Université de Montréal

The impact of social bots on public COVID-19 perceptions during the 2020 U.S. presidential election.

Par
Anne Imouza

Département de science politique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention
du grade de maîtrise (M.Sc.) en science politique

Juillet 2022

Université de Montréal
Faculté des arts et des sciences, Département de science politique

Ce mémoire intitulé:

The impact of social bots on COVID-19 perceptions during the 2020 US presidential election.

Présenté par Anne Imouza

A été évalué par un jury composé des personnes suivantes :

Frédérick Bastien
Président-rapporteur

André Blais
Co-directeur de recherche

Reihaneh Rabbany
Co-directeur de recherche

Laurie Beaudonnet
Membre du jury

Résumé

Plusieurs études ont démontré que les contenus nuisibles et perturbateurs en ligne sont en partie produits par des acteurs communément appelés robots sociaux. Ils représentent des entités autonomes ou semi-autonomes capables de partager, aimer et poster des messages à des fins préjudiciables. Plusieurs auteurs ont mis en évidence une stratégie utilisée par ces acteurs, l'utilisation du cadrage conflictuel des enjeux. Dans ce mémoire, j'examine les caractéristiques et le potentiel rôle des robots sociaux sur la perception de la COVID-19 en période de forte polarisation au moment de l'élection présidentielle américaine de 2020. Je m'appuie sur plusieurs méthodes en science computationnelle pour analyser les caractéristiques (stratégies et comportements) des robots sociaux ainsi que leur portée politique en utilisant des données Twitter durant l'élection présidentielle de 2020. Les résultats de cette étude montrent que les robots sociaux conservateurs envoient plus de tweets de conspiration que leurs homologues libéraux. Cependant, en termes d'émotion liée à la COVID-19, les humains et les robots ont tous les deux un sentiment positif à l'égard de cet enjeu. Finalement, aucune évidence ne suggère que le contenu négatif et la proportion des robots sociaux ont un effet sur la perception de la COVID-19 par les utilisateurs.

Mots-clés : Robots sociaux, élection présidentielle, idéologie, théories du complot, modèles de sentiment, annotation manuelle, COVID-19, Twitter.

Abstract

Increasing evidence suggests that a growing amount of disruptive and harmful content is generated by rogue actors known as malicious social bots. They are autonomous entities that can share, like, or post messages for detrimental purposes. Several authors have highlighted one strategy employed by those automated actors, the use of a conflicting frame of issues, employed throughout this paper. In this work, I present a framework to depict their potential role in online discussions related to COVID-19 topics around the 2020 U.S. presidential election. I leverage different computational methods to look into their online characteristics and potential impact on the users' COVID-19 perception using Twitter data during the 2020 U.S. presidential election. The results of this study show that conservative bot users send more conspiracy tweets, but human and bot users talk positively about COVID-19. Social bots do not send more negative tweets or retweets over time than human users. Additionally, no evidence suggests that the negativity of bots' content, as well as their online proportion, will cause a change in users' COVID-19 perception.

Keywords: Social bots; election; user ideology; conspiracy theories; sentiment models; manual annotation; covid-19; Twitter.

Table of Contents

List of Tables

List of Figures

List of tables - Appendix

List of Figures – Appendix

List of Abbreviations

API: Application Programming Interface

BERT: Bidirectional Encoder Representations from Transformers

CAP: Complete Automation Probability scores

CCP: Chinese Communist Party

COVID-19: Coronavirus disease

DV: Dependent Variable

EU: European Union

ESC: Ensemble of Specialized Classifier

FEIR: Model Forecasting Error Impulse Response

GOP: Grand Old Party

IRF: Impulse Response Functions

IV: Independent Variable

JSON: JavaScript Object Notation

KE-MLM: Knowledge Enhanced Masked Language Model

LIWC: Linguistic and Word Count

MAGA: Make America Great Again

OANN: One America News Network

OSNs: Online Social Networks

PHEIC: Public Health Emergency of International Concern

RoBERTa: Robustly Optimized BERT Pre-training Approach

TGR: The Great Reset

ROC: Receiver Operating Characteristics

UTC: Coordinated Universal Time

U.S.: The United States of America

VADER: Valence Aware Dictionary for Sentiment Reasoner

VAR: Vector AutoRegressive

WHO: World Health Organization

WEF: World Economic Forum

Remerciements

I would like first to thank my co-directors, Rabbany Reihaneh and André Blais. Thank you for your availability, generosity, and opportunities that you offered me. Your advice was precious throughout my master's degree. I would also like to thank all those who provided feedback and comments from the Quebec Artificial Intelligence Institute (MILA) team, Professor Jean-François Godbout, Aarash Feizi, Gabrielle Desrosiers-Brisebois, Jacob Tian, Jiewen Liu, Kelline Pelrine, Sacha Lévy and Zachary Yang. Thank you to the participants of the 2022 Graduate conference of the CSDC, as well as the financial support from the FRQSC and la Faculté des Arts et des Sciences, which highly contributed to the success of this project.

To my parents, sister, and brother, who gave everything to enable me to come to Canada and pursue my intellectual passion. Thank you for teaching me hard work and helping me surpass my limits in believing in myself. A special thanks to my close friends, Julie and Khaoula, with whom I discussed my paper and were present when I needed to breathe fresh air.

Finally, Ryan, I would like to thank you for being my rock. I am grateful to have you by my side for all these years and thank you for helping me become the person I am today. No word can tell how much I am grateful for your help and love.

Chapter 1. Introduction

1.1 What is this about?

To understand contemporary political communication, we must now investigate the politics of algorithms and automation - Woolley and Howard (2016).

In recent years, scholars have brought awareness of the emergence of bots on social media. In the United States (U.S.), levels of automated bot accounts have risen (Ferrara et al. 2016; Woolley 2018) and reached unprecedented heights in recent presidential elections. Malicious political bots are automated social media accounts that can automatically produce content and interact with humans on digital platforms to propel the spread of hyper-partisanship or false information (Woolley et Howard 2016). The aftermath of the 2016 U.S. presidential election was shaken by the Mueller Report detailing Russian-based Internet Research Agency (IRA) activities during the campaign in online discussions (Hanson et al. 2019). This report disclosed over fifty thousand inauthentic accounts with Russian ties sharing content during this period. This phenomenon is not isolated, and many governments try to prevent foreign interference from impacting the integrity of democracy. More recently, it has been said that the amplified levels of political conflict and the presence of social bots[1] targeted toward a particular group may have affected how individuals responded to the ongoing COVID-19 pandemic. Indeed, some recent studies have found that Republicans were more

[1] The term "social bot" is employed throughout this master's thesis to refer to human actors who are behind the production of bots and their activities in social media. It is essential to mention that social bots are the production of human actors that translate their political intentions into the Twittersphere (Hajli et al. 2022, 1238). This expression makes the text lighter to read. I was not able to identify the actors behind the bots. Hence, it is crucial to keep in mind that when looking at the behaviors and actions of "social bots" in this master's thesis, I refer to the actions of the humans developing these bots.

likely to be exposed to deceptive bots (Badawy et al. 2018), and that automated accounts are involved in sharing disbeliefs related to vaccines (Ferrara 2020; Shi et al. 2020; Marx et al. 2020). Hence, this master's thesis questions a possible relationship between users' ideology, their exposition to social bots, and their feelings towards COVID-19 measures. This question highlights the importance of understanding how automated accounts interact with public opinion in polarized times. Since the politicization of the COVID-19 pandemic primarily unfolded on social media platforms, it is crucial to investigate how these rogue actors reach out to users around public health measures. Across the U.S., some states, counties, or cities implemented strict lockdown orders and mask mandates, while others refused to limit social distancing. These different restrictions rapidly became politicized (Jiang et al. 2021). They have encountered a barrage of intense reactions from their supporters and opponents, both Republicans and Democrats, on Online Social Networks (OSNs). On that matter, many have argued that social bots manifest themselves during election times and other polarized periods associated with state-wide health measures (Uyheng et Carley 2020; Himelein-Wachowiak et al. 2021).

This master's thesis aims to determine whether social bots are involved in those politicized discussions and investigate the use of a conflicting frame by automated accounts as a strategy to disturb online debate. More precisely, two fundamental questions are addressed. The first is descriptive: 1) how much or little did social bots intervene during the 2020 U.S. presidential election about the COVID-19 pandemic? The second question is causal: 2) Did social bots influence users' opinions about COVID-19 measures during the 2020 U.S. presidential election. To address these questions, we need to understand how social bots are engaged in online debates and how they frame polarized issues.

1.2 Presentation of the research

Some recent works suggest a general strategy operated by social bots: the use of a conflictual frame in online content (Parra-Novosad 2020; Entman et Usher 2018). Indeed, Parra-Novosad (2020) demonstrates that social bots have political aims, such as manipulating narratives or disturbing the public sphere. In doing so, social bots attempt to divide users by presenting a conflictual frame through the share of negative content or false information. This theoretical frame allows to explain how the 2020 U.S. presidential election and the ongoing COVID-19 crisis may have reinforced division in American society. In this master's thesis, it is considered that the polarization exerted by the COVID-19 and the election is conductive of the accumulation and the presence of social bots online, which may reinforce two antagonist groups, pro-measures vs. anti-measures. From this perspective, several works have investigated numerous behaviors that can be explained by the use of a conflicting frame of issues (Shao et al. 2017; Stella et al. 2018; Marx et al. 2020; Uyheng et Carley 2020). Since social bots employ different strategies, examining how they behave with humans through a conflictual frame is critical.

In this context, four descriptive hypotheses emerge. First, it is assumed that social bots may have a political leaning side to share their ideas, ideologies, and content in favor or disfavor toward liberal and conservative users (Ferrara et al. 2020; Badawy et al. 2018). Indeed, by being part of a cluster of users, social bots can directly communicate with users having similar homophily toward topics such as COVID-19 measures' implementation. On this matter, the first hypothesis suggests that:

- *Conservative social bots are generally more active among conservative human users (H1)*.

Secondly, other works have demonstrated that bots share misinformation (Marx et al. 2020; Ferrara 2020; Shao et al. 2018), through the conflicting frame, to potentially disturb the online sphere. The following hypotheses assume that:

- *Bot users share more conspiracy theories related to COVID-19 than their counterparts during the 2020 U.S. presidential election (H2),* and that
- *Conservative bot-users share conspiracy theories at a greater rate than liberal-bot users (H3)*.

The fourth assumption tested relates to the share of negative content. Indeed, a growing number of works emphasize the use of negativity in social bots' tweets to create emotion (Stella et al. 2018; Shi et al. 2020). Hence, the hypothesis assumes that:

- *Bot users are more negative in their tweet content than human users (H4).*

These first analytical steps will provide a global landscape of social bots' behaviors and actions in a polarized context. Lastly, a causal analysis is performed to investigate a fifth hypothesis stating that:

- *The number of social bots and the volume of harmful content they share (as a conflict frame) produce a higher proportion of users that talk negatively about COVID-19 measures in the U.S. (H5).*

This second analysis entails assessing how much/little social bots influence the evolution of users' opinions over time.

An investigation of a three-month period on Twitter is performed to test the hypotheses. This empirical analysis involves 37,960 users and 1,466,218 tweets from October 9th, 2020, to January 4th, 2021, dealing with COVID-19 and U.S. election-related topics. A collaborative team that includes political scientists from the University of Montreal[2], computational scientists[3] from the Quebec Artificial Intelligence Institute (MILA), and I worked closely to collect the data and categorize Twitter discussions related to COVID-19 and the U.S. election.

1.3 Main contributions

This master's thesis offers five main contributions, summarized below:

- I find a significant proportion of online social bots. 45% of the users studied in this study were social bots. Bot accounts are too numerous to be ignored, and more scholarly research needs to depict their behaviors, characteristics, and complexity that might affect democracy's integrity.
- I shed light and find evidence of specific narratives discussing COVID-19 conspiracy theories from conservative bot users.
- I highlight that not all social bots are malicious since some bot users were not sharing any misinformation or negative content posts during the period studied.
- I provide a rigorous examination of the actual influence of social bots on users' opinions related to COVID-19 subjects.

[2] Professor André Blais, Professor Jean-François Godbout, and the student Gabrielle Desrosiers-Brisbois.
[3] Professor Reihaneh Rabbany, and the following students: Aarash Feizi, Jacob Tian, Jiewen Liu, Kelline Pelrine, Sacha Lévy, Zachary Yang.

- I present how challenging sentiment analysis models are and recommend proper evaluation when applying these models in social science.

1.4 Project plan:

This master thesis starts with a literature review (Chapter 2) on social bots' political strategies and behaviors. This chapter displays their characteristics and how they have been studied. It is followed by a presentation of the theoretical frame employed throughout this master's thesis: the conflictual frame of issues. Additionally, this chapter unveils related works divided into two branches. The first focuses on their potential political impact on users' opinions during election times, while the second focuses on their involvement during national sanitary crises. This chapter reveals different mechanisms by which social bots contribute to users' perceptions. Lastly, the research hypotheses are formulated.

The third chapter, called *Methodology*, displays and justifies the data used and the topics of the 2020 U.S. presidential election and COVID-19. It highlights the collection process and the construction of the bot score. Lastly, it is followed by the presentation and the justification of different classification models (e.g., ideology and bot-like of users, sentiment, and conspiracy tweets) to test the hypotheses. This chapter ends with the production of several descriptive statistics.

The next chapter, *Results,* exposes the empirical findings. It is divided into four parts. First of all, descriptive results are showed to explore bots' activities. Then, a content analysis and an examination of conspiracy theories shared by bots are performed. Furthermore, an exploration of the sentiment of the content at the tweet and word levels is achieved. This chapter ends with lag regression models to evaluate the influence of the negative content shared by social bots on users' perceptions of COVID-19 measures.

The chapter *Discussions* presents several explanations of the results, the limits of the study, and an ethical statement. Lastly, the chapter *Conclusion* summarizes the main results and the theoretical and practical implications. The contribution to the advancement of knowledge and suggestions for future research are addressed at the end.

Chapter 2. The literature on social bots

It is critical to understand how social bots have been studied previously. A large body of literature is dedicated to descriptive analyses examining their specificities, how they frame issues and how they correlate with users' opinions. However, few scholars have focused on explaining how this technology may impact users' views. The following section proposes a definition of social bots, their implications, and why it is essential to analyze them as our primary explicative variable (2.1). A theoretical perspective is presented in a second subsection, based on a framing approach, to explore social bots' impact on public opinion (2.2). Lastly, a literature review of social bots' actions and how they correlate with users' views during elections (2.3) and national sanitary crises (2.4) is offered.

2.1 Defining social bots

Social bots refer to "computer algorithms designed to mimic human behavior and interact with humans in an automated fashion" (Yuan et al. 2019, 2). The rising presence of social bots on digital platforms has increased interest in numerous fields to examine their characteristics. This subsection presents a detailed description of social bots' behaviors and highlights the need to study their actions in political science.

Not all bots are built to be malicious (Khaund et al. 2022). In the literature, distinguishing a malicious from a benign bot is fundamental. A social bot could be harmless or neutral by sharing neutral or specific information, such as news bots, promotional bots, chatbots, or suicide helpline bots (Khaund et al. 2022, 532). Their common point is their positive or neutral intervention on social media since they do not put any threat to the community. Indeed, some bots have been created to simply retweet or share posts from a political institution or a firm, such as *@big_ben_clock* (Yang et al. 2019). Another example is the creation of the *Botovist* bot that encourages users to take actions to favor participation (Savage et al. 2016).

Nonetheless, some scholars argue that malicious automated actors are being developed by humans to manipulate, emulate and alter the behavior of users (Himelein-Wachowiak et al. 2021). Specifically, this type of bot is generally conceived during controversial political events such as international crises, elections, and political campaigns, fostering a polarized public opinion (Khaund et al. 2022, 532; Howard et Kollanyi 2016). On this matter, some scholars qualify them as influential (Subrahmanian et al. 2016) since they can shape users' behaviors, or as propaganda bots since they disseminate political information by covering dissenting beliefs (Williamson III et Scrofani 2019). Governments, media outlets, and political parties also use automated accounts to communicate political information. As a result, the presentation of social bots' behaviors is vital in understanding their strategies to intervene on OSNs.

One of the traits developed by humans behind bots' actions is highlighted by previous studies which is their political purpose. Indeed, some works argue that social bots are developed for political aims. Hegelich et Janetzko (2016) demonstrate that the actors developing these automated accounts have a **political agenda**. After collecting 1,740 automated accounts via *Twifarm*[4], they indicate that social bots hide their identity and promote topics by pushing political hashtags and re-echoing retweets. Other scholars emphasize another behavior: the **amplification of misinformation** or low credibility sources online, mainly through various shared hashtags

[4] Twifarm is a program that manages significant amount of data (social bots) on Twitter by following URLs.

(Khaund et al. 2018), replies, and mentions (Shao et al. 2018). Similarly, Himelein-Wachowiak et al. (2021) find evidence of COVID-19 misinformation shared by bots. Adding to that, other works have focused on their **sophistication and mimic behavior** (mimicry) toward humans on OSNs (Al-Khateeb et Agarwal 2016; Luceri et al. 2019). These behaviors are treacherous since mimicking humans on social networks improves their ability to influence online discussions without being noticed. From the literature review, other works present the **misdirection** behavior, which is the use of context-related hashtags without mentioning the topic discussed in a specific online conversation (e.g., human users talk about mountains in Canada, and a hashtag #ChinaVirus pops up in the discussion). Abokhodair et al. (2015) show evidence of this behavior by qualitatively coding almost 3,000 tweets from Syrian social bots on Twitter. They find that hashtags are not aligned with the networks' discussions. In addition, other behaviors, such as **hashtags latching**,[5] **thread-jacking**,[6] and **reverberation**, defined as the amplification of selected tweets and retweets, are studied to understand their scope (Khaund et al. 2022). This list of actors' behaviors behind bots is not exhaustive. They can be mutual depending on the political purpose and information they want to disseminate.

Generally, textual and social network analyses capture most of these behaviors (Khaund et al. 2018; 2022; Al-Khateeb et Agarwal 2016). Even though previous works have analyzed such behaviors, other scholars attempted to consider numerous technical features to indicate whether an account is a human or a bot. Indeed, Himelein-Wachowiak et al. (2021) have listed several technical features of bots from related work to detect them. In the first place, they categorize social

[5] Definition: "Social bots associate trending hashtags to their narrative to get a bigger crowd exposure" (Khaund et al. 2022, 532)
[6] Definition: "Social bots alter discussion in a comments thread by interjecting unrelated topics" (Khaund et al. 2022, 532)

bots with **network properties.** Some authors have demonstrated that bots could be core bots, meaning that they are strongly connected and generate their content, or peripherical bots, which are more isolated in disseminating information (Khaund et al. 2022, 532). Moreover, accounts are classified as social bots depending on their **account activity and temporal patterns**. Indeed, they usually have fewer original tweets but tend to retweet others' tweets more frequently than humans. Besides, the interval between tweets is relatively short, contrary to human users. Lastly, **the profile and tweets' content** can predict the probability that an account is a bot, for which the age of the account is usually lower and the username longer.

These behaviors demonstrate that the people developing rogue actors create specific entities in OSNs that behave and act in particular ways and may be sophisticated and complex to capture since they are the production of human actors' intentions (Hajli et al. 2022, 1238). Still, more importantly, their political aspect may potentially influence deeply social media interaction. It is essential to first apprehend the numerous behaviors previously studied by scholars to understand how researchers can capture actors' activities and behaviors behind the bots. Several measures have been listed and can be combined between them to study the different behaviors exposed. This list of behaviors and measures presented is not exhaustive. **Table 1** summarizes the behaviors and features to categorize bots' characteristics.

Table 1: Bots' behaviors and measures.

Behaviors	Measures
Political agenda	Networks properties
Misinformation amplification	Account activities
Sophisticated and mimicry	Temporal patterns
Misdirection	Content of the profiles
Hashtag-latching	Content of the tweets
Thread-jacking	
Reverberation	

The political aspect of social bots is the core of this research. Many scholars have studied social bots due to their potentially harmful consequences (Woolley et Howard 2016; Broniatowski et al. 2018; Uyheng et Carley 2020). Besides bots' behaviors, certain political entities, such as governments and elites, have the financial capabilities to obtain these tools to frame online issues (Woolley et Howard 2016). Indeed, Broniatowski et al. (2018) demonstrate that Russian trolls can stimulate online discord via politically divisive messages. Since social bots have been said to manipulate the public (Woolley et Howard 2016), it appears imperative to identify how this technology works, impacts users, and find new ways to prevent their actions.

This master's thesis attempts to explain the extent of social bots' characteristics in descriptive and causal fashions through one mechanism, the conflictual frame of issues. Even though social bots' behaviors have been well analyzed, few articles explore the framing of issues as a specific strategy from social bots. Consequently, it is key to present previous works exploring theoretical approaches that examine social bots' frames. The following subsection presents several theoretical frameworks and the conceptual approach employed in this study.

2.2 Theoretical approaches to study social bots' frames

Scholars have long studied the framing of issues around social bots. One prominent definition comes from Entman (2003), who defines a frame as a selection of "some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, a treatment recommendation" (Entman 2003, 55). Indeed, in Entman's (2003) cascading activation approach, media and public opinion have crucial roles in framing issues from the political elites. In other words, framing constitutes a path to communicate specific translations of the realities. As the last subsection explains, social bots employ several behaviors to skew and potentially disturb the overall digital discussion. Indeed, the spread of fake news or the expression of negative stories can be part of a larger strategy employed by automated accounts to present an issue in a specific and intentional way: so-called *framing*. As a result, an overview of how social bots' framing is theorized is essential.

Entman et Usher (2018) present a cascading network model of frame activation by considering new features, called "pump-valves," that play significant roles in the issue framing. These features are online platforms, data analytics, algorithms, ideological media, and rogue actors. The first one, online platforms, is also called OSNs and is defined as a "group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (Kaplan et Haenlein 2010, 61). One example is Twitter. Digital analytics embody the second feature, representing a "cluster of technologies that allow organizations to monitor online sentiment, test and refine communications, and quantify opinion and engagement" (Karpf 2016, 11). An example is the A/B testing, which is an experiment where two similar groups interact with a similar message, but one of them has a variation (Siroker et

Koomen 2013). Ultimately, this data analytics would shed light on whether one marketing change would modify user or customer behavior (Karpf 2017). In other words, it represents the data related to the users' online behavior collected for economic/political purposes, such as a partisan website or a political campaign variation (Siroker et Koomen 2013). Then, the paper underlines a third feature, algorithms representing "procedures for turning input into output based on a series of calculations and ordered steps" (Entman et Usher 2018, 301). An example is the recommendation algorithm or targeted advertising (Patino 2019, 75), where the future content of a user will depend on every previous user's choice. Thus, Twitter's algorithm can shape users perceived content on their thread. Besides, *ideological media* represents media outlets with clear ideological identification, such as Breitbart.com, as an extreme right media outlet. Finally, the work presents a last feature, rogue actors (e.g., social bots).

The difference from Entman's (2003) previous cascading network activation model is that mainstream institutional media are not the only channels to frame issues from/for elites and the public. Those new features may have a more significant impact on users. Indeed, platforms have significantly shifted how the information is transferred among elites, the public, and the media, which are faster and more continuous information. Regarding algorithms, this feature is powerful since they have the faculty to shape users' experience and what they see since it frames users in a specific direction. As such, users do not encounter information that goes against their opinion. In addition, digital analytics enable the building of tools for elites to understand who is receiving their messages and how to customize them. Turning to ideological media, they can frame issues in a certain ideological way with a low degree of fact-checking. Finally, and most importantly for this project, the work considers *rogue actors* as entities that intensify conspiracy theories, spread misinformation, and influence elections through issue framing. Low-quality information is transferred and explicitly framed to groups that only see the information they agree on, even if it is inaccurate. It is essential to mention that these five actors depend on each other, interact together, and impact the circulation of information.

From this new model, social bots deliberately pollute the communication path among citizens and media, curtailing citizens' access to high-quality information. This model helps understand how rogue actors alter the quality of the information in online spheres and validate the potential negative impact of this technology on users' perceptions. However, this model does not present a precise mechanism of how social bots use specific frames (e.g., conflict, morality, responsibility frames), even if it highlights a new channel of information in which these automated actors are involved.

Hence, some authors have investigated other paths to grasp how social bots frame topics. One of them is emotional contagion. It is described as a "phenomenon in which certain individuals' emotions propagate to others and trigger similar ones" (Shi et al. 2020, 3). Indeed, Yu (2020) explored the emotional effect on online users' sentiment reactions. To this extent, he uses the Instagram account of the first human-robot *Sophia* (@realsophiarobot) and collected likes and comments for each of her posts and pictures. Likes and comments were proxies for users' interest and engagement. The emotional expression was allowed by a facial recognition software called *Microsoft Azure*. The primary result is that the emotions of fear and disgust expressed online are the prominent factors affecting users' interest and engagement. Indeed, fear is significantly associated with a higher number of comments. This interesting research is a departure point in considering human-robot interaction online.

However, this theoretical approach is problematic for the research presented. It would be challenging to capture bot users' facial emotions since social bots collected post few images of themselves. Thus, this theoretical approach may not be appropriate for methodological reasons to investigate specific frames employed by social bots as part of this research.

Lastly, Parra-Novosad (2020) sought to examine how bots framed online discussions around the US-Mexico border wall in 2019. She relies on Entman et Usher (2018) 's revised framing model and five prominent types of frames developed by other scholars (Semetko et Valkenburg 2000; Neuman et al. 1992) to capture how bots may frame the online conversation. She first exposes the morality and responsibility frames. While the first frame places an event or issue in a religious context, the second gives the responsibility of an event to a specific group or person (Semetko et Valkenburg 2000). Then, she presents the human interest and economic consequences frames applied to social bots. The human-interest frame uses personal experience to communicate specific emotions, such as compassion related to an issue, whereas the last frame expresses the consequences of an issue to a particular group (Semetko et Valkenburg 2000; Neuman et al. 1992). Lastly, she refers to the conflict frame (Semetko et Valkenburg 2000; Neuman et al. 1992). It mainly highlights the conflict between antagonists in news media, including stories presenting the conflicting groups as "us versus them" viewpoint. As illustrated with the human-interest frame, this frame evokes emotions towards groups.

Ultimately, she scrutinizes how social bots frame the debate around the US-Mexico border wall. She employs a case study, Trump's border wall campaign, and a content analysis of tweets. The evaluation of tweets includes coding their political leaning, valence, and frame type. The author uses the well-known *Botometer* detection, which is a bot detection model. Even though she obtains a sparse number of bots in her data set (e.g., 0.5%), she finds that bots can frame content through their posts and that the frame is consistent with the ones employed by human users. The two main frames were conflict and morality. Lastly, bots' tweets tend to be more negative (78% of them) than human tweets (54%) and promote more political right learning frames than left-leaning frames.

Nonetheless, the article has some limitations. Firstly, the period studied is very short (e.g., 48 hours). Indeed, the lack of temporal analysis makes it challenging to assess how bot users frame the topic differently over time. The same limitation can be drawn for human users. Besides, the low number of observations (tweets) can be problematic when generalizing the results. The considerable number of tweets inherent in the research presented will allow to compare the conclusions with this work. In contrast with this work, several machine learning methods to detect tweets' sentiment and users' ideology will be employed in this master's thesis.

Although Parra-Novosad's (2020) paper has some limits, the conflictual frame of issue is appropriate and will be employed in this study since I expect that malicious bots will share negative content as well as misinformation to divide online groups. As mentioned above, two research questions are addressed: 1) how much or little did social bots intervene during the 2020 U.S. presidential election about the COVID-19 pandemic? 2) Did social bots influence users' opinions about COVID-19 measures during the 2020 U.S. presidential election? The following subsections present in the first place an overview of the social bots' engagements and behaviors during elections and in the second place during national sanitary crises. This literature review will lead to the formulation of specific hypotheses linked to the conflictual frame of issue by social bots.

2.3 Engagements of social bots with public opinion during elections

To better apprehend social bots' frame intervention, this subsection reviews scientific papers on automated accounts' behaviors and their potential impact on users and public opinion. While some scholars have investigated the party affiliation of bots as a possible factor, other scholars have focused on the proliferation of disinformation and conspiracy theories during election times. The first factor is rooted in Entman et Usher's (2018) theoretical approach, where two new actors interact together: rogue actors (social bots) and ideological media (hyper-partisan media outlets). It is also rooted in Parra-Novosad's (2020) approach since bots are said to promote a more political right-leaning frame. Hence, if bots in this research share more right-leaning (left-leaning) articles, it would indicate that bots are conservatives (liberals). Lastly, as expressed in Entman et Usher's (2018) work, social bots may intensify conspiracy theories and spread misinformation since they follow a political agenda (Hegelich et Janetzko 2016).

2.3.1 Social bots and ideology

A body of literature has focused on determining social bots' political actions during elections. Ferrara et al. (2020) characterize social media manipulation through automated accounts in the context of the 2020 U.S. presidential election. They use the Botometer *v4* algorithm to detect social bots. They measure the political leaning of users by looking at the political media outlets endorsed by users. The study examines four million tweets users post from June 20th to September 9th, 2020. Results show that automated accounts eminently use hashtags related to the Trump campaign and conspiracy theories, and retweet more. This result echoes to two strategies employed in the conflicting frame approach, the share of false information and reverberation. Additionally, the study demonstrates a link between bot users' ideology and hashtags. Indeed, if the hashtags are more liberal (conservative), social bots will share with more liberal (conservative) media outlets.

Other authors, such as Badawy et al. (2018), have drawn similar conclusions about social bots and online discussions during the 2016 U.S. presidential election. They use a similar methodology as Ferrara et al. (2020) to classify users' ideology based on the media outlets shared and add a semi-supervised network-based algorithm to identify users' ideology for those who did not share any media outlets. The period studied is from September 16th to October 21st, 2016, during which they collect about 5.7 million users. Additionally, they work with a list of 2,752 bots (Russian trolls) released by the U.S. Congress. Furthermore, they use the Botometer detection on 2,126 accounts. The results show that conservative users were more engaged with Russian Trolls than liberals. They demonstrate that conservative bots were more prominent than liberal bots and produced more tweets and retweets. The sophistication and the mimic behavior of trolls shown in this paper align with the results presented by Al-Khateeb et Agarwal (2016) and Luceri et al. (2019), for which bot users are sophisticated and disguised. However, the density of the bot score turns out to be similar for liberal and conservative users.

Another study from Bessi et Ferrara (2016) explores the relation between automated accounts and political discussion surrounding the 2016 U.S. presidential election. They analyze over 20.7 million tweets posted by nearly 2.8 million distinct users during the 2016 U.S. presidential election from September 16th and October 21st, 2016. Again, the Botometer detection algorithm is employed. In contrast, sentiment analyses are employed (*SentiStrength)* to understand how bot and human users are discussing the election. Finally, they use Trump and Clinton supporting hashtags to infer the partisanship of individuals. One noticeable result is that nearly 15% of the total population emerged as bots and were responsible for almost 19% of the tweets.

Similar conclusions surfaced from these studies when looking at the overall actions of social bots during elections. Conservative social bots are more frequent and interact more extensively with conservative users. Notwithstanding, it is unclear how they interact with conservative human users when discussing a topic. Indeed, these studies do not mention whether users are more engaged with election-specific issues, such as abortion or the COVID-19 handling regarding their ideology. My study will contribute to the literature by delving into the actions of bot users during an election for specific issues related to COVID-19. Moreover, these studies describe how conservative (liberal) social bots retweet and share more conservative (liberal) posts. Still, there is no analysis of how social bots frame a specific issue during the election (e.g., COVID-19 measures such as masks, vaccines, or lockdown). The analytical frame from Parra-Novosad (2020) found evidence that social bots promoted more political right-learning frames than left-leaning frames. From these previous results, the expected pattern is that conservative bot users should mimic conservative human users with a conflicting frame toward Joe Biden. Hence, the first hypothesis is that conservative social bots will be more active among conservative users (H1).

2.3.2 Social bots and disinformation

While a body of literature has shown that social bots can be characterized according to their political leaning, other scholars have investigated another component: the spread of disinformation and fake news by social bots. Disinformation is defined as "the intentional spread of inaccurate information" (Fetzer 2004), whereas fake news is seen as "distorted signals uncorrelated with the truth" (Allcott et Gentzkow 2017). Indeed, social media have been the place of a fast and quick spread of intentionally misleading information for the past years. A published S*cience* paper by Grinberg et al. (2019) investigates how American online users interacted with fake news during the 2016 U.S. presidential election. They find that right-wing users are more inclined to be exposed to fake news, and the sources they share stem from the right and the extreme right-leaning entities. Many authors have endeavored to investigate the link between misleading information and the proportion of bot accounts in reaction to this finding. The proliferation of disinformation, fake news, and conspiracy theories can be seen as the intention to divide and express a conflict between at least two groups. This echoes bot users' perceived strategic behavior, using a conflict frame (Parra-Novosad 2020) to manipulate users' opinions and reinforce false beliefs.

The article of Ferrara et al. (2020), presented in a previous subsection, also explains how bots can be linked to conspiracy theories and hyper-partisan media outlets during a political campaign. They distinguish tweets from users posting hashtags related to QAnon, -gate, COVID-19 conspiracy, and non-conspiracy. The results show that the first three groups of tweets have a higher median bot score than the latter group. The analysis of hyper-partisan media outlets shows that users who share URLs from media like Infowars or One America News Network (OANN) have the highest bot scores but exhibit a low volume of tweets. This study highlights the political agenda of social bots and their amplification of misinformation on the COVID-19 topics. Finally, users using 'QAnon' keywords are highly associated with a high Botometer score. Additionally, the COVID-19 conspiracy group is relatively large and focuses on false claims using three keywords: #plandemic, #scamdemic, and #fakevirus. This method can be problematic as online users discussed many other essential conspiracy theories during the period studied (e.g., June 20^{th} to

September 9th, 2020), such as the inefficiency of masks: #Burnyoumaskchallenge, and vaccines: #StopWearingMask. Contrary to this study, the current research contributes to the literature by investigating whether the conflicting frame employed by social bots, through the spread of conspiracy topics about Covid-19, correlates with users' opinions.

Other authors, such as Shao et al. (2017), show that social bots play a prominent role in spreading fake news. This paper aims to understand if social bots spread fake news and manipulate human users. They analyze 14 million tweets that spread 400,000 true and false claims on Twitter during and after the 2016 U.S. presidential election run-up. They use the *Hoaxy* platform to trace back fake news and fact-check news on Twitter from May 2016 to March 2017. The Botometer detection algorithm is employed to detect rogue actors. The main results indicate that a handful of accounts share many misleading articles, generally highly active bots. They also discover that one strategy social bots use is to mention and reply to influential persons such as Donald Trump to expose them and their followers to misleading information. This behavior is also known as hashtag latching (Khaund et al. 2022). The most concerning result is that human users are as likely to retweet bots as other human users, indicating their inability to distinguish fake news spread by automated accounts (Al-Khateeb et Agarwal 2016; Luceri et al. 2019).

Contrary to Ferrara et al. (2020), Shao et al. (2017) do not examine the COVID-19 component during an election. In the light of the results presented in this subsection and the findings, the aim is to determine whether bot users share more conspiracy theories related to COVID-19 than their counterparts during the 2020 U.S. presidential election (H2). Several hashtags related to COVID-19 and conspiracy theories will be added to the analysis to contribute to the results of Ferrara et al. (2020). Supposing this hypothesis turns out to be validated, it would confirm that the strategy of social bots to use a conflict frame, through the share of conspiracy theories, may disrupt the online sphere and correlates with Ferrara et al. (2020) study.

2.4 Engagement of social bots and public opinion during pandemics

As presented in the previous section, scholars have studied social bots' political actions during election times. Notwithstanding, a growing body of literature examines social bots' potential impact and actions during national sanitary crises (Yuan et al. 2019; Shi et al. 2020; Broniatowski et al. 2018; Parra-Novosad 2020; Marx et al. 2020; Ferrara 2020). This section presents scientific articles that deepen the research on automated accounts by looking at national sanitary crisis events. Two main avenues are drawn when looking at the impact of social bots' behaviors and strategies during crises. Indeed, while some scholars investigate how social bots impact human users by looking at the sentiment of the tweets, other academists focus on the share of misinformation. This last part of the literature review will enable to grasp whether social bots employ a conflict frame when discussing COVID-19 topics such as mask-wearing, lockdowns, or vaccines. The expectation is to see more negative content, as suggested by Parra-Novosad (2020), and more conspiracy theories shared by social bots.

2.4.1 Social bots and tweets' sentiments and stances:

A body of literature has focused on ascertaining social bots' political actions during sanitary crises by looking at tweets' sentiments and stances. Yuan et al. (2019) investigated communication between pro- and anti-vaccine tweets from human and bot users related to the MMR vaccine (measles, mumps, rubella) in 2015. The methodology employed is a classification to detect the stance of each user (e.g., sentiment analysis) and the DeBot detection algorithm to identify users with a high bot score (higher probability of having automated activities). They also create a retweet network and a community detection to ascertain who is retweeting whom and whether retweets come from users with a similar stance. They discover that 1.45% of users within the data set were identified as likely bots and engendered 4.59% of all tweets. The main results indicate that retweets for pro- and anti-vaccine threads are high within similar opinion groups (e.g., echo chamber). Additionally, bots are prone to be hyper-social by initiating retweets in the same group opinion. This strategy is known as reverberation, amplifying selected tweets and retweets. One limitation of this paper is the collection time span. Indeed, tweets were not collected during highly polarized periods such as a presidential election, which may have potentially underestimated the role of social bots in online debates. Indeed, following what Woolley (2018) demonstrated, the deployment of social bots is more extensive when polarized events happen. The present research will analyze the proportion of social bots and the conflictual frame employed during a national sanitary crisis: the COVID-19 pandemic.

Additionally, Broniatowski et al. (2018) endeavor to grasp the role of social bots and trolls regarding vaccination content through an observational analysis. They collect 793,690 tweets from July 14th, 2014, through September 26th, 2017. The Botometer is once again used for its bot-detection property. Manual annotation is performed to classify tweets' stances as anti-vaccine, pro-vaccine, or neutral. The results convey that bot accounts are more inclined to share vaccination-related content than human users, while attention to pro- and anti-vaccination sentiment is relatively equal.

Rather than looking at the stances, other authors have looked at the sentiments of the tweets during the COVID-19 crisis. Shi et al. (2020) investigate the differences between inauthentic actors and human users by analyzing sentiment through Linguistic and Word Count (LIWC) and structuring topic modeling. They collect tweets from three periods of the COVID-19 crisis. First, on January 22nd, the World Health Organization (WHO) announced the high level of virus propagation. Then, on January 31st, the WHO labeled coronavirus a public health emergency of international concern (PHEIC). Finally, on March 11th when the COVID-19 was officialized as a pandemic. The primary result indicates that while the proportion of social bots contributing to COVID-19 discussion is 9.27%, they share a similar sentiment as humans. However, social bots are more likely to amplify emotions. Indeed, they manage to instill anger and express more sadness toward health risks. Social bots' ability to provoke feelings through a conflicting frame can be rooted in their sophistication and capacity to mimic human online behavior (Al-Khateeb et Agarwal 2016; Luceri et al. 2019). While this study contributes to the growing literature on the role of bot users during specific times (e.g., sanitary crises), there is no certainty that the high volume of social bots is solely due to the announcements of the WHO.

Other studies have reached similar conclusions using sentiment analyses but not related to sanitary crises. Indeed, Stella et al. (2018) investigate social bots' activities and how they contribute to exacerbating social conflict online during the 2017 Catalan referendum. They collect over 3.6 million tweets, for which bots posted 24%. They use sentiment analysis to characterize bot behavior in online discussions. They find evidence supporting the hypothesis that bots accentuate "the exposure to negative, hatred-inspiring, inflammatory content, thus exacerbating social conflict online" (Stella et al. 2018). In other words, the results demonstrate that social bots generated and exacerbated harmful content aimed at the Independentists group (Catalan Independence supporters).

For most of the studies presented above, even though the tweets' sentiment/stance is equivalent for both humans and bots (Yuan et al. 2019; Broniatowski et al. 2018; Shi et al. 2020), the conflict frame seems to be used prominently by social bots. From these results, another hypothesis is formulated: bot users are more negative in their tweet content than human users (H4).

2.4.2 Social bots and disinformation during Covid-19

On top of the studies related to fake news, other academics have tackled the issue of COVID-19 'Infodemic' by social bots, which refers to the spread of disinformation related to the COVID-19 pandemic (Marx et al. 2020). Indeed, Marx et al. (2020) investigate how social bots spread misinformation through a manual content analysis. They could detect 78 bots out of 542,345 users. They propose a novel method to detect social bots by considering active users, tweet uniqueness, tweet frequency, and friend-follower ratio. Their main result indicates that social bots' tweets disseminate misinformation posts but share at the same time news from accurate sources. This research shows that some tweets and retweets published by social bots conveyed misleading information. However, there is no clear evidence of the impact of misleading information on users' opinions.

Additionally, Ferrara (2020) characterizes the activity of social bots online during the COVID-19 pandemic. He uses the general Botometer algorithm in conjunction with content analysis to understand how bots are engaged with political issues and conspiracy theories. He collects data related to COVID-19 from January 21 to March 12, 2020. His time-series analysis indicates that social bots can be a tool for common goods by bringing political issues that China censored to light. At the same time, he characterizes social bots as political means to distort online narratives by disseminating political conspiracy theories.

Thus, social bots frequently participate in online discussions surrounding national sanitary crises. Their role is to amplify existing discourses rather than create new ones. Besides, social bots tend to share misleading information by boosting them (Shao et al. 2018). As Howard et Kollanyi (1, 2016) suggest, "political bots tend to be developed and deployed in sensitive political moments when public opinion is polarized."

The results from these articles related to COVID-19's 'Infodemic' are consistent with the second hypothesis, which states that bot users tend to share more conspiracy theories associated with Covid-19 than their counterparts during the 2020 U.S presidential election (H2). Thus, the expectation is that conservative social bots share at a more significant rate conspiracy theories related to COVID-19 than liberal-bot users (H3) since they will employ a conflictual frame by highlighting a division between the pro-Biden and the pro-Trump for the COVID-19 crisis handling.

Most studies have focused on documenting social bots' presence and how they frame content through emotion, misinformation, and political leaning. Nonetheless, they do not demonstrate any causal inference. Few articles investigate the causal impact of automation on the public and users' opinions during the COVID-19 pandemic (Uyheng et Carley 2020; Duan et al.

2022). The study of Uyheng et Carley (2020) is one of them. They explore how hate speeches may be linked to bot-driven activities by performing network and cluster analyses of tweets in the U.S and the Philippines for 75 days. Contrary to previous studies, they use the BotHunter detection algorithm. To analyze hate speech and how humans behave, they compute a predicted hate score at the community level. The multi-level regression analyses suggest that social bots predict a significantly high level of hate speeches in dense community groups in the U.S. and the Philippines. This result could be partially explained by social media's propensity to amplify echo chambers and trap individuals in similar communities into identical information channels and sources through a conflictual framing of issues.

In the same vein, Duan et al. (2022) present time-series analyses to investigate how algorithmic agents (e.g., social bots) predict partisan media outlets' attention allocation to COVID-19-related topics. They collect 1,657,551 COVID-19-related tweets from March 1st to May 31st, 2020, and 50,356 COVID-19 news stories. They apply the Botometer v4 to get users' bot scores and topic modelings to highlight the topics from tweets and news stories. Then, time-series analyses are performed using Impulse Response functions (IRFs) to investigate whether changes in human activities impact bots' activities and media coverage of COVID-19 topics. IRFs are lag regression models that capture the immediate and long-term outcomes of "shocks" (change in variables) in specific variables in a noisy system (social media). Recently, several works have applied this methodology to investigate users' online attention to public policy or international crises (Barbera et al. 2019; Polyzos 2022). The results indicate a relatively small proportion of 8.98% active bot accounts, but these users tend to amplify and retweet more human tweets. In addition, they find a positive relationship between media outlets and bot users for which liberal media positively respond to a shock in bots' activity. In other words, topics amplified by bot users impact the attention of the liberal media outlets related to COVID-19 in the last ten days. Nonetheless, this relationship is weaker with conservative media outlets.

Few works investigate a clear causal impact of social bots on public opinion. This research project strives to contribute to this field of research by proposing lag regression models, as presented in Duan et al. (2022), to understand whether the content shared, and the proportion of social bots positively predict change in public opinion related to COVID-19 around the 2020 U.S. election. To this extent, the last hypothesis poses that the number of social bots and the volume of harmful content they share (as a conflict frame) will produce a higher proportion of users that talk negatively about COVID-19 measures in the U.S. (H5). **Table 2** resumes the related work presented and how my study contributes to the literature. Contrary to Uyheng et Caley (2020) and Duan et al. (2022), the data collection and the causal analyses involved in my research will look at not only at COVID-19 issues, but also election-related topics.

Table 2: Related work: Social bots and public opinion.

Papers	Elections	Sanitary crises	Disinformation or negative content	Causal inference
Badawy et al. (2018)	✓			
Bessi et Ferrara (2016)	✓			
Broniatowski et al. (2018)		✓		
Duan et al. (2022)		✓	✓	✓
Ferrara (2020)		✓	✓	
Ferrara et al. (2020)	✓	✓	✓	
Marx et al. (2020)		✓		
Shao et al. (2017)	✓		✓	
Shi et al. (2020)		✓	✓	
Stella et al. (2018)	✓		✓	
Uyheng et Carley (2020)			✓	✓
Yuan et al. (2019)		✓		
Study presented	✓	✓	✓	✓

Chapter 3. Methodology

3.1 Data

To empirically investigate bot and human users' behaviors and COVID-19 perceptions, an empirical analysis of 37,960 users and 1,466,218 tweets from October 9th, 2020, to January 4th, 2021, is conducted. This period includes the run-up and the aftermath of the 2020 US presidential election and the COVID-19 crisis. The collection was made possible by a multidisciplinary collaboration between political scientists at the University of Montréal and computational scientists at the Quebec Artificial Intelligence Institute (MILA), which studies online polarization. Even though the methodology relies on users' characteristics, it is critical to keep in mind that the unit of analysis in the results section is users' tweets.

3.1.1 Collection process

Four aspects of the collection process are considered to collect the data.

1) **The period of study.** The data came from an extensive data set of 387,090,097 real-time tweets collected from 23,758,112 users. At first, the data was collected to understand online polarization during the 2020 U.S. presidential election. Since some studies have demonstrated that bots' activities are prominent in polarized times (Uyheng and Carley 2020), this extensive data set fits well with the research. Indeed, the master's thesis aims to examine any behavior change in users when political bots intervene in online activities. Additionally, several authors have shown that bot

users have participated in online discussions during the 2020 U.S. presidential election (Chang et al. 2021, 15) and the 2016 U.S. presidential election (Bessi et Ferrara 2016), which makes the extensive data set more relevant for this study.

2) **The political nature of the tweets collected.** The central goal of the study is to investigate *political* tweets from the 2020 U.S. presidential election. As such, the collaborative team, including me, looked for political keywords related to the 2020 U.S. presidential election: 'JoeBiden,' 'DonaldTrump,' 'Biden,' 'Trump,' 'vote,' 'election,' '2020Elections,' 'Elections2020,' 'President-ElectJoe,' 'MAGA,' 'BidenHaris2020,' 'Election2020.' This list was first chosen to investigate political discussions regarding American polarization. Hence, tweets that included one of those key terms were automatically collected in real-time. Since the research analyzes political discussions on COVID-19, this first filtering is necessary.

3) **Filtering down tweets discussing COVID-19 topics.** Since this project investigates how bot users intervene politically in COVID-19 debates, two expert coders, including me, looked at keywords in the literature and on Twitter. This search includes lockdown, mask, vaccine, and conspiracy debates.[7] Then, from the extensive data set, we filtered down tweets that used at least one of the key terms chosen. Detailed and meticulous research of keywords was performed to include most of the discussions related to this international sanitary crisis. Numerous articles that picked and justified the use of keywords were considered (Petersen et Gerken

[7] The full list of keywords can be found here in Appendix H.

2021; Kouzy et al. 2020; Chen et al. 2020; Dimitrov et al. 2020; Al-Ramahi et al. 2021; Ahmed et al. 2020; Memon et Carley 2020) and trending hashtags at the time of the collection were examined.

4) **Considering users who were given a bot score and wrote in English**. The last step was to sample 40,000 users from the extensive data set that have tweeted on COVID-19 topics and who already have a bot score from an automated detection method called Botspot.[8][9][10] The Botometer *pro version 4*[11][12] detection model was performed to obtain users' bot scores. The definition and the process of extracting the score are explained in the following subsection (3.1.2).[13]

3.1.2 Estimating the bot score

Numerous bot detections have been developed to estimate users' bot probability. Generally, there are three common ways to detect bots (Alothali et al. 2018). One way is through a *graph-based* method focusing on networks and relations between users (e.g., trust propagation, clustering) (Jia et al. 2017; Mehrotra et al. 2016). The second technique is *crowdsourcing,* whereby experts look at users' online actions and characteristics and identify general patterns through the labeling

[8] The *botspot* is an automated algorithm developed by a computational scientist student from the MILA collaboration. it replicates a bot detector model (Botometer) presented in Yang et al. (2020) and Rheault et Musulan (2021). This model relies "on users' metadata to detect social bots" (Rheault et Musulan 2021, 8). Several parameters such as the lexical characteristics of usernames, the growth rates of tweets, followers, or friends, are used to produce a score between 0 and 1. A score close to 0 means that the user has a low probability of having automated activities.
[9] 165,089 users discussing COVID-19 were given a botspot score.
[10] This detection model was dropped and replaced by the *Botometer* model since it performed poorly when I manually evaluated the performance.
[11] https://botometer.osome.iu.edu
[12] https://rapidapi.com/OSoMe/api/botometer-pro
[13] Appendix A detailed the process to obtain users' bot scores from the Botometer detection model.

and annotation of the users studied (Subrahmanian et al. 2016; Alarifi et al. 2017). In this category, text-based methods are also employed using natural language processing to ascertain the user's bot-like probability. Lastly, the third way is to detect bots with *machine learning* through *text-based* and *feature-based* methods, which implies the development of algorithms and statistical probabilities that base their outputs on specific features to ascertain the likelihood of accounts being bots (Cai et al. 2017; Chavoshi et al. 2016; Davis et al. 2016). The standard features considered when producing the bot score are usually the hashtags, the number of tweets and retweets, the number of mentions, the age of the account, or the screen name (Alothali et al. 2018). These features represent patterns of "timing, test use, sentiment, automation, and clickstream behavior" that are specific to the automated activities of bots (Alothali et al. 2018, 178).[14]

This master's thesis employs the *Botometer Pro version 4* detection model (Yang et al. 2022) to detect users with bot activities on Python. This well-known *machine learning* detection model regroups different *feature-* and *text-based* methods. This bot detection version was first developed by Davis et al. (2016), at the Indiana University Network Science Institute and the Center for Complex Networks and Systems Research (Davis et al. 2016). This method is classified as a supervised learning-based detection approach, which means that a classifier learns to identify accounts as bots based on prominent features that have been previously trained. As explained earlier, this detection algorithm is based on classifications of "selected features to sort accounts into either legitimate or bot accounts" (Alothali et al. 2018, 177). Indeed, it can extract more than 1,000 features from a Twitter account. Davis et al. (2016) mention four significant categories of features. The first category is the *network* features, such as networks of mentions or retweets. The

[14] Appendix D presents an extensive list of features considered when producing bot scores with the Botometer detection model from Yang et al. (2020).

second category is the *user's* features, that is, the geographical location or the account creation time. The third category is related to the *friends' characteristics*: the number of followers, followees, and posts of each account. The fourth category is *temporal attributes*, such as the tweet rate time of an account.[15] These categories are similar as Himelein-Wachowiak et al.'s (2021) list of technical features. As a result, it evaluates the extent to which a Twitter account presents a similar characteristic to social bots (Davis et al. 2016). In the end, a score from 0 to 1 represents the user's account evaluation output. The closer the score is to 1, the more likely this user is a bot.[16]

Many users could not be given a score since they are unauthorized users, private or suspended accounts. Additionally, since the content and language features from the *Botometer* classifier are based on English (Yang et al. 2022, 4), it was necessary to remove any non-English accounts. Consequently, those restrictions reduced the sample of users to 37,960. Finally, the last step was to merge users with their tweets from the extensive data set collected from October 9th, 2020 ("2020-10-09 00:27:53 UTC") to January 4th, 2021 ("2021-01-04 23:52:40 UTC"), in the run-up and the aftermath of the 2020 U.S. presidential election on Twitter, which represents 1,466,218 tweets.

3.2 Classification

This subsection highlights four distinct classification measures to test the five hypotheses presented in Chapter 2. It includes a classification of the bot-like (bot vs. human), the ideology

[15] The *Botometer* classifier requires the account's most recent 200 tweets and mentions from other users (Yang et al. 2022).
[16] Appendix A presents the practical steps to get the Botometer score for the study with Python, the Twitter API, the Rapid API, and the Botometer API.

(conservative vs. liberal), the sentiment (positive vs. negative), and the conspiracy (conspiracy vs. non-conspiracy).

3.2.1 Users and social bots

The primary variable is the probability of users being bot accounts. Users are classified in a dichotomous fashion to capture their bot-like. Users with a score equal to or above 0.5 were classified as having a high probability of automation activities (bot). Users below 0.5 were classified as users with a low chance of automation activities (human). To analyze and investigate bot-like users, some authors have compared users' mean bot scores with t-tests (Yang et al. 2022). In contrast, many other authors dichotomize bot scores and only consider an account with a higher bot score (Shao et al. 2018). In the literature, the most common threshold is 0.5 (Vosoughi et al. 2018; Shao et al. 2018; Bessi et Ferrara 2016). Indeed, some academics argue that transforming bot scores into a binary classification using a threshold of 0.5 is a conservative choice that will minimize false negatives and positives (Shao et al. 2018, 24).

Additionally, some studies have shown that using a binary assessment using this threshold maximizes accuracy (Varol et al. 2017). Indeed, Varol et al. (2017) present a framework to detect online bots and test different thresholds that "best discriminate between humans and bots" (Varol et al. 2017, 280). In other words, they computed classification accuracy for a set of different thresholds considering "all accounts scoring below each threshold as human" (Varol et al. 2017, 285). The maximum accuracy threshold was 0.5 (Varol et al. 2017, 286). Nonetheless, other works have argued for a threshold of 0.7 (Grinberg et al. 2019) or 0.8 (Broniatowski et al. 2018). Indeed, since bot detection is a challenging task, Broniatowski et al. (2018) compared users' accounts with which the Botometer algorithm was highly certain (score between 0.8 and 1 vs. score between 0 and 0.2) to accounts for which the Botometer system was more uncertain (from 0.21 to 0.79). Since my research wants to compare users with a high and low probability of being automated, the threshold is 0.5.

3.2.2 Users and ideology

The second main variable is the ideology of users. The MILA team, including me, has performed the classification of user ideology (Yang et al. 2021). To obtain users' ideology via the training and the evaluation of automated models, we classify a sample of users according to their party affiliation and ideology based on their profile description (Yang et al. 2021, 895). First, we classify users as 'conservative,' 'liberal,' or 'unknown' based on identifiers' description. For 'conservative', we use: [conservative, GOP, republican, trump]. For "liberal," we use: [liberal, progressive, democrat, biden] (Yang et al. 2021, 895).

We labeled users as "conservative'" ("liberal") if the description contains at least one of the conservative (liberal) identifiers and does not include any of the liberal (conservative) identifiers. The rest of the users remain "unknown." We combine concepts related to ideology and partisanship to label liberal and conservative users. This classification is considered 'weak' since user keywords may not match their party affiliation or ideology. For example, instead of a president's name indicating support, they could say "I hate Trump" or "I hate Pelosi." We then classified 1000 general public Twitter users from each side to validate the overall performance of these labels (Yang et al. 2021, 895). This 'strong' classification either confirms the weak labels or indicates the presence of a coding error. Note that a small number of these users can also be independent or apolitical. Therefore, we used the strong labels to train a classifier to generate more accurate labels.

The computer scientists from the MILA collaboration fine-tuned a RoBERTa-large (Liu 2019) (Robustly Optimized BERT Pre-training Approach) model[17] to predict the party each user "is closest to from their profile description" (Yang et al. 2021, 895). Computer scientists employ this model based on the Transformer model[18] (Vaswani et al. 2017) and a BERT architecture[19] (Devlin et al. 2019), with modifications designed to improve the training process.

3.2.3 Users and tweets' sentiment

I am interested in determining whether a given tweet expresses a positive or negative sentiment toward COVID-19 measures. Two sentiment models are presented in Chapter 4. The first is a sentiment model at the tweet level called **Stancov-19** (3.2.3.1), and the second is at the word level, called **Bing** (3.2.3.2). These methods are employed to test the last two hypotheses (H4 and H5). The first states that bot users are more negative in their tweet content than human users (H4). While the second states that the presence of social bots and the negative content shared online will produce a higher number of users talking negatively about Covid-19 measures in the U.S. (H5).

---

[17] This model is a pretrained language model that optimizes the training of a model architecture, called BERT, for which the goal is to take less time during pre-training.
[18] This model relies on an attention mechanism that takes longer sequence but with shorter training time (Vaswani et al. 2017, 6)
[19] It stands for Bidirectional Encoder Representations from Transformers and represents a natural language processing model.

3.2.3.1 Sentiment at the tweet level

The **Stancov-19** classification model is presented in this subsection to investigate how human and bot users employ sentiment to discuss COVID-19 topics. From the collaborative MILA team, a classification of the posts about COVID-19 (e.g., mask, vaccine, lockdown, miscellaneous (anything else) were produced. Three stances were possible: neutral, negative, or positive. Additionally, these stances indicate whether a post is about misinformation (e.g., anti-vaccine rhetoric is classified as negative). It gives a taxonomy with 12 categories (4 topics × 3 stances) (Yang et al. 2021, 895). For this classification, we manually classified 18k+ popular hashtags from our datasets and keywords explained in subsection 3.1.1.[20] More specifically, a post is “positive if it has at least one positive and no negative keywords” (Yang et al. 2021, 895). Negative posts contain at least one negative keyword. The rest are neutral. A post is classified as "Lockdown," "Mask," "Vaccine," or "Miscellaneous" if it contains the relevant keywords. "Miscellaneous" express any other keywords that could not be included in one of the three key topics. Then, a case-insensitive text search was performed to classify all posts (Yang et al. 2021, 895).

At first, this classification represented tweets' stance related to COVID-19 measures. In other words, a tweet classified as negative would mean that the tweet has a negative stance toward COVID-19 measures. However, since the accuracy was low, the validation from a manual labeling performed for this study considers tweets classified as negative in the **Stancov-19** classification as tweets talking negatively about COVID-19 measures (not necessarily against COVID-19 measures). The same logic is considered for positive tweets. A detailed explanation of the choice

[20] The full list of keywords can be found here in Appendix H.

of labeling and the interpretation of the **Stancov-19** classification is presented in Appendix C. In the end, the **Stancov-19** classification had a higher association with the Manual sentiment labeling than the Manual stance labeling classification performed on a random test set of 50 tweets.

3.2.3.2 Sentiment at the word level

A second sentiment model, **Bing**, is performed to compare the results from the first method. This model assesses the sentiment of tweets by looking at the sentiment of words in a tweet (Hu et Liu 2004). Since I only look at the word in each tweet, a corpus of words is created using the *tidy text* package in R. This model includes a lexicon of words with a classification sentiment (positive vs. negative). It has 6,789 words with annotations as positive or negative (Kiritchenko et al. 2014, 737). Thus, words analyzed in this master's thesis only match words from the lexicon. This way, I could investigate the count of positive and negative words overall and over time.

3.2.4 Users and conspiracy tweets

From the extensive database, when filtering down COVID-19 topics, two manual coders, including me, had to define a category called 'conspiracy.' The literature defines this term as an explanation of some event or practice by referencing the machinations of powerful individual(s) who conceal their role (Sunstein et Vermeule 2008, 4). After manually analyzing 18k hashtags talking about themes related to COVID-19 (e.g., mask, lockdown, vaccine), we also looked at other articles that use COVID-19 and conspiracy keywords to merge them with the ones we found, such as in Ahmed et al. (2020), Al-Ramahi et al. (2021) and Kouzy et al. (2020). Tweets labeled

'Conspiracy' were those that discussed COVID conspiracy theories (either supporting or opposing).[21]

3.3 Reliability of the classification measures

Users and the Botometer score. I am confident that the Botometer classification is reliable since this method is considered the state-of-the-art algorithm for detecting online bot activities. Firstly, the accuracy of the Botometer model has been evaluated through "5-fold cross-validation" on numerous manual annotated datasets (Yang et al. 2022, 4). Compared to those datasets, the latest version of Botometer had an "area under the receiver operating characteristic curve" of 0.99. Thus, the model can accurately distinguish bot and human accounts from the annotated datasets. Additionally, other accounts not related to those in the annotated datasets had very high accuracy (Yang et al. 2022, 4).

Furthermore, as presented in the literature review section, almost all political science works look at the impact of social bots on public opinion using this detection model (Bessi et Ferrara 2016; Broniatowski et al. 2018; Ferrara 2020). It has been well documented and employed by computer scientists and political science researchers over the last few years. Even a website was built to allow the public to look at the bot probability of any account.[22] Finally, using 0.5 as the threshold is commonly used in the literature (Vosoughi et al. 2018; Shao et al. 2018; Bessi et Ferrara 2016).

[21] A list of conspiracy keywords can be found in Appendix G.
[22] https://botometer.osome.iu.edu

Furthermore, a validation check is performed on 50 random accounts. Those accounts were divided into two categories depending on their bot score. Varol et al. (2017) checked 3,000 accounts manually to estimate whether the bot detection model employed is reliable. A similar methodology was followed to validate the Botometer version 4 performed in this study. After randomly sampling 25 bot accounts for each category,[23] 50 accounts were manually labeled based on different features.[24] Several features analyzed manually were considered, such as the users' profiles and activities. Appendix B details which features were thought to classify accounts as bots or humans. Appendix B also presents the output of this robustness check and F1 scores. This accuracy test represents the harmonic mean between the precision and the recall (Lipton et al. 2014, 227). It enables to compare the performance of the Botometer classifier with the manual classification of the bot score. The first bot-score category has an accuracy of 0.81 (F1-score), while the second category has an accuracy of 0.71 (F1-score). Thus, this research considers the Botometer detection model since the overall F1 scores are high.

Users and ideology. The classifier provides a reasonably accurate classification of ideological labels. Even though the accuracy of the users' ideology is not perfect, we are confident of the reliability of the classification. We found that the accuracy of users' ideology is around 97.7% and 96.8% for conservative and liberal respectively.[25]

---

[23] Varol et al. (2017) proceed by decile to evaluate manually the bot score of users. By proceeding this way, they could evaluate the spectrum of human and bot accounts "without being biased by the distribution of bot score" (Varol et al. 2017, 283).
[24] First category: score below 0.5, second category: score above or equal to 0.5
[25] The ideology classification is related to a work in progress from the MILA collaboration. This master's thesis was presented on June 9, 2021 at the Polarization & Politics section of the Canadian Political Science Association. https://cpsaevents.ca/2021/sessions_details.php?id=155

Users and sentiment. Different models are tested to validate the sentiment analyses. I examinate their associations with different manual annotations built for this research. These manual annotations indicate what should represent a positive, negative, supportive, or opposing tweet. The two models chosen are the most strongly associated with the manual labeling performed for this project, which are **Stancov-19,** and **Bing**.[26]

At the tweet level, two methods were tested: the **Stancov-19** and the Valence Aware Dictionary for Sentiment Reasoner (VADER) model. This last model was developed by Hutto et Gilbert (2014) and looks at the polarity as well as the intensity of the sentiment of a tweet. It is a well-known model since it is more sensitive to sentiment expressions in social media contexts (Elbagir et al. 2019, 2) and easy to interpret since it gives a single unidimensional measure of sentiment score for each tweet.[27] For both methods, a test set of 50 tweets was built with the output for each model. Then, two new columns were added for the manual annotations. The first indicates whether the tweet supported or opposed COVID-19 measures. If a tweet explicitly indicates that it favors respecting the COVID-19 measures, it will be written 0; otherwise, 1. This column is called '**Manual Stance.'** The second column indicates whether the tweet is positive or negative. If the tweet explicitly indicates a positive emotion, it will be written 0; otherwise, 1. This column is called '**Manual Sentiment**.' Phi coefficient correlations are employed to evaluate the association between the models and the manual labelings. The Phi coefficient[28] is a method for determining the strength

[26] These different tests show that detecting the tweets' stance and sentiment remains challenging.
[27] The mathematical formula is:

$$x = \frac{x}{\sqrt{x^2 + \alpha}}$$

[28] Phi Coefficient means Mean Square Contingency Coefficient.

of an association between two categorical variables, each of which is measured as binary, that is, they only have two groups.[29]

It turns out that the correlation between the **Manual sentiment** and the **Stancov-19** has the strongest positive association, with 0.76. Hence, this study proceeds with the **Stancov-19** classification due to its high association with the manual sentiment labeling. However, the output of the **Stancov-19** classification translates the sentiment (negative vs. positive) and not the stance (opposing vs. supportive) of the COVID-19 tweets. Another test was reported in Appendix E, which looks at a stance detection model from Kawintiranon et Singh (2021). However, when reproducing their results on the data test set, the association was low.[30]

Turning to the method at the word level: **Bing**, a similar robustness check was performed. However, only one manual labeling was completed to classify words as positive or negative since it was impossible to label words as supportive or opposing COVID-19 based on only one term. The result of the association is high, at 0.93. Hence, the two models used in this project are the **Stancov-19** classification and the **Bing** model.

Users and conspiracy tweets. Finally, two expert coders, including me, randomly sampled and labeled 400 tweets to test whether conspiracy tweets were talking about conspiracy theories. It has given a high average accuracy of 90%.

[29] The Phi correlations table can be found in Appendix C.
[30] A detailed explanation is given in Appendix E.

The reliability of the four classifications is key to this research since most of the results relied on these computational methods. **Table 3** presents a summary of the variables employed in this project.

Table 3: Variables of interest.

Name	Description	Indicator	Range
Users' bot like	Probability of having automated activities	Continuous	From 0 to 1
Users ideology	Political leaning	Binary	Conservative or liberal
Conspiracy tweets	Tweet discussing conspiracy theory(ies)	Binary	Conspiracy or non-conspiracy
Tweet sentiment	Emotion of a tweet	Binary	Positive or negative
Word sentiment	Emotion of a word	Binary	Positive or negative

3.4 Descriptive analyses

Before testing the five hypotheses from Chapter 2, this subsection presents some descriptive statistics related to users and tweets' number in **Table 4**.

Table 4: Users and tweets' descriptive statistics.

	Human	Bot	Conservative	Liberal	Total
Users	28,798	9,162	10,837	27,122	37,960
Tweets	835,418	630,800	216,865	1,249,353	1,466,218

Users with a bot score were 98% English speakers. Overall, 37,960 users could be given a Botometer score, for which 28,798 are liberals, and 9,162 were conservatives. From **Table 4**, even though the number of bot accounts is low, their tweets are significant. Indeed, while human accounts represent three-fourths of the data set (75.86%), bot users count for almost one-fourth (24.14%). On one hand, this finding correlates with what scholars demonstrate, where the number of bots is lower than the number of human users on social media (Yuan, Schuchard, et Crooks

2019, 6; Badawy, Ferrara, et Lerman 2018) (Varol et al. 2017, 288). On the other hand, the number of tweets' bot accounts remains significant in this data set.[31] Other works have found similar results when looking at the number of tweets from bots. Varol et al. (2017) found that bots accounts could represent between 9% and 15% in online activities. Luceri et al. (2019, 1008) found 21,1% of users with bot activities.

Turning to the descriptive statistics regarding tweets, **Table 4** indicates that overall, there are 1,466,218 tweets, 56.98% of them human-like tweets and 43.02% bot-like tweets. Users with a low bot score are numerous in tweets. Most users posting and retweeting are liberal users. The number of tweets from bot users is significant, almost 45%. Generally, related works found few tweets, such as Luceri et al. (2019) with 30.6% or Bessi et Ferrara (2016, 5) with 18.45%.

[31] Appendix A presents the distribution of users with a bot score and an ideology.

Chapter 4. Results

Following the analytical and methodological strategy presented in Chapter 3, the next chapter empirically documents the role played by social bots in users' opinions on COVID-19 around the 2020 U.S. presidential election. Results are divided into five sections reflecting the analytical questions and hypotheses.

4.1 Frequency tweets and retweets analysis

One avenue to perceive user bots' frame is by investigating their online proportion. This subsection investigates tweets and retweets from users given a bot score from the period of study: October 9th, 2020, to January 4th, 2021. To validate the first hypothesis that social bots are generally more active among conservative users (H1), users' tweets proportion is investigated. In doing so, I compute their proportion following practical steps. At first, liberal (conservative) users are kept, and the number of tweets and retweets posted by humans and bots are counted daily. After taking the total number of tweets and retweets from liberal (conservative) bots and humans, the percentage of liberal (conservative) bot-like users' tweets is multiplied by 100 and divided by the total number of liberal (conservative) tweets and retweets each day.

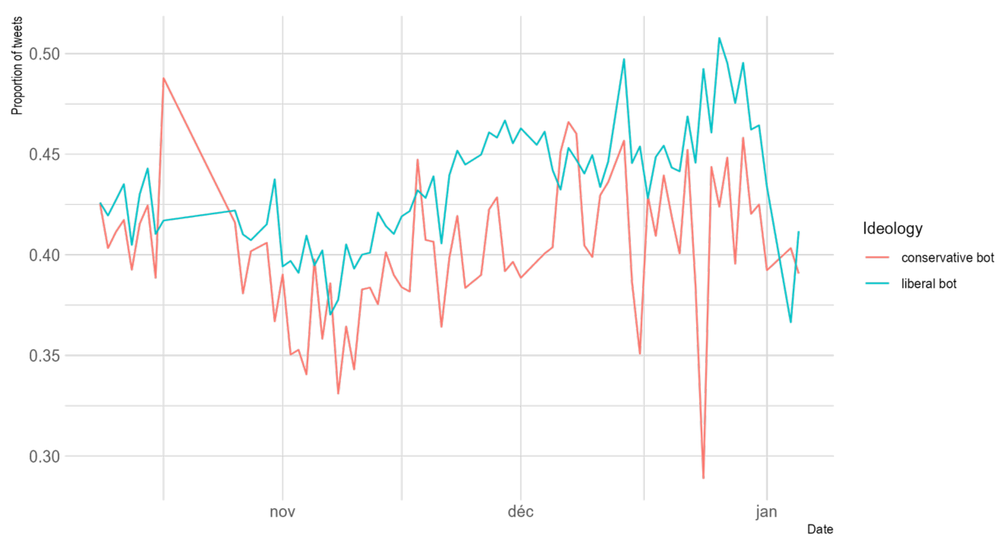
Table 5 presents the overall proportion of tweets per user's ideological and bot category. The mean proportion of tweets and retweets from conservative bot-like is 40.1%, while the overall mean proportion of tweets and retweets from liberal bot-like is 43.5%. Among conservatives, 59.9% are human tweets, and among liberals, 56.5% are human tweets. In other words, neither liberal nor conservative bot users have a higher proportion among liberal and conservative, respectively. Thus, the hypothesis that conservative social bots are generally more active must be rejected (H1). Since tweets' bots are more liberal than conservative, this contradicts what has been reported in the literature, where social bots are deemed to be more conservative during the U.S. presidential elections (Badawy et al. 2018).

Table 5: The proportion of tweets and retweets (%).

	Conservative	**Liberal**
Bot	40.1	43.5
Human	59.9	56.5

Figure 1 investigates whether conservative or liberal bot users tend to be higher in proportion among their ideological group during a specific day.

Figure 1: The proportion of bots' tweets and retweets according to their ideology.



The blue line represents the proportion of tweets and retweets from liberal bots among liberals. In contrast, the red line presents the proportion of tweets and retweets from conservative bots among conservatives. The general picture indicates that a couple of days after the general U.S. presidential election held on November 8th, 2020, there is an upward trend over time in the proportion of bot users, reaching a peak at the end of the year on December 26th for liberal bots. Indeed, on December 26th, the proportion of liberal bot-like tweets was 51%. It is closely related to the release of a statement from President-elect Joe Biden asking the incumbent Donald Trump to sign as soon as possible the COVID-19 relief bill passed by the U.S. Congress (Grayer et Luhby 2020). Another peak is seen on December 29th, 2020, for conservative bots, but it never surpasses the proportion of conservative human tweets and retweets.

The graph shows that liberal bots' tweets and retweets may be as many as liberal humans' tweets and retweets (e.g., December 26th). Notwithstanding, the content shared by social bots may be a more helpful strategy to disturb the online sphere, as suggested by several authors (Entman et Usher 2018; Parra-Novosad 2020). The following subsection looks at another specific strategy documented by numerous authors; the share of misinformation related to COVID-19 (Marx et al. 2020; Ferrara 2020; Shao et al. 2018). This subsection examines whether social bots use a conflictual frame by sharing false information when interacting with humans.

4.2 Conspiracy tweets and retweets analysis

Do bots harness a conflicting frame and share conspiracy theories? In this second section, two hypotheses presented in Chapter 2 are tested. As discussed by Neuman et al. (1992), the media are seen as means of attracting attention where polarized forces are central themes in their news presentation (Neuman et al. 1992, 64). This subsection hypothesizes that bot users on social media are seen as means of attracting attention with two polarized forces discussed at the end of the year 2020: the incumbent president, Donald Trump, and the president-elect Joe Biden. More specifically, bot users will emphasize stories that offer clashing interpretations (Neuman et al. 1992, 65), such as pro-mask vs. anti- mask or pro-lockdown vs. anti-lockdown narratives. One avenue to attract users' attention is using conspiracy theories related to COVID-19, which has already been broadly discussed in previous studies (Ferrara 2020; Marx et al. 2020). Hence, the first hypothesis tested here is whether bot users share more conspiracy theories related to COVID-19 during the 2020 U.S. presidential election than humans (H2). The second hypothesis is whether conservative social bots share at a more significant rate conspiracy theories related to COVID-19 than liberal-bot users (H3).

As discussed in Chapter 3, conspiracy tweets were categorized when classifying tweets by topic. The total percentage of conspiracy tweets and retweets is 0,54% in the sample studied. **Table 6** presents the proportion and number of conspiracy tweets and retweets according to the bot-like of users. The number of conspiracy discussions is higher for users with a low bot score with 4,733 tweets and retweets discussing conspiracy. The number of conspiracy tweets and retweets for bot-like users is up to 3,190. In the end, the proportion of bots tweeting and retweeting about conspiracy is lower than human users, with a proportion of 0.51% vs. 0.57% of tweets and retweets about conspiracy theories.

Table 6: The proportion of conspiracy tweets and retweets among bots and human users.

Bot level	Conspiracy (%)	Conspiracy	Non-conspiracy	Total
Bot	0.51	3,190	627,609	630,799
Human	0.57	4,733	830,668	835,401

This analysis invalidates the second hypothesis since bot users do not share more conspiracy theories related to COVID-19 during the 2020 U.S. presidential election (H2). The spread of conspiracy theories between bot and human users is still concerning. Indeed, as Shao et al. (2017) discussed in their paper, this result means that human users are as likely to retweet and tweet conspiracy theories as bots without being able to distinguish them.

Even though bots do not share at the same rate as human conspiracy theories, one question remains: do conservative bots share more conspiracy theories than liberal bots? Hence, H3 states that conservative social bots share at a more significant rate conspiracy theories related to COVID-19 than liberal-bot users (H3).

Table 7 shows the proportion and the number of conspiracy tweets and retweets from bot users' ideological categories (conservative vs. liberal bots' tweets and retweets).

Table 7: The proportion of conspiracy tweets among conservative and liberal bot users.

Bot ideology	Conspiracy (%)	Conspiracy	Non-conspiracy	Total
Conservative	2.83	2,456	84,421	86,877
Liberal	0.13	734	543,188	543,922

This table shows that 2.8% of conservative tweets refer to conspiracy, and 0.1% of liberal tweets refer to conspiracy. Conservative bot users tend to retweet and tweet conspiracy tweets at a higher rate than liberal bot users. In fact, conservative bot-like users posted tweets and retweeted about conspiracy three more times (2,453 tweets and retweets) than liberal bot-like users (734 tweets and retweets) in the sample. In testing this assumption, it is impossible to determine whether a liberal (conservative) bot user shares a counter-misinformation toward a conservative (liberal) conspiracy tweet. Notwithstanding, this finding validates the third hypothesis that conservative bot-like users share more conspiracy theories than liberal-bot users related to COVID-19 during the 2020 U.S. presidential election (H3).

The conspiracy tweets are analyzed over time to support these last findings. A content analysis of the days where conservative bot users refer to conspiracy tweets in a higher proportion is performed. The word clouds were calculated by looking at the proportion of conspiracy tweets from conservative bot users among liberal and conservative bot users. These descriptive results give insights into bot users' strategy, the share of misinformation as a conflicting frame. Hence, **Figures 2a, 2b, and 2c** present word clouds from October 31st, November 10th, and December 27th, 2020.

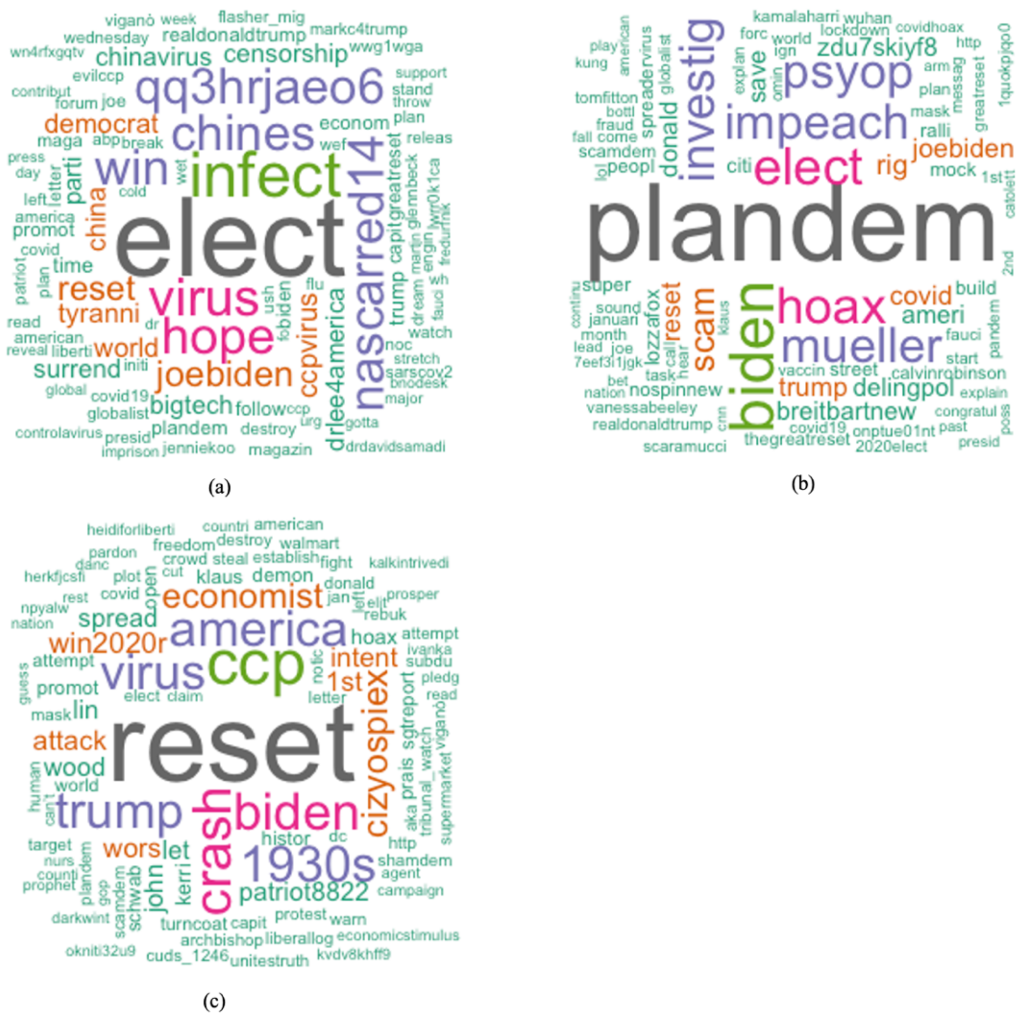


Figure 2: a, b, c: represent word clouds on the days with the highest peaks. Each word cloud contains several discussions related to well-known conspiracy theories. (a) Words frequency on October 31st, 2020; (b) Words frequency on November 10th, 2020; (c) Words frequency on December 27th, 2020.

Word cloud 2a presents terms expressed on October 21st, 2020. The bigger the term is, the more frequent it appears on tweets and retweets on a specific day. The total number of tweets was 74 for this day. Since 'Joe Biden' and Donald 'Trump' often appeared, those terms were removed from the analysis. This analysis brings four misinformation topics related to the COVID-19 pandemic: the virus as a bioweapon created by the Chinese Communist Party (CCP), the Great Rest (TGR), the pandemic as an organized plan, and the pandemic as a mean to establish a new world order. These conspiracy theories are almost exclusively conservative narratives. All four conspiracies on this day refer to a division between two entities, the Chinese party vs. the world and the leaders of developed countries vs. the others. Indeed, the first conspiracy theory reflects misinformation stating that the *CCP has intentionally caused the virus* and disturbed the commercial and political world order (Havey 2020, 320). This conspiracy theory was expressed through different keywords and hashtags, such as ‘china virus,’ ‘ccp virus,’ or ‘evil ccp.’ The second conspiracy refers to a theory on the rise in online COVID-19 discussions: *The Great Reset* (TGR), which is the name given to the structural reforms discussed during the World Economic Forum (WEF) meeting in May 2020 (Schuller 2021, 195). Many leaders met to discuss the COVID-19 crisis, its aftermath, and ways to build a sustainable economy. However, conspiracy theorists have manipulated the narrative detaching TGR from its true WEF meaning. As a result, many conspiracy theorists indicate that TGR is one of the final steps by the elites to control the economy and social life (Schuller 2021, 196), and users in online discussions picked this story up. The terms that represent this narrative are: ‘great reset,’ ‘Davos,’ and ‘forum’ on the **word cloud 2a,** and some examples of the tweets discussing this conspiracy are presented below:

“RT @user: This is in regard to the #GreatReset, and the meeting about which occurred in Davos in January of this year. The only th…”

“RT @user: Time Magazine is now promoting the World Economic Forum's Great Reset of capitalism. But that's not what it is. It would ush..”

Additionally, **word clouds 2a** and **2b** present terms linked to another conspiracy theory, the *Plandemic*. It refers to an expression that seeks to delegitimize public health safety measures.

Especially, this theory delegitimizes Dr. Fauci's competencies (Kearney et al. 2020, 3). This theory comes up with words like 'plandem.' Finally, another conspiracy closed to the TGR is the *globalists*. This term was re-used over time by claiming a new world order governed by "evil foreigners" with a progressive political standpoint (Santini et al. 2022, 14). The term 'globalist' appears a couple of times.

In addition to the first four misinformation topics on November 10th, 2020, another conspiracy topic overlapped with the topics discussed above, *the 2020 U.S. presidential election fraud* as presented in **word cloud 2b.** A total of 156 tweets are reported on this day discussing conspiracy theories. Several users argued that the election was rigged through the mail-in-box process. Indeed, the conspiracy theory related to the election translates efforts from a specific group to "use illegal means to alter election outcomes" (Alvarez et al. 2009, 149). The terms echoing this conspiracy are: 'rig' and 'fraud.' This conspiracy overlapped with the COVID-19 crisis perceived as a hoax. Indeed, some users were arguing about the unnecessity of wearing masks since COVID-19 is a hoax while discussing the 2020 U.S. presidential fraud. Two examples are presented as follow:

"Taking off mask, then swig from a bottle around & putting mask back on & passing to the next clown 🤡 LOL. #CovidHoax"

"RT @user: The Mueller investigation was a scam. Impeachment was a hoax. The plandemic was a psyop. The election was rigged. Biden w… »

Finally, **word cloud 2c** presents an overlapping misinformation topic that discusses *election fraud* and *TGR* using terms such as; 'Schwab,' 'Klaus' to talk negatively about the creator of the term *TGR*, Klaus Schwab, and 'steal' as well as 'fraud' to talk about the *election fraud*.

This content analysis explains how conservative social bots have been discussing and sharing misleading information. The central themes exposed on Twitter always express a duality between two groups. It may be Trump vs. Democrats, the Chinese party vs. the rest of the world, or mask vs. anti-mask as suggested in the **word cloud 2b**. Furthermore, most are conservatives when looking at the ideology of users employing those terms. This analysis shows a clear conflictual representation shared by social bots online. This dichotomous vision of tweets highlights a conflict frame (Parra-Novasad 2020) employed by conservative social bots since they disagree with the COVID-19 measures and the election result. As conspiracy theories are easily shared on social media such as Twitter, this may directly impact democratic systems since conspiracy individuals call into question fundamental democratic structures such as the election integrity or the role of politicians and doctors in the severity of the COVID-19 disease.

While conspiracy tweets are more prevalent among human users than social bots, the following section tackles whether the sentiment is similar between bot and human users and investigates if one of them is more negative. Some works found that bot users are more negative in their content, which can be a strategy to disturb the online sphere. The fourth assumption is tested in the next section, stating that bot users are more negative in their tweet content than human users (H4).

4.3 Sentiment analysis

Several authors have demonstrated that social bots and humans share a similar sentiment, but some studies have found that rogue bots tend to be more negative in their content. Thus, the share of negative content is perceived as a proxy of the conflicting frame presented by Parra-Novosad (2020). In this section, the fourth hypothesis is tested with two sentiment detection models presented in Chapter 3: the **Stancov-19** classification at the tweet level and the **Bing** model (Hu et Liu 2004) at the word level.

4.3.1 Stancov-19 classification sentiment analyses.

Table 8 presents the sentiment mean of human and bot users with the **Stancov-19** classification model. The mean scale goes from 0 to 1, where 0 means a positive mean sentiment while 1 means a negative mean sentiment. The overall picture indicates that bot and human users have a similar sentiment in their tweets, as demonstrated in Shi et al.'s (2020) study. The mean sentiment for bot users is 0.012, while the mean for humans is 0.014. Both are talking more positively about COVID-19 topics. Hence, H4 is invalidated.

Table 8: The mean sentiment for bot and human users.

Bot level	Mean sentiment
Bot	0.012
Human	0.014

4.3.2 Bing sentiment analysis.

The same hypothesis is tested in this subsection but at the word level. The second method, **Bing**, inspects whether the first result at the tweet level is consistent when looking at the word level. This model assesses the sentiment of words in a set of tweets in a dichotomous fashion (positive vs. negative).

Table 9 present the overall proportion of negative words employed by human and bot users. The proportion of negative words among positive and negative words is given. Neutral words are removed from the analysis. The proportion scale goes from 0 to 100%. The overall picture indicates that bot and human users have a similar sentiment again in the terms employed. This result correlates with the first method at the tweet level. However, the overall negative sentiment mean is 0.73 for bot users and 0.72 for human users. Both are using more negative words.

Table 9: The overall proportion of negative words.

Bot level	**Negative words (%)**
Bot	0.73
Human	0.72

The differences in terms of sentiment between the **Stancov-19** and the **Bing** models are twofold. First, the set of annotated words (positive vs. negative) is probably different from the data set with the **Stancov-19** classification. Indeed, the lexicon presented in the **Bing** model includes 6,789 words that may be more matched with words in this project that are negative rather than positive. Secondly, the accuracy of each method may impact the output presented. In fact, when testing a test set of 50 tweets, the accuracy at the word level was higher than at the tweet level, this experimentation is presented in Appendix C.

Hence, the sentiment of bot and human users are similar over time, regardless of the model chosen. However, humans are more negative than bots at the tweet level, and bots use a bit more negative words at the word level. Notwithstanding, at the tweet level, the result is not substantial and both groups talk positively about COVID-19 in general. It is impossible to validate that bot users tend to be more negative. This result does not sustain the conflicting conceptual frame used by Parra-Novosad (2020).

4.5 Time-Series Analyses. Impulse Response Functions.

Since the Phi correlations have indicated that the **Stancov-19** classification is highly associated with the **Manual sentiment** coding of tweets (positive vs. negative), this last subsection tests the fifth hypothesis with the **Stancov-19** as a measure of tweets' sentiment. The last hypothesis states that the number of social bots and the volume of harmful content they share (as a conflict frame) will produce a higher proportion of users that talk negatively about COVID-19 measures in the U.S. (H5). It is critical to study the real-time effects (short and long runs) of the explanatory variables (sentiment and proportion of bots) on the dependent variable (sentiment tweets). The social media activities collected enable to determine sentiment responses to the proportion of bots and their content as they occur.

A general way to model the immediate and long-term outcomes of shocks in specific variables in a noisy system (social media) is the Impulse Response Functions (IRF) based on Vector

AutoRegressive (VAR) models (Barberá et al. 2019). This technique observes the impact of any variable on others and can be used in empirical causal analysis (Lin 2006, 1). Since the Twitter discussions have been monitored for a period of three months, data series are generated to reflect individuals' perception of COVID-19 topics over time. Hence, IRFs are built to investigate the immediate and long-term effects of the proportion and the sentiment content of social bots on the sentiment of human users. The period studied is the same as for the last hypotheses tested, from October 9th, 2020, to January 4th, 2021. At first, it is essential to perform stationary tests to validate whether the data has any trend or seasonable trends over time, which is not the case for these data series. Indeed, the t-statistic indicated a p-value = 0.01, <0.05 for each variable. After this, a 10-day window was chosen to measure the immediate and long response changes in Twitter sentiment.[32] I built a Vector Autoregressive model and examined the IRF, following Polyzos (2022), Duan et al. (2022) and Barberá et al.'s (2019) methodology. Thus, first, unrestricted VAR models, which take the following form, are estimated:

$$Y_t = \alpha + \sum_{p=1}^{n} \beta_p X_{t-p} + \varepsilon_t$$

Where α is the intercept (constant), and β_p is the coefficient of the lags of Y. X is the lag independent variable representing the proportion of bot users with a scale from 0 to 1. Y is the lag dependent variable that represents the mean sentiment at day t, going from -1 (negative) to 1 (positive). Lastly, ε is the residual element.

[32] Normality and heteroskedasticity tests were performed. The results are presented in Appendix F.

Second, since all variables in this model may depend on each other, individual coefficient estimates "cannot provide useful information on the reaction of the system to a shock" (Polyzos 2022, 8). To resolve this problem, several economists and scientists have then calculated the IRF by following the Model Forecasting Error Impulse Response (FEIR), also called the moving average (ϕ_i) below:

$$\phi_i = \sum_{p=1}^{i} \phi_{i-j} X_j$$

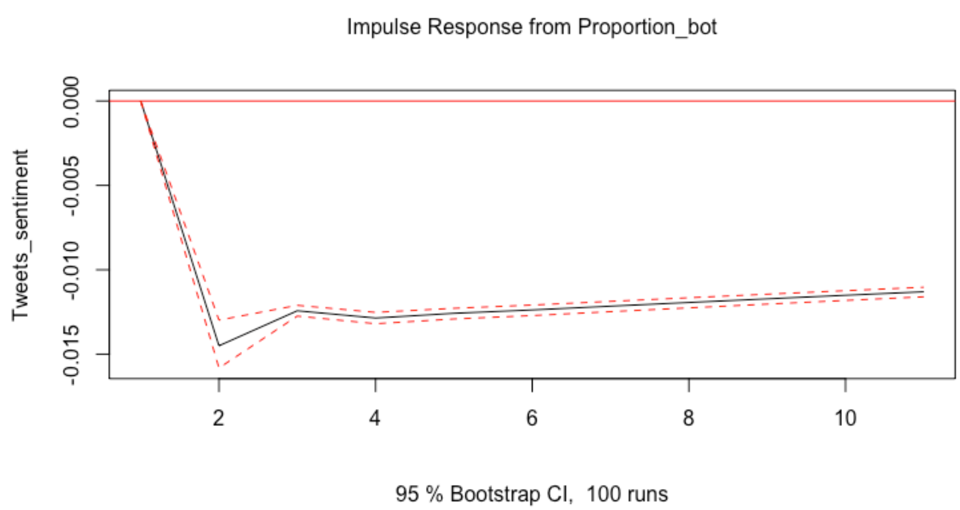
Where i is going from 1 to 10 lags and with $X_j = 0$ for $j > p$ which represents the lag order of the VAR model. More specifically, it represents the response of a variable (mean sentiment of a tweet with a scale from -1 to 1) to a unit impulse in another variable (proportion of bot and sentiment of their content which goes from 0 to 1) occurring in previous days. The proportion of bots is a continuous variable, while the tweet's sentiment is a binary variable.

Shock to Twitter indicator (sentiment of the tweets) reflects shock in the public perception of the COVID-19 measures around the 2020 U.S. presidential election. In this sense, a negative sentiment indicates a negative discussion related to COVID-19 measures. The sentiment variable represents the sentiment of the tweets. Thus, a positive shock means that users talk positively about the COVID-19 measures, while a negative shock to the sentiment means that users talk negatively about COVID-19 measures.

The findings are shown in **Figures 3 and 4**. **Figure 3** presents the result of the IRF between the proportion of social bots (Independent variable (IV)) and the tweets' sentiment (Dependent

Variable (DV)). The ordinate line (Tweets_sentiment) represents the scale sentiment for which a negative score would mean that users talk negatively about COVID-19 measures in their tweets while a positive score would mean that users talk positively about these measures. The scale goes from -1 to 1. The abscissa line represents the effect of the proportion of bots (called "Proportion_bot" that goes from 0 to 1) over a ten-day period (0 to 10). The dark line in the figure represents the level of the IRF while the red lines around the dark one represents the two standard deviation bands. Finally, the straight red line at 0 means no variation at t time.

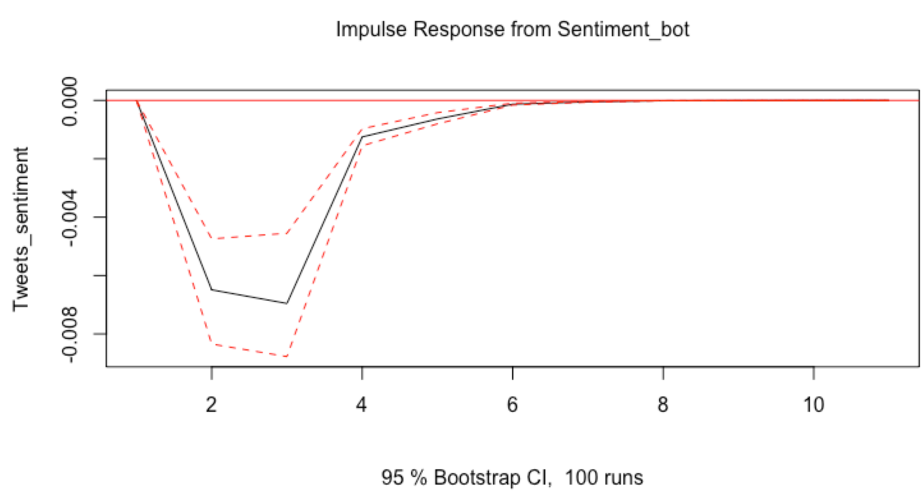
Figure 3: The Impulse Response Functions to bot users proportion.



This first result demonstrates only the impact of the proportion of bot users (Proportion_bot) on users' sentiment in the short (1 day) and long terms (10 days). The effect of a growing proportion of social bots does follow the expected pattern for the sentiment of the tweet indicator, where a response in the opposite direction is registered between two and ten days after the "shock." Even though it follows the expected pattern, this result is not substantial since its maximum is -0.014 during period t = 2.

When looking at the proportion of bot users and their negative sentiment in tweets (Sentiment_bot), **Figure 4** presents the expected pattern, but the effect is again relatively small over time and affects only the short term. A response in the opposite direction is registered in a very short time (t = 3), but the effect is relatively small at -0.07 since the scale of the proportion of bots goes from 0 to 1 and the scale of the mean sentiment of tweet is between -1 to 1. In the long run, there is no significant result since the pattern presented is not different from 0. These results implicate that the proportion of bot users does impact the sentiment expressed in tweets by the human users, but this effect is relatively small. Hence, the fifth hypothesis cannot be validated (H5). The conflicting conceptual frame explains partially bot users' intervention in online discussions in the context of COVID-19 and the 2020 U.S. presidential election.

Figure 4: The Impulse Response Functions to bot tweet sentiment from bots.



Chapter 5. Discussions

5.1 Interpretation of the results:

Social bots' activities are a fast-changing phenomenon. In light of the ongoing COVID-19 pandemic, numerous discussions on this topic and conspiracy theories have abounded online. Going beyond previous studies on social bots' activities, this study provides an empirically informed analysis of the interplay of social bots' behaviors and their potential effect on users' perceptions of COVID-19 measures.

Several results have emerged from this research. From the perspective of the conflict frame (Parra-Novosad 2020), bots use heterogeneous strategies to reach human users. Still, bots do not share as much negative content (e.g., negative and conspiracy Twitter posts) as it is expected, and their impact is not yet clear. Regarding the presence of social bots, evidence reveals an unbalanced presence of social bots compared to human users even if bots' proportion remains significant. Nonetheless, when differentiated between conservative and liberal bot users, their number is almost balanced among their counterparts, even over time. Thus, conservative bot users do not display distinct behaviors compared to human and liberal bot users. This result contradicts the findings of other scholars, such as Luceri et al. (2009), who report disparate behavior according to partisan affiliation. One explanation of this finding is that even though conservative and liberal bots have similar tweets' activities, they may efficiently employ other strategies through a conflicting frame, such as the share of conspiracy theories.

Accordingly, this master thesis reveals a variety of prevalent conspiracy explanations circulating on Twitter through social bots: COVID-19 is a hoax, COVID-19 is a creation of the Chinese communist party, the 2020 U.S. presidential election is a fraud, and progressive elites want to control the economic and social life of citizen through the adoption of TGR. While most of the related conspiracies are directed against the institutions and elites (e.g., TGR, the 2020 U.S. presidential fraud), they challenge scientific consensus and delegitimize the risk of the COVID-19 disease (e.g., COVID-19 is a hoax and a creation of the Chinese communist party) (Mahl et al. 2021, 9). Conservative bot users mainly shared these narratives. Most strikingly, it highlights the capability of social bots to combine different conspiracy theories in the same tweet, even if their proportion is small.

However, the conflict frame does not seem to apply when looking at the sentiment of the content shared by social bots. Indeed, similar sentiment for both human and bot users is observed, which correlates with previous studies. This similar behavior between humans and bots may also underscore social bot's ability to easily mimic human behavior to be part of online discussions without being noticed. However, the proportion of positive discussions at the tweet level during the last trimester of 2020 can be explained by the first announcement of a possible end to the COVID-19 pandemic. Especially with the application for an emergency use authorization of COVID-19 vaccine by the Pfizer Industry to the Food and Drug Administration (FDA) on November 20th, 2020 (Pfizer Inc. 2020), and its approvement on December 11th, 2020 (Hahn 2020). Lastly, the first vaccine transportation throughout the U.S. was on December 14th, 2020 (BBC News 2020). Those political events could explain the positivity of the tweets at the end of the year, where liberal bot users' activity was higher.

Finally, the causal analysis presents no evidence of the impact of bots' conflicting frames on users' perception of COVID-19 measures. Only a tiny impact is found for the relationship between bots' sentiment and humans' sentiment in the short term. This finding echoes the French 2022 presidential election for which the candidate of *La Reconquête* has tremendously employed social media to divide citizen perceptions on immigration but did not reach the second turn. Indeed, Abdine et al. (2022) looked at candidates' online clusters during the 2022 French presidential election and detected bot activities. They found that even if the Eric Zemmour cluster was small compared to the other candidates, he sent the most tweets and retweets and embodied the cluster with the most active bots (Abdine et al. 2022, 7). An emerging outcome is that human users are not influenced by the significant proportion of content shared by bot users since they do not react as much to their negative online content. Instead, the impact of the social network structure of social bots on human users (Luceri et al. 2019) should be more considered.

5.2 Limitations of the study:

Numerous limitations to this study ought to be highlighted. Firstly, most of the works presented in this master's thesis primarily focus on the U.S.. Very few studies have investigated the role of social bots in other countries, such as Howard et Kollanyi (2016) in Great Britain and Stella et al. (2018) in Spain. Howard et Kollanyi (2016) argue that social bots had a small but decisive impact during the UK Referendum. By investigating social media related to Brexit, they discovered that social bots frequently used a family of hashtags associated with the departure from the European Union (EU), amplifying those stances by retweeting content about *StrongerInBrexit*. Thus, bots' strategies studied in the U.S. may be replicated abroad.

Another noteworthy limitation of the analysis lies in the representativeness of the data set. Although only tweets written in English and discussing American politics were evaluated, some users may tweet, share, and post from other countries. One possibility would have been to check users' self and unmasked geo-localization. However, few people self-geolocate their accounts, and many hide this information in their parameters.

A further limitation relates to the time when the Botometer score was given and the collection of users' accounts. Since the scores were not given at the time of the collection of the accounts and tweets, the score may be biased since the algorithm fetches the 200 last tweets as one of the parameters to create the score (Yang et al. 2022, 3). Hence, the score may be underestimated since the score was collected after the campaign and the election. In other words, it was collected two months after the tweets' collection of the same users. We should then be careful when drawing descriptive conclusions as the activities of the bot users may have been higher during the 2020 U.S. presidential election.

As presented in the master's thesis and the Appendix, while the research presents and harnesses numerous machine learning models to classify stances and sentiments, the model selection is based on how accurate they are compared to other models tested. In this regard, accuracy remains far from perfect.

Turning to the use of IRF, omitting variables may significantly impact the results. Indeed, it may lead to significant distortions in the results (Lin 2006). It would have been helpful to control the ideology, likes', and mentions' proportions. Lin (2006) also points out the need to correctly order the variables since the model is sensitive to the variable ordering in the model.

Lastly, we should note that the effectiveness of the Botometer is to be nuanced, as underlined by Rauchfleisch et Kaiser (2020). It may have several issues detecting the probability of an account as a bot. Indeed, they assess the overall performance of the Botometer model using the Receiver Operating Characteristics (ROC) curve to assess whether the model accurately distinguishes between human and bot users. They also look at numerous data sets applying this model and test them. Even though the Botometer model performs well on a test data set, the results become unreliable when looking at a specific context or use case (Rauchfleisch et Kaiser 2020, 14). Indeed, they highlight the difficulty for the Botometer detection algorithm to identify false positives and negatives in their data collection. Lastly, they mention the model's vulnerability when looking at a temporal element or different languages tweets (Rauchfleisch et Kaiser 2020, 15). Even though Botometer v4 is different from the v3 and should perform better by considering new features and additional training data, the false positive problem may still be challenging.

5.3 Ethical statement:

The data used in this observational study are available to the public without any restrictions since it has been collected through Twitter's official streaming API. According to Twitter's policy, researchers can obtain a list of tweet IDs and information if they have a verified Twitter developer account. The data used in this study is for research purposes only. Furthermore, tweets presented in this research were anonymized. In doing so, anonymized data lose its "character as personal data, therefore informed consent is no longer necessary to handle the data" (Weinhardt 2021, 7).

Chapter 6. Conclusion

6.1 <u>Summary of the research</u>

In this work, social bots and humans' interactions are investigated around COVID-19 and the 2020 U.S. Presidential election. Numerous machine learning procedures are performed to investigate how social bots employ a conflicting frame to disturb the Twittersphere (Parra-Novosad 2020). These techniques answer two research questions: 1) How much or little did social bots intervene during the 2020 U.S. presidential election about the COVID-19 pandemic? 2) Did social bots influence users' opinions about COVID-19 measures during the 2020 U.S. presidential election? Five hypotheses are presented. The evidence show that social bots sent almost as many tweets as human users (H1). Additionally, the master's thesis highlights that social bots are not the ones talking the most about conspiracy theories (H2); however, conservative bot users shared conspiracy tweets at a greater rate than liberal bot users, validating H3. Furthermore, the sentiment analyses demonstrate that social bots express a similar sentiment in their tweets as humans, and the overall sentiment is positive at the tweet level (H4). Lastly, no evidence suggests that the negativity of social bots' content impact users' perceptions of COVID-19 measures (H5).

6.2 Contributions:

This study sheds light on the challenges presented in explaining users' bot online behaviors through the specter of the conflicting conceptual frame. The overall sentiment analysis produces surprising findings since it shows how positively similar bots and humans are when talking about COVID-19 around the 2020 U.S. presidential election. Indeed, we would logically assume more negative discussions since these topics are mainly polarized and have been demonstrated to stir conflictual discussions between clusters and groups. This result emphasizes the need to study the difference between conservative and bot users' behaviors and discover other strategies that may be particular to liberal bot users. Indeed, we can anticipate that liberal bot users are essentially mimicking human user behavior to remain invisible but participate significantly by amplifying specific (liberal) messages.

Another contribution from this master's thesis is the growing opening of political science research overlapping computational science methods. Indeed, to study public opinion related to COVID-19 around the 2020 U.S. presidential election, six automatic detection models are used to conduct this study.[33] This new innovative way to study political attitude and behavior is crucial in today's world since many aspects of everyday life are moving into the digital sphere (e.g., communication, expressing an opinion on platforms, purchases, etc.). Additionally, the COVID-19 pandemic has reduced physical discussions since many employments moved to remote mode, and the public debate was restrained and also moved online. Hence, it is critical to understand how issues are debated online and how the debate is affected by the involvement of automated actors. However, several experimentations presented in the Appendix reveal that classification tasks remain challenging, and scientists must be cautious when manipulating and interpreting the model outputs.

In addition, this work highlights the need to study social bots due to their significant number online. Other conceptual frames should be investigated to explain social bots' behaviors. Indeed,

[33] Botometer detection algorithm, ideology classifier, Vader sentiment model, Stancov-19 classifier, KE-MLM detection model, and Bing sentiment model.

this study has highlighted that the conflicting frame from social bots may not be the most appropriate frame to understand all bot's behaviors and strategies. Different frames should be considered to depict bots' behaviors, such as the economic consequences frame (Semetko et Valkenburg 2000), in which a bot exposes an issue's economic impact.

6.3 Research perspectives

These results ease further investigation and interpretation. Why do liberal bot users outnumber conservative bot users? Most studies analyzing bot users looked at how conservative bots behave online and how they may impact conservative human users (Freelon et al. 2020). Even though behaviors may be similar between liberal and conservative bot users, the political aim is probably different. Indeed, few studies look at the behavior of liberal bot users (Luceri et al. 2019). Hence, this research gave first insights into their potential behavior among liberal human users.

Differently, future research should employ a network analysis rather than simply looking at the textual and the proportion of tweets and bots over time. Especially when looking at the distribution of counter-conspiracy and conspiracy tweets among liberal and human users. This way, we could have a clearer understanding of how misinformation is spread and from which ideological group.

Social bots do exist and interact with human users without being noticed. Hence, to understand how they intervene in the Twittersphere, a poll analysis could examine human reactions when knowing that automated entities interact with them and investigate how this information shapes their perceptions and how much they trust social network platforms. In fact, since social bots have been rising, future studies must look at ways to teach users to identify automated accounts and false misinformation. In doing so, an experimental design study could have two treatment groups for which one receives 1) information from a bot user, the second, 2) misinformation from a bot user, and a control group that receives information from a human. The dependent variable could be the level of trust in the information received. Thus, it would be an approach to capture how social bots may impact users' opinions and trust toward information received online.

References

Abdine, Hadi, Yanzhu Guo, Virgile Rennard, et Michalis Vazirgiannis. 2022. "Political Communities on Twitter: Case Study of the 2022 French Presidential Election". *arXiv preprint,* 1-10. https://doi.org/10.48550/arXiv.2204.07436.

Abokhodair, Norah, Daisy Yoo, et David W. McDonald. 2015. « Dissecting a Social Botnet: Growth, Content and Influence in Twitter ». In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 839-51. https://doi.org/10.1145/2675133.2675208.

Ahmed, Wasim, Francesc López Seguí, Josep Vidal-Alaball, et Matthew S. Katz. 2020. « COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data ». *Journal of Medical Internet Research,* 22 (10):1-8 https://doi.org/10.2196/22374.

Al-Khateeb, Samer, et Nitin Agarwal. 2016. "Understanding Strategic Information Manoeuvres in Network Media to Advance Cyber Operations: A Case Study Analysing Pro-Russian Separatists' Cyber Information Operations in Crimean Water Crisis". *Journal on Baltic Security* 2 (1): 6-27. https://doi.org/10.1515/jobs-2016-0028.

Allcott, Hunt, et Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election". *Journal of Economic Perspectives* 31 (2): 211-36. https://doi.org/10.1257/jep.31.2.211.

Alothali, Eiman, Nazar Zaki, Elfadil A. Mohamed, et Hany Alashwal. 2018. "Detecting Social Bots on Twitter: A Literature Review". In *2018 International Conference on Innovations in Information Technology (IIT)*, 175-80. https://doi.org/10.1109/INNOVATIONS.2018.8605995.

Al-Ramahi, Mohammad, Ahmed Elnoshokaty, Omar El-Gayar, Tareq Nasralah, et Abdullah Wahbeh. 2021. "Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Dat". *Journal of Medical Internet Research Public Health and Surveillance* 7 (4): e26780. https://doi.org/10.2196/26780.

Alvarez, R. Michael, Thad E. Hall, et Susan D. Hyde. 2009. *Election Fraud: Detecting and Deterring Electoral Manipulation*. Brookings Institution Press.

Badawy, Adam, Emilio Ferrara, et Kristina Lerman. 2018. "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign". In *2018 Institute of Electrical and Electronic Engineers/Association for Computing Machinery (IEEE/ACM) International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258-65. https://doi.org/10.1109/ASONAM.2018.8508646.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, et Joshua A. Tucker. 2019. "Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data". *American*

Political Science Review 113 (4): 883-901. https://doi.org/10.1017/S0003055419000352.

Bessi, Alessandro and Ferrara, Emilio. 2016. "Social Bots Distort the 2016 US Presidential Election Online Discussion". *First Monday*, 21 (11):1-14. https://ssrn.com/abstract=2982233

Broniatowski, David A, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, et Mark Dredze. 2018. "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate". *American Journal Public Health,* 108 (10): 8.

Cai, Chiyu, Linjing Li, et Daniel Zengi. 2017. "Behavior enhanced deep bot detection in social media". In *2017 Institute of Electrical and Electronic Engineers International Conference on Intelligence and Security Informatics (ISI)*, 128-30. https://doi.org/10.1109/ISI.2017.8004887.

Chavoshi, Nikan, Hossein Hamooni, et Abdullah Mueen. 2016. *DeBot: Twitter Bot Detection via Warped Correlation*. https://doi.org/10.1109/ICDM.2016.0096.

Chen, Emily, Kristina Lerman, et Emilio Ferrara. 2020. "Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set". *Journal of Medical Internet Research Public Health and Surveillance* 6 (2): e19273. https://doi.org/10.2196/19273.

Davis, Clayton Allen, Onur Varol, Emilio Ferrara, Alessandro Flammini, et Filippo Menczer. 2016. "BotOrNot: A System to Evaluate Social Bots". In *Proceedings of the 25th International Conference Companion on World Wide Web*, 273-74. WWW '16 Companion. Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872518.2889302.

Dimitrov, Dimitar, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, et Stefan Dietze. 2020. "TweetsCOV19 - A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic". In *Proceedings of the 29th Association for Computing Machinery International Conference on Information & Knowledge Management*, 2991-98. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3340531.3412765.

Duan, Z., J. Li, J. Lukito, K. C. Yang, F. Chen, D. V. Shah, & S. Yang. 2022. "Algorithmic Agents in the Hybrid Media System: Social Bots, Selective Amplification, and Partisan News about COVID-19". *Human Communication Research*. 48: 516-42.

Entman, Robert M. 2003. "Cascading Activation: Contesting the White House's Frame After 9/11". *Political Communication* 20 (4): 415-32. https://doi.org/10.1080/10584600390244176.

Entman, Robert M, et Nikki Usher. 2018. “Framing in a Fractured Democracy: Impacts of Digital Technology on Ideology, Power and Cascading Network Activation”. *Journal of Communication* 68 (2): 298-308. https://doi.org/10.1093/joc/jqx019.

Ferrara, Emilio. 2020. “What Types of COVID-19 Conspiracies Are Populated by Twitter Bots?”, avril. https://doi.org/10.5210/fm.v25i6.10633.

Ferrara, Emilio, Herbert Chang, Emily Chen, Goran Muric, et Jaimin Patel. 2020. “Characterizing Social Media Manipulation in the 2020 U.S. Presidential Election”. *First Monday*, octobre. https://doi.org/10.5210/fm.v25i11.11431.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, et Alessandro Flammini. 2016. “The rise of social bots”. *Communications of the Association for Computing Machinery* 59 (7): 96-104. https://doi.org/10.1145/2818717.

Fetzer, James H. 2004. “Disinformation: The Use of False Information”. *Minds and Machines* 14 (2): 231-40. https://doi.org/10.1023/B:MIND.0000021683.28604.5b.

Freelon, D., A. Marwickn et D. Kress. 2020. False equivalencies: Online activism from left to right. *Science*. 369 (6508): 1197-1201.

Grayer, Annie et Tami Luhby. Biden urges Trump to sign Covid-19 relief bill: "This abdication of responsibility has devastating consequences". *CNN*. https://www.cnn.com/world/live-news/coronavirus-pandemic-vaccine-updates-12-26-20/h_91f646f292ac4442261325e3ab343d0d (consulted page on May 3rd, 2022).

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, et David Lazer. 2019. “Fake news on Twitter during the 2016 U.S. presidential election”. *Science* 363 (6425): 374-78. https://doi.org/10.1126/science.aau2706.

Hajli, Nick, Usman Saeed, Mina Tajvidi, and Farid Shirazi. 2022. "Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence." *British Journal of Management* 33 (3): 1238-1253.

Havey, Nicholas Francis. 2020. “Partisan Public Health: How Does Political Ideology Influence Support for COVID-19 Related Misinformation?”. *Journal of Computational Social Science* 3 (2): 319-42. https://doi.org/10.1007/s42001-020-00089-2.

Hegelich, Simon, et Dietmar Janetzko. 2016. “Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet”. In *Tenth International Association for the Advancement of Artificial Intelligence Conference on Web and Social Media*: 579-82.

Himelein-Wachowiak, McKenzie, Salvatore Giorgi, Amanda Devoto, Muhammad Rahman, Lyle Ungar, H. Andrew Schwartz, David H. Epstein, Lorenzo Leggio, et Brenda Curtis. 2021. “Bots and Misinformation Spread on Social Media: Implications for COVID-19”. *Journal of Medical Internet Research* 23 (5): e26933. https://doi.org/10.2196/26933.

Howard, Philip N., et Bence Kollanyi. 2016. "Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum" (juin). *Social Science Research Network*: 1-6. https://doi.org/10.2139/ssrn.2798311.
`

Hu, Minqing, et Bing Liu. 2004. "Mining and summarizing customer reviews". In *Proceedings of the tenth Special Interest Group on Knowledge Discovery and Data Mining. Association for Computing Machinery. International conference on Knowledge discovery and data mining*, 168-77. New York, USA: Association for Computing Machinery. https://doi.org/10.1145/1014052.1014073.

Jia, Jinyuan, Binghui Wang, et Neil Zhenqiang Gong. 2017. "Random Walk Based Fake Account Detection in Online Social Networks". In *2017 47th Annual Institute of Electrical and Electronic Engineers International Forum for Invention Promotion International Conference on Dependable Systems and Networks (DSN)*, 273-84. https://doi.org/10.1109/DSN.2017.55.

Jiang, Julie, Xiang Ren, et Emilio Ferrara. 2021. "Social Media Polarization and Echo Chambers in the Context of COVID-19: Case Study". *Journal of Medical Internet Research* 2 (3): e29570. https://doi.org/10.2196/29570.

Kaplan, Andreas M., et Michael Haenlein. 2010. «"Users of the World, Unite! The Challenges and Opportunities of Social Media". *Business Horizons* 53 (1): 59-68. https://doi.org/10.1016/j.bushor.2009.09.003.

Karpf, David. 2016. *Analytic Activism: Digital Listening and the New Political Strategy*. Oxford University Press.

———. 2017. "Will the Revolution Be A/B-Tested?". In *Analytic Activism*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190266127.003.0001.

Kawintiranon, Kornraphop, et Lisa Singh. 2021. "Knowledge Enhanced Masked Language Model for Stance Detection". In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4725-35. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.376.

Kearney, Matthew D., Shawn C. Chiang, et Philip M. Massey. 2020. "The Twitter Origins and Evolution of the COVID-19 "Plandemic" Conspiracy Theory". *Harvard Kennedy School Misinformation Review* 1 (3). https://doi.org/10.37016/mr-2020-42.

Khaund, Tuja, Samer Al-Khateeb, Serpil Tokdemir, et Nitin Agarwal. 2018. "Analyzing Social Bots and Their Coordination During Natural Disasters". In *Social, Cultural, and Behavioral Modeling*, édité par Robert Thomson, Christopher Dancy, Ayaz Hyder, et Halil Bisgin, 207-12. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93372-6_23.

Khaund, Tuja, Baris Kirdemir, Nitin Agarwal, Huan Liu, et Fred Morstatter. 2022. “Social Bots and Their Coordination During Online Campaigns: A Survey”. *Institute of Electrical and Electronic Engineers (IEEE) Transactions on Computational Social Systems* 9 (2): 530-45. https://doi.org/10.1109/TCSS.2021.3103515.

Kiritchenko, S., X. Zhu, et S. M. Mohammad. 2014. “Sentiment Analysis of Short Informal Texts”. *Journal of Artificial Intelligence Research* 50 (août): 723-62. https://doi.org/10.1613/jair.4272.

Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, et Khalil Baddour. 2020. “Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter”. *Cureus* 12 (3). https://doi.org/10.7759/cureus.7255.

Lipton, Z. C., Elkan, C., & Naryanaswamy, B. 2014. Optimal thresholding of classifiers to maximize F1 measure. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*: 225-39). Springer, Berlin, Heidelberg.

Luceri, Luca, Ashok Deb, Adam Badawy, et Emilio Ferrara. 2019. “Red Bots Do It Better:Comparative Analysis of Social Bot Partisan Behavior”. In *Companion Proceedings of The 2019 World Wide Web Conference*, 1007-12. San Francisco USA: ACM. https://doi.org/10.1145/3308560.3316735.

Mahl, Daniela, Jing Zeng, et Mike S. Schäfer. 2021. “From “Nasa Lies” to “Reptilian Eyes”: Mapping Communication About 10 Conspiracy Theories, Their Communities, and Main Propagators on Twitter”. *Social Media + Society* 7 (2): 20563051211017480. https://doi.org/10.1177/20563051211017482.

Marx, Julian, Felix Brünker, Milad Mirbabaie, et Eric Hochstrate. 2020. “Conspiracy Machines -- The Role of Social Bots during the COVID-19 Infodemic”. arXiv:2012.09536. arXiv. https://doi.org/10.48550/arXiv.2012.09536.

Mehrotra, Ashish, Mallidi Sarreddy, et Sanjay Singh. 2016. “Detection of fake Twitter followers using graph centrality measures”. In *2016 2nd International Conference on Contemporary Computing and Informatics*, 499-504. https://doi.org/10.1109/IC3I.2016.7918016.

Memon, Shahan Ali, et Kathleen M. Carley. 2020. “Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset”. arXiv. https://doi.org/10.48550/arXiv.2008.00791.

Neuman, W. Russell, Russell W. Neuman, Marion R. Just, et Ann N. Crigler. 1992. *Common Knowledge: News and the Construction of Political Meaning*. University of Chicago Press.

Parra-Novosad, Nathalie. 2020. *Social Bots versus Real Humans: The Framing of “Trump’s Wall” on Twitter*. University of Missouri-Columbia.

Petersen, Kai, et Jan M. Gerken. 2021. “#Covid-19: An Exploratory Investigation of Hashtag Usage on Twitter”. *Health Policy* 125 (4): 541-47. https://doi.org/10.1016/j.healthpol.2021.01.001.

Polyzos, Efstathios. 2022. “Escalating Tension and the War in Ukraine: Evidence Using Impulse Response Functions on Economic Indicators and Twitter Sentiment”. *Social Science Research Network* (*SSRN*) Scholarly Paper 4058364. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4058364.

Rauchfleisch, Adrian, et Jonas Kaiser. 2020. “The False Positive Problem of Automatic Bot Detection in Social Science Research”. *PLOS ONE* 15 (10): e0241045. https://doi.org/10.1371/journal.pone.0241045.

Santini, Rose Marie, Débora Salles, et Carlos Eduardo Barros. 2022. “We Love to Hate George Soros: A Cross-Platform Analysis of the Globalism Conspiracy Theory Campaign in Brazil”. *Convergence*, mai, 13548565221085832. https://doi.org/10.1177/13548565221085833.

Savage, Saiph, Andres Monroy-Hernandez, et Tobias Höllerer. 2016. “Botivist: Calling Volunteers to Action Using Online Bots”. In *Proceedings of the 19th Association for Computing Machinery (ACM) Conference on Computer-Supported Cooperative Work & Social Computing*, 813-22. San Francisco California USA: ACM. https://doi.org/10.1145/2818048.2819985.

Schuller, Sebastian. 2021. “World Conspiracy Literature and Antisemitism”. *TRANSIT* 13 (1). https://doi.org/10.5070/T713153441.

Semetko, Ha, et Pm Valkenburg. 2000. “Framing European Politics: A Content Analysis of Press and Television News”. *Journal of Communication* 50 (2): 93-109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, et Filippo Menczer. 2018. “The Spread of Low-Credibility Content by Social Bots”. *Nature Communications* 9 (1): 4787. https://doi.org/10.1038/s41467-018-06930-7.

Shi, Wen, Diyi Liu, Jing Yang, Jing Zhang, Sanmei Wen, et Jing Su. 2020. “Social Bots’ Sentiment Engagement in Health Emergencies: A Topic-Based Analysis of the COVID-19 Pandemic Discussions on Twitter”. *International Journal of Environmental Research and Public Health* 17 (22): 8701. https://doi.org/10.3390/ijerph17228701.

Siroker, Dan, and Pete Koomen. 2013. *A/B testing: The most powerful way to turn clicks into customers*. John Wiley & Sons.

Stella, Massimo, Emilio Ferrara, et Manlio De Domenico. 2018. “Bots increase exposure to negative and inflammatory content in online social systems”. *Proceedings of the*

National Academy of Sciences 115 (49): 12435-40. https://doi.org/10.1073/pnas.1803470115.

Stieglitz, Stefan, Florian Brachten, Davina Berthelé, Mira Schlaus, Chrissoula Venetopoulou, et Daniel Veutgen. 2017. "Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with Those of Humans". In *Social Computing and Social Media. Human Behavior*, édité par Gabriele Meiselwitz, 379-95. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58559-8_30.

Subrahmanian, V.S., Amos Azaria, Skylar Durst, Vadim Kagan, Aram Galstyan, Kristina Lerman, Linhong Zhu, Emilio Ferrara, Alessandro Flammini, et Filippo Menczer. 2016. "The DARPA Twitter Bot Challenge". *Computer* 49 (6): 38-46. https://doi.org/10.1109/MC.2016.183.

Sunstein, Cass R., et Adrian Vermeule. 2008. "Conspiracy Theories". *Social Science Research Network* (SSRN) Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.1084585.

Uyheng, Joshua, et Kathleen M. Carley. 2020. "Bots and Online Hate during the COVID-19 Pandemic: Case Studies in the United States and the Philippine". *Journal of Computational Social Science* 3 (2): 445-68. https://doi.org/10.1007/s42001-020-00087-4.

Varol, Onur, Emilio Ferrara, Clayton Davis, Filippo Menczer, et Alessandro Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization". *Proceedings of the International AAAI Conference on Web and Social Media* 11 (1): 280-89.

Vosoughi, Soroush, Deb Roy, et Sinan Aral. 2018. "The spread of true and false news online". *Science* 359 (6380): 1146-51. https://doi.org/10.1126/science.aap9559.

Weinhardt, Michael. 2021. "Big Data: Some Ethical Concerns for the Social Sciences". *Social Sciences* 10 (2): 36. https://doi.org/10.3390/socsci10020036.

Williamson III, William et James Scrofani. 2019. "Trends in Detection and Characterization of Propaganda Bots". *Proceedings of the 52nd Hawaii International Conference on System Sciences*:7118-23.

Woolley, Samuel. 2018. "The Political Economy of Bots: Theory and Method in the Study of Social Automation". In *The Political Economy of Robots: Prospects for Prosperity and Peace in the Automated 21st Century*, édité par Ryan Kiggins, 127-55. International Political Economy Series. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-51466-6_7.

Woolley, Samuel C., et Philip N. Howard. 2016. "Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents — Introduction". *International Journal of Communication* 10 (0): 9.

Yang, Kai-Cheng, Emilio Ferrara, et Filippo Menczer. 2022. "Botometer 101: Social Bot Practicum for Computational Social Scientists". *arXiv preprint:1-14.* https://doi.org/10.48550/arXiv.2201.01608.

Yang, Kai-Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, et Filippo Menczer. 2019. "Arming the Public with Artificial Intelligence to Counter Social Bots". *Human Behavior and Emerging Technologies*1 (1): 48-61. https://doi.org/10.1002/hbe2.115.

Yang, Kai-Cheng, Onur Varol, Pik-Mai Hui, et Filippo Menczer. 2020. "Scalable and Generalizable Social Bot Detection through Data Selection". *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)Conference on Artificial Intelligence* 34 (01): 1096-1103. https://doi.org/10.1609/aaai.v34i01.5460.

Yang, Zachary, Anne Imouza, Kellin Pelrine, Sacha Lévy, Jiewen Liu, Gabrielle Desrosiers-Brisebois, Jean-François Godbout, André Blais, et Reihaneh Rabbany. 2021. "Online Partisan Polarization of COVID-19". In *2021 International Conference on Data Mining Workshops (ICDMW)*, 893-901. https://doi.org/10.1109/ICDMW53433.2021.00117.

Yu, Chung-En. 2020. "Emotional Contagion in Human-Robot Interaction". *E-Review of Tourism Research* 17 (5). https://ertr-ojs-tamu.tdl.org/ertr/index.php/ertr/article/view/561.

Yuan, Xiaoyi, Ross J. Schuchard, et Andrew T. Crooks. 2019. "Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter". *Social Media + Society* 5 (3): 2056305119865465. https://doi.org/10.1177/2056305119865465.

Appendix

APPENDIX A. Practical explanation to get users' Botometer score (Davis et al. 2016) and descriptive analyses:

Straightforward technical steps were performed in constructing users' bot scores. A Twitter Developer account was created to retrieve users' data to get the Botometer score. The Twitter Developer account is accessible to students, professionals, and academics. It contains a large set of tools used to manage access and collect tweets through the Twitter API. Application Programming Interface (API) is what enables connectivity and interactivity. In other words, API represents a " channel" (Twitter API) that will send requests to another API (Botometer) and returns a response. One of the main tools is the possibility to manage projects and apps that contain authentication keys and tokens provided to collect tweets or obtain a bot score for each user addressed through APIs.

After setting up a Twitter Developer account with the authentications and tokens, several lists of usernames (40,000 users separated into seven lists) were created on Python. These lists of usernames represent users that obtained a Botometer score. These lists are sent to the Rapid server API in JSON format. The Rapid API platform is the "channel" discussed in the previous paragraph. This platform is one of the world's largest API marketplaces, in which numerous other APIs, such as the Botometer API, are implemented and enable developers to access them. Furthermore, Rapid API helps developers to manage API rate limits and users' subscriptions (Yang et al. 2022, 6). Concretely, the Rapid API (channel) takes a request from a developer (a list of usernames) sent to a system (Botometer pro API), that will analyze the users and score them.

Afterward, the Rapid API returns the response to the developer on a programming software in JSON format. Additionally, every API has a rate limit since thousands of developers make daily requests. These limits usually help API owners to provide reliable and scalable APIs.[34] The Botometer API presents different plans to best match the developer's application needs. For this study, 50$ were purchased to have the Ultra Botometer Pro plan, which checked 17,280 accounts per day. **Table 1-A** presents the seven lists sent to the Botometer Pro API. Seven lists were created since the computer used could not handle the daily collection of 17,280 accounts.

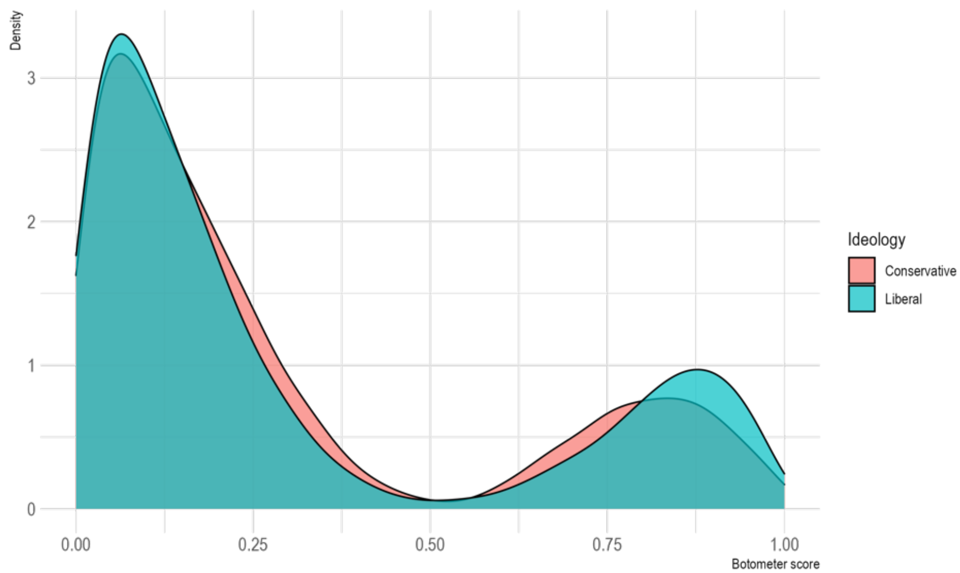
Table 1: (A) Chronology of the list of accounts (screen names) sent to the Botometer API through the Rapid API.

Dates	Lists	Number of accounts analyzed
9.02.2022	a	335
11.02.2022	b	908
14.02.2022	c	1,275
15.02.2022	d	6,582
16.02.2022	e	14,217
17.02.2022	f	3,669
18.02.2022	g	13,014

Turning to descriptive statistics, **Figure 1-A** presents the distribution of users discussing COVID-19 topics scored with the Botometer detection model. Most users have a low probability of having automated activities. Furthermore, the low number of users, around 0.5, is explained by the fact that the Botometer classifier has few confusing cases. Indeed, for example, if a bot user generates similar content to human users, the classifier will automatically classify this user with a score of 0.5 (Yang et al. 2022, 4).

[34] https://developer.twitter.com/en/docs/rate-limits

Figure 1: (A) The distribution of users with a Botometer score.



APPENDIX B. Validation of the performance of the Botometer score

This section presents the validation of the Botometer score with manual checking of 50 accounts. This step is essential to validate the model employed. Varol et al. (2017)'s work was considered to present a manual classification task. To manually validate that a user is a bot, there is no "simple set of rules to assess whether an account is a human or bot" (Varol et al. 2017, 283). Indeed, even though more than a thousand features are considered, it is difficult for a human coder to classify and give a score by looking at all of these parameters. However, they are few rules that are accepted when classifying an account as a bot, human or undecided (Varol et al. 2017, 283).

Several authors have looked at some specific features to manually code users (Varol et al. 2017; Khaund et al. 2022; Ferrara et al. 2016; Hegelich et Janetzko 2016) which form two distinct categories. On one side, numerous scholars emphasize the inspection of the Twitter profile of an account (Varol et al. 2017; Ferrara et al. 2016). On the other side, other works mention the inspection of the users' activities (Khaund et al. 2022; Ferrara et al. 2016; Hegelich et Janetzko 2016). For the first category, several indicators must be investigated from the Twitter profile of users such as the age of the account (at the time of the 2020 US presidential election), which is usually short, and the length of the username that is longer than human users (Ferrara et al. 2016). Furthermore, Varol et al. (2017) highlight the user image profile and background on Twitter. As an example, stock profile images are used by bot users (Varol et al. 2017). The second indicator includes different types of activities. Khaund et al. (2022) and Varol et al. (2017) mention that bots use a lot of retweets and few original tweets. Varol et al. (2017) add that a user may have a higher probability of being a bot if they retweet every message of another account within a second. Other features are considered such as the use of hyperlinks (URLs), ideas posted that are not supported in the majority of the discussion, and whether they have fewer followers and mentions than humans (Stieglitz et al. 2017). Lastly, the high share of false information is another feature of the probability of users being bots (Khaund et al. 2022). However, this last feature may be difficult to use since some conspiracy tweets have already been removed from the Twitter platform.

To validate the bot detection, 50 Twitter accounts were checked by dividing users' bot scores into two categories as suggested by Varol et al. (2017). First, 25 bot accounts with a score below 0.5 and 25 bot accounts with a score above 0.5 were randomly sampled, which results to 50 accounts being manually labeled. Users' profile and their activities were analyzed. As explained in Varol et al.'s (2017) paper, there is no precise set of instructions to classify accounts as bots or humans, but the investigation of the previous categories will help indicate the bot-like of users. Hence, accounts are classified as bots or humans for each category.

Table 2-B presents a Confucius matrix for each bot category investigated, and **Table 3-B** presents the accuracy results for bots and humans. The number of bots that are actually bots is 14, while the number of humans that are actually human is 25. In other words, the test set of bots has 56% of them that are true positives, and the test set of humans has 100% true positives has a precision of 1 since the number of accounts manually labeled was all actually human users. As suggested in Rauchfleisch et Kaiser's (2020) paper, even if the Botometer task is a well-known detection model that is vastly used in different fields, it may suffer from false positive identifications.

Table 2: (B) Botometer manual evaluation: Confucius matrix.

		Predicted	
		Bot	Human
True	Bot	14	0
	Human	11	25

Table 3: (B) Evaluation of the classification performance between the Botometer (Predicted) and the manual classification of bots (True) : (n=50).

	Precision	Recall	F1-score
BOT	0.56	1	0.71
HUMAN	1	0.69	0.81

APPENDIX C. Phi correlation table with the different sentiment models tested:

Table 4-C presents Phi coefficients of the two manual coding and sentiment models. The correlation between the **Manual Sentiment** and the **Stancov-19** is the strongest positive association with 0.76. Whereas the **Manual Stance** and the **Stancov-19** classification was lower with 0.44. This result is surprising since the highest association should have been between the **Manual Stance** and the **Stancov-19** since they should measure the same conception (in support or in opposition of COVID-19 measures). In other cases, the correlation coefficients remain low. Hence, the **Stancov-19** classification is used throughout the master's thesis due to its high association with **Manual sentiment** labeling.

Table 4: (C) Phi correlation coefficients between sentiment models (n=50).

	Manual Sentiment	Stancov-19	Vader model	Manual Stance
Manual Sentiment	1			
Stancov-19	0.76	1		
Vader model	0.09	0.24	1	
Manual Stance	0.18	0.44	0.15	1

APPENDIX D. Features considered in creating automatically users' bot score from the Botometer detection model presented in Yang et al.'s (2020) paper:

Table 5: (D) Features to get automatically users' Botometer scores.

user metadata		derived features		
feature name	type	feature name	type	calculation
statuses_count	count	tweet_freq	real-valued	statuses_count / user_age
followers_count	count	followers_growth_rate	real-valued	followers_count / user_age
friends_count	count	friends_growth_rate	real-valued	friends_count / user_age
favourites_count	count	favourites_growth_rate	real-valued	favourites_count / user_age
listed_count	count	listed_growth_rate	real-valued	listed_count / user_age
default_profile	binary	followers_friends_ratio	real-valued	followers_count / friends_count
profile_use_background_image	binary	screen_name_length	count	length of screen_name string
verified	binary	num_digits_in_screen_name	count	no. digits in screen_name string
		name_length	count	length of name string
		num_digits_in_name	count	no. digits in name string
		description_length	count	length of description string
		screen_name_likelihood	real-valued	likelihood of the screen_name

APPENDIX E. Tests and results of the stance detection model from Kawintiranon et Singh (2021).

Since the first purpose of the fifth hypothesis was to see whether users are supportive or against COVID-19 measures, another model of stance detection was examined. While a majority of the studies use sentiment models to depict users' opinions, one article from Kawintiranon et Singh (2021) looks at the stance of tweets related to Joe Biden and Donald Trump. Even though the stance detection only considering tweets discussing Biden or Trump, many tweets from their pre-trained data discussed COVID-19 related topics. Hence, this subsection attempted to reproduce their results.

Kawintiranon et Singh (2021) endeavor to automatically predict the stance of users discussing political manners during the 2020 U.S. presidential election with the Knowledge Enhanced Masked Language Model (KE-MLM) (Kawintiranon et Singh 2021). For each leading political actor (Biden and Trump), they manually code 2,500 tweets as supportive or opposing the candidates. Two random samples of 100 tweets from the master's project data were filtered to obtain test sets. A first test set included 50 tweets that mentioned the term "Biden" and another test set of 50 tweets that mentioned "Trump". I then manually look at two components for each set: the stance toward Donald Trump or Joe Biden, and the stance toward the COVID-19 measures. Three options were possible when labeling the tweets: supportive, opposite, or neutral.

While Kawintiranon et Singh (2021)'s work reported a high prediction accuracy, the accuracy was low when testing the model with the COVID-19 measures stance. The results show that the KE-MLM model does not perform well on the two test sets with the two manual labelings performed. **Tables 6-E et 7-E** present Pearson correlations between the two manual labelings (actual output) and the predicted output for both test sets (Biden and Trump test sets). We can see that the association is low between the manual labeling of the stance toward the candidates and the KE-MLM prediction. The association is also low between the manual labeling of the stance toward COVID-19 measures and the KE-MLM model. In both cases, the association is below 0.3.

Table 6: (E) Pearson correlations for tweets with the word "Biden". It indicates whether there is an association between the KE-MLM model given a stance toward j. Biden, the manual stance annotation toward J. Biden, and the manual stance annotation toward COVID-19 measures (n=50).

	KE-MLM detection	Manual Biden stance	Manual COVID-19 stance
KE-MLM detection	1		
Manual Biden stance	0.25	1	
Manual COVID-19 stance	0.29	0.63	1

Table 7: (E) Pearson correlations for tweets with the word "Trump". It indicates whether there is an association between the KE-MLM model given a stance toward D. Trump, the manual stance annotation toward D. Trump, and the manual stance annotation toward COVID-19 measures (n=50).

	KE-MLM stance	Manual Trump stance	Manual COVID-19 stance
KE-MLM stance	1		
Manual Trump stance	0.23	1	
Manual COVID-19 stance	0.05	0.58	1

Furthermore, both test sets' precision, recall, and F1 scores are reported. F1 score is a measure of accuracy on a data set. It combines precision and recall (Lipton et al. 2014). On one hand, precision is the number of true positives divided by the number of false positives plus the true positives. On the other hand, recall is the number of true positives divided by the number of true positives plus false negatives. The mathematical form is as follows:

$$F1\ score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

In **Table 8-E**, the overall accuracy for the test set of Joe Biden when considering the stance of the COVID-19 measures is 0.44. Only the F1 score for support toward the COVID-19 stance reaches 0.5. In **Table 9-E**, the overall accuracy for the test set of Donald Trump when considering the stance of the COVID-19 measures is almost inexistence, with an overall accuracy of 0.1.

Table 8: (E) Evaluation of the classification performance between the KE-MLM model (predicted) and the manual labelling related to COVID-19 stance (actual): (n=50) **Biden and COVID-19 stance**. **Overall accuracy: 0.44**

	Precision	Recall	F1-score
AGAINST	0.25	0.20	0.22
NONE	0.27	0.89	0.41
FAVOR	0.81	0.36	0.50

Table 9: (E) Evaluation of the classification performance between the KE-MLM model (predicted) and the manual labelling related to COVID-19 stance (actual): (n=50) **Trump and COVID-19 stance**. **Overall accuracy: 0.1**

	Precision	Recall	F1-score
AGAINST	0	0	0
NONE	0.20	0.71	0.31
FAVOR	0	0	0

To sum up, it is a hard task to detect users' stances toward Joe Biden, Donald Trump, and COVID-19 measures when the model is used in a different context. Indeed, here by changing the context (e.g., add COVID-19 discussions), the model struggles to correctly obtain the stance of each user. In other words, this may indicate that Donald Trump's support tweets are not linked to COVID-19's support tweets. The accuracy is better for the support of Joe Biden and COVID-19 measures.

APPENDIX F. Diagnostics test for H5. Normality and heteroskedasticity tests were performed.

The last hypothesis (H5) is tested with two lag models. The first one represents the impact of the proportion of bot users on human tweets sentiment, and the second model represents the impact of bot users' negative content on human tweets sentiment. Unrestricted standard Vector Autoregressions are estimated, and two diagnoses are performed: a normality and heteroskedastic test before testing the relationships.

The heteroskedastic test is a statistical test that analyzes volatility in time series[35]. It is assumed when the test statistic has a p-value below 0.05. Hence, the null hypothesis of homoskedasticity is rejected. In my analyses, the first model has a p-value less than 2.2e-16. Hence, the null hypothesis of homoskedasticity is rejected, and heteroskedasticity is assumed. The second model presents the same output. These results mean that they are systematic changes in the variance of residuals. Hence the regression output may not be reliable to interpret.

[35] https://www.rdocumentation.org/packages/aTSA/versions/3.1.2/topics/adf.test

The second test is the normality test. This test assesses whether the distribution is normally distributed.[36] Both models are not normally distributed since the p-value is far from 0 and it is below 0.05. Hence, it is essential to be cautious when interpreting the results from the last hypothesis.

APPENDIX G. List of conspiracy theory keywords and hashtags

Table 10: (G) List of keywords and hashtags related to conspiracy theories.

Hashtags and keywords found manually	Scamdemic ; Scamdemic2020 ; Scamdemic 2020 ; ScamdemicIsOver ; Scamdemic Is Over; Shamdemic; electioninfection ; election infection; Covidhoax; Covid hoax ; ConstitutionOverCoronavirus ; Constitution Over Coronavirus; Plandemic; Fakepandemic; Fake pandemic; Controlavirus ; Covid 1984 ; Covid1984; GreatReset; Great Reset ; TheGreatReset; The Great Reset ; CCPVIRUS; CCP VIRUS ; corona hoax; coronahoax; WhatCOVID; What COVID
Petersen et Gerken (2021)	WHOHoax; WHO Hoax; FakeCovid19; Fake
Chen et al. (2020)	Kungflu ; Kung flu
Dimitrov et al. (2020)	Coronials ; dr. fraud fauci

[36] https://www.rdocumentation.org/packages/vars/versions/1.5-6/topics/normality.test

APPENDIX H. List of positive, negative and neutral hashtags and keywords employed for the filtering of tweets and the Stancov-19 classification.

Table 11: (H) List of keywords and hashtags for the stance of users.

	Lockdown	Vaccine	Mask	Misc.
Neutral	Quarantine, secondlockdown, lockdownDC, californialockdown, 2ndLockdown, Lockdown3, lockdowns, coronavirusshutdown, covidlockdown, covidshutdown, shutdowns, coronaviruslockdown, lock down, Wuhanlockdown, lockdownextension, homequarantine, lockdownTrump	vaccination, vaccines, CovidVaccine, Covid19Vaccine, Covid19Vaccination, Astrazeneca, Moderna, Modernavaccine, Pfizer, Pfizervaccine, PfizerBionTech, BioNTech, Covaxin, Coronavaccine, SputnikV,CoronaVac, PfizerBioNTech, Operationwarpspeed, Warpspeed, CovidVaccination, Vaccinating, CoronavirusVaccines, Covisdhield, Astrazenecavaccine, Janssen, Janssenvaccine, JohnsonandJohnson, JohnsonandJohnsonvaccine, immunization, SputnikVVaccine, CoronavacVaccine, SinovacVaccine, Sinovac, Herdimmunity	mask, masks, facemask, facemasks, ppe, n95, kn95, CoronavirusMask, surgicalmasks, clothmasks, n95facemas, kn95facemask, facecover, brown, cloth mask, ffp2mask, ffp3mask, ffp3, ffp1, ffp2, kn95 mask, n95 mask, surgical mask, faceshield	SARSCoV2, Pandemic, COVID19, Covid_19, COVID, COVIDSecondWave, CovidCases, Covid19usa, Covidusa, COVID20, CovidReliefBill, COVIDReliefBill, COVIDRelief, COVID Relief, StimulusChecks, COVIDReliefPackage, 2000StimulusCheck, Virus, Regeneron, DrFauci, Secondwave, longcovid, covid19pandemic, CDCgov, FDA, GlobalPandemic, CovidVariants, coronaviruspandemic, Sars-cov-2, corona, coronavirusoutbreak, 2019ncov, COVID-19, coronavirus, CoronavirusUpdate, CoronaOutbreak, CDC, Epidemic, corona virus covd, oronavirusimpact, covid19 epidemic, covid19 pandemic, covid19update, covidlife, covidresearch, koronavirus, outbreak, pandemic2020, sars-cov-2 virus, centre for disease control, covidiot, covidiots
Positive	Stayhome, StayHomeStaySafe, lockdownlife,StayHomeSaveLives,nationallockdown, TogetherAtHome, Prolockdown, Proshutdown, LockdownWorks, AvoidGatherings, stay home challenge, safe at home, stay at home, stay home, sheltering in place, quarantine life, 14DayQuarantine,inmyquarantinesurvivalkit, quarantine shelter, shelteringinplace, stay home challenge, stayathome, stay_home_safe, stayhometosavelives, workfromhome, stayhome!, saferathome, safe at home	VaccinesSaveLives, ThisIsOurShot, Vaccinated, getvaccinated, BidenVaccine, Trumpvaccine, VaccinesWork, VaccinationWorks, TakeYourShot, BestShot, vaccinationdone, CovidVaccineFacts, SecondDose, GetYourShot, VaxUp, VaccineSelfie, CrushCOVID, MyCOVIDVax, IGotTheShot, VaccinesAreSafe, vaccine against coronavirus	masksSaveLives, wearamask, maskup, wearadamnmask, MaskYourKids, MaskMandates, WearMaskProtectLife, Wear a mask protect a life, WearAMaskSaveALife, maskon, Doublemasking, Doublemask, MaskOnAmerica, MaskSelfie, MasksWork, MaskWorks, mandatorymask, GetMePPE, masks4all, wear face mask, CoronavirusCoverup	FlattentheCurve, StopTheSpread, crushthevirus, BidenWillCrushCovid, StopCOVID19, FauciHero, COVIDWise, Protectyourself, Protectothers, GetTested, BreakTheChain, washyourhands, fightagainstcorona, Socialdistancing, Social Distancing, Social Distancing Now, Dont be a spreader, don't touch your face, fightagainstcorona, fightagainstcoronavirus,fightcoronatogether, fightcovid19, fighttogether, slow the spread of covid19, standtogether, treat coronavirus, wehealasone, uniteagainstcovid19, togetherwecan, togetherwecandoit

	Lockdown	Vaccine	Mask	Misc.
Negative	endthelockdown, endlockdowns, NoShutdown, NoMoreLockdown, NoMoreShutdown, ReopenAmerica, OpenAmericaNow , Antilockdown, LockdownsKill, Breakthelockdown, LockdownsAreNotACur e, nolockdown2, BreakTheLockdowns, LockdownChaos, LockdownsDontWork, StopThelockdowns, LockdownFraud, Bidenlockdown, NoMoreLockdowns, CancelTheLockdown, freetheUSA2020, NoLockdowns, NoLockdown, Antishutdown, Anti shutdown, endtheshutdown	NoVaccine, MedicalFreedom, Medical freedom, AstrazenecaKills, Astrazeneca kills, AntiVaccine, AntiVacc, AntiVaxx, KnowtheRisk, BewaretheNeedle, FuckAstrazeneca, FuckPfizer, PfizerKills, FuckModerna, ModernaKills, FuckJohnsonandJohnson, JohnsonandJohnsonkills, Deathbyvaccine, VaccinesKill, NeverVaccine, SayNoToVaccines, hydroxycloroquine, saynobillgatesvaccine	maskdontWork, Nomasks, Nomask, MasksOff, MaskOff, antimasker, antimaskers, NoMaskMandate, Nomoremasks, UnMaskAmerica, Maskless, IWillNotWearAMask, SheepNoMore, unmask, MaskOffAmerica, Talesoftheunmaskedpatriot, takeoffthemask, maskburning, Burnyourmask, Burnyourmaskchallenge, nomaskonme, nomaskselfie, maskhoax, nomaskEVER, facefreedom, masksmakemesweaty, MasksAreDangerous, TakeMaskOff, Stopforcingmaskonme, takeoffyourmask, refusemask, NeverMasker, StopWearingMask, StopWearingtheDamnMasks, MasksdontMatter, stopmasking, stopthestupidmask, maskingchildrenischildabuse, MomsAgainstMasks, MasksRUnhealthy, SheepWearMasks,	Scamdemic, Scamdemic2020, ScamdemicIsOver, Shamdemic, electioninfection, Covidhoax, ConstitutionOverCoronavirus, chinesevirus, chinavirus, Plandemic, Fakepandemic, TrumpPandemic, TrumpCovid, TrumpCovid19, Controlavirus, Covid1984, TrumpCovidHoax, GreatReset, TheGreatReset, CCPVIRUS, TrumpVirus, TrumpVirus2020, wuhanvirus, WuhanFlu, coronahoax, AmyCovidBarrett, TrumpVirusCatastrophe, SuperSpreaderEvents, TyphoidTrump, WhatCOVID, TrumpPlague, OmnibusCovidReliefBill, WHOHoax, FakeCovid19, KongFlu, Kungflu, Wuhancoronavirus, Trump pandemic, coronapocalypse, china virus, coronials, dr. fraud fauci