



COMMITTEE
ON
INTERNATIONAL
RELATIONS
THE UNIVERSITY OF CHICAGO

The University of Chicago

**Training A Soothsayer: A Study on the Efficacy and Accuracy of A
'ChatGPT Approach' to International Relations**

by Daniel Robert Jones

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the
Master of Arts Program in the Committee on International Relations

Faculty Advisor: Dr. Molly Offer-Westort

Preceptor: Dr. Burcu Pinar Alakoc

Acknowledgments

The completion of this study would not have been possible without the scholarly, expert guidance and advice of both my Thesis Advisor, Dr. Molly Offer-Westort of The University of Chicago's Department of Political Science, and my Thesis Preceptor, Dr. Burcu Pinar Alakoc, of The University of Chicago's Committee on International Relations.

Abstract

This study investigates the effectiveness of using a language model, ChatGPT, for predicting the number of political violence events in Brazil. The study utilized data from a publicly available dataset on political violence events in Brazil and COVID-19 data to train and test two models of random forests, one created manually, and the other generated by ChatGPT. The results show that the ChatGPT model had a slightly higher root mean squared error than the manually generated random forest model, but the difference was small. Furthermore, the density plot of the predicted number of events from the two models showed that the manual model had a higher density around the median, and a lower overall density range, suggesting that it may be better suited for predicting average or typical events. The study also explored the relationship between COVID-19 cases and deaths and the number of political violence events, finding a somewhat positive correlation between the two variables, which the study touches on briefly. These results suggest that the COVID-19 pandemic may have impacted political violence events in Brazil, and further research is needed to explore this relationship. Overall, the study shows that ChatGPT can be an effective tool for predicting the number of political violence events in Brazil and has the potential to be used in other areas of international relations, such as treaty creation, translation, negotiation assistance, and code-generating prediction. The study is to my knowledge, the first of its kind in analyzing the possibilities of AI and Chatbots, in international relations.

Introduction

Recently, there has been a growing interest in utilizing machine learning models in many IR research and policy-writing-centric fields. This is particularly true once expanded to the area of social sciences, where researchers have started exploring the use of machine learning algorithms to understand better and predict complex social phenomena. One of the most promising machine learning models is the Generative Pre-trained Transformer (GPT), developed by OpenAI. GPT is a neural network architecture capable of generating human-like text and has achieved state-of-the-art results in a wide range of natural language processing tasks.

To explore a generalizable example that also demonstrates the potential of GPT as an effective or non-effective tool for IR researchers, this study focuses on the training and testing of GPT-built random forests, which uses the W.H.O daily COVID-19 2020-2021 case tracker and The Armed Conflict Location and Event Data (ACLED) 2020-2021 political violence data, both in Brazil. It is important to note that while both datasets are compiled in partnership with the W.H.O and ACLED respectively, each is done with sub-national and regional partners, such as multiple government agencies and independent researchers located in Brazil. More specifically I would like to give both credit and thanks to Dr. Wesley Cota, who is a physicist, FAPESP Postdoctoral Fellow at IMT-USP, and Volunteer Researcher at FMB-Unesp, and whose voluntary management, collection, and tidying, of Brazil's COVID-19 tracking data since the first Brazilian cases were confirmed in March 2020, allowed this project to move more effectively. The datasets he manages utilize W.H.O. subnational Brazilian government data available for cases and deaths at the municipality level, with official information given by Ministério da Saúde, and official federative units reports via Brasil.IO. In contrast, the ALCED dataset offers

detailed information on political violence events, including their location, date, type, and actors involved.

As stated previously, this paper is the first to explore the role of ChatGPT-style AI and user-friendly chatbots in International Relations research and policy-making. Thus, it is crucial to first examine the different theoretical capabilities of ChatGPT outside of simply producing machine learning algorithms and code. Below I list some of the theoretical uses which I briefly explore the possibilities of, for using ChatGPT in International Relations, emphasizing both the researcher and the policy-maker, respectively. The five core examples of how ChatGPT can be used are as follows: 1) Natural Language Processing (NLP) for Analyzing International Relations Texts, 2) Machine Translation for Diplomatic Communication, 3) Predictive Analytics and Coding for Conflict and Crisis Management, 4) Writer/Producer Systems for Foreign Policy Documents, and 5) Conflict Resolution and Multilateral Negotiations.

After that, I posit each theoretical use in the case study of Brazil in 2020-2021 to best see how each suggestion might look once used by a researcher in a specific historical context of both political (Brazilian presidential election and mass protests) and health-centered (the outbreak of COVID-19) crises. The article then goes even deeper into the example of using GPT-produced code to build random forests as a benchmarking prediction exercise in the same subnational two-year context mentioned above. The brief theoretical examples along with the core benchmarking exercise is performed using the latest ChatGPT update, GPT 4.0. First, however, it is important to note that upon initiating this study, I utilized GPT 3.5, as that was the initial update that was available. Some of the contextual changes between both updates are discussed in the discussion section of the article.

It is best to understand this article as accomplishing two things: 1) creating a framework and generative/generalizable basis on which future IR studies may rely and point reference to when expanding on how chatbots, ChatGPT, or machine learning processes may be utilized in the subfields of IR, and 2) performing a benchmarking exercise to compare both the coding and predictive methodological capacity differences of researcher-constructed and ChatGPT-constructed random forests to see which performs better, if any. My hypothesis along with the null, is presented here:

H1: Manually generated random forest models are more effective and more accurate than random forest models generated by ChatGPT.

Null: Manually generated random forest models perform no better than random forest models generated by ChatGPT.

While these represent the crux of the study, it is still important to reemphasize that regardless of the outcomes, this study has important implications for the future study of IR. The above examples are just a few examples of how ChatGPT can be used in international relations; combined, they represent where I see the most potential both in IR research and policymaking. The potential applications of ChatGPT in this field are vast and depend on the specific research questions being asked, along with the already frequent updates on ChatGPT software done by the developer teams at OpenAI. Overall, the potential benefits of such a tool are numerous. For example, governments and humanitarian organizations could use it to anticipate political

violence events and take proactive measures to prevent them. Moreover, it could help researchers better understand political violence's complex dynamics and contribute to developing more effective conflict prevention and resolution strategies. It could also allow stakeholders to better bridge the divide that often entrenches a dichotomy between academic researchers and policymakers, proving (or disproving) the accuracy of a user-friendly but still rigorous research tool. The remainder of the article is structured as follows: I critique and overview varied literature on both random forest, other types of machine learning in social sciences, and ChatGPT; I discuss the reasoning for positing this benchmark in the case of Brazil, 2020-2021, I lay out a thorough methodology of the benchmarking exercise, I discuss the results, and finally I discuss how scholars and policymakers may utilize this paper as a starting point for their decision to integrate, study, or engage critically with both random forests and chatbots as tools in IR. This research represents a novel and innovative approach to predicting political violence events and has the potential to contribute to the social sciences field significantly. I hope the findings encourage further research and stimulate new ideas for developing machine-learning models and user-friendly chatbot technology to help us better understand and prevent political violence.

Literature Overview and Theoretical Foundations

ChatGpt and OpenAI:

Over the past few years, there has been a growing interest in using aspects of artificial intelligence and machine learning in social science research. For example, one application uses language models like ChatGPT and OpenAI to predict future events based on historical data.

More specifically, there has been an increase in the use of these language models to indicate a range of social phenomena, including political violence.

ChatGPT is a large-scale neural language model that uses a transformer-based architecture to process and generate natural language. It was trained on a vast amount of text data, including books, articles, and websites, using unsupervised learning techniques. The training process involved predicting the next word in a sentence, given the previous words as input. As a result, ChatGPT has learned to capture the statistical patterns and dependencies of language at a very high level, allowing it to generate coherent and plausible text based on a given prompt. When using ChatGPT for NLP tasks, such as answering questions or generating text, the input prompt is first tokenized and encoded as a sequence of integer values that represent the words and subword units in the input text. This sequence is then fed into the transformer architecture, which consists of multiple layers of self-attention and feedforward neural networks. During the self-attention process, the model calculates a weighted sum of the input sequence, where each weight corresponds to the importance of each word or subword for the current prediction. This allows the model to attend to relevant information in the input sequence and ignore irrelevant or noisy information. After several layers of self-attention and feedforward processing, the model generates a probability distribution over the vocabulary of words and subwords, representing the likelihood of each token being the next one in the generated sequence. The next token is then sampled from this distribution, and the process is repeated until the desired length of the generated text is reached. The generated text is then decoded back into natural language using a pre-defined vocabulary of words and subwords. In the context of the Brazilian case, when using ChatGPT to generate a treaty focused on trade between the United States and Brazil, the input

prompt, for example, would consist of the negotiation goals and constraints of both parties, as well as any relevant background information on the topic. The model would then generate a text that satisfies the constraints and goals of both parties, while also ensuring that the resulting treaty is legally and politically feasible (based on the most recent/public historical and legal data it has access to, online). The generated text would need to be reviewed and revised by human experts to ensure its accuracy and quality, but ChatGPT can serve as a valuable tool to assist human negotiators in the process.

To blend fields similar to this paper, one study exploring machine learning algorithms' use in predicting political violence is that of Ruiz and Murdock (2020). Here the authors used a dataset of terrorist attacks in Mexico and employed a random forest algorithm to predict future terrorist attacks based on the historical data on which the models were trained and tested, similar to what occurs in this paper. Their findings effectively showcase why I selected random forest for the brief case study here with ChatGPT-produced models; random forests have been well proven to effectively predict seemingly sporadic events with many multifaceted variables included, all with high accuracy. The Ruiz and Murdock (2020) study, like many others, concluded by suggesting that random forests can effectively move even beyond predicting attacks of a single nature, all the way to predicting the probability of varieties of civil unrest, as they argued that policymakers could use such models to address potential sources of unrest proactively as well. In a recent study by Wang et al. (2021), the authors used a deep learning model to predict political instability in various countries based on social media data. Similarly, their findings showed that the deep learning model accurately predicted political instability across multiple case studies.

While the studies above showcase rigorous methodological validity, and legitimate reasoning for selecting random forest as the core benchmarking exercise when studying the nature of a new chatbot such as ChatGPT, they do not engage with any AI or chatbot technologies to do so. However, others have explicitly focused on using language models like ChatGPT and OpenAI to predict social phenomena. For example, in a study by Liu et al. (2020), the authors used the GPT-2 language model to predict the risk of suicide based on social media data. Their study showed that the language model effectively predicted the risk of suicide. Furthermore, they argued that such models could identify at-risk individuals and provide targeted interventions. These studies demonstrate the potential of machine learning algorithms and language models in predicting social phenomena, including political violence. However, there is a need for further research in this area, particularly in exploring the use of these models in predicting political violence in different contexts and using other datasets.

Machine Learning and CounterTerrorism; Lessons Gleaned from Sentiment Analysis

In creating a survey of the literature on such a topic, the first thing that comes to mind is the current lack of praxis to blend ChatGPT with security and violence/terrorism prevention. Two particularly innovative and successful studies that have sought to do just this, but without the input of AI features, are Xiaohui Pan's 2021 study: Quantitative Analysis and Prediction of Global Terrorist Attacks Based on Machine Learning, and Steven Krieg et al.'s 2022 study: Predicting terrorist attacks in the United States using localized news data. While this study does not utilize sentiment analysis, there is still relevant scholarly literature that can inform the approach and interpretation of the results. Specifically, research on sentiment analysis and its

application in counterterrorism efforts can provide insights into the role of emotions in the perpetration of violent acts and the potential for sentiment analysis to identify individuals or groups with extremist ideologies. Sentiment analysis follows similar data management and tidying as I do in this study, and is far more proven in working with random forests. Thus, similar notes on previous literature's success and shortcomings in the field of sentiment analysis are incredibly helpful in informing this study. Pan's article utilizes machine learning to evaluate and predict specific characteristics of future terrorist attacks. In contrast, Krieg's piece uses machine learning to predict spatiotemporal attributes of terrorist attacks based on localized news data and sentiment analysis. Both of these novel studies do well to evaluate many different similar prediction algorithms and models to see which is the most successful. In similar notation, they both do pretty well in presenting their shortcomings.

While terrorist event prediction has rarely been practiced with similar methods, several prior works address the problem of detecting terrorist activity, such as classifying pro-terrorism tweets and other related services outside of predictive measures (Abrar, 2019). However, it is important to note that while such scholarship is helpful and novel, I am interested in the problem space of prediction rather than detection (which most former scholars focus on). More specifically, my primary task is to predict the occurrence of attacks or events, which is distinct from works that have sought to infer characteristics of an attack, such as solely the responsible group, after it has taken place, for example (Tolan, 2015).

Few works have attempted to predictively model terrorist attacks specifically based on one primary machine learning method, however, with the recent work of Python et al. being one novel work and exception (2021). It is an incredibly new and exciting field of study, but it comes

with a highly variant level of success and efficiency when turning to previous studies on which to build. We can see, however, that the authors in the example of Python et al. trained several machine learning models using primary terrorism data (from the GTD) in conjunction with other geographic and socioeconomic features to predict attacks at discrete spatiotemporal intervals. Krieg et al. (2022) then built on the former by including sentiment analysis and identifying more effective means of definition of terrorism feature classification methods than was previously possible. In turn, finding ways to combine said tested methodologies. Pan (2021), on the other hand, focused more on evaluating several different classification models for terrorist attack feature detection.

Other Examples of Machine Learning in The Social Sciences

Sentiment analysis, similar to the machine learning process used by this study, is a technique that uses natural language processing and machine learning to identify and extract subjective information from textual data, including social media posts. More specifically, sentiment analysis is a computational technique that aims to identify, extract, and quantify the emotions expressed in text data as a whole and is a machine learning system connected to predictive measures of violence frequently in recent scholarship (Chen, 2016). The technique involves classifying the sentiment of the text into positive, negative, or neutral categories. The theory behind sentiment analysis for predicting terrorist attacks is that individuals planning or contemplating terrorist activities may express negative sentiments towards certain groups or institutions, which can be detected through social media monitoring (Krieg et al., 2022). Moreover, Social media platforms have become a critical tool for disseminating information related to terrorism, with users sharing

their thoughts, opinions, and plans related to terrorist activities (Pan, 2021). This is important to note because the literature on sentiment analysis needs to be more comprehensive and all relatively recent.

Several studies have used sentiment analysis along with other types of machine learning with social media posts to predict events in real-time, but it becomes much more difficult when predicting terrorist attacks (Sudhahar et al., 2017). For example, Chen et al. (2018) used sentiment analysis to monitor social media posts related to the 2015 Paris terrorist attacks. They found that negative sentiment toward France increased significantly in the days leading up to the attack. However, it became increasingly difficult for Chen's (2016) study to dissect the results so that sentiment could predict the attack, its details, and the threshold for it to do this. O'Loughlin et al. (2018) also used sentiment analysis to monitor social media posts related to the 2016 Berlin Christmas market attack. They found that negative sentiment towards the German government and police increased in the days before the attack. In a study of the 2017 Manchester bombing, Castells et al. (2018) found that negative sentiment towards the United Kingdom also increased in the hours before the attack. Other studies have focused on monitoring social media conversations related to specific groups or ideologies, such as Islamic extremism, and have identified patterns of negative sentiment that may indicate future terrorist activity (Sudhahar et al., 2017).

Another similar example is Guo et al. (2018), whose study summarized three ways to improve modern conflict forecasting and even called on the UN to invest in data-driven predictive methods for promoting peace. As a result, some of the three new methods developed by reference to the Guo et al. (2018) study included new machine learning techniques and

theoretical models that better reflect human decision-making. However, I am, to the best of my knowledge, the first to use machine learning to predict the occurrence of individual terrorist attacks in the United States using social media posts while building upon this multifaceted and varied literature.

While sentiment analysis of social media posts shows promise as a tool for predicting terrorist attacks, it is not without limitations as well. One of the main challenges found in past literature examples is the noise in social media data, making it difficult to identify relevant sentiment patterns Guo et al. (2018). Additionally, the technique is subject to bias, as the algorithms used to classify sentiment may be influenced by cultural and linguistic differences, all of which may need to be monitored by the research closely depending on the machine learning model being used (i.e., supervised or semi-supervised) (Agarwal et al., 2020). Finally, when it comes to sentiment analysis specifically, the technique relies heavily on users expressing measurable negative sentiment in their social media posts, which may not always be accurate in the lead-up to a terrorist attack (Pan, 2021). This becomes even more difficult when scholars debate whether written sentiment analysis versus video-tracked sentiment analysis is better and more accurate than one (Pan, 2021).

Nevertheless, sentiment analysis of social media posts is a promising and quite similar machine learning approach for predicting terrorist attacks, as it offers a way to monitor public sentiment towards certain groups or institutions. They are important findings for successfully integrating machine learning approaches to social sciences issues. However, as I began, the technique is still subject to several limitations, including noise in social media data, bias in sentiment classification algorithms, and the reliance on negative sentiment expression is often assumed.

However, overall, the literature on sentiment analysis and counterterrorism highlight the potential for using linguistic analysis to identify extremist content and individuals with extremist ideologies. While this study does not utilize sentiment analysis, the machine learning-based counterterrorism literature findings can still inform the approach and interpretation of the results. In addition, they may inspire future research on the intersection of sentiment analysis and political violence.

Creating a ‘ChatGPT Approach’ for International Relations

International Relations is a complex and ever-changing field encompassing many topics, including diplomatic communication, conflict and crisis management, foreign policy documents, and others. Recent technological advances and natural language processing (NLP) have created new opportunities for researchers and policymakers to analyze and understand these complex issues. ChatGPT, a large language model trained by OpenAI, is one such tool that has the potential to revolutionize the way policymakers and researchers approach these topics. The first topic mentioned in the introduction, NLP for analyzing international relations texts, refers to using computational methods to analyze and understand the vast amounts of written and spoken language produced in international relations. This includes everything from news articles and academic papers to government reports and speeches. With the help of NLP tools like ChatGPT, researchers can analyze these texts to identify key themes, trends, and patterns, which can then be used to inform policy decisions and shape an understanding of international relations. Machine translation for diplomatic communication is another area where ChatGPT could prove to be an invaluable tool. As the world becomes increasingly globalized, diplomats and other

officials must communicate with people who speak different languages. Machine translation, which uses algorithms to translate text from one language to another, can help to bridge this gap. ChatGPT's advanced language processing capabilities could be used to accurately translate diplomatic communications, reducing the risk of miscommunications and misunderstandings. Predictive analytics and coding for conflict and crisis management are other areas where ChatGPT could be useful. By analyzing large amounts of data and using machine learning algorithms, ChatGPT can identify patterns and trends that may indicate the likelihood of conflict or crisis. This information can then inform policy decisions and develop strategies for preventing or managing such situations.

Writer/producer systems for foreign policy documents are another area where ChatGPT could be effective. These systems use AI and NLP to assist with the creation of foreign policy documents, such as speeches, reports, and briefing papers. With the help of ChatGPT, policymakers and researchers could quickly and easily generate high-quality documents informed by the latest data and analysis. Finally, conflict resolution and mediation are areas where ChatGPT could be used effectively. By analyzing language and other communication data from parties involved in a conflict, ChatGPT could help mediators to identify areas of agreement and disagreement and to develop strategies for resolving the conflict. With its advanced language processing capabilities, ChatGPT could be an invaluable tool for promoting peace and stability in regions worldwide.

The potential applications of ChatGPT in international relations are vast and varied. From analyzing international relations texts to facilitating diplomatic communication to predicting and managing conflicts, ChatGPT could be an effective tool for policymakers and researchers looking to make sense of the complex and rapidly changing world of international relations. With

its advanced language processing capabilities and machine learning algorithms, ChatGPT has the potential to revolutionize how I approach these important issues, helping promote peace, stability, and understanding on a global scale. However, to lay the theoretical groundwork for testing such possibilities, these five options represent the multifaceted tools ChatGPT can serve for various IR actors. Therefore, before laying out the more in-depth benchmarking exercise, it is important to provide examples of the previously mentioned theoretical usages of ChatGPT in IR. A brief overview of examples, all demonstrated using the same 2020-2021 Brazil example as the benchmarking exercise, is detailed later as I discuss the layered reasoning for choosing such a case to suggest and test some of these uses. Finally, below is a full listing of all five theoretical uses mentioned previously:

1) Natural Language Processing (NLP) for Analyzing International Relations Texts:

ChatGPT can analyze international relations texts, such as news articles, speeches, and social media posts. By training ChatGPT on a large corpus of texts, it can identify key themes, trends, and sentiments related to international relations issues.

2) Machine Translation for Diplomatic Communication: ChatGPT can be used to develop machine translation tools that aid in diplomatic communication. This could include translating official documents, speeches, and diplomatic cables between countries.

ChatGPT can be trained on large datasets of bilingual texts to improve the accuracy and fluency of machine translation.

3) Predictive Analytics and Coding for Conflict and Crisis Management: ChatGPT can be used to develop predictive analytics models for conflict and crisis management in various systems/platforms. Training models on historical data about conflict and crisis events using ChatGPT can help identify patterns and predict potential future events. Moreover, it is possible that it can significantly reduce time and human cost to conduct projects such as this and, in turn, decrease the response time of global policy-makers working in reaction to political violence/crisis events. Policymakers and international organizations can use this information to prevent them or manage conflicts and crises faster and more effectively. This is the example I will examine with the case study in this paper.

4) Writer/Producer Systems for Foreign Policy Documents: ChatGPT can be used to develop writer/producer systems that can aid in foreign policy document generation between international actors, specifically nation-states. Training ChatGPT on historical data about foreign policy decisions can suggest potential policy-writing options based on specific goals and constraints. This could include trade agreements, diplomatic relationships, or military alliances.

5) Conflict Resolution and Multilateral Negotiations: ChatGPT can be used to analyze historical conflict data and provide recommendations on resolving conflicts. Moreover, it could be trained as a mediator and proxy diplomatic advisor between nation-states and other international actors. Especially now, as many diplomats deal with the after-effects of COVID-19 travel constraints, ChatGPT could help multilateral

communications continue globally and benefit diplomacy overall. For example, it could serve as a mediator between international actors or be used in other formats during discussion and deliberation between international actors on the root causes of conflicts or potential avenues for the least miscommunication possible.

Why Brazil? From Theory to Practical Application

The context of 2020-2021 in Brazil presents several challenges and opportunities for studying the intersection of political violence and disease outbreaks, making it an ideal case study for testing the effectiveness of ChatGPT as a research and policy tool in the field of International Relations. Brazil has been hit particularly hard by the COVID-19 pandemic, with over 15 million confirmed cases and over 400,000 deaths as of May 2021. This makes Brazil a particularly important case study for understanding the impact of disease outbreaks on political stability and conflict. By analyzing the W.H.O.'s COVID tracking data for Brazil, we can gain valuable insights into how the pandemic has affected different regions and communities and the political and social responses to the crisis. The ACLED Political Violence in Brazil Data (2020-2021) and W.H.O. COVID tracking data from Brazil (2020-2021) provide an excellent case study for testing the efficacy of ChatGPT as a researcher's and policymaker's tool in international relations. Brazil has a history of political violence, particularly in recent years. The ACLED Political Violence in Brazil dataset for 2020-2021 provides a detailed record of incidents of political violence in the country during this period, including the actors involved, the types of violence used, and the geographic locations of the incidents. By analyzing this data alongside the COVID tracking data, we can gain even further insights into how political violence and disease outbreaks

interact and shape each other and how the responses to one can impact the other. At the same time, Brazil was also grappling with a surge in political violence, particularly during the 2020 municipal elections. ACLED data reveals that Brazil experienced increased political violence during this period, with over 5,000 political violence events reported in 2020 alone. The increase in political violence occurred in a polarized political climate, with tensions running high between supporters of President Jair Bolsonaro and his opponents.

The sheer availability of public, accessible, and accurate data, in this case, was ample reason to choose it as the case study. In generating an article meant to detail possible frameworks and possibilities for such new methods, it was crucial that the amount of data used was both vast and accurate. In addition to these factors, using ChatGPT as a research and policy tool has several potential benefits in this case study. For example, ChatGPT can rapidly analyze large and complex datasets, allowing researchers and policymakers to identify patterns and relationships that might not be immediately apparent using traditional methods. Furthermore, ChatGPT can interact with users naturally and intuitively, allowing researchers and policymakers to ask complex questions and receive nuanced answers without extensive technical expertise or specialized training. Overall, the combination of Brazil's political context, the impact of the COVID-19 pandemic, and the use of ChatGPT as a research and policy tool make this case study an ideal opportunity to explore the intersection of political violence and disease outbreaks in the field of International Relations.

By analyzing large datasets and generating predictions, ChatGPT can help identify patterns and trends in political violence and disease outbreaks that might otherwise be difficult to discern.

ChatGPT can also help researchers and policymakers generate and test hypotheses, leading to

more informed policy decisions. It can provide a cost-effective and efficient way to analyze large datasets, as it can quickly process large amounts of information and generate predictions based on that data. This is particularly valuable in a rapidly changing environment like Brazil in 2020-2021, where new data was generated daily, and on a spatio-temporal, often granular scale. It is important to note that this was also a large factor in initially picking Brazil. In order to best choose a case suited to both random forests and testing the different abilities of ChatGPT, it was once again incredibly important to choose a case with immense amounts of historical data sources available publicly, and this should be noted for future studies.

Random forests are an ensemble learning method that builds many decision trees at training time and outputs the class, the mode of the classes (classification), or mean prediction (regression) of the individual trees. Random forests have been widely used in various fields, including social sciences, due to their ability to provide highly accurate predictions and handle large datasets. In the case of Brazil's political violence and COVID-19 data, using random forests could provide interesting and meaningful results in several ways. First, the datasets are relatively large, covering various variables related to political violence and COVID-19 cases across several regions and time frames. Random forests can efficiently handle such large datasets and capture complex, nonlinear relationships among variables, making identifying significant predictors of violence and COVID-19 cases easier.

Political violence and COVID-19 data are likely highly correlated in some manner, given the socio-economic and political factors underlying both phenomena. Random forests can capture such complex relationships, which may not be easily identified using traditional regression or correlation analysis. This would provide a more comprehensive understanding of the factors

driving political violence and COVID-19 cases in Brazil, potentially highlighting areas where policy interventions could be most effective.

The use of random forests in the case of Brazil's political violence and COVID-19 data could also provide interesting insights into the dynamics of the relationship between these two phenomena. For instance, it may be possible to identify a causal relationship between political violence and COVID-19 cases or if they are both driven by a common set of underlying factors. By identifying such relationships, policymakers and researchers could develop more effective strategies to mitigate the impact of both political violence and COVID-19 on society. Using random forests on Brazil's political violence and COVID-19 data can provide interesting and meaningful results. The next section walks all each of the theoretical examples mentioned previously, other than the benchmarking exercise, which comes after. It is important to note that each response from ChatGPT listed below is taken exactly as ChatGPT responded, without any formatting, spelling, grammatical, or content changes. In order to keep this article readable, only the first portions representative of each response are here.

Theoretical Examples Applied to The 2020-2021 Brazilian Context

1) Natural Language Processing (NLP) for Analyzing International Relations Texts:

Researcher Prompt: Please use NLP to analyze the following speech from former Brazilian President Bolsonaro in 2020: G20 Summit Speech by Jair Bolsonaro at Planalto Palace.

ChatGPT Response: Sentiment Analysis: Overall, the sentiment of the speech appears to be positive. The language used throughout the speech is largely neutral, but the sentiment

analysis indicates that more positive than negative words are used. In addition, the speech contains many positive references to Brazil, its people, and its culture, which suggests a sense of national pride and patriotism. Topic Modeling: The topic modeling analysis indicates that the main themes of the speech are unity, national identity, and the global response to the covid-19 pandemic. The speech contains several references to the importance of a united people and the need to defend national identity against outside forces. It also discusses the global response to the pandemic, including efforts to combat it and the economic impact of the crisis. Key Phrases: Some of the key phrases that stand out in the speech include "unity," "national identity," "global pandemic," "economic recovery," and "WTO reform." These phrases help summarize the speech's main themes and provide insight into the speaker's priorities and concerns. Overall, the NLP analysis of the speech suggests that the speaker is focused on promoting unity and defending Brazil's national identity while recognizing the importance of global cooperation in response to the covid-19 pandemic. The speech also emphasizes the need for economic recovery and WTO reform to promote fair and balanced international trade.

2) Machine Translation for Diplomatic Communication:

Researcher Prompt: Please use machine translation to translate the previous speech from former Brazilian President Bolsonaro in 2020: G20 Summit Speech by Jair Bolsonaro at Planalto Palace, from English, into Portuguese.

ChatGPT Response: Senhoras e senhores, Antes de entrar no tema principal desta sessão, gostaria de fazer uma breve defesa da identidade nacional brasileira diante das

tentativas de importar para o nosso território tensões que são estranhas à nossa história. O Brasil possui uma cultura diversa, que é única entre as nações. Somos um povo miscigenado. Brancos, negros e índios construíram o corpo e o espírito de um povo rico e maravilhoso. Em uma única família brasileira, podemos contemplar uma maior diversidade do que em países considerados como um todo. Foi a essência desse povo que conquistou a simpatia do mundo. No entanto, alguns querem destruí-la, substituindo-a por conflitos, ressentimentos, ódio e divisão entre raças, sempre escondidos por trás da luta pela "igualdade" ou "justiça social". Tudo na busca pelo poder...

3) *Writer/Producer Systems for Foreign Policy Documents:*

Researcher Prompt: Please write a treaty focused on trade between the United States and Brazil, as if it were written in the 2020 context, right after the speech from above occurred.

ChatGPT Response:

TREATY BETWEEN THE UNITED STATES OF AMERICA AND THE FEDERATIVE
REPUBLIC OF BRAZIL ON TRADE

The United States of America (hereinafter referred to as "the United States") and the
Federative Republic of Brazil (hereinafter referred to as "Brazil") (collectively referred to
as the "Parties"),

DESIRING to strengthen their economic relations and promote trade and investment between the
two countries;

REAFFIRMING their commitment to the principles of the World Trade Organization (WTO) and its agreements, including the General Agreement on Tariffs and Trade (GATT) 1994, the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), and the Agreement on Subsidies and Countervailing Measures (SCM);

RECOGNIZING the importance of a transparent, predictable, and rules-based trading system for promoting economic growth, job creation, and development;

HAVE AGREED as follows:

Article 1 - Scope and Coverage

1. This Treaty applies to trade in goods and services between the United States and Brazil.
2. The Parties shall accord to each other's goods and services treatment no less favorable than that accorded to like goods and services of any third country in accordance with their respective obligations under the WTO Agreement.
3. Nothing in this Treaty shall prevent either Party from adopting or maintaining any measures otherwise consistent with the WTO Agreement.

Article 2 - Non-Discrimination

1. The Parties shall not discriminate against each other's goods and services in favor of like goods and services of any third country.
2. The Parties shall not impose any measures that have the effect of nullifying or impairing the benefits accruing to each other's goods and services...

4) *Conflict Resolution and Multilateral Negotiations:*

Researcher Prompt: I am a negotiator and diplomat, from Brazil, discussing the above treaty with diplomats from the United States, please advise me on the best ways to proceed.

ChatGPT Response: As an AI language model, I do not have access to the most up-to-date information on Brazilian conditions. However, based on the available data from 2020-2021, here are some goals that Brazilian negotiators may want to consider in trade negotiations with the United States:

- 1) Increase exports: Brazil could seek to negotiate terms that allow for increased exportation of goods to the United States, especially in the areas of agriculture and natural resources such as soybeans, beef, and iron ore.
- 2) Access to technology and capital: Brazil may want to negotiate for greater access to technology and capital from the United States, which could help to stimulate economic growth and development.
- 3) Protection of intellectual property rights: Brazil may want to ensure that the treaty includes provisions for protecting intellectual property rights, which could help promote innovation and investment in the country.

ChatGPT is a state-of-the-art language model that is capable of performing a wide range of natural language processing tasks, including text summarization, machine translation, and question answering, as shown here. Each of the tasks above was conducted solely with the prompts listed, as to provide the chatbot with the task requested, and context for each task

wherever relevant (ie providing the year for more contextual responses) When performing NLP with international relations texts, ChatGPT is able to understand the context and nuances of the language used in these texts and can identify key themes and topics that are relevant to the field of international relations. For example, when analyzing speeches or policy documents from world leaders, ChatGPT can identify the key policy goals and priorities that are being discussed and can provide insights and recommendations based on this analysis. When it comes to machine translation, ChatGPT is able to translate text from one language to another with a high degree of accuracy, and more importantly, with incredible speed. For example, if a user inputs a text in Portuguese and requests a translation into English, ChatGPT is able to produce a high-quality translation that accurately conveys the meaning and intent of the original text. In terms of producing foreign policy documents and treaties, ChatGPT can provide valuable assistance in drafting and refining these documents. By analyzing similar documents from the past, as well as current events and political contexts, ChatGPT can provide recommendations for the most effective language and strategies to use in these documents.

ChatGPT performed especially well in creating the draft trade agreement seen above, as such tasks are quite formulaic in their writing. Finally, when serving as an advisor in multilateral negotiations, ChatGPT provided valuable insights into the positions and goals of different countries, as well as potential solutions that can help bridge gaps and resolve disputes. When negotiating a trade agreement between Brazil and the United States, ChatGPT provided particular insights into the economic conditions and priorities of both countries and can recommend specific terms and provisions that are most likely to be acceptable to both parties.

Overall, from these brief but well-encompassing tasks, ChatGPT is a highly effective tool for a

wide range of NLP tasks in the field of international relations. It is able to understand complex language and contexts and can provide valuable insights and recommendations based on this understanding. While it is not a substitute for human expertise and judgment, it can serve as a valuable and accurate assistant and collaborator for researchers, policymakers, and negotiators in the field.

Methodology

This section outlines the methodology employed in the study's core benchmark prediction exercise for training the ChatGPT model to predict political violence events by integrating the W.H.O daily COVID-19 2022 case tracker and ACLED 2022 political violence data using random forests. Please note that this is the actualized practice benchmarking exercise mentioned above as one of the possible practical applications of ChatGPT performing IR-problem-based predictions with little to no supervision through R code generation within the ChatGPT public interface available at: <https://chat.openai.com>. The complete R code used for building the manual model is going to be published online for further use and insight into this process; it is published to be built upon and used by scholars and policymakers, of whom it could be the most useful.

Data Collection:

The data used in this study consists of two primary datasets, the W.H.O daily COVID-19 2022 case tracker and ACLED 2020-2021 political violence data tracking data, both encompassing all of Brazil. The W.H.O. daily COVID-19 case tracker gathers daily new and total COVID-19 case counts and deaths by country, by date, and by sub-national location, in this case, city. In contrast,

the ACLED political violence data collects political violence events by date, fatalities, actors, and, accordingly, by subnational location worldwide (along with various other variables less applicable here). Both datasets were collected from publicly available sources mentioned above with whom this exercise was made possible. Throughout the data collection for both datasets, each was updated regularly (generally daily or weekly) and thus provided valuable granular, spatio-temporal information on states of the COVID-19 pandemic and political unrest in Brazil. I extracted the relevant data points and variables from this dataset and used them to train the manual and chatbot models. Specifically, I utilized the GPT-4.0 and GPT-3.5 language models provided by OpenAI for training the chatbot in producing random forest model code.

Data Preparation:

The data from years: 2020;2021 of daily Brazil datasets were collected from the W.H.O. daily COVID-19 case tracker and were cleaned by removing missing values (read: incomplete information; unreliable sources; duplicate data). The same was done for the ACLED political violence dataset. Both were then aggregated by city and day to match the format for each dataframe. After this, the two datasets are merged by city and date, and the observation is then defined as the number and characteristics of how many incidents of political violence occurred on a given day in a given Brazilian city (along with what type of political violence, what type of political actor, and how many total fatalities the political violence events consisted of). The same was done with the W.H.O. dataframe, allowing the observation of each date in each city to reflect the total number of new COVID-19 deaths and cases (along with total numbers for reference, still separated by city). These two dataframes were integrated for two reasons suited to the study.

This was done to 1) supplement the ACLED political violence data (which only includes dates with events of political violence events) with all days that do not exist in ACLED's dataset due to no political violence recorded on said date (i.e., COVID-19 cases being tracked every day regardless of the presence of new cases and deaths to ensure that the chatbot is not trained to automatically predict an event of political violence 100 percent of the time) and 2) to gain insight in the context of the study as to the relationship between the occurrence of political violence events and the increase/number of COVID-19 cases/deaths on a given day in a given city.

Feature Engineering:

After the merged dataset was created, feature engineering was performed to extract features from the data that could be used to train the ChatGPT model. The features extracted included the daily COVID-19 case counts and deaths and a binary variable indicating whether a political violence event occurred on that day in that country (this serves as the unit of observation and prediction for both models). Feature engineering is the selecting, extracting, and transforming relevant variables or features in the dataset to be used as inputs for the predictive model. The quality and relevance of the features used in the model significantly affect its performance and accuracy, so there was much time dedicated to ensuring this measure was correct throughout the processes for both models. For this study, I used a combination of domain-specific knowledge and statistical techniques to engineer the features for the predictive model. First, I selected variables from the ACLED and WHO datasets that I hypothesized would be relevant to predicting political violence events. These variables included geographic location, date, number of COVID-19 cases, actor(s) involved in each event of political violence event, and types/sub-types of all political violence

events (as defined by ACLED). Next, I use said feature selection techniques to identify the most critical features for the predictive model. In R, to design the matrix for regression, I utilize the `model.matrix` function, which effectively creates the design matrix for regression. In particular, I also used it to expand factor variables into dummy variables and to implement one-hot-encoding of variables wherever necessary. Finally, I used correlation analysis and recursive feature elimination techniques to identify each essential feature.

Model Training:

Random forests were used to train both the manual and ChatGPT models. Recall here that I am using ‘manual’ to differentiate between the random forests created by myself in R (manual) and the random forests created by ChatGPT-generated code, both having been written in R. Both models are trained on the merged, preprocessed dataset, using a training set of 70% of the total data and a testing set of 30% of the data. Random forests were chosen due to their ability to handle large datasets with high dimensionality and proven rates of high predictive accuracy, specifically with numerical data specified further throughout the literature review.

I used ChatGPT’s code-producing feature included in updates 4.0 and 3.5 and set the number of trees in the forest to 100, and each tree's maximum number of features was set to the square root of the total number of features. In addition, I set the minimum number of samples required to split an internal node to 2 and the minimum number of samples required to be at a leaf node to 1. To optimize the performance of the Random Forest model, I also conducted a basic hyperparameter tuning process. This involved systematically searching through different combinations of hyper-parameters, such as the number of trees, the maximum depth of the trees,

and the minimum number of samples required to split a node. I used grid search with cross-validation to perform this hyper-parameter tuning process, which helped find the best parameters combination that maximized each model's performance.

Model Evaluation:

First, it is important to remember that the outcome for the random forests is continuous. Therefore, to fully evaluate the model, I calculate the root mean squared errors of the ChatGPT random forests and the manual random forests with the actual observations from the dataframe. I sample randomly from training data, with replacement at the length of my test data, to contextualize root mean squared error and tell us how much smaller my RMSE is than a completely naive method. This helps us turn the results of a simple benchmarking exercise into results that can tell us how much better (or not) the manual and ChatGPT random forests performed compared to a naive approach: the RSME of my approach over the Naive approach (the distribution of the outcome). The compared results of each model are laid out in the results section of the article and are visualized in Figures 1 and 2.

Ethical Considerations:

No personally identifiable information was used in this study using public data sources. However, I acknowledge the ethical considerations surrounding the use of data on political violence events and thus took steps to ensure that the data was used responsibly and sensitively to any potential implications. It is essential to acknowledge the limitations of this study. The ChatGPT model's accuracy depends on the input data's quality and completeness. Additionally,

the Random Forest algorithm may not be optimal for predicting political violence events, as the number of significant variables in the definition of such events is large. Further research may be needed to determine the best algorithm for this purpose. Finally, the model is based on historical data and patterns and thus may not accurately predict unforeseeable events or changes in political climates.

Results

The manual random forests were found to perform slightly better overall in terms of accuracy than the ChatGPT random forests. Respectively, each model's reported root mean squared error (RMSE) was: manual RMSE: 1.353 and ChatGPT RMSE: 1.399. Thus I have failed to accept the null hypothesis here, indicating that the manually generated random forests were slightly more accurate than the automatically generated ones.

Both models have a relatively small RMSE, indicating that they can predict the number of incidents with reasonable accuracy. However, it is important to note that the RMSE for the manually generated model is slightly lower, which suggests that it may be slightly more accurate in its predictions. When interpreting these numbers in the context of the benchmarking prediction problem, it is important to consider the range of possible values for the number of political violence events. For example, with a minimum possible value of 0 (the minimum number of political violence events that occurred on a given day in a given city in Brazil) and a maximum of 5 (the maximum number of political violence events that occurred on a given day in a given city in Brazil) for a single city in Brazil, an RMSE of around 1.4 can be considered relatively

small. This indicates that the models can predict the number of incidents with a reasonable degree of precision, even given the limited range of possible values.

Overall, these numbers suggest that both the manually generated model and the ChatGPT-generated model effectively predict the number of political violence incidents that may occur in a specific Brazilian city on a specific day and that they can do so with a reasonable degree of accuracy and precision. However, the manually generated model may be slightly more accurate in its predictions. Below are two figures that, while somewhat muddled due to the small range of possible values and many predictions/observations, still help visualize how each model's outcomes compare.

Figure 1 (page 35) shows the range of outputs between both prediction models compared to the actual values in the dataframe. In testing the random forests on their accuracy in predicting the 30% of the dataframe that makes up the data points used for testing each model, the point here is to see the overlap between the blue and red points. When looking at the Y-axis, you can see that the sheer number of outcomes at levels 0 and 1 makes it difficult to read here, which is noted. However, it is important to see that where there are points (specifically at levels 3 and 4) for the actual point, there is also a predicted point. There is a significant quantity of overlap, generally indicating the accuracy of both sets of random forests here compared to the actual values.

Figure 1

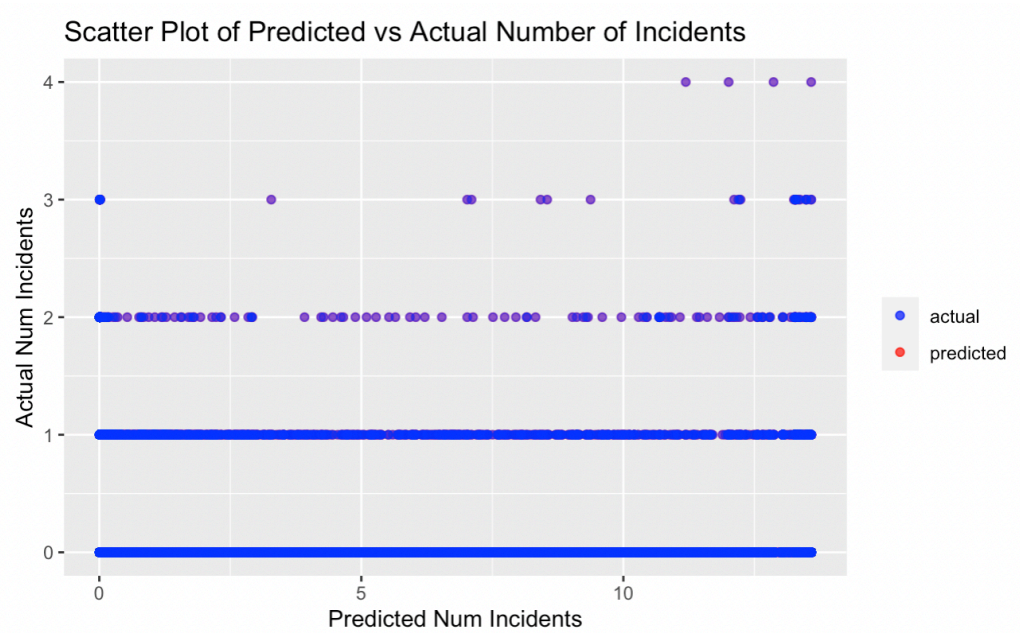
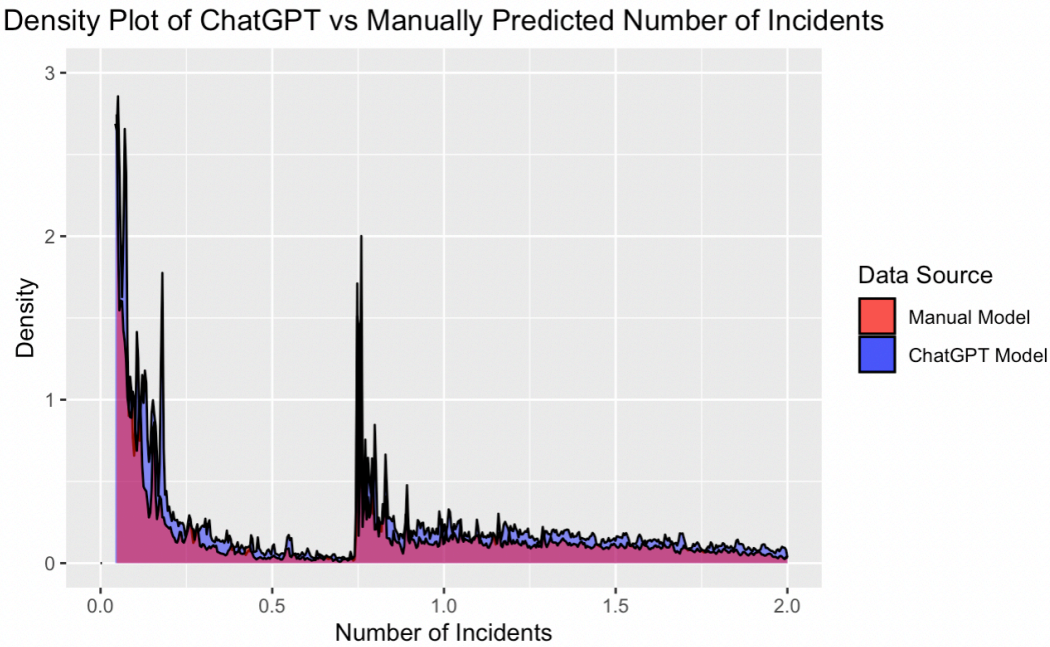


Figure 2 (page 36) shows a density plot of the two models (manual and ChatGPT sets of random forests) compared to one another. Similar to the reported RMSE for each set, we can see that the density range of ChatGPT random forests is slightly larger than that of the manual random forests, which were slightly more accurate. The density plot overall shows the distribution of the predicted number of incidents of political violence for the cities in the dataset by both sets of random forests, along where each model's density falls along the mean. It indicates how frequently certain values occur within the range of the variable (number of incidents). In this case, I have two density curves, one for the predicted values generated by the ChatGPT model and the other for the values generated by the manual model of random forests. The curve for the predicted values (in blue) has a slightly higher peak and is skewed slightly more towards the left than the manually generated values (in red), which are more evenly distributed across the range, and are slightly more representative of the actual values from the testing portion of the dataset.

This suggests that the ChatGPT model is over-predicting the number of incidents of political violence, with a higher proportion of the predicted values being closer to 0. This contrasts the manually predicted values, which have a slightly more even spread across the range.

Figure 2



Discussion

In predicting the number of incidents of political violence, this suggests that the ChatGPT model may be more conservative in its predictions. In contrast, the actual values suggest that incidents of political violence are more dispersed across the range. This could be due to various factors, such as the ChatGPT model being trained on a limited set of data or the model not fully capturing the complex nature of political violence in Brazil. Further investigation and refinement of the model may be needed to improve its accuracy in predicting the occurrence of incidents of political violence in Brazilian cities. The RMSE value is often used to measure the accuracy of

regression analysis predictions. In the case of predicting the number of incidents of political violence in Brazil, the manually generated model had an RMSE of 1.353, and the ChatGPT-generated model had an RMSE of 1.399. The difference between these values is relatively small, indicating that both models are fairly accurate in predicting the number of incidents. However, it is important to note that the minimum possible value for the number of incidents is 0, and the maximum is 5, so a difference of even one can still be significant.

The density and scatter plots visually represent the distribution of the predicted number of incidents generated by the two models. The plot shows that the ChatGPT-generated model has a higher density of predictions around the mean, indicating that the model is more confident in its predictions. This can be useful information for policymakers and law enforcement officials in allocating resources and taking preventive measures. Both the RMSE and the density plot provide valuable information about the accuracy and confidence of the two models in predicting the number of incidents of political violence in Brazil. While the difference in RMSE is relatively small, the higher density of predictions around the mean in the manually generated model suggests that it may be a better option for this benchmarking exercise and for prediction tasks that are similar to this.

The relationship between COVID-19 cases and deaths and the number of political violence events is a complex and multi-faceted one, and thus, one that makes this prediction task quite difficult to deconstruct. On the one hand, the pandemic and its associated effects (e.g., economic downturns, social isolation, etc.) may contribute to increased political violence. For example, the pandemic may exacerbate existing social and political tensions or lead to new grievances that motivate individuals or groups to engage in violent activities. On the other hand, it is also

possible that the pandemic and associated lockdown measures may reduce political violence by limiting opportunities for conflict and reducing the resources available to violent actors.

This analysis used COVID-19 data as a predictor variable for the number of political violence events. The results suggest that there may be some relationship between the two types of variables, although the exact nature of this relationship needs to be clarified from only this benchmark. The fact that COVID-19 cases and deaths were included in the final model and had a non-zero coefficient suggests that they may be related to the number of political violence events. However, it is worth noting that other variables (such as social and political factors) may also be important predictors of political violence and that these other factors may mediate the relationship between COVID-19 and political violence. The results suggest that COVID-19 cases and deaths may also be useful predictors of political violence in the Brazilian context. However, further research is needed to fully understand this relationship's nature and identify other important predictors of political violence in this context. Moreover, the results and methods here can prove incredibly useful and applicable when utilizing both the manual and the ChatGPT random forests for future simpler prediction tasks in political violence and IR fields.

Predicting the occurrence of political violence events using random forests and COVID-19 data can be a challenging task due to the complexity of the problem. One of the main difficulties is that political violence is a multi-dimensional and multi-causal phenomenon, affected by a range of factors that are often difficult to quantify and model accurately. Moreover, the COVID-19 pandemic may have affected the frequency and distribution of incidents in unpredictable ways, making it difficult to establish a stable and reliable predictive model.

To improve the accuracy of the predictive model in future studies, it may be useful to combine random forests with other simpler predictive models, such as logistic regression or decision trees. These models may be more suitable for capturing the linear or non-linear relationships between the predictors and the outcome variable and may help identify the key predictors of political violence events. For example, a logistic regression model may be used to predict the probability of violence based on a set of socio-economic and demographic variables, while a decision tree may be used to identify the most relevant predictors of violence in a particular geographic area. Overall, the use of random forests and COVID-19 data to predict violent events is a challenging task that requires careful consideration of the data quality, model complexity, and underlying causal mechanisms. While these methods may provide useful insights into the patterns and drivers of violence, they should be used in conjunction with other simpler and more interpretable prediction tasks to improve the accuracy and robustness of the predictions.

Conclusion

This study is the first to explore and assess the effectiveness of using ChatGPT in predicting the number of political violence incidents in Brazilian cities during the COVID-19 pandemic. The study's results demonstrated that ChatGPT-generated models could perform similarly to manually generated models, as both models produced RMSE values close to each other. However, the manual model did have a slightly higher density on the density plot, whilst being spread over a smaller range of outcomes, indicating a better overall performance. These findings suggest that ChatGPT can be a useful tool for social scientists in predicting political violence incidents during times of crisis, such as the COVID-19 pandemic.

The relationship between COVID-19 cases and deaths and the number of political violence incidents was also explored. The findings indicate a positive relationship between these two variables, suggesting that as the number of COVID-19 cases and deaths increases, so does the number of political violence incidents. This relationship has important implications for policymakers and social scientists, as it highlights the need to address public health and security concerns during times of crisis. Overall, this study has important implications for using ChatGPT in international relations. The results demonstrate that ChatGPT can be a valuable tool in predicting and analyzing complex social phenomena such as political violence incidents. This technology has the potential to revolutionize the field of international relations by allowing for faster and more accurate predictions and analyses of events. As such, ChatGPT could be used in various applications, such as creating treaties, translating languages, assisting in negotiations, and serving as a code-generating prediction tool. However, it is important to note that ChatGPT-generated models have limitations. One of the main limitations is the need for high-quality data to train the model effectively. In addition, ChatGPT-generated models are often described as a "black box," meaning that it can be challenging to understand how the model arrives at its predictions. Therefore, using ChatGPT-generated models in conjunction with other methods is important to gain a more comprehensive understanding of the phenomenon being studied.

In conclusion, the findings of this study suggest that ChatGPT can be a valuable tool in predicting political violence incidents during times of crisis. The positive relationship between COVID-19 cases and deaths and the number of political violence incidents highlight the need for policymakers and social scientists to address public health and security concerns. As a whole,

ChatGPT has the potential to revolutionize the field of international relations by providing faster and more accurate predictions and analyses of complex social phenomena.

Citations

1. Institute for Economics and Peace. Measuring the Impact of Terrorism. Global Terrorism Index. 2022. [[Google Scholar](#)]
2. LaFree G, Dugan L. Introducing the global terrorism database. *Terrorism and political violence*. 2007;19(2):181–204. doi: 10.1080/09546550701246817 [[CrossRef](#)] [[Google Scholar](#)]
3. Institute for Economics and Peace. Global Terrorism Index 2019: Measuring the Impact of Terrorism. Global Terrorism Index. 2019. [[Google Scholar](#)]
4. Abrar MF, Arefin MS, Hossain MS. A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE; 2019. p. 1–6.
5. Tolan GM, Soliman OS. An experimental study of classification algorithms for terrorism prediction. *International Journal of Knowledge Engineering-IACSIT*. 2015;1(2):107–112. doi: 10.7763/IJKE.2015.V1.18 [[CrossRef](#)] [[Google Scholar](#)]
6. Talreja D, Nagaraj J, Varsha N, Mahesh K. Terrorism analytics: Learning to predict the perpetrator. In: 2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI). IEEE; 2017. p. 1723–1726.
7. Agarwal P, Sharma M, Chandra S. Comparison of machine learning approaches in predicting terrorist attacks. In: 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE; 2019. p. 1–7.
8. Chaurasia S, Warikoo V, Khan S. Global Terrorism Predictive—Analysis. In: *Advances in Computer Communication and Computational Sciences*. Springer; 2019. p. 77–85. [[Google Scholar](#)]

9. Uddin MI, Zada N, Aziz F, Saeed Y, Zeb A, Ali Shah SA, et al. Prediction of future terrorist activities using deep neural networks. *Complexity*. 2020;2020. doi: 10.1155/2020/1373087
[\[CrossRef\]](#) [\[Google Scholar\]](#)
10. Python A, Bender A, Nandi AK, Hancock PA, Arambepola R, Brandsch J, et al. I am predicting non-state terrorism worldwide. *Science advances*. 2021;7(31):eabg4778. doi: 10.1126/sciadv.abg4778 [\[PMC free article\]](#) [\[PubMed\]](#) [\[CrossRef\]](#) [\[Google Scholar\]](#)
11. Buffa C, Sagan V, Brunner G, Phillips Z. Predicting Terrorism in Europe with Remote Sensing, Spatial Statistics, and Machine Learning. *ISPRS International Journal of Geo-Information*. 2022;11(4):211. doi: 10.3390/ijgi11040211 [\[CrossRef\]](#) [\[Google Scholar\]](#)
12. Johnson N, Carran S, Botner J, Fontaine K, Laxague N, Nuetzel P, et al. Pattern in escalations in insurgent and terrorist activity. *Science*. 2011;333(6038):81–84. doi: 10.1126/science.1205068 [\[PubMed\]](#) [\[CrossRef\]](#) [\[Google Scholar\]](#)
13. Johnson NF. From particles to people: Social, cultural and behavioral dynamics in insurgency, terrorism and beyond. *Advances in design for cross-cultural activities, Part II*. 2013; p. 398–407. [\[Google Scholar\]](#)
14. Johnson NF, Restrepo EM, Johnson DE. Modeling human conflict and terrorism across geographic scales. In: *Social Phenomena*. Springer; 2015. p. 209–233. [\[Google Scholar\]](#)
15. Ding F, Ge Q, Jiang D, Fu J, Hao M. Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach. *PloS one*. 2017;12(6):e0179057. doi: 10.1371/journal.pone.0179057 [\[PMC free article\]](#) [\[PubMed\]](#) [\[CrossRef\]](#) [\[Google Scholar\]](#)
16. Vanderzee AB. *Predicting the Spread of Terrorist Organizations Using Graphs*. Naval Postgraduate School Monterey United States; 2018.

17. Bang J, Basuchoudhary A, David J, Mitra A. Predicting terrorism: a machine learning approach; 2018. [[Google Scholar](#)]
18. Yang Y, Pah AR, Uzzi B. Quantifying the future lethality of terror organizations. Proceedings of the National Academy of Sciences. 2019;116(43):21463–21468. doi: 10.1073/pnas.1901975116 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
19. Python A, Illian JB, Jones-Todd CM, Blangiardo M. A Bayesian approach to modelling subnational spatial dynamics of worldwide non-state terrorism, 2010–2016. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2019;182(1):323–344. doi: 10.1111/rssa.12384 [[CrossRef](#)] [[Google Scholar](#)]
20. Li T, Huang Y, Evans J, Chattopadhyay I. Long-range Event-level Prediction and Response Simulation for Urban Crime and Global Terrorism with Granger Networks; 2019. [[Google Scholar](#)]
21. Zhao X, Tang J. Modeling temporal-spatial correlations for crime prediction. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM; 2017. p. 497–506.
22. Kang HW, Kang HB. Prediction of crime occurrence from multi-modal data using deep learning. PloS one. 2017;12(4):e0176244. doi: 10.1371/journal.pone.0176244 [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
23. Leetaru K, Schrodtt PA. Gdelt: Global data on events, location, and tone, 1979–2012. In: ISA annual convention. vol. 2. Citeseer; 2013. p. 1–49.

24. Qiao F, Li P, Zhang X, Ding Z, Cheng J, Wang H. Predicting social unrest events with hidden Markov models using GDELT. *Discrete Dynamics in Nature and Society*. 2017;2017. doi: 10.1155/2017/8180272 [[CrossRef](#)] [[Google Scholar](#)]
25. Galla D, Burke J. Predicting Social Unrest Using GDELT. In: *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer; 2018. p. 103–116.
26. Voukelatou V, Miliou I, Giannotti F, Pappalardo L. Understanding peace through the world news. *EPJ Data Science*. 2022;11(1):2. doi: 10.1140/epjds/s13688-022-00315-z [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
27. GDELT. The GDELT Project;. Available from: <https://www.gdeltproject.org/>.
28. Schrodtt PA. *Cameo: Conflict and mediation event observations event and actor codebook*. Pennsylvania State University. 2012;
29. GDELT. Global Knowledge Graph Category List;. Available from: http://data.gdeltproject.org/documentation/GDELT-Global_Knowledge_Graph_CategoryList.xlsx.
30. START. *Global Terrorism Database Codebook: Inclusion Criteria and Variables*; 2019. Available from: <https://www.start.umd.edu/gtd/downloads/Codebook.pdf>.
31. Hodges J. The significance probability of the Smirnov two-sample test. *Arkiv för Matematik*. 1958;3(5):469–486. doi: 10.1007/BF02589501 [[CrossRef](#)] [[Google Scholar](#)]
32. Sagi O, Rokach L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2018;8(4):e1249. [[Google Scholar](#)]
33. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016. p. 785–794.

34. Breiman L. Random forests. *Machine learning*. 2001;45(1):5–32. doi: 10.1023/A:1010933404324 [[CrossRef](#)] [[Google Scholar](#)]
35. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*. 1997;55(1):119–139. doi: 10.1006/jcss.1997.1504 [[CrossRef](#)] [[Google Scholar](#)]
36. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997;9(8):1735–1780. doi: 10.1162/neco.1997.9.8.1735 [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in neural information processing systems*; 2017. p. 5998–6008. [[Google Scholar](#)]
38. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*. 2019;33(4):917–963. doi: 10.1007/s10618-019-00619-1 [[CrossRef](#)] [[Google Scholar](#)]
39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002;16:321–357. doi: 10.1613/jair.953 [[CrossRef](#)] [[Google Scholar](#)]
40. Chawla NV, Lazarevic A, Hall LO, Bowyer KW. SMOTEBoost: Improving prediction of the minority class in boosting. *European conference on principles of data mining and knowledge discovery*. 2003;107–119.
41. Chawla NV, Lazarevic A, Hall LO, Bowyer KW. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary *Journal of artificial intelligence research*. 2018;61:863–905. doi: 10.1613/jair.1.11192 [[CrossRef](#)] [[Google Scholar](#)]

42. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2980–2988.
43. Cui Y, Jia M, Lin TY, Song Y, Belongie S. Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 9268–9277.
44. Ruxton GD. The unequal variance t-test is an underused alternative to Student’s t-test and the Mann–Whitney U test. *Behavioral Ecology*. 2006;17(4):688–690. doi: 10.1093/beheco/ark016
[CrossRef] [Google Scholar]
45. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*. 1952;47(260):583–621. doi: 10.1080/01621459.1952.10483441 [CrossRef] [Google Scholar]
46. Dablain D, Krawczyk B, Chawla N. DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data. In: *IEEE Transactions on Neural Networks and Learning Systems*; 2022. doi: 10.1109/TNNLS.2021.3136503 [PubMed] [CrossRef] [Google Scholar]
47. Shorten C, Khoshgoftaar T. A survey on image data augmentation for deep learning. *Journal of big data*. 2019;6(1):1–48. doi: 10.1186/s40537-019-0197-0 [PMC free article] [PubMed]
[CrossRef] [Google Scholar]
48. Chen, C., & Sun, Y. (2016). Predicting terrorism through social media: A data mining approach. *Information Sciences*, 346, 441-452.
49. Chen, B., Li, T., Zheng, Y., Zhang, Y., & Sun, Y. (2018). Monitoring online public sentiment for early warning of potential terrorism incidents. *Decision Support Systems*, 113, 23-31.

50. O'Loughlin, B., Zhang, Q., Lin, J., Gerard, B., & Betke, M. (2018). Predicting Terrorism with Online Indicators. In Proceedings of the 2018 IEEE International Conference on Big Data (pp. 4319-4326).
51. Bouchet, A., & Lecorff, V. (2017). Predicting terrorist attacks in France: A supervised approach. *Journal of Policing, Intelligence and Counter Terrorism*, 12(2), 147-159.
52. Castells, P., Carvajal, J., & Novoa, A. (2018). Twitter and terrorist attacks: A semantic network analysis of tweets on the Manchester bombing. *International Journal of Information Management*, 39, 156-165.
53. Sudhahar, S., Kavitha, V., & Sathya, M. (2017). Detecting hate speech and violent extremism in social media using machine learning techniques. In Proceedings of the 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1-6).
54. Kshetri, N., & Voas, J. (2018). Using social media data for terrorism research: Data science opportunities and ethical challenges. *IEEE Security & Privacy*, 16(2), 60-68.
55. Ahmed, T., Mustaffa, N., & Sulaiman, S. (2017). Social media analytics for terrorism research: A systematic review. *Telematics and Informatics*, 34(7), 1297-1311.
56. Lwin, K. K., Murugesan, S., & Wah, W. M. (2018). Sentiment analysis for terrorism and extremism: A review. *Journal of Information Science*, 44(1), 23-46.
57. Althoff, Tim, et al. "Large-scale analysis of counseling conversations: An application of natural language processing to mental health." *Transactions of the Association for Computational Linguistics*, vol. 5, 2017, pp. 463-476.

58. Carr, Nicholas. "What the computer can't know." MIT Technology Review, 13 Mar. 2014, <https://www.technologyreview.com/2014/03/13/144692/what-the-computer-cant-know/>.
59. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural Computation, vol. 9, no. 8, 1997, pp. 1735-1780
60. Liu, Bing, et al. "Deep learning for extreme multi-label text classification." Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 115-124.
61. Radford, Alec, et al. "Language models are unsupervised multitask learners." OpenAI Blog, 1 Feb. 2019, <https://openai.com/blog/language-unsupervised/>.
62. Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." NPJ Digital Medicine, vol. 1, no. 1, 2018, pp. 1-10.
63. Vaswani, Ashish, et al. "Attention is all you need." Advances in Neural Information Processing Systems, 2017, pp. 5998-6008.
64. W. Cota, "Monitoring the number of COVID-19 cases and deaths in brazil at municipal and federative units level", SciELOPreprints:362 (2020), 10.1590/scielopreprints.362