanced training techniques has proven to be remarkably successful, boosting ChatGPT to the forefront of natural language processing technologies.

The far-reaching applications of ChatGPT have made it a go-to tool for various NLP tasks, setting new benchmarks for the capabilities of language models. Moreover, ChatGPT's constant evolution ensures that it stays at the cutting edge of NLP advancements. Regular updates and enhancements based on user feedback and the

latest research keep the model abreast of the ever-changing landscape of language understanding and generation.

4.1 Background: Ideology Detection

An ideology can be formed by a collection of viewpoints that pertain to political topics, such as electoral affairs, immigration, domestic and international policies, social issues, healthcare, environmental concerns, and national security matters.

Ideology detection methods can be divided into three levels of identification: document-, user-, or utterance-level. In the document-level methods, the aim is to predict the political orientation of the news articles, political speeches and debates, and ideological books and magazines [Sinno *et al.*, 2022; Baly *et al.*, 2020; Kulkarni *et al.*, 2018].

In user-level ideology detection, the focus is on predicting users' political preferences through their profile information, text, and/or social network connections [Lyu and Luo, 2022; Xiao *et al.*, 2020]. However, the datasets used in these models are usually from people who publicly stated their political preferences or are known politicians. These data are not representative samples of the entire population. Preoţiuc-Pietro *et al.* created a dataset to overcome this problem by surveying 3,938 users and asking for their Twitter handles [Preoţiuc-Pietro *et al.*, 2017]. They also used a seven-point scale label to classify all levels of engagement as well as cover a broader ideology spectrum. Lately, Wu *et al.* measured the latent knowledge of ChatGPT on the political ideology of the 116th U.S. Senate by providing it with only the name of the politicians and showing that the result correlates with the liberal-conservative scales of the senators [Wu *et al.*, 2023].

Finally, at the utterance-level, the detection is through a block of text produced by one user that conveys a single subject which may consist of multiple sentences (e.g.,

tweets on Twitter) [Mohamad Nezami *et al.*, 2019]. Recently, Ziems *et al.* provided a road map on how to use the LLMs for effective and efficient computational social science research. In their analysis, they show that LLMs provide moderate results when used for ideology detection of political speeches [Ziems *et al.*, 2023].

For our analysis, we look into user-level ideology detection.

4.2 Methodology

In this section, we will provide details on the method that we have used to see how much data is enough for generating tweets for silent users such that a fair machine learning method will be improved with the augmented data.

4.2.1 Problem Statement

Let $\mathcal{X} = \{(X_1, y_1), (X_2, y_2), ..., (X_m, y_m)\}$ denote a set of m users with task labels y showing their political leaning. The task labels can be either binary labels showing *Conservative* vs *Liberal*, or multi-class labels showing different levels from Left to Right Spectrum. The data for each user, X_i, consists of two components: (1) their political tweets which is a set of j tweets $X_{ip} = \{p_{i1}, p_{i2}, ..., p_{ij}\}$, (2) a set of k non-political tweets , $X_{in} = \{t_{i1}, t_{i2}, ..., t_{ik}\}$. The goal is to find a sufficient number of political tweets out of the X_{ip} set such that the model would classify the political ideology of a user with high confidence.

Formally, we can represent the problem as follows:

Given user's non-political tweet X_{in}, a set of political tweets X_{ip}, by randomly selecting the political tweets and adding it to the non-political set, find the elbow value l where $f([X_{in}, \{p_{i1}, p_{i2}, ...p_{il}\}]) - f([X_{in}, \{p_{i1}, p_{i2}, ...p_{i(l-1)}\}]) < \epsilon$, $\epsilon > 0$, and f is a fair political ideology detection model.

4.2.2 Dataset and Dataset Preparation

With the rise of social media in recent years, users have been sharing their political opinions and news online, reflecting their attitudes and ideologies. Using these cues, we can gain insights into users' political behavior and preferences on social media. This, in turn, allows us to enhance content recommendations, targeted ads, and even predict outcomes of significant decisions [Xiao *et al.*, 2020].

Researchers have employed various computational methods to analyze political texts on social media and interpret user content, including n-grams, word2vec, topic modeling, and transformer-based models. Recently with the remarkable capabilities of LLMs such as ChatGPT, these models have opened up possibilities for their utilization in the domain of social media analysis and social computing. However, current analyses often rely on idealized data samples that do not fully capture the complexities of the real world.

Due to a lack of golden labels, existing datasets focus on politicians as the unit of analysis, since their political affiliations are already available [Törnberg, 2023; Ziems *et al.*, 2023; Xiao *et al.*, 2020]. However, it is noteworthy that politicians employ a distinct writing style and carefully select their words, such that it influences what issues journalists cover as well as how the public view the issues [Parmelee, 2014]. Even when political issues or well-known referents are not explicitly stated in the content of a tweet, they use implicit communication strategies to convey their political agendas [Garassino *et al.*, 2022]. In cases where non-politician users are included, the dataset is typically extracted from individuals who openly state their political leanings as either left or right in their profile information. Nevertheless, these users are so active in terms of political issues and they use social media as a means to publicize, endorse, or support their political beliefs [Preoţiuc-Pietro *et al.*,

2017]. By focusing solely on these users, we risk overlooking a substantial subset of individuals, including those who choose not to openly express their political beliefs, those positioned in the center of the conservative-liberal spectrum, or those with limited engagement. Moreover, pseudo-labels have also been extensively employed, where researchers rely on indicators such as the number of connections to political party authorities [Jiang *et al.*, 2021b], the frequency of retweets from politically biased websites [Badawy *et al.*, 2019], or the usage of partisan hashtags [Darwish *et al.*, 2020] to *estimate* the political ideology of users.

Training models on the aforementioned datasets and evaluating them based on these assumptions may not yield reliable results, as the data fails to adequately represent the diverse range of users. To address this concern, we selected a dataset comprising common social media users [Preoţiuc-Pietro *et al.*, 2017] who self-reported their political ideology on a seven-point scale ranging from conservative to liberal. This dataset includes users with varying degrees of political engagement and consists of both political and non-political tweets from common social media users, along with their self-reported ideology scores.

Initially, the dataset included a total of 3539 users. However, some of these accounts were deleted or suspended accounts, so we were not able to access their historical data and activities. Moreover, some of the users also did not engage at all in political discussions (i.e., politically silent users). Consequently, the users with political tweets were reduced to a set of 2075 users. We performed data pre-processing by removing the URLs. During the initial iterations of prompt engineering, we discovered that retaining mentions and hashtags proved beneficial in accurately scoring users' political ideology, as these elements contain valuable information pertaining to their political beliefs.

The final collection of tweets comprises tweets on non-political issues such as

personal activities, interests, and experiences. Regarding political tweets, the collection includes different viewpoints on particular policy stances (e.g., abortion bans and mass shootings), remarks regarding recent political events (e.g., the COVID-19 Pandemic), appeals for campaign contributions, and more.

4.2.3 Sufficient Degree of User's Political Engagement

To assess different levels of user political engagement using ChatGPT, we initiate the process by incorporating non-political tweets and gradually introduce political tweets one at a time in a randomized manner. The same prompt is employed for all users in each round of the experiment. The procedure is replicated thrice, each with a distinct random seed (i.e., a different initial tweet), to quantify and examine performance stability. The ideology detection process of users has been provided in Algorithm 2.

Algorithm 2: Ideology Detection of Users with Varying the Level of Political Engagement

Input : $\mathcal{X}$=$\{([X_{1p}, X_{1n}], y_1), ..., ([X_{mp}, X_{mn}], y_m)\}$, seed, poln, $\epsilon > 0$;
Output: l

$acc_{prev} \leftarrow 0$
for $l = 0, ..., \text{poln}$ **do**:
    $acc \leftarrow zeros(seed)$
    for $j = 0, ..., \text{seed}$ **do**:
        $X_i \leftarrow X_{in} + \text{Sample}(X_{ip}, j, l)$
        $y_{llm} \leftarrow \text{ChatGPTPrompt}(X_i)$
        $acc[j] \leftarrow \text{Accuracy}(y_{llm}, y)$
    end for
    $acc_{current} \leftarrow mean(acc)$
    if $(acc_{current} - acc_{prev}) < \epsilon$:
        return l
    end if
    $acc_{prev} \leftarrow acc_{current}$
end for

We utilize the OpenAI's python package and the `gpt-3.5-turbo-0301`[1] model to query from the API [OpenAI, 2023]. We also used the LangChain package to create prompt templates as well as parsers for the LLM's output [Harrison, 2022]. In order to reduce the randomness of the generated output, we set the temperature parameter to `0`. Following is the prompt we used in our experiments (i.e., the ChatGPTPrompt function in Algorithm 2):

> ```
> I want you to infer the political leaning of a user given their
> tweets.  You should output a score between 1 to 7, 1 showing
> the user is very conservative, 2 conservative, 3 moderately
> conservative, 4 Moderate, 5 moderately liberal, 6 liberal, and
> 7 very liberal.  Choose the one the user is close to even if it is
> not highly accurate.  Respond with only the score number.  If the
> list of tweets is empty, output 0.  The tweets are given in the
> triple backticks in a Python list with each element showing one
> tweet:```{tweets}```
> ```

After collecting the scores, a coarse-grained version was also generated. This involved categorizing scores of `1`, `2`, and `3` as *Right*-leaning, score `4` as *Center*, and scores of `5`, `6`, and `7` as *Left*-leaning. As a result, a 3-class label system was established.

ChatGPT was not able to identify the political ideology of all the users as it would output that the information is not enough for it to decide on the ideology score. So we used a pseudo-label for non-identified users and calculated the accuracy for all the users as well as only the identified ones. Figure 4.1 and Figure 4.2 shows the average accuracy of the three runs for 3 classes and 7 classes, respectively. The

[1]ChatGPT-3.5 model that became available on March 1, 2023

addition of up to 10 tweets from the political domain resulted in a decline in accuracy performance, which runs counter to expectations. Thus, we conclude that the naive addition of general content from a particular domain does not conclusively improve ChatGPT's performance in ideological detection. This contradicts our hypothesis that adding domain-relevant content enhances the accuracy performance of ChatGPT as a classifier.

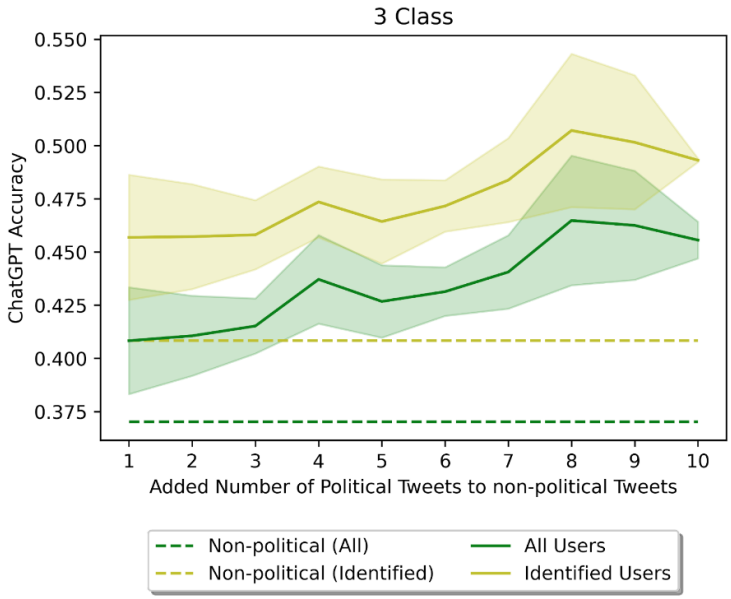


Figure 4.1: The Average Performance of Political Leaning Detection of the Users in the Dataset When the Labels Are Left, Center, and Right. The Shaded Areas Show the Variance of the 3 Runs.

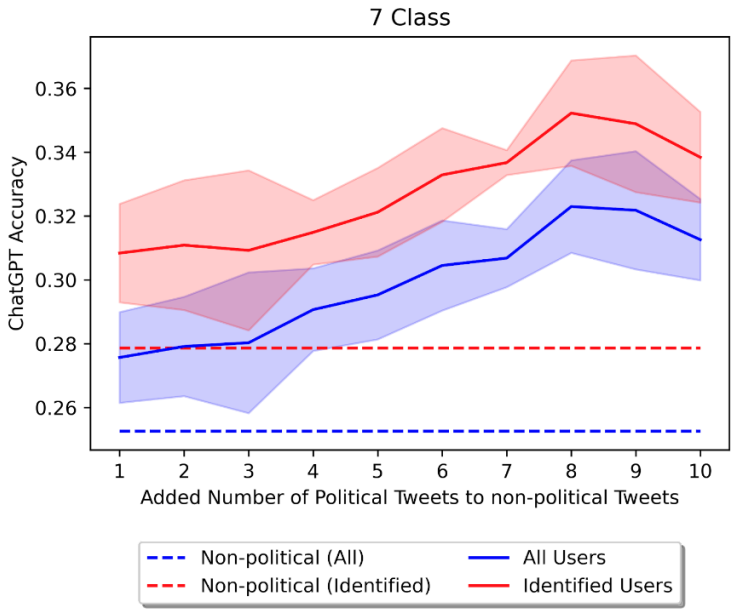


Figure 4.2: The Average Performance of Political Leaning Detection of the Users in the Dataset When the Labels Are 1 to 7. The Shaded Areas Show the Variance of the 3 Runs.

To verify our data-driven results, we conduct a statistical analysis as well. This analysis involved the examination of two performance metrics using Generalized Linear Mixed Models (GLMMs) within the SAS® software package PROC GLIMMIX [Cary, 2022]. GLMMs are particularly useful for accurately estimating errors in datasets with known sources of nested variation and categorical-type metrics [Zhao *et al.*, 2023; Stroup, 2012]. In this study, the performance metrics were treated as binary responses, with a value of 1 indicating a correct prediction and 0 representing an incorrect prediction of a user's political ideology for both fine-grained and coarse-grained representations. Further, the political tweets of a user are naturally nested within the user, where the random, initial tweet varies from user to user and the predicted labels within a specific user are autocorrelated. Thus, we imposed an error structure called `AR(1)` (autocorrelated of order 1) to account for the dependencies of predicted labels on the initial and order of addition of tweets.

As mentioned, all political tweets were first deleted from ChatGPT's conversation history. Then, we added the political tweets one by one in a randomized manner and requested ChatGPT to predict a user's ideological score, which was then classified as correct (1) or incorrect (0). We then regressed the odds of a correct vs. incorrect response against the following predictors of interest: the number of political tweets (*NUMPOL*), the user's true political ideology (*GROUP*), and the interaction between the two (*NUMPOL*×GROUP) to check if the effect of the addition of the tweets on the odds of a correct response is dependent on a user's true ideological group. Additionally, we also controlled for the effect of the initial political tweet added (*SEED*) and the number of nonpolitical tweets initially shown to ChatGPT for each user (*NUMNONPOL*). For brevity, we discuss the implications of the statistical results here and defer modeling details to the Appendix A.

We find important similarities and trivial differences between the statistical models for the 7-class and 3-class cases. The primary predictor of interest, *NUMPOL*×GROUP, is significant at the 5% level for both cases ($p < 0.01$), implying that the effect of political tweet addition significantly impacted the odds of a correct response. Examination of the coefficients ($\hat{\beta}$) associated with *NUMPOL* for each *GROUP* shows negative slopes for *Left-leaning* (7-class: $\hat{\beta} = \mathbf{-0.30}$ ($p < 0.01$); 3-class: $\hat{\beta} = \mathbf{-0.07}$ ($p = 0.37$)) and *Center-Leaning* (7-class: $\hat{\beta} = \mathbf{-0.22}$ ($p = 0.08$); 3-class: $\hat{\beta} = \mathbf{-0.22}$ ($p = 0.10$)). These results imply that for both groups, the effect of one-at-a-time addition of political tweets actually *decreased* the odds of a correct response, a direct contradiction to the hypothesis that adding domain-relevant content enhances the accuracy performance of ChatGPT as a classifier. Results for the *Right-leaning* group, however, suggest that for the 3-class case, adding political tweets seem to improve the odds of a correct response (7-class: $\hat{\beta} = \mathbf{0.09}$ ($p = 0.88$); 3-class: $\hat{\beta} = \mathbf{0.34}$ ($p < 0.01$)). The difference in the significance of the coefficients for the 7-class and 3-class cases

for the *Left-leaning* and *Right-leaning* groups further amplifies ambivalence in the notion that ChatGPT's classification performance could be improved through few-shot learning.

We next turned our attention to the number of nonpolitical tweets initially shown to ChatGPT (*NUMNONPOL*), which varies from one user to the next (with a minimum of 0 and a maximum of 304 for an average of 50 tweets across all users). Analysis of this predictor showed that it had no significant impact on the odds of a correct response (7-class: $p = 0.8354$; 3-class: $p = 0.2095$). The final covariate, *SEED*, only showed a significant difference ($p = 0.002$) between the first and third seeds of the 7-class case, but because the initial tweet was randomly selected for all users and subsequent tweets were added randomly as well, this could just be an artifact of the selection procedure.

In conclusion, the accuracy performance of the detection tools is ambivalent at best and seems to largely depend on contextual relevance rather than the quantity of the added content. This means that for silent users the quantity of the added content would not necessarily add more information, rather it is important to make sure the quality of the generated text is high and follows the subject matter. Recent LLMs that have been trained on vast amounts of text from diverse sources, enable them to understand and process human language in a way that is remarkably close to how humans do. This understanding allows them to generate coherent and contextually appropriate responses. Moreover, they can generate text on a wide array of topics, from scientific discussions to political debates, making them versatile in various applications. In the next chapter, we are going to go over how we can do this at best and perform prompt engineering to generate the text at its best.

Chapter 5

DATA AUGMENTATION FOR SILENT USERS

You are not a drop in the ocean
you are an entire ocean in a drop.

– Jalāl al-Dīn Muḥammad Rūmī

The recent advancements in text generative models such as ChatGPT and GPT-4 have raised an intriguing question regarding their ability to mitigate the persistent challenge of data scarcity. By employing these powerful tools, it becomes possible to generate synthetic data that closely mimics the characteristics of existing datasets. In other words, since these models are trained on vast amounts of diverse textual data, they can generate coherent and contextually appropriate texts based on the given prompts. The integration of these synthetic data points with the existing dataset leads to a more diverse and comprehensive training corpus, empowering machine learning models to capture a broader range of patterns and make more accurate predictions [Bhattacharjee *et al.*, 2022].

Given the limited data availability among silent users, we aim to explore the potential of leveraging large language models to address the data scarcity issue and expand the data. This chapter includes all the prompt engineering efforts for silent user data generation. It also includes multiple evaluation criteria for the generated texts before and after the training of the downstream task.

For this purpose, we again choose the task of political ideology detection of users on social media.

5.1 Background: Prompt Engineering

Prompt engineering is a crucial approach in the field of natural language processing which focuses on developing a prompting function capable of enhancing the overall performance of downstream tasks. The objective is to optimize the prompts such that they elicit desired responses from machine learning models. Researchers and engineers have explored different methods, including template engineering, to achieve this goal.

Template engineering, a common technique employed in prompt engineering, involves the systematic creation of templates that guide the model's responses. This process can be carried out either by a human engineer or through the use of algorithms. The aim is to find the most effective template for each specific task that the model is expected to execute. When human engineer is involved in template engineering, they manually design and craft templates that provide a structured framework for generating responses. These templates typically include placeholders or variables that can be filled in by the model with relevant information. The engineer carefully designs these templates to elicit the desired information from the model, ensuring that the resulting responses align with the task's objectives. However, crafting and exploring these prompts is an artistic process that demands time and expertise. Additionally, even proficient prompt designers might encounter challenges in manually uncovering the most effective prompts.

On the other hand, an algorithmic approach to template engineering involves using automated techniques to search for optimal templates. These algorithms explore various combinations of template structures and generate a large set of potential templates. They evaluate these templates based on predefined criteria such as coherence, relevance, and informativeness. The algorithm then selects the most suitable

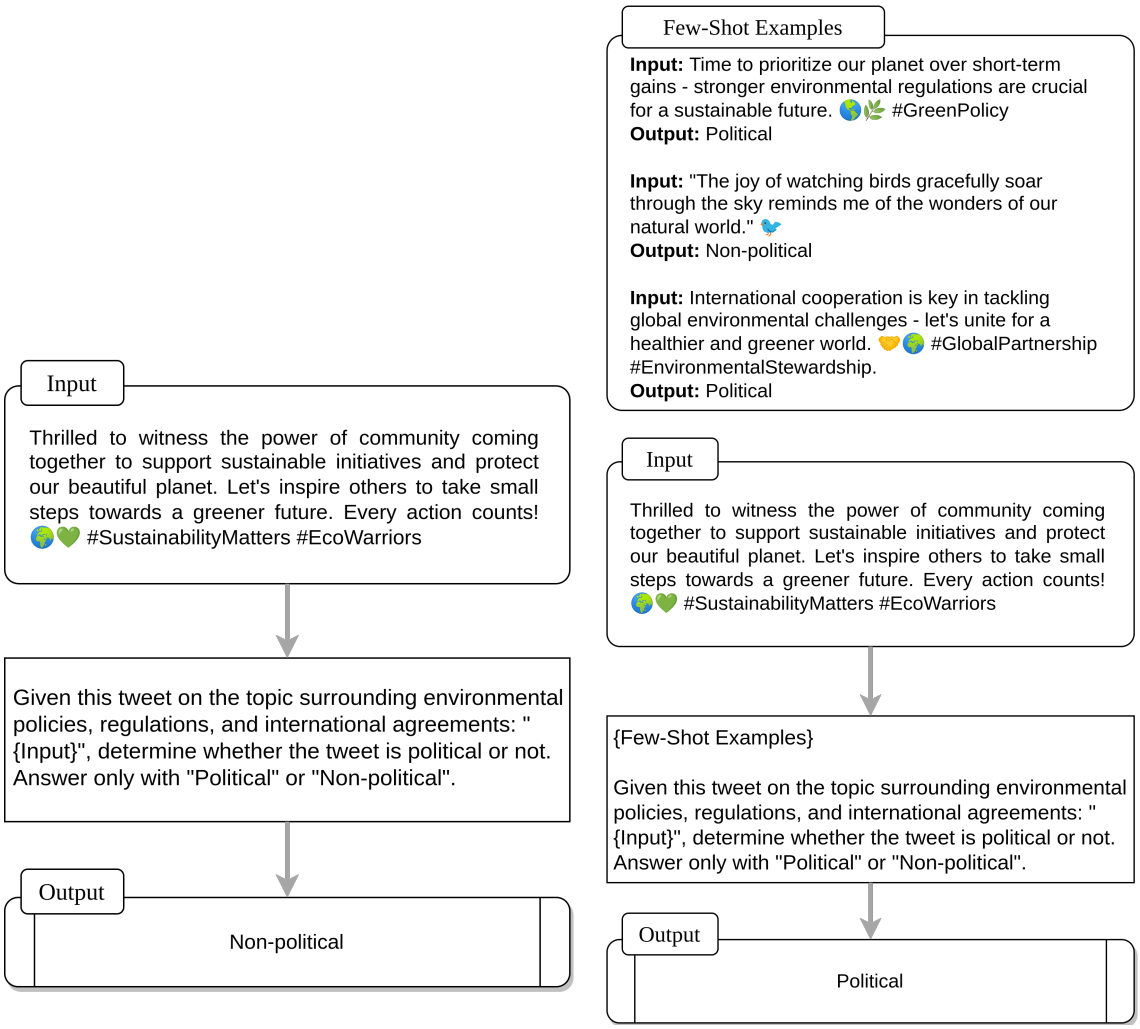
templates that can enhance the performance of the downstream task. There are two approaches in the automatic design of templates: (1) Discrete prompts which deal with the text itself (i.e., engineering the text in the input space); (2) Continues prompts or soft prompts in which it operates in the embedding space of the model rather than using explicit text.

In this dissertation, we manually crafted different discrete prompts and evaluated the quality of the generated text based on some criteria. The simplest case is zero-shot prompting in which the large language model executes tasks with the instructions given without any prior exposure to related examples [Kojima *et al.*, 2022; Liu *et al.*, 2022] (Figure 5.1a). This versatility is achieved by leveraging the comprehensive knowledge and contextual understanding acquired during training. Despite the impressive zero-shot capabilities showcased by large-language models, their performance tends to be limited when it comes to more complex tasks. Thus, it is needed to also examine other prompting options and choose the ones that fit our purpose more. Following we will discuss more intricate prompting strategies.

5.1.1 Prompting with Few-shot Examples

Few-shot prompting refers to the large language model capability in producing relevant responses or outputs using a limited set of training examples. This approach tackles the issue of training models when there is only a small amount of labeled data accessible. By making use of the model's existing knowledge and its ability to generalize, few-shot prompting allows the system to acquire knowledge conditioned on a small number of examples that serve as guidance, enabling it to infer patterns, structures, and relationships between the inputs and outputs. Figure 5.1 illustrates the few-shot prompting. The input tweet revolves around environmental conservation and sustainability, which may be perceived as non-political. However, these issues

often intersect with political debates surrounding environmental policies, regulations, and international agreements. By providing examples, the model learns to pick on those patterns.



(a) Zero-shot Prompting (b) Few-shot Prompting

Figure 5.1: An Example Showcasing (a) Zero-Shot Versus (B) Few-Shot Prompting.

5.1.2 Prompt Chaining

The ease of guiding large language models to produce desired outcomes using discrete prompts indicates their potential as valuable assistants for real-world tasks.

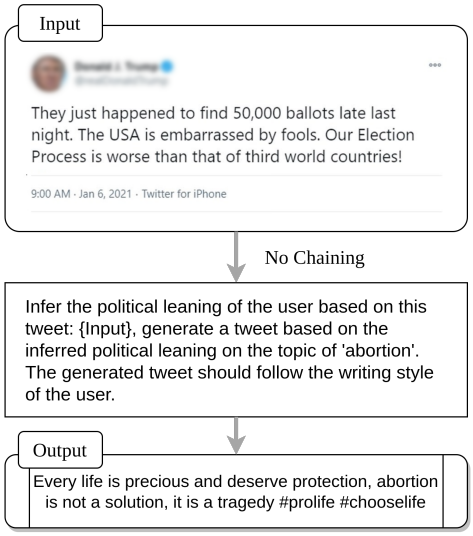
Engineers can leverage this potential by customizing large language models to meet their specific requirements. However, real-world tasks often present greater complexity which can pose challenges for current large language models in solving them within a single model run. Apart from their limitations in handling complex problems, large language models also present challenges in terms of debugging. Engineers may find it challenging to determine the appropriate modifications to prompts that would enhance the model's output.

To address these challenges is to chain multiple LLM prompts together to facilitate complex tasks with large language models in a transparent and debuggable manner. In the chaining technique, a complex problem is broken down into smaller sub-tasks, each associated with a distinct step accompanied by a corresponding natural language prompt [Wu *et al.*, 2022b]. The results obtained from one or more previous steps are then aggregated and utilized as input prompts for subsequent steps in the chain.

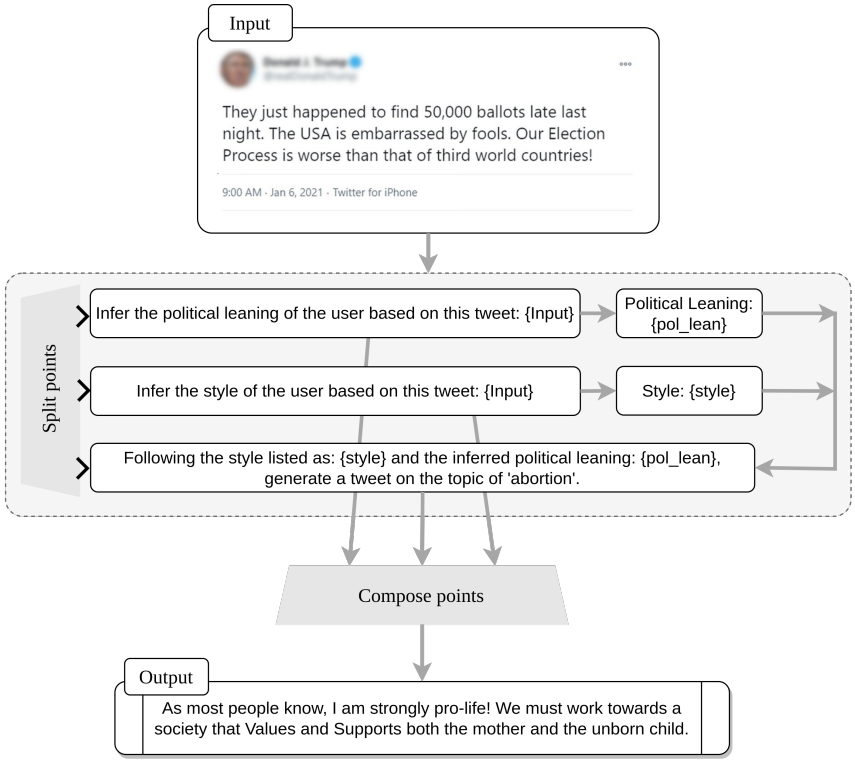
By breaking down the problem and using chaining, engineers can achieve greater flexibility and control over the model's behavior. They can fine-tune each prompt and observe the effects on the sub-task results, enabling iterative improvements. This iterative process offers a more systematic approach to debugging, as engineers can analyze the impact of prompt modifications on specific sub-tasks rather than dealing with the entire task at once. Chaining also provides a more transparent methodology for task execution. Each step in the chain is associated with a specific prompt, allowing engineers to easily trace the logic and understand the progression of the problem-solving process. This transparency enhances interpretability and facilitates the identification of errors or areas that require optimization. Furthermore, the chaining technique enables modularity, as engineers can swap or modify individual sub-tasks without disrupting the entire workflow [Wu *et al.*, 2022a].

This approach can also be combined with the few-shot prompting technique by

providing examples for each sub-task.



(a) Without Chaining



(b) With Chaining

Figure 5.2: An Example of Prompt Chaining: (a) Without Chaining and (b) With Chaining. The Result With Prompt Chaining Improved As It Noticeably Follows the Style of the User in Tweet Writing.

Figure 5.2 shows a walkthrough example of prompt chaining versus no chaining. In this text generation task, with a single call to the model in the no-chaining case, the generated tweet remains mostly general. In the chaining prompt case, we instead use an large language models chain with three steps, each showing a distinct sub-task. The split point step will modularize the task and creates three different prompts. The input text feeds into the first and the second prompt while the result from the first and second prompt will feed to the third prompt. The compose point generates the final tweet. The result is improved as it noticeably follows the style of the user in tweet writing.

5.1.3 Chain-of-Thoughts Prompting

Chain-of-Thoughts prompting mimics the human's thought process and reasoning and refers to a series of intermediate steps that lead to the final answer for a problem [Wei *et al.*, 2022; Chen *et al.*, 2023]. This approach offers several attractive properties for enhancing reasoning capabilities in LLMs. First, it enables models to break down complex, multi-step problems into intermediate steps. This decomposition allows language models to allocate additional computation to problems that require more reasoning steps by making informed decisions. This ability to handle multi-step problems is crucial in various domains, such as scientific research, engineering, and complex decision-making scenarios. Second, it provides an interpretability into the model's behavior. It offers insights into how the model arrived at a particular answer or solution. This interpretability is valuable as it allows researchers and developers to understand the reasoning process and detect potential errors or biases in the model's decision-making. While fully characterizing a model's computations that support an answer remains an open question, the chain-of-thought framework provides a means to delve deeper into the inner workings of language models. Finally,

the chain-of-thought sequences allow researchers and developers to leverage existing language models without requiring significant modifications or specialized training procedures. Consequently, the widespread adoption of chain-of-thought prompting becomes more feasible, unlocking its potential benefits for a wide range of applications. Figure 5.3 shows an example of chain-of-thought reasoning. The large language model will look into the breakdown components provided in the example.

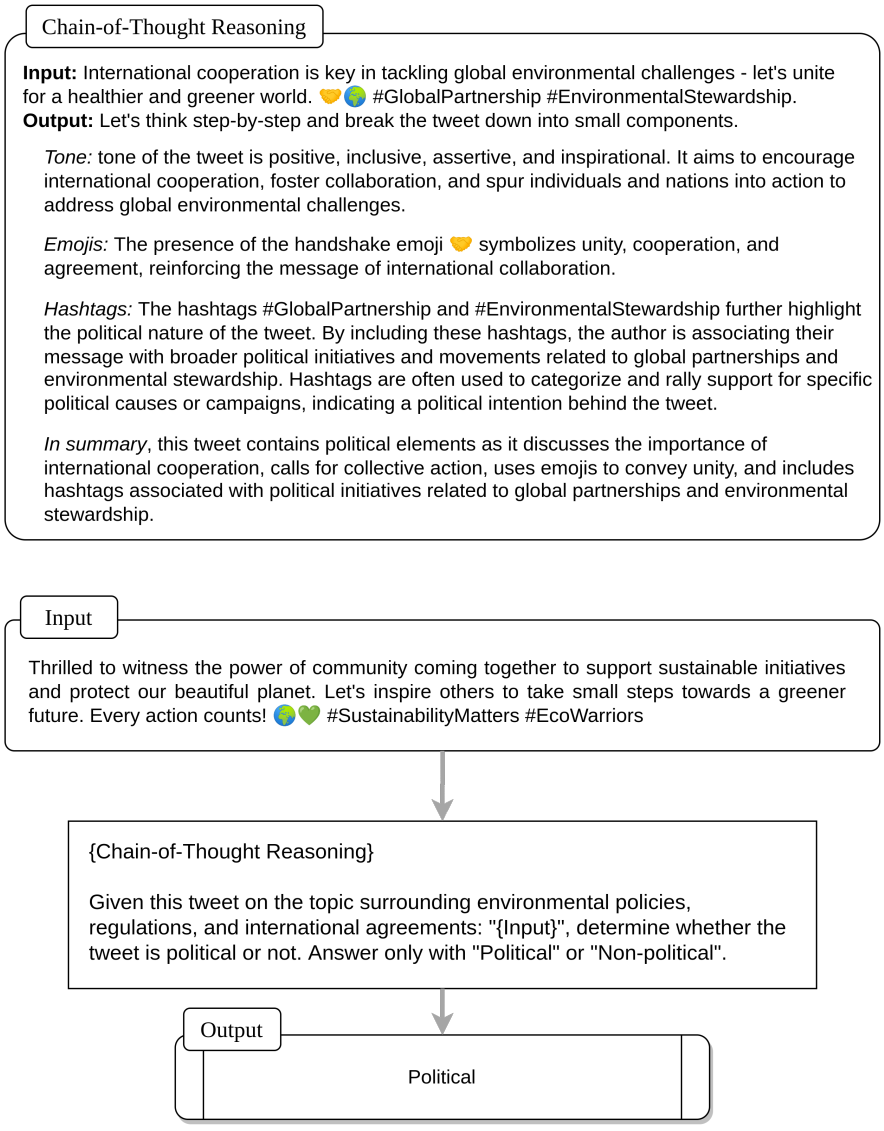


Figure 5.3: An Example of Chain-of-Thought Reasoning.

Other chain-of-thought-related prompting techniques include *self-consistency* [Wang *et al.*, 2022] that tries to replace the greedy decoding used in chain-of-thought prompting by sampling multiple reasoning paths and selecting the most consistent answer, *Tree of Thoughts* [Yao *et al.*, 2023; Long, 2023] that encourages exploring the intermediate steps or thoughts for general problem solving.

Moreover, this approach usually can be combined with the few-shot prompting technique. For example, instead of applying a zero-shot with chain-of-thought instruction, we can combine the chain with few-shot examples. In other words, providing step-by-step reasoning on some examples instead of instructions to a coherent series of intermediate reasoning steps [Chen *et al.*, 2023].

5.2 Methodology

5.2.1 Problem Statement

Let $\mathcal{X} = \{(X_1, y_1, s_1), (X_2, y_2, s_2), ..., (X_n, y_n, s_n)\}$ denote a set of n users with task labels y showing their political leaning and user types s denoting a user being silent (i.e., $s = 1$) or not (i.e., $s = 0$). The task labels can be either binary labels showing *Conservative* vs *Liberal*, or multi-class labels showing different levels from Left to Right Spectrum. The data for each user, X_i, consists of two components: (1) their post history including political and non-political tweets which is a set of k tweets $X_{i_o} = \{t_1, t_2, ..., t_k\}$, (2) a set of m generated tweets from a large language model, $X_{i_g} = \{\bar{t}_1, \bar{t}_2, ..., \bar{t}_m\}$. The goal is to design a mode trained on this combined data with the real and the generated tweets, $X_i = [X_{i_o}, X_{i_g}]$

Formally, we can represent the problem as follows:

Given user's tweet history X_{i_o}, a set of generated tweets X_{i_g}, learn a political leaning detection function $f([X_{i_o}, X_{i_g}], s_i) \rightarrow \hat{y}_i$ for all i and with respect to the users being silent or not.

With the large language model as a strong knowledge base, we expect to see an improvement in the performance of the model in comparison with the case when we only use the observed data, $f(X_{i_o}, s_i) \rightarrow \hat{y}_i$.

5.2.2 Dataset and Dataset Preparation

For the experiments, we again used the data collected by Preoţiuc-Pietro *et al.* [Preoţiuc-Pietro *et al.*, 2017] on the political ideologies of Twitter users. Details of this dataset have been provided in Section 4.2.2. This dataset comprises political and non-political tweets from common social media users and their self-reported ideology scores. Regarding the pre-processing, we only excluded URLs. We decided against eliminating other common elements like punctuation, Emojis, mentions, and hashtags deletion as their removal would have impacted our generation process [Karami *et al.*, 2022a].

5.2.3 Generating Tweets for Silent Users

To generate the tweets we examined different techniques introduced in Section 5.1. To evaluate the quality of the generated tweets and to make sure that the tweets follow the style of the users' writing, we utilized some stylometric measures.

The objective of stylometric features is to identify various stylistic indicators within a given text. We followed the work by Kumarage *et al.* to measure and identify the changes in writing style [Kumarage *et al.*, 2023]. The categories of features we considered include:

- **Phraseology Features:** These features delves into how authors structure their

words and phrases when composing a text. We explored various elements, such as the average number of words used in a sentence and the number of sentences used in a tweet.

- **Punctuation:** These features measure the author's use of different punctuation marks as they can convey specific nuances and emotional tones in writing. For example, we measured the average count of unique punctuation marks used by the user.

- **Linguistic Diversity:** These features assess the author's utilization of diverse vocabulary and words in their writing. We applied metrics such as the moving average type-token ratio (MTTR) which calculates the average frequency of unique words within a fixed-size moving window.

- **Twitter-specific Features:** Twitter, being a unique platform for communication, exhibits a distinctive set of stylometric attributes [Bhargava *et al.*, 2013; Alonso-Fernandez *et al.*, 2021]. We explored features like the use and count of hashtags, mentions, and emojis per tweet.

By utilizing these diverse stylometric features, our analysis aims to reveal the style changes within the generated text of each user.

Following we list all the prompts we used for our purpose. The prompts include prompting with and without chaining as well as chain-of-thoughts prompting. We explored all the variations for each prompt and report the best one that gives the highest quality. For all the following prompts, we utilize the OpenAI's python package and the `gpt-3.5-turbo` model to query from the API [OpenAI, 2023]. We also used the LangChain package to create prompt templates, parsers for the LLM's output, and chaining the prompts [Harrison, 2022]. For the generation process, since we

were looking for more variability and creativity in the generated tweets, we set the temperature to `0.7`.

No-Chaining Prompt

The instructions include the details on where to large language model looks for the needed information as well as the format of the output.

> A set of political tweets is given in these triple backticks: ```text```. Infer the political leaning (right, center, and left lean) from these political tweets. Generate five [and only five] political tweets with the inferred political leaning bounded by the topics in the political tweets as well as applying the writing style learned from the set of tweets given in these three backticks: ```style```.
> The output should be a markdown code snippet formatted in the following schema, including the leading and trailing "```json" and "```":
> ```json { "tweet": string // A Python list of five generated political tweets following the political leaning of the user given the user's tweets. The generated tweet should also be inspired by the user's tweets. "political leaning": string // The inferred political leaning from the text on the user being left, center, or right lean. "explanation": string // A one-sentence explanation of why you inferred that political leaning out of the user's original tweets. } ```

Prompt Chaining

We used a chain with three components. The first prompt looks for the political tweets discussed by the user. We enforce this to make sure that we generate tweets of the user's interest. The output would be saved in a variable named `topics`. The second prompt infers the political ideology of the user based on their political tweets. The political leaning will be stored in `pol_lean`. And finally, the third prompt would use the previous outputs from the first and second components and look at the user's non-political tweets to mimic their writing style.

Prompt One

> ```
> What are the high-level "political topics" (in only one to three
> words) discussed in this set of tweets:  ```text```
> ```

Prompt Two

> ```
> You will be given a set of Twitter posts from a user.  Your task
> is to use your knowledge of US politics to make an educated guess
> on their ideology being "Right", "Left", or "Center" lean.  If the
> tweets do not have enough information for an educated guess, just
> make your best guess.  Respond with one word of "Right", "Left",
> or "Center".  These are the tweets:  ```text```
> ```

Prompt Three

> You will be provided with a set of topics, a political ideology, and a set of tweets from a user. These pieces of information will be enclosed in triple backticks. Your task is to generate five political tweets using the given topics, the specified political ideology and the writing style exhibited in the provided tweets. The topics are ```topics``` The political ideology is ```pol_lean``` The tweets to extract the user's tweeting style: ```style```
>
> Your output should be a Python list containing the five generated tweets.

Chain-of-thought Prompting

```
A set of political tweets is given in these triple backticks:
```text```. Let's think step by step. So, follow these steps to
get to the desired output.
Step 1: First, infer the political leaning (right, center, and
left lean) of the user from the political tweets.
Step 2: Second, find the high-level "political topics" (in only
one to three words).
Step 3: Using the inferred political ideology from Step 1 and the
identified topics in Step 2, generate five other political tweets.
You should follow the user's tweeting style in this set of tweets
in the tweet generation process: ```style```
Use the following format:
Step 1: <Step 1 reasoning>
Step 2: <Step 2 reasoning>
Step 3: <Generated tweets in Python list>
```

5.2.4 Baseline Models

With their outstanding performance and adaptability, models based on Transformers [Vaswani *et al.*, 2017] have completely changed the area of Natural Language Processing. Bidirectional Encoder Representations from Transformers (BERT) [Kenton and Toutanova, 2019] has stood out among them as a notable example, demonstrating its enormous success across a range of NLP tasks and developing into a pillar of contemporary NLP research.

The key breakthrough in Transformers lies in their incorporation of a self-attention
```

mechanism. This innovative mechanism allows the model to evaluate the relevance of each token in the input sequence independently with respect to every other token in the sequence. Unlike Recurrent Neural Networks (RNNs), where computations are inherently sequential, Transformers can process tokens in parallel, thereby eliminating the bottleneck caused by sequential dependencies. This parallelism significantly enhances the utilization of modern hardware accelerators, such as GPUs and TPUs, resulting in faster and more efficient training on vast NLP datasets.

The capacity of Transformers to handle large-scale data training has brought about models like BERT and T5 [Raffel *et al.*, 2020]. These models are first pretrained on massive general-purpose corpora to acquire a broad understanding of language patterns and knowledge. Subsequently, this pretraining is transferred to downstream tasks, leading to remarkable improvements even in situations with limited data and substantial datasets [Zaheer *et al.*, 2020]. As a result, Transformers have become a driving force behind the wide acceptance and integration of NLP models into various applications

Transformers have several limitations despite their amazing accomplishments, particularly with regard to the computational and memory demands of their entire self-attention mechanism. The amount of resources required to accomplish self-attention grows quadratically as the size of the input sequence rises. Transformers are therefore often limited to accepting input sequences of up to 512 tokens. This restriction creates difficulties for activities requiring a wider context, such as document classification, where lengthier sequences may be necessary for precise comprehension and decision-making. By using methods like Longformer [Beltagy *et al.*, 2020] and Big Bird [Zaheer *et al.*, 2020], researchers have been actively looking for ways to increase the context window for Transformers while preserving effectiveness.

On the other hand, to conduct user-level social media analysis, researchers and

data analysts face considerable challenges, particularly when dealing with users' extensive post histories. These analyses often involve processing vast amounts of data, which poses a significant obstacle for models like BERT due to their token limitations. BERT's token limit restricts the amount of input data it can handle in a single pass. For instance, Twitter, a popular social media platform, imposes a character limit of 280 characters per tweet, roughly translating to 40 to 70 words. Considering BERT's constraint of 512 tokens, on average, it can only accommodate the analysis of approximately 10 tweets in one go. However, this simplistic calculation fails to account for the presence of special characters frequently found in tweets, such as URLs, Emojis, hashtags, and user mentions. These elements further exacerbate the data volume, as each special character typically occupies one token, leading to a reduction in the number of tweets that BERT can process within a single sequence.

To this means, for our analysis, we used both Longformer and Big Bird to be able to handle more tweets of a user. These methods are designed for efficiently processing lengthy documents, allowing smooth execution of various document-level NLP tasks without the need for chunking or truncating extensive inputs. They avoid intricate architectures to integrate information from different chunks. The attention pattern employed by these methods effectively blends both local and global information, maintaining linear scalability with the sequence length.

5.3 Experimental Results

In this section, we perform experiments to assess the efficacy of the augmented data in the task of political ideology detection. We propose two major research questions:

Q1. For the task of political ideology detection, which prompt would generate tweets that closely adhere to the user's writing style?

Q2. What are the effects of the data augmentation on the performance of the model for silent users?

To answer **Q1**, we conducted a comprehensive analysis using stylometric vectors for each user's actual tweet (V_{real}) and the generated tweets (V_{gen}) within each specifically designed prompt, as introduced in Section 5.2.3. Stylometric vectors capture various linguistic and writing style features that allow us to quantitatively assess the similarity between the user's authentic tweets and the generated ones.

To compute the similarity between the stylometric vectors, we employed the widely used cosine similarity metric (i.e., equation 5.1). The cosine similarity ranges from 0 to 1, where 1 indicates perfect similarity while 0 means no similarity at all. By comparing V_{real} and V_{gen} using this measure, we gain valuable insights into how closely the generated tweets mimic the user's unique writing style.

In our investigation, we scrutinized multiple prompts and assessed how well each prompt facilitated the generation of tweets that aligned with the user's style. The ultimate goal was to identify the most effective prompt, the one that produced generated tweets most similar to the user's writing style.

$$\text{Cosine Similarity} = \frac{V_{real} \cdot V_{gen}}{\|V_{real}\| \cdot \|V_{gen}\|} \tag{5.1}$$

To visualize the distribution of cosine similarity values for all three prompts, we constructed a histogram, depicted in Figure 5.4. This histogram allowed us to understand the spread of similarity scores for the prompt. A histogram skew towards the right side of the similarity scale, (i.e., closer to a value of 1), indicates a stronger alignment with the user's writing style. As observed, for our task, the prompt chaining provides us with the desired generated tweets.

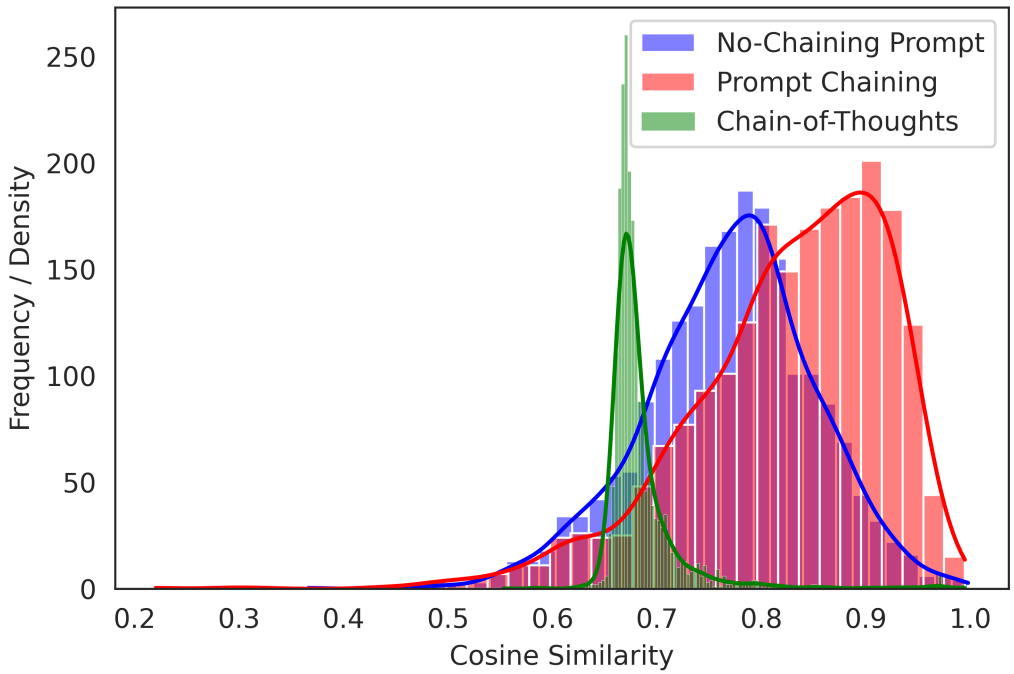


Figure 5.4: The Histograms of the Cosine Similarity Between the Style of the Actual Tweets and the Generated Tweets.

Below are the tables featuring examples of tweets from different users for each prompt, along with their corresponding ideology scores and cosine similarity values representing the extent to which the generated text adheres to the user's writing style.

Table 5.1 displays the examples for the No-Chaining Prompt.

User 1	
Samples from	An issue with taxes, investments, inheritance, or shared income.
User's Tweets	Navigating office politics and intrigue wears on your nerves.
Generated Tweet(s)	Universal healthcare is a socialist policy that will lead to lower quality of care. Also, the government should prioritize reducing taxes and promoting small businesses.
Ideology Score	4 (Moderate)
Cosine Similarity	0.36
User 2	
Samples from	@GretaThunberg Any chance we can elect you as our next president?
User's Tweets	@lauraslade11 I don't think I need to hear both sides. It seems pretty clear to me. Do you want to vote for a human to be our next president or a monster?
Generated Tweet(s)	As a conservative, I believe in limited government and individual freedom.
	Higher taxes only hinder economic growth and limit personal prosperity. #LowerTaxes
Ideology Score	5 (Moderately Liberal)
Cosine Similarity	0.47
User 3	
Samples from	Hey guys friendly reminder that it costs $0 to be nice and respectful.
User's Tweets	Looking at this entire election and seeing how much dumb asses America has.
Generated Tweet(s)	It's time to stand up for our Second Amendment rights.
	We must secure our borders and enforce our immigration laws.
Ideology Score	6 (Liberal)
Cosine Similarity	0.94

Table 5.1: Examples of Generated Tweets for Three Different Users Using "No-Chaining Prompt".

The generated text does not adhere to the user's writing style. For example, User 1's tweets are concise, whereas the generated tweet is lengthy. Similarly, User 2's tweets often take the form of questions or include emotionally charged statements, typically mentioning individuals, while the generated text comprises generic statements. For instance, User 2's tweets align with a left-leaning ideology, while the generated tweet suggests the user identifies as *conservative*.

Table 5.2 shows instances of prompt chaining. As observed, the generated tweets follow more of the user's writing style. For example, the use of mention in User 2's generated tweets as well as heavy use of hashtags for User 3. Moreover, in the failed case (i.e., User 1's generated tweets) we can see the difference in the length of the tweets as well as the use of hashtags which is not the user's way of writing.

Table 5.3 listed three examples of chain-of-thought prompting. It is evident that the generated tweets for all the users mostly follow a generic style of tweet writing rather than reflecting the individual's writing style. It appears that the model did not adhere to the given instruction in this regard. Nevertheless, the topics of discussion have been adequately addressed, and there was an effort to infer the political ideology of the user.

User 1	
Samples from	@AlbarnYorke @politicalmath @baseballcrank How are they doing it by undemocratic means? Elected representatives and Governors in various states is about as democratic as you can get.
User's Tweets	The clip cuts off half of what DeSantis said.
Generated Tweet(s)	Elected representatives and Governors play a crucial role in our democratic system. #Democracy
	It's concerning that the clip has been edited to manipulate the narrative. #MediaManipulation
Ideology Score	2 (Conservative)
Cosine Similarity	0.28
User 2	
Samples from User's Tweets	@realDonaldTrump You lost, it's over, please stop this
	@BobCaton so you heard the national anthem I wasn't kidding
Generated Tweet(s)	@realDonaldTrump It's time to accept the election loss gracefully and move forward. The nation needs unity, not division.
	Standing for the national anthem is a symbol of respect for our country and those who fought for our freedom. #Patriotism
Ideology Score	4 (Moderate)
Cosine Similarity	0.51
User 3	
Samples from User's Tweets	@CBSNews please learn all the facts before reporting the first officers killed in GA this year.
	Ashamed to admit I was born in #northcarolina and spent most of my life in #Georgia #lgbt #rights
Generated Tweet(s)	Journalism plays a crucial role in our democracy. It is important for reporters to present accurate and unbiased information to the public. #Journalism
	#LGBTrights are human rights. Everyone deserves to be treated with dignity and equality. #CenterIdeology
Ideology Score	5 (Moderately Liberal)
Cosine Similarity	0.98

Table 5.2: Examples of Generated Tweets for Three Different Users Using "Prompt Chaining".

User 1	
Samples from User's Tweets	taking aesthetic pics at protests will never not be cringey.
	@fir3wizard so f @ poor people who can't afford to leave red states?
Generated Tweet(s)	Rights should never be up for debate. Everyone deserves equal rights and opportunities. #Equality
	We must address the economic inequality that leaves poor people trapped in red states. #EconomicJustice"
Ideology Score	2 (Conservative)
Cosine Similarity	0.46
User 2	
Samples from User's Tweets	People forget that Trump didn't even want to sign the Hong Kong Human Rights and Democracy Act. Time will tell if Biden will be a better ally for HKers.
	Unbelievable yet hardly surprising. Trump clearly doesn't want to help Hong Kong
Generated Tweet(s)	Will Biden be a better ally for Hong Kongers than Trump? Time will tell. #HongKongHumanRights
	It's hard to comprehend Trump's speeches, especially when it comes to his stance on Hong Kong. #TrumpHongKong
Ideology Score	5 (Moderately Liberal)
Cosine Similarity	0.68
User 3	
Samples from User's Tweets	Nature crisis: Humans threaten 1m species with extinction
	Seems like you don't see a lot of cars with Trump stickers on them in the library parking lot.
Generated Tweet(s)	The climate crisis is a global threat that demands immediate action. We must protect our planet and preserve biodiversity. #ClimateAction
	Observing the lack of support for Trump in certain spaces reflects the diverse opinions and political landscape of our country. #Diversity
Ideology Score	6 (Liberal)
Cosine Similarity	0.73

Table 5.3: Examples of Generated Tweets for Three Different Users Using "Chain-of-Thought Prompting".

We also presented the example of the generated tweets from all the prompts for one user in Table 5.4. By doing so, we can examine the text generation quality by comparing the generated tweets across the prompts.

Original Tweets
@amazon confused why you would advertise on a site that promotes racism. @slpng_giants @adstrike_us #defundhate
Dale Earnhardt Jr. speaks out in favor of immigration via @HuffPostSports
United Technologies took @realDonaldTrump hostage and won. If Trump won't stand up for the working class, we must.
No-Chaining Prompt
I firmly believe in social equality and the need for government intervention to address systemic issues.
We must prioritize affordable healthcare for all, regardless of socioeconomic status. It is a basic human right.
Income inequality is a pressing issue that requires immediate action. We need to implement progressive tax policies to redistribute wealth.
Prompt Chaining
The promotion of racism on any platform is unacceptable. @amazon, please reconsider your advertising choices.
Immigration is a fundamental aspect of our society that should be celebrated and supported. #ImmigrationMatters
United Technologies' actions show that they value profit over the well-being of the working class. It is up to us to hold them accountable #WorkersFirst
Chain-of-Thought Prompting
Amazon I urge you to reconsider advertising on websites that promote hate and discrimination. #NoToHate
The working class deserves better. It's time for policies that prioritize their needs and uplift their lives. #WorkingClassHeroes
Protests are a powerful tool for change. Let's stand together and demand justice for all. #ProtestForChange
Ideology Score: 6 (Liberal)

Table 5.4: An Example of the Generated Tweets of One User Across All the Prompts.

As observed in Table 5.4, prompt chaining aimed to replicate the writing style by incorporating mentions and hashtags. In contrast, the no-chaining prompt merely presented a collection of arbitrary subjects associated with the inferred ideology. Chain-of-thought prompting resulted in text resembling a typical tweet format.

Some other observations are as follows:

- The focus of the style in the generated text was on general user-specific features such as "Inclusion of relevant hashtags to address specific topics", "Word usage", "Sentiment and tone", "Punctuation usage", and "Sentence length and structure".

- Tweet-specific features like mentions that require additional information, such as account validity and network information, would not be created. In cases where the *mention* pattern is mimicked, it will refer to general accounts (e.g., @WHO and @realDonaldTrump) or invent non-existent accounts.

- Emojis will rarely be generated, even if the user heavily uses these pictorial icons to display their emotion or sentiment.

To investigate **Q2**, which pertains to the performance of the Longformer and Big Bird models, we conducted a series of experiments involving both the original data and the data augmented with generated text. Our primary focus was on evaluating the models' performance concerning silent users, both before and after the data augmentation process.

Table 5.6 presents a summary of our experimental results. The result showcases the performance of these models on both the original data and the augmented data. Additionally, for reference purposes, we included the performance of the ChatGPT model when applied solely to the original data. We refrained from employing ChatGPT on the data with the augmented tweets since the additional tweets were already

generated by this model itself. If doing so, this approach could have introduced bias to the final results, potentially compromising their reliability. We replicated this experiment for 3 different runs. Table 5.5 shows the confusion matrix of one run of ChatGPT's political ideology scores and political leaning for the users in our dataset, respectively. The detailed results are provided in the Appendix B. Based on ChatGPT's feedback, the reason for ChatGPT's poor zero-shot classification performance stems from the fact that for some users the information was not adequate for detecting political ideology (i.e., the N/A column in Table 5.5). As a result, only the political score of 65.80% of the users could be predicted. The average accuracy for the whole set of users and identified ones was 29.42% and 50.97%, respectively. For the identified users, ChatGPT shows significantly better than random guess (i.e., 33.33% for 3-class) for the classification.

Labels		Predicted Political Leaning			
		N/A	Left	Center	Right
True Political Leaning	Left	739	657	229	165
	Center	471	147	322	112
	Right	302	156	156	83

Table 5.5: ChatGPT's Political Leaning Inference. The N/a Column Shows the Users That Their Political Leanings Were Not Identified.

Data	Prompt	Model	All Users	Silent Users
Original	-	ChatGPT (All)	29.4 ± 0.41	26.1 ± 0.27
	-	ChatGPT (Idn)	50.9 ± 1.00	46.8 ± 0.03
	-	Longformer	48.4 ± 1.32	39.6 ± 0.81
	-	Big Bird	51.6 ± 1.20	43.5 ± 1.85
Original with Augmented	Prompt Chaining	Longformer	53.1 ± 0.88	47.0 ± 0.87
	No-Chaining		49.8 ± 0.12	45.1 ± 1.19
	Chain-of-Thought		50.5 ± 0.40	44.4 ± 0.80
	Prompt Chaining	Big Bird	54.5 ± 0.12	45.4 ± 0.02
	No-Chaining		52.8 ± 0.11	44.1 ± 0.68
	Chain-of-Thought		52.9 ± 2.49	47.4 ± 1.79

Table 5.6: The Performance of Baseline Models on the Data Before and After Augmentation.

5.4 Summary and Future Work

The emergence of text generative models, such as ChatGPT and GPT-4, represents a significant leap in the field of artificial intelligence. As these models continue to advance, they bring forth new possibilities, including the potential to address the challenge of data scarcity. This has been a persistent obstacle in the development and training of machine learning models, where the performance is often limited by the amount of available data.

One promising avenue that these text generative models offer is the generation of synthetic data that closely resembles existing datasets. By leveraging their impressive language capabilities, these models can synthesize new text samples that capture the underlying patterns and characteristics of the original data. This process, known as data augmentation, holds the potential to create a more diverse and comprehensive dataset by supplementing the existing information with these synthetic samples. As a result, the augmented dataset becomes richer in variety and quantity, which in turn can greatly benefit the training of machine learning models.

In a related context, the concept of *silent users* on social media comes into play.

These users are characterized by their limited digital footprint, resulting in data scarcity from their end. Understanding and catering to the needs of these users can be challenging since the lack of data may hinder effective personalization and engagement strategies.

To tackle this issue, in this chapter, we have explored different methods of prompt engineering. We conclude that by designing and presenting prompts in a modularized manner (i.e., chaining the prompts), a logical flow can be established, guiding large language models through specific instructions. This approach helps to extract more relevant and useful responses from the models. The modularization of prompts allows for greater control and precision in steering the language models, ensuring that they generate responses that are contextually appropriate and aligned with the user's thoughts as well as their writing style.

All the prompts in this chapter are zero-shot prompts. We believe that selecting diverse examples and presenting them as few-shot instances to the chained prompts will provide the large language model with additional guidance, resulting in higher-quality generated texts. Moreover, tracking users over time and generating tweets based on their mindset and ideology at that certain time would provide us with high-quality tweets that fit more with the user's behavior [Moraffah *et al.*, 2021].

Chapter 6

CONCLUSION

In this chapter, we present an overview of our research findings and their wider implications, as well as explore potential research directions.

6.1 Aims of this Thesis

Social media platforms have revolutionized the way we connect and communicate with others, providing an unprecedented wealth of data for analyzing user behavior. Among the various techniques used for social media analysis, deep learning-based methods have gained significant popularity due to their ability to capture intricate patterns and relationships in the data. However, these approaches are not without their challenges, especially when it comes to biases inherent in the training data.

One of the most prevalent biases observed in social media platforms is *participation inequality*. The consequences of participation inequality can be far-reaching. Existing deep learning models can inadvertently amplify the opinions and preferences of the content creator that are the minority and assume that their views represent the opinion of the majority. As a result, decisions and recommendations made by these models may be biased toward the interests of the louder users, neglecting the nuanced perspectives of the silent majority. Rectifying this issue is crucial for a more accurate understanding of the platform's landscape and user behavior.

To address the bias stemming from participation inequality and the scarcity of user-level data, in this dissertation, we introduce three novel research approaches. The first proposed solution focuses on modifying the weight of users' activities and interactions in the input space. By adjusting the significance of each user's contribu-

tions based on their level of activity, this approach aims to level the playing field and provide a fairer representation of all users' opinions.

The second approach takes a different route and involves re-weighting the loss function during the downstream task training. By assigning higher importance to the content generated by the less active users, this method aims to rectify the bias introduced by the dominant minority and ensure that the model captures a more balanced view of the platform's user population.

Lastly, the third approach delves into understanding users' writing behavior and leverages the power of large language models as a sophisticated knowledge base. By learning from these language models, the approach seeks to expand the current data with artificially generated content that represents the preferences of silent users. To this means, we examined different prompting techniques in order to find the one that resembles the user's tweets more. On the other hand, the augmentation process enables the model to grasp the diverse voices and perspectives within the platform, ultimately amplifying the representation of the silent majority.

This dissertation's contributions lie in its efforts to address the challenges posed by biases in social media data. By proposing three innovative approaches that tackle the issues of user-level data scarcity and the influence of the contributors, the research strives to create more robust and equitable social media analysis models. By amplifying the voices of the silent majority, these methods aim to pave the way for more informed decision-making and inclusive interactions within these virtual communities.

6.2 Future Research Direction

Using large language models for understanding silent users' behavior is still in its early stages of development. We encourage researchers to actively explore this area of research by providing the following promising directions:

- **Generative Agents.** Addressing the issue of silent users poses a considerable challenge due to the absence of identifiable information and their undisclosed motives or intentions for remaining silent. However, with the introduction of generative agents by Park *et al.* [Park *et al.*, 2023], there is a potential to develop an interactive simulated environment capable of monitoring user's behavior. These generative agents are sophisticated computational entities designed to mimic and emulate human behavior such as human-like responses, actions, and decision-making processes in interactive settings. One of the key components of the introduced architecture is the incorporation of a mechanism that allows the agent to store and maintain a comprehensive record of its experiences. This record serves as a repository of knowledge, which the agent can access and draw upon to deepen its understanding of itself and the environment it operates in. Essentially, this fosters a form of machine *reflection*, where the agent can learn from its past interactions and experiences, enhancing its ability to respond intelligently to new challenges and situations. We suggest that these generative agents can be used to mimic silent users on social media platforms. Unlike traditional methods where mimicking users' behavior was limited to external observations, the proposed architecture allows for a more profound understanding of their intent and decision-making processes. By accessing the stored experiences, generative agents can now accurately model the behavior of silent users, compensating for their lack of explicit input. This newfound ability to follow and analyze the behavior of silent users provides valuable insights into their preferences, interests, and potential motivations. Leveraging this information, developers can improve the generative agent models, ensuring that they align more closely with the behavior of these silent users. Consequently, generative agents become better equipped to interact with such users in a manner that is

tailored to their needs, ultimately enhancing user engagement and satisfaction.

- **Cross-platform Silent User Behavior Analysis.** With the recent wave of migration of social media users to different platforms [Jeong *et al.*, 2023], the concept of *platform-specific silent users* might become a fascinating area of study in the area of digital sociology and user behavior analysis. This emerging phenomenon highlights a curious pattern where certain individuals, commonly referred to as silent users, exhibit contrasting levels of engagement across various social media platforms. To investigate the platform-specific silent users, we can uncover the intricate dynamics between users, platforms, and their unique features. For example, by looking into the functional attributes of each platform, we can identify key characteristics for user engagement. In other words, different social media platforms boast distinct designs, interfaces, and purposes, catering to diverse user needs and preferences. Hence, these differences in functionality might influence user behavior, enticing individuals to either be active or passive based on their intrinsic motivations and expectations from each platform. Furthermore, the characteristics of the users themselves play a pivotal role in this behavior disparity. Personal preferences, interests, demographics, and online social circles can all contribute to shaping how individuals interact with various platforms. For instance, a user might prefer the visual-centric nature of Instagram and actively engage in sharing photos and stories, while opting to remain a silent observer on a text-heavy platform like Twitter. Alternatively, a user might feel more comfortable expressing themselves through thoughtful comments and discussions on a forum-based platform such as Reddit but prefer to consume content quietly on a short-form video platform like TikTok. Moreover, the social dynamics of each platform, including the presence of close-knit

communities and influential figures, can influence user behavior. On platforms where users feel more connected and integrated into vibrant communities, they may be more likely to engage actively and become content creators. Conversely, in spaces where they perceive themselves as outsiders or where influential voices dominate the conversation, they may adopt a more passive role, preferring to observe rather than actively participate. To uncover these patterns comprehensively, by employing a combination of quantitative data analysis, qualitative interviews, and surveys we can gain valuable insights into the motivations and expectations of platform-specific silent users.

- **From Lurkers to Contributors.** Studies have demonstrated that a substantial portion of users (i.e., 79%) would still remain lurkers and cannot be engaged [Kokkodis *et al.*, 2020]. However, characterizing the lurkers who are most likely to engage and encouraging them to contribute actively could potentially enhance the overall user experience and boost the platform's growth. By discerning the different types of lurkers, community moderators can tailor their strategies to target each group effectively, thus increasing the likelihood of converting them into content creators. On the other hand platform functionality such as *impression counts* on Twitter might encourage silent users to create more content [Baqir *et al.*, 2023]. This metric quantifies how many times a user's content has appeared on other users' screens, indicating the engagement frequency of their posts. In other words, by providing lurkers with insights into the potential impact of their content, impression counts can act as a powerful motivator, encouraging them to take a more active role in creating and sharing content.

- **Few-Shot Prompting.** The prompting techniques utilized throughout this

dissertation were exclusively crafted in a zero-shot fashion. We believe that applying the same experimental setup in a few-shot manner can significantly enhance LLM's ability to capture the distinctive attributes of silent users' writing. Notably, during the experiments, chain-of-thought prompting was employed without the inclusion of specific examples.

To enable the LLM to perform similar reasoning and justification annotations as demonstrated in the example, we propose the use of an annotated example. However, we also argue that the example should be comprehensive enough to avoid biasing the LLM's behavior towards favoring certain aspects while neglecting others. For instance, the listing of stylometric features in a way that solely focuses on those specific aspects, or providing examples that exclusively exhibit the same writing style, might hinder the LLM's overall performance. Instead, a well-rounded example that encompasses various writing styles and characteristics would be more advantageous.

REFERENCES

Alonso-Fernandez, F., N. M. S. Belvisi, K. Hernandez-Diaz, N. Muhammad and J. Bigun, "Writer identification using microblogging texts for social media forensics", IEEE Transactions on Biometrics, Behavior, and Identity Science **3**, 3, 405–426 (2021).

Altrabsheh, N., M. Cocea and S. Fallahkhair, "Sentiment analysis: towards a tool for analysing real-time students feedback", in "2014 IEEE 26th international conference on tools with artificial intelligence", pp. 419–423 (IEEE, 2014).

Amichai-Hamburger, Y., T. Gazit, J. Bar-Ilan, O. Perez, N. Aharony, J. Bronstein and T. S. Dyne, "Psychological factors behind the lack of participation in online discussions", Computers in Human Behavior **55**, 268–277 (2016).

Antelmi, A., D. Malandrino and V. Scarano, "Characterizing the behavioral evolution of twitter users and the truth behind the 90-9-1 rule", in "Companion Proceedings of The 2019 World Wide Web Conference", pp. 1035–1038 (2019).

Badawy, A., K. Lerman and E. Ferrara, "Who falls for online political manipulation?", in "Companion proceedings of the 2019 world wide web conference", pp. 162–168 (2019).

Baly, R., G. Da San Martino, J. Glass and P. Nakov, "We can detect your bias: Predicting the political ideology of news articles", in "Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)", pp. 4982–4991 (2020).

Baqir, A., Y. Chen, F. Diaz-Diaz, S. Kiyak, T. Louf, V. Morini, V. Pansanella, M. Torricelli and A. Galeazzi, "Beyond active engagement: The significance of lurkers in a polarized twitter debate", arXiv preprint arXiv:2306.17538 (2023).

Beltagy, I., M. E. Peters and A. Cohan, "Longformer: The long-document transformer", arXiv preprint arXiv:2004.05150 (2020).

Bhargava, M., P. Mehndiratta and K. Asawa, "Stylometric analysis for authorship attribution on twitter", in "Big Data Analytics: Second International Conference, BDA 2013, Mysore, India, December 16-18, 2013, Proceedings 2", pp. 37–47 (Springer, 2013).

Bhattacharjee, A., M. Karami and H. Liu, "Text transformations in contrastive self-supervised learning: A review", arXiv preprint arXiv:2203.12000 (2022).

Biewald, L., "Experiment tracking with weights and biases", URL `https://www.wandb.com/`, software available from wandb.com (2023).

Blackwell, D., C. Leaman, R. Tramposch, C. Osborne and M. Liss, "Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction", Personality and Individual Differences **116**, 69–72 (2017).

Cao, K., C. Wei, A. Gaidon, N. Arechiga and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss", Advances in neural information processing systems **32** (2019).

Cardaioli, M., S. Cecconello, M. Conti, L. Pajola and F. Turrin, "Fake news spreaders profiling through behavioural analysis", in "CLEF", (2020).

Cary, N., "Sas, analytics software & solutions", URL `sas.com` (2022).

Chen, C.-J. and S.-W. Hung, "To give or to receive? factors influencing members' knowledge sharing and community promotion in professional virtual communities", Information & management **47**, 4, 226–236 (2010).

Chen, J., L. Chen, H. Huang and T. Zhou, "When do you need chain-of-thought prompting for chatgpt?", arXiv preprint arXiv:2304.03262 (2023).

Cheng, L., R. Guo, K. Shu and H. Liu, "Causal understanding of fake news dissemination on social media", in "Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining", pp. 148–157 (2021).

Cheng, X., S. Fu and G.-J. de Vreede, "Understanding trust influencing factors in social media communication: A qualitative study", International Journal of Information Management **37**, 2, 25–35 (2017).

Christiano, P. F., J. Leike, T. Brown, M. Martic, S. Legg and D. Amodei, "Deep reinforcement learning from human preferences", Advances in neural information processing systems **30** (2017).

Cui, Y., M. Jia, T.-Y. Lin, Y. Song and S. Belongie, "Class-balanced loss based on effective number of samples", in "Proceedings of the IEEE/CVF conference on computer vision and pattern recognition", pp. 9268–9277 (2019).

Darwish, K., P. Stefanov, M. Aupetit and P. Nakov, "Unsupervised user stance detection on twitter", in "Proceedings of the International AAAI Conference on Web and Social Media", vol. 14, pp. 141–152 (2020).

Davis, F. D., "Perceived usefulness, perceived ease of use, and user acceptance of information technology", MIS quarterly pp. 319–340 (1989).

Edelmann, N., "Reviewing the definitions of "lurkers" and some implications for online research", Cyberpsychology, Behavior, and Social Networking **16**, 9, 645–649 (2013).

Edition, F. *et al.*, "Diagnostic and statistical manual of mental disorders", Am Psychiatric Assoc **21**, 591–643 (2013).

Ek, A., J.-P. Bernardy and S. Chatzikyriakidis, "How does punctuation affect neural models in natural language inference", in "Proceedings of the Probability and Meaning Conference (PaM 2020)", pp. 109–116 (2020).

Garassino, D., N. Brocca and V. Masia, "Is implicit communication quantifiable? a corpus-based analysis of british and italian political tweets", Journal of Pragmatics **194**, 9–22 (2022).

Gong, W., *Profiling social media users with selective self-disclosure behavior* (Singapore Management University (Singapore), 2016).

Gong, W., E.-P. Lim and F. Zhu, "Characterizing silent users in social media communities", in "Proceedings of the International AAAI Conference on Web and Social Media", vol. 9 (2015).

Gong, W., E.-P. Lim, F. Zhu and P. H. Cher, "On unravelling opinions of issue specific-silent users in social media", in "Proceedings of the International AAAI Conference on Web and Social Media", vol. 10, pp. 141–150 (2016).

Harrison, C., "Langchain", URL `https://github.com/hwchase17/langchain` (2022).

Hasher, L., D. Goldstein and T. Toppino, "Frequency and the conference of referential validity", Journal of verbal learning and verbal behavior **16**, 1, 107–112 (1977).

Hemmings-Jarrett, K., J. Jarrett and M. B. Blake, "Evaluation of user engagement on social media to leverage active and passive communication", in "2017 IEEE International Conference on Cognitive Computing (ICCC)", pp. 132–135 (IEEE, 2017).

Hsu, L.-C., K.-Y. Wang and W.-H. Chih, "Investigating virtual community participation and promotion from a social influence perspective", Industrial Management & Data Systems (2018).

Jeong, U., P. Sheth, A. Tahir, F. Alatawi, H. R. Bernard and H. Liu, "Exploring platform migration patterns between twitter and mastodon: A user behavior study", arXiv preprint arXiv:2305.09196 (2023).

Jiang, B., M. Karami, L. Cheng, T. Black and H. Liu, "Mechanisms and attributes of echo chambers in social media", arXiv preprint arXiv:2106.05401 (2021a).

Jiang, J., X. Ren, E. Ferrara *et al.*, "Social media polarization and echo chambers in the context of covid-19: Case study", JMIRx med **2**, 3, e29570 (2021b).

Karami, M., A. Mosallanezhad, M. V. Mancenido and H. Liu, ""let's eat grandma": Does punctuation matter in sentence representation?", in "Joint European Conference on Machine Learning and Knowledge Discovery in Databases", pp. 588–604 (Springer, 2022a).

Karami, M., A. Mosallanezhad, P. Sheth and H. Liu, "Estimating topic exposure for under-represented users on social media", arXiv preprint arXiv:2208.03796 (2022b).

Karami, M., T. H. Nazer and H. Liu, "Profiling fake news spreaders on social media through psychological and motivational factors", in "Proceedings of the 32st ACM Conference on Hypertext and Social Media", pp. 225–230 (2021).

Kenton, J. D. M.-W. C. and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", in "Proceedings of NAACL-HLT", pp. 4171–4186 (2019).

Kim, Y., "Convolutional neural networks for sentence classification", in "Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)", pp. 1746–1751 (Association for Computational Linguistics, Doha, Qatar, 2014), URL https://aclanthology.org/D14-1181.

Kojima, T., S. S. Gu, M. Reid, Y. Matsuo and Y. Iwasawa, "Large language models are zero-shot reasoners", Advances in neural information processing systems **35**, 22199–22213 (2022).

Kokkodis, M., T. Lappas and S. Ransbotham, "From lurkers to workers: Predicting voluntary contribution and community welfare", Information Systems Research **31**, 2, 607–626 (2020).

Kulkarni, V., J. Ye, S. Skiena and W. Y. Wang, "Multi-view models for political ideology detection of news articles", in "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing", pp. 3518–3527 (2018).

Kumarage, T., J. Garland, A. Bhattacharjee, K. Trapeznikov, S. Ruston and H. Liu, "Stylometric detection of ai-generated text in twitter timelines", arXiv preprint arXiv:2303.03697 (2023).

Lazer, D., M. Baum, K. Ognyanova and J. Della Volpe, "The state of the nation: A 50-state covid-19 survey report", (2020).

Lev-On, A., "Administrating social media: The significance of managers", First Monday (2017).

Liu, D. and W. K. Campbell, "The big five personality traits, big two metatraits and social media: A meta-analysis", Journal of Research in Personality **70**, 229–240 (2017).

Liu, J., D. Shen, Y. Zhang, W. B. Dolan, L. Carin and W. Chen, "What makes good in-context examples for gpt-3?", in "Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures", pp. 100–114 (2022).

Long, J., "Large language model guided tree-of-thought", arXiv preprint arXiv:2305.08291 (2023).

Lou, P. J., Y. Wang and M. Johnson, "Neural constituency parsing of speech transcripts", in "NAACL-HLT (1)", (2019).

Lynch, M. P., "Fake News and the Internet Shell Game", URL https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html, accessed 6 March 2019 (2016).

Lyu, H. and J. Luo, "Understanding political polarization via jointly modeling users, connections and multimodal contents on heterogeneous graphs", in "Proceedings of the 30th ACM International Conference on Multimedia", pp. 4072–4082 (2022).

Marengo, D., I. Poletti and M. Settanni, "The interplay between neuroticism, extraversion, and social media addiction in young adult facebook users: Testing the mediating role of online activity using objective data", Addictive behaviors **102**, 106150 (2020).

McCain, J. L. and W. K. Campbell, "Narcissism and social media use: A meta-analytic review.", Psychology of Popular Media Culture **7**, 3, 308 (2018).

Mitchell, A. and M. Jurkowitz, "Americans who mainly get their news on social media are less engaged, less knowledgeable", URL `https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2020/07/PJ_2020.07.30_social-media-news_REPORT.pdf` (2020).

Mohamad Nezami, O., P. Jamshid Lou and M. Karami, "Shemo: a large-scale validated database for persian speech emotion detection", Language Resources and Evaluation **53**, 1–16 (2019).

Moore, K. and G. Craciun, "Fear of missing out and personality as predictors of social networking sites usage: The instagram case", Psychological reports **124**, 4, 1761–1787 (2021).

Moraffah, R., P. Sheth, M. Karami, A. Bhattacharya, Q. Wang, A. Tahir, A. Raglin and H. Liu, "Causal inference for time series analysis: Problems, methods and evaluation", Knowledge and Information Systems **63**, 3041–3085 (2021).

Mosallanezhad, A., M. Karami, K. Shu, M. V. Mancenido and H. Liu, "Domain adaptive fake news detection via reinforcement learning", in "Proceedings of the ACM Web Conference 2022", pp. 3632–3640 (2022).

Mousavi, S., S. Roper and K. A. Keeling, "Interpreting social identity in online brand communities: Considering posters and lurkers", Psychology & Marketing **34**, 4, 376–393 (2017).

Nguyen, T.-M., "Four-dimensional model: a literature review on reasons behind lurking behavior", VINE Journal of Information and Knowledge Management Systems (2020).

Nguyen, T.-M., L. V. Ngo and W. Paramita, "Turning lurkers into innovation agents: An interactionist perspective of self-determinant theory", Journal of Business Research **141**, 822–835 (2022).

Nielsen, J., "Participation inequality: The 90-9-1 rule for social features", `https://www.nngroup.com/articles/participation-inequality/`, accessed on 21 Sept 2021 (2006).

Noelle-Neumann, E., "The spiral of silence a theory of public opinion", Journal of communication **24**, 2, 43–51 (1974).

Nonnecke, B., D. Andrews and J. Preece, "Non-public and public online community participation: Needs, attitudes and behavior", Electronic Commerce Research **6**, 1, 7–20 (2006).

Nonnecke, B. and J. Preece, "Why lurkers lurk", (2001).

Nonnecke, B. and J. Preece, "Silent participants: Getting to know lurkers better", in "From usenet to CoWebs", pp. 110–132 (Springer, 2003).

OpenAI, "Gpt-4 technical report", (2023).

Osatuyi, B., "Is lurking an anxiety-masking strategy on social media sites? the effects of lurking and computer anxiety on explaining information privacy concern on social media platforms", Computers in Human Behavior **49**, 324–332 (2015).

Ouyang, L., J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback", Advances in Neural Information Processing Systems **35**, 27730–27744 (2022).

Park, J. S., J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior", arXiv preprint arXiv:2304.03442 (2023).

Parmelee, J. H., "The agenda-building function of political tweets", New media & society **16**, 3, 434–450 (2014).

Popovac, M. and C. Fullwood, "The psychology of online lurking", in "The Oxford handbook of cyberpsychology", pp. 285–305 (Oxford University Press, 2019).

Preece, J., B. Nonnecke and D. Andrews, "The top five reasons for lurking: improving community experiences for everyone", Computers in human behavior **20**, 2, 201–223 (2004).

Preoţiuc-Pietro, D., Y. Liu, D. Hopkins and L. Ungar, "Beyond binary labels: political ideology prediction of twitter users", in "Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers)", pp. 729–740 (2017).

Qian, F., C. Gong, K. Sharma and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence.", in "IJCAI", vol. 18, pp. 3834–3840 (2018).

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer", The Journal of Machine Learning Research **21**, 1, 5485–5551 (2020).

Ruchansky, N., S. Seo and Y. Liu, "Csi: A hybrid deep model for fake news detection", in "Proceedings of the 2017 ACM on Conference on Information and Knowledge Management", pp. 797–806 (2017).

Shu, K., A. Bhattacharjee, F. Alatawi, T. H. Nazer, K. Ding, M. Karami and H. Liu, "Combating disinformation in a social media age", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**, 6, e1385 (2020a).

Shu, K., L. Cui, S. Wang, D. Lee and H. Liu, "defend: Explainable fake news detection", in "Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining", pp. 395–405 (2019a).

Shu, K., D. Mahudeswaran, S. Wang, D. Lee and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media", Big data **8**, 3, 171–188 (2020b).

Shu, K., A. Mosallanezhad and H. Liu, "Cross-domain fake news detection on social media: A context-aware adversarial approach", in "Frontiers in Fake Media Generation and Detection", pp. 215–232 (Springer, 2022).

Shu, K., S. Wang and H. Liu, "Beyond news contents: The role of social context for fake news detection", in "Proceedings of the twelfth ACM international conference on web search and data mining", pp. 312–320 (2019b).

Shu, K., X. Zhou, S. Wang, R. Zafarani and H. Liu, "The role of user profiles for fake news detection", in "Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining", pp. 436–439 (2019c).

Sinno, B., B. Oviedo, K. Atwell, M. Alikhani and J. J. Li, "Political ideology and polarization: A multi-dimensional approach", in "Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies", pp. 231–243 (2022).

Stroup, W. W., *Generalized linear mixed models: modern concepts, methods and applications* (CRC press, 2012).

Sun, N., P. P.-L. Rau and L. Ma, "Understanding lurkers in online communities: A literature review", Computers in Human Behavior **38**, 110–117 (2014).

Tajfel, H., "Social identity and intergroup behaviour", Social science information **13**, 2, 65–93 (1974).

Törnberg, P., "Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning", arXiv preprint arXiv:2304.06588 (2023).

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, "Attention is all you need", Advances in neural information processing systems **30** (2017).

Wang, H., L. Liu, W. Song and J. Lu, "Feature-based sentiment analysis approach for product reviews", Journal of software **9**, 2, 274–279 (2014).

Wang, X., J. Wei, D. Schuurmans, Q. V. Le, E. H. Chi, S. Narang, A. Chowdhery and D. Zhou, "Self-consistency improves chain of thought reasoning in language models", in "The Eleventh International Conference on Learning Representations", (2022).

Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models", Advances in Neural Information Processing Systems **35**, 24824–24837 (2022).

Weir, K., "Why We Believe Alternative Facts", Monitor on Psychology **48**, 5, 34–39, URL `https://www.apa.org/monitor/2017/05/alternative-facts` (2017).

Wu, J.-J., Y.-H. Chen and Y.-S. Chung, "Trust factors influencing virtual community members: A study of transaction communities", Journal of Business Research **63**, 9-10, 1025–1032 (2010).

Wu, P. Y., J. A. Tucker, J. Nagler and S. Messing, "Large language models can be used to estimate the ideologies of politicians in a zero-shot learning setting", arXiv preprint arXiv:2303.12057 (2023).

Wu, T., E. Jiang, A. Donsbach, J. Gray, A. Molina, M. Terry and C. J. Cai, "Promptchainer: Chaining large language model prompts through visual programming", in "CHI Conference on Human Factors in Computing Systems Extended Abstracts", pp. 1–10 (2022a).

Wu, T., M. Terry and C. J. Cai, "Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts", in "Proceedings of the 2022 CHI conference on human factors in computing systems", pp. 1–22 (2022b).

Xiao, Z., W. Song, H. Xu, Z. Ren and Y. Sun, "Timme: Twitter ideology-detection via multi-task multi-relational embedding", in "Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining", pp. 2258–2268 (2020).

Yao, S., D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models", arXiv preprint arXiv:2305.10601 (2023).

Yen, C., "How to unite the power of the masses? exploring collective stickiness intention in social network sites from the perspective of knowledge sharing", Behaviour & Information Technology **35**, 2, 118–133 (2016).

Zaheer, M., G. Guruganesh, K. A. Dubey, J. Ainslie, C. Alberti, S. Ontanon, P. Pham, A. Ravula, Q. Wang, L. Yang *et al.*, "Big bird: Transformers for longer sequences", Advances in neural information processing systems **33**, 17283–17297 (2020).

Zhao, J., Y. Wang, M. V. Mancenido, E. K. Chiou and R. Maciejewski, "Evaluating the impact of uncertainty visualization on model reliance", IEEE Transactions on Visualization and Computer Graphics (2023).

Zhou, X. and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities", ACM Computing Surveys (CSUR) **53**, 5, 1–40 (2020).

Ziems, C., W. Held, O. Shaikh, J. Chen, Z. Zhang and D. Yang, "Can large language models transform computational social science?", arXiv preprint arXiv:2305.03514 (2023).

APPENDIX A

A NOTE ABOUT THE STATISTICAL METHODS USED

As black-box technologies, large language models (LLMs) require extensive and judicious experimentation and subsequent analyses to ensure that results and conclusions are in fact sustained signals and not just artifacts of random variation. This is partly why we opted to replicate the number of runs to three per user – we hypothesized that the initial political tweet added to the collection of non-political tweets *could* be a source of variation; and subsequently, whichever subset of political tweets that comprised the collection at any point in time *could* affect the predicted labels. In short, the order in which tweets were added *could* matter. Replicating the runs also gave us a measure of "expected experimental noise due to random variation." If appropriately estimated, this could be interpreted as a measure of ChatGPT's uncertainty in its prediction for a specific user. The estimated quantities for both 7-class and 3-class cases were similar (around 1.0, with almost identical standard errors). Of course, this quantity is not meaningful unless compared to other estimates of variance or covariance, such as those resulting from the autocorrelation parameter.

The `AR(1)` structure imposed allowed for the estimation of the strength of autocorrelation of predicted labels across time (or more specifically, across the number of added content) for one user. In the GLMMs used in this study, the `AR(1)` variance-covariance matrix has the following form:

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \dots & \rho^{10} & \rho^{11} \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \dots & \rho^9 & \rho^{10} \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \dots & \rho^8 & \rho^9 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^7 & \rho^8 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \dots & \rho^6 & \rho^7 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \dots & \rho^5 & \rho^6 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \dots & \rho^4 & \rho^5 \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & \dots & \rho^3 & \rho^4 \\ \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \dots & \rho^2 & \rho^3 \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \dots & \rho & \rho^2 \\ \rho^{10} & \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \dots & 1 & \rho \\ \rho^{11} & \rho^{10} & \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \dots & \rho & 1 \end{bmatrix}$$

		Number of Political Tweets Added to the Set										
		0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
All	Run 1	25.26	27.33	25.95	29.75	29.41	28.37	32.17	29.41	34.25	34.60	32.87
	Run 2	-	25.95	28.02	24.91	27.33	28.71	28.71	31.14	31.18	31.83	29.75
	Run 3	-	29.41	29.75	29.41	30.44	31.48	30.44	31.48	30.79	30.10	31.14
	Mean	-	27.56	27.91	28.02	29.06	29.52	30.44	30.68	32.29	32.17	31.25
	Std	-	1.42	1.55	2.20	1.29	1.39	1.41	0.90	1.44	1.85	1.27
		0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
Idn	Run 1	27.86	30.26	28.40	32.69	31.71	30.59	35.22	33.46	37.35	37.31	35.84
	Run 2	-	29.29	31.51	27.37	30.15	31.80	31.67	34.22	34.98	35.24	32.82
	Run 3	-	32.94	33.33	32.69	32.59	33.95	32.95	33.33	33.33	32.10	32.84
	Mean	-	30.83	31.08	30.92	31.48	32.11	33.28	33.67	35.22	34.88	33.84
	Std	-	1.54	2.03	2.50	1.00	1.38	1.46	0.39	1.65	2.14	1.42

Table A.1: The Performance (% Accuracy) of the ChatGPT's Political Ideology Score for Three Different Random Seeds When Political Tweets Are Added One by One to the User's Non-Political Content.

		Number of Political Tweets Added to the Set										
		0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
All	Run 1	37.02	38.40	39.10	41.86	44.63	40.48	44.63	42.21	48.09	49.13	45.32
	Run 2	-	39.79	40.48	39.79	40.83	42.90	41.86	46.36	49.13	46.71	44.63
	Run 3	-	44.29	43.59	42.90	45.67	44.63	42.90	43.59	42.21	42.90	46.71
	Mean	-	40.83	41.06	41.52	43.71	42.67	43.13	44.05	46.48	46.25	45.55
	Std	-	2.51	1.88	1.29	2.08	1.70	1.14	1.72	3.04	2.56	0.86
		0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
Idn	Run 1	40.84	42.52	42.80	46.00	48.13	43.65	48.86	48.03	52.45	52.98	49.43
	Run 2	-	44.92	45.52	43.72	45.03	47.50	46.18	50.95	53.99	51.72	49.23
	Run 3	-	49.61	48.83	47.69	48.88	48.13	46.44	46.15	45.69	45.75	49.27
	Mean	-	45.68	45.72	45.80	47.35	46.43	47.16	48.37	50.71	50.15	49.31
	Std	-	2.94	2.46	1.62	1.66	1.98	1.20	1.97	3.6	3.15	0.08

Table A.2: The Performance (% Accuracy) of the ChatGPT's Political Leaning for Three Different Random Seeds When Political Tweets Are Added One by One to the User's Non-Political Content.

For both cases, the estimated `AR(1)` parameter ρ is similar at around 0.35, with similar standard errors. By modeling the error with this structure, we ensured that sustained trends were captured in the modeling procedure.

APPENDIX B

CHATGPT'S ZERO-SHOT PERFORMANCE

Following is the confusion matrix of the three runs for both the political ideology scores and political leaning labels.

Scores		Predicted Ideology Score							
		N/A	1	2	3	4	5	6	7
True Ideology Score	1	102	18	6	35	16	10	8	1
	2	172	24	14	92	45	27	25	3
	3	197	24	9	100	51	33	36	4
	4	302	23	12	121	83	57	92	7
	5	205	13	4	79	50	63	68	20
	6	290	8	6	66	74	84	138	27
	7	244	8	4	41	41	51	173	33

Table B.1: ChatGPT's 7-Scale Ideology Score Inference. The N/a Column Shows the Users That Their Political Ideologies Were Not Identified. Run 1.

Scores		Predicted Ideology Score							
		N/A	1	2	3	4	5	6	7
True Ideology Score	1	100	16	5	46	6	11	12	0
	2	174	18	8	106	20	30	42	4
	3	201	14	8	117	30	32	45	7
	4	308	24	9	145	50	59	89	13
	5	211	10	6	86	35	64	80	10
	6	307	9	6	86	35	74	154	22
	7	256	10	1	63	28	49	162	30

Table B.2: ChatGPT's 7-Scale Ideology Score Inference. Run 2.

Scores		Predicted Ideology Score							
		N/A	1	2	3	4	5	6	7
True Ideology Score	1	101	15	3	48	6	7	15	1
	2	173	20	6	106	20	27	25	3
	3	194	13	8	123	29	31	50	6
	4	308	22	8	150	44	63	89	13
	5	211	10	3	84	33	66	84	11
	6	305	8	6	96	33	68	156	21
	7	252	8	2	67	25	49	162	30

Table B.3: ChatGPT's 7-Scale Ideology Score Inference. Run 3.

Labels		Predicted Political Leaning			
		N/A	Left	Center	Right
True Political Leaning	Left	739	657	229	165
	Center	471	147	322	112
	Right	302	156	156	83

Table B.4: ChatGPT's Political Leaning Inference. The N/a Column Shows the Users That Their Political Leanings Were Not Identified. Run 1.

Scores		Predicted Political Leaning			
		N/A	Left	Center	Right
True Political Leaning	Left	774	641	277	95
	Center	457	183	338	56
	Right	308	161	178	50

Table B.5: ChatGPT's Political Leaning Inference. Run 2.

Scores		Predicted Political Leaning			
		N/A	Left	Center	Right
True Political Leaning	Left	768	647	284	91
	Center	468	187	342	55
	Right	308	165	180	44

Table B.6: ChatGPT's Political Leaning Inference. Run 3.

	Acc	Run 1	Run 2	Run 3
Political Ideology Scores	All	12.69%	12.25%	12.43%
	Idn	22.15%	22.05%	22.06%
Political Leaning	All	30.01%	29.08%	29.19%
	Idn	52.39%	50.35%	50.18%
Political Leaning (Silent Users)	All	26.56%	25.90%	26.09%
	Idn	46.87%	46.80%	46.82%

Table B.7: ChatGPT's Performance (Accuracy) of Political Ideology Score and Political Leaning Performance of the Three Runs. "All" Shows the Performance of All the Users in the Experiment, While "Idn" Shows the Identified Users.

APPENDIX C

OTHER RESEARCH CONTRIBUTIONS

This dissertation focuses on investigating the role of silent users in user behavioral analysis on social media. In addition to studying silent users, I have also had the opportunity to collaborate on various other problems related to user behavior and text analysis. A summary of my research as a graduate student is provided in Figure C.1. The research discussed in my dissertation is centered around the analysis of silent users' behavior, prompt engineering, and data augmentation. In this appendix, I will present a summary of some additional contributions I have made.

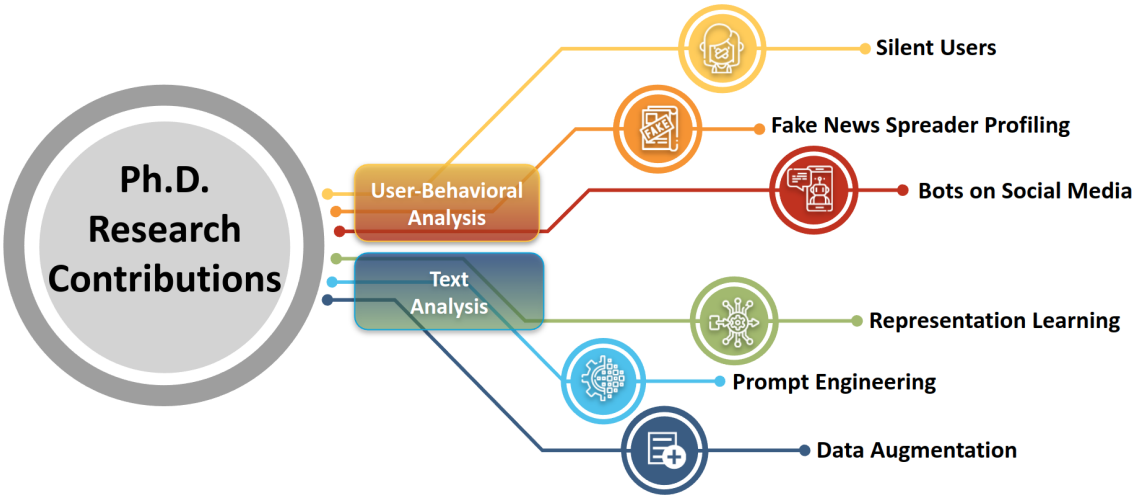


Figure C.1: An Overview of My Ph.D. Research Contribution.

Fake News Spreader Profiling

The prevalence of fake news in the last decade has resulted in various ramifications, such as influencing elections and causing uncertainty during pandemics. Existing methods to combat disinformation primarily concentrate on fake news content and the individuals responsible for its generation. However, the extent of fake news's spread largely relies on the users who disseminate it. A more profound comprehension of these users can aid in creating a framework to identify potential spreaders of fake

news.

As a result, we conducted an investigation to determine whether psychological traits observed in users who propagate fake news in behavioral studies on human subjects are also apparent in social media users who spread fake news. For this purpose, we identified five categories of features based on psychological theories, which can be quantified for social media users (Figure C.2). Our analysis of two real-world datasets revealed that (i) social media users who spread fake news differ significantly in terms of the majority of these features, and (ii) these features hold predictive power in detecting new and previously unobserved spreaders of fake news.

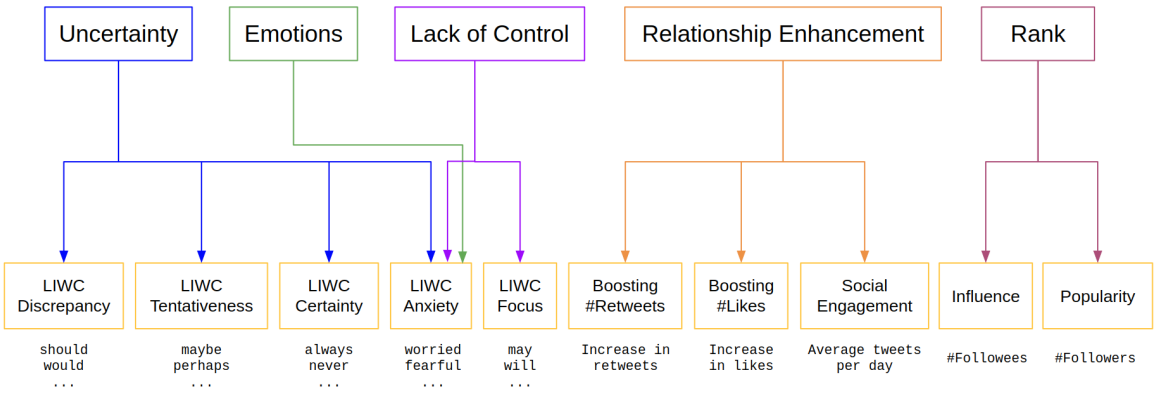


Figure C.2: Summary of the Features of the Users Who Spread Fake News Along With the Metrics Used To Measure Them.

Enhancing the Text Representations

The recent shift in NLP models towards pre-training with language modeling has been highly successful in various downstream tasks. Nevertheless, pre-trained language models often handle punctuation as a regular word or a predefined token, and sometimes, they are filtered out during pre-processing [Karami *et al.*, 2021; Mosallanezhad *et al.*, 2022]. The lack of considerable attention to punctuation in NLP models stems from the fact that punctuation has long been considered as cues that only aid text's readability, thus not providing additional semantic value to the sen-

tence's coherence [Ek *et al.*, 2020]. However, studies indicate that misplacing or removing punctuation can alter the original meaning or obscure the implicit sentiment of a text [Lou *et al.*, 2019; Altrabsheh *et al.*, 2014; Wang *et al.*, 2014], as it contains valuable information about the structural relationships within the text. For example, "What is this thing called love?" and "What? Is this thing called love?" have drastically different meanings and implications (Figure C.3).

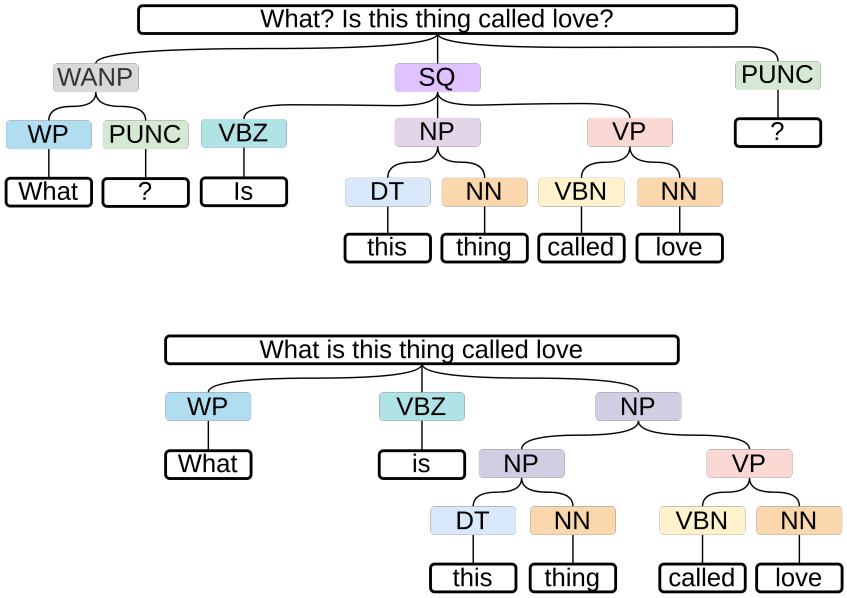


Figure C.3: The Constituency Tree of a Text With and Without Punctuation, "What Is This Thing Called Love" Versus "What? Is This Thing Called Love?"

But BERT, as a representation tool, will assign a fixed predefined token to the punctuation treating it as an ordinary word in the data; under BERT, the vector representations of these two sentences are nearly the same. Additionally, approaches that consider punctuation are often specific to certain models and cannot be seamlessly integrated into state-of-the-art representation models.

Thus, we propose that trivializing the role of punctuation in text analysis tasks leads to lower-quality representations, consequently affecting traditional classifier performance metrics. To support our hypothesis, we have developed a model-agnostic

module that represents the syntactic and contextual information derived from punctuation. Our method involves an encoder that integrates structural and textual embeddings, accurately capturing sentence-level semantics through parsing trees.