# FROM POSTS TO VOTES: USING XLM-RoBERTa (XLM-R) AND AGENT-BASED MODELING (ABM) TO ANALYZE AND PREDICT ELECTION WINS IN THE PHILIPPINES THROUGH X

A Thesis

Presented to the

Department of Information Systems

and Computer Science

Ateneo de Manila University

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science

by

**Michael Gavin N. Del Castillo**

**Mary June Aubrey C. San Jose**

**Shaira Jane L. Sy**

2026

## ABSTRACT

Recent election periods have seen the rise of social media platforms as tools through which political ideologies are spread, and support for candidates are garnered. This increase of reliance on social media for political interaction allows for data analysts to collect, follow and study human behavior online during election periods. With developed masked language models (MLMs) and modeling tools, researchers are able to track and categorize sentiments expressed in social media posts.

The main motivation for this research is to determine the effectiveness of social media as an indicator of electoral wins. Through multilingual MLMs like XLM-RoBERTa and Agent-based modeling (ABM), this paper aims to perform sentiment analysis on posts from X, a popular social media platform for public discussions on politics in the Philippines, to draw insights on how citizens perceived the electoral candidates during pre- and proper election periods and see if it can be used as an indicator of an election win to serve as a tool for predicting future elections. Posts related to the 2019 and 2025 midterm and 2022 presidential elections will be collected, annotated, and used to define an environment and agent for an Agent-Based model.

## TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

## ABSTRACT

Recent election periods have seen the rise of social media platforms as tools through which political ideologies are spread, and support for candidates are garnered. This increase of reliance on social media for political interaction allows for data analysts to collect, follow and study human behavior online during election periods. With developed masked language models (MLMs) and modeling tools, researchers are able to track and categorize sentiments expressed in social media posts.

The main motivation for this research is to determine the effectiveness of social media as an indicator of electoral wins. Through multilingual MLMs like XLM-RoBERTa and Agent-based modeling (ABM), this paper aims to perform sentiment analysis on posts from X, a popular social media platform for public discussions on politics in the Philippines, to draw insights on how citizens perceived the electoral candidates during pre- and proper election periods and see if it can be used as an indicator of an election win to serve as a tool for predicting future elections. Posts related to the 2019 and 2025 midterm and 2022 presidential elections will be collected, annotated, and used to define an environment and agent for an Agent-Based model.

# CHAPTER I

# INTRODUCTION

## 1.1 Context of Study

In recent years, social media has had a major role as a platform through which political ideologies are spread and political discussions occur. This change, while gradual, makes it more convenient for data analysts to collect and analyze how people behave, specifically using Sentiment Analysis and Agent-based Modeling (ABM), and track the effectiveness and utility of social media as a campaign tool.

When it comes to social media campaign regulations, the Philippines, along with Malaysia and Indonesia, are generally a laissez-faire and it has come to be a vital part of campaigns to maximize awareness on a candidate, as proven by the "first social media election"— the 2016 Philippine Presidential election [34, 33]. Cases, however, have occurred in which online support did not directly translate to election wins. The 2022 Philippine presidential elections saw the Angat Buhay campaign of former Vice President Leni Robredo garner the attention of young audiences on social media. Rallies in support of Robredo alongside Robredo's track record as a politician made her a popular choice for millions as a capable presidential candidate [20]. Despite massive online support, Robredo lost the elections, garnering some 14.8 million votes (as opposed to Ferdinand "Bongbong" Marcos, Jr, who gathered 31.1 million) [1].

Given the possible disparity between social media popularity and election votes, the aim of this research is to provide a data-driven analysis on the effectiveness of social media as an indicator of election wins by observing social media trends at the time of both 2022 and 2024 elections, as well as comparing and contrasting these elections in terms of said trends.

Currently, many newer transformer models and variations of BERT, a masked language model initially developed by Google, have been released, like RoBERTa, DeBERTa and XLM-RoBERTa. When it comes to analyzing political science papers and studies, the newer models greatly outperform BERT, although XLM-RoBERTa shines more when it comes to cross-lingual applications [35]. This makes it an ideal transformer model to use when analyzing the political landscape on social media for a multilingual nation like the Philippines. Additionally, for a chaotic and constantly changing environment, Agent-based Modeling and Simulation (ABMS) excel at modeling systems full of autonomous, interacting agents, like social media users, that are able to change and adapt new behaviors [27].

Through Natural Language Processing (NLP) Algorithms, this paper aims to analyze the conversations on the aforementioned presidential candidates that had transpired in online spaces during pre-election seasons, namely X (formerly Twitter). Previous research endeavors have already shown the effectiveness of sentiment analysis in determining key themes behind social media posts, especially in the context of events such as elections [31]. Thus, this research would like to push this idea further by not only contextualizing the data within a single setting. Rather, this paper aims to compare and contrast different election periods of the Philippines to create an Agent-based model (ABM) to, given the proper datasets, simulate and serve as a tool to predict future (Philippine) elections.

This paper intends to determine whether or not social media support directly translates to election success, or if other factors were present which had contributed to the losses of Robredo and worked to raise Duterte and Marcos Jr's popularity.

## 1.2 Research Questions

1. Can Agent-Based Modeling effectively predict election results by way of simulating interactions between users on social media that are centered around electoral candidates?

    (a) The XLM-RoBERTA model can be used to perform sentiment analysis on multilingual Filipino and English posts. Can the sentiments classified by XLM-RoBERTA extract the necessary sentiments needed to define social media users as agents in a simulation?

(b) Can the parameterized model not only simulate the recent 2025 Philippine senatorial elections, but also predict winning candidates if a Presidential election were to be held in the current year as of writing (2025)?

(c) Does the analysis produced by conducting an Agent-Based simulation of the elections reflect actual winning candidate statistics of Philippine elections?

## 1.3 Research Objectives

1. An Agent-Based Model will be created using sentiments extracted from posts on social media platform X, for the purpose of simulating not only candidate-user interactions but also user-user interactions in election campaigning periods.

   (a) Through XLM-R, perform sentiment analysis on social media posts about both the 2019 and 2022 Philippine elections and create a list of parameters with which Agents and interactions in the simulation are defined.

(b) Through our baseline model, perform an Agent-Based Simulation to evaluate the efficacy of the model to predict winning candidates of the 2025 Philippine senatorial elections, using information diffusion to represent the amount of social interactions done in the simulation with respect to a given candidate. Then, after verifying the accuracy of the Agent-Based model, perform a simulation to predict potential winning candidates of the upcoming 2028 Philippine presidential elections, as of the political landscape in the current year of writing (2025).

(c) Compare the results of the model with actual election results to determine if there is a statistically significant difference between them, i.e. through t-testing simulated results with actual election results.

## 1.4 Scope and Limitations of the Study

The scope of the study is to create a virtual environment that simulates the Philippine social media during an election season, with both users and candidates, through ABM.

Data collection will be limited to three Philippine election seasons for training and evaluation of the model: namely, the 2019 midterm and 2022 presidential elections for training purposes, and the 2025 midterm elections for validation. Scraping of social media posts will be restricted within X due to the short character limit of 280 for each post, which makes performing sentiment analysis more manageable than on other platforms. The range of dates for posts collected depends on the last day of filing a Certificate of Candidacy (COC), and the last day before the election day proper. Posts made on the election days themselves will not be included in our dataset.

As for the number of candidates, all presidential candidates for the 2022 elections will be the main agents; for senatorial elections, on the other hand, only posts about the top 20 candidates will be collected from their official pages. This is so that the model is still able to capture the contribution of those outside the winning 12 candidates to online public discourse on which candidates to vote for.

### 1.5 Significance of the Study

This paper aims to contribute to the field of computational social science by analyzing the behavior of users and political candidates in online spaces during election periods, especially in the Philippine context where such a kind of study has yet to be pursued in-depth. The interaction of ABM and cross-lingual transformers like XLM-RoBERTa can computationally create a pipeline from analyzing a large-scale X dataset to simulating real-world interactions by: (a) scraping posts from X in bulk, (b) extracting the necessary sentiments from said posts about electoral candidates to define the necessary models for the simulation, (c) running and refining the simulation to accurately reflect past elections, and finally (d) extending the simulation by predicting future elections. The combined usage of ABM and LLM in the local context is significant for future researchers who plan to explore further the interaction between transformer models and agent-based modeling from the perspective of social events other than elections.

# CHAPTER II

# REVIEW OF RELATED LITERATURE

The sentiment analysis subsection discusses its definition and how it is essential in analyzing social media engagement. Next, tools used for analyses will be discussed, especially the transformer model Bidirectional Encoder Representations for Transformers (BERT)— its architecture, its descendants like RoBERTa, how effective BERT is for sentiment analysis using various studies as evidence, and the capabilities of XLM-RoBERTa. Lastly, the Social Media and Elections subsection discusses how social media shapes the general public, especially in the context of Philippine politics.

The sentiment analysis subsection discusses its definition and how it is essential in analyzing social media engagement. Next, tools used for analyses will be discussed, especially the transformer model Bidirectional Encoder Representations for Transformers (BERT)— its architecture, its descendants like RoBERTa, how effective BERT is for sentiment analysis using various studies as evidence, and the capabilities of XLM-RoBERTa. Lastly, the Social Media and Elections subsection discusses how social media shapes the general public, especially in the context of Philippine politics.

## 2.1 Sentiment Analysis for Social Media and Elections

According to Liu [24], the study of people's views, sentiments, assessments, appraisals, attitudes, and emotions about goods, services, organizations, people, problems, events, subjects, and their characteristics is commonly referred to as sentiment analysis or opinion mining. With the explosive growth of social media, it has become a hotspot of opinions, shaping our decisions, especially in an important political event like elections. In the field of social computing, election seasons are one of the widely researched topics, especially on how the interaction in social media affects society in terms of making decisions on who to vote for. With the recent rise in popularity of many large language models (LLMs) like ChatGPT, LLMs have also been considered to perform sentiment analysis tasks. LLMs have been measured and evaluated to have satisfactory performance in simpler tasks, lag behind in more complex tasks requiring structured sentiment information, and have a potential when annotation resources are limited [25].

There are examples of studies using sentiment analysis to analyze social media activity. In a study by Macrohon, et al. [28] and Demillo, et al. [14], they used the Naïve Bayes classifier, a probabilistic learning method, to determine the probability of a post belonging to the best class-applicable in determining the polarity of a post. Then, previous studies showed the usage of bidirectional encoder representation from transformers (BERT) models, modified to handle emojis and Tagalog language posts. Aquino, et al. [5] introduced the emotion-infused BERT-GCN model for sentiment analysis, which includes emoji semantics into the models, treating them as sentiment representation; meanwhile, Cruz, et al. [13] used the RoBERTa-tagalog-cased model to get the vectorized version of Tagalog embeddings, essential to map echo chambers on X via K-Means modeling. Lastly, the Support Vector Machines (SVM) Classifier model was used by Demillo, et al. [14] to handle binary classification of data, classifying them as either a negative or positive sentiment.

For reasons discussed more in-depth in the next section, BERT was chosen for the study's methodology given its capabilities of performing nuanced analyses and classification of social media posts.

### 2.1.1 Development Pipeline of BERT, RoBERTa, and XLM-RoBERTa

Recent developments in devising models for NLP tasks have ensured that models are updated to be more context-aware, being able to provide a more holistic and nuanced analysis of certain texts. One such model is BERT, short for Bidirectional Encoder Representations from Transformers (referred to henceforth as BERT). Developed by the Google AI Language Laboratory, the main advantage provided by BERT is its ability to analyse text in a bidirectional manner, as opposed to more traditional machine learning models, such as GPT, which only analyse text left-to-right or vice versa [21]. Bidirectional analysis of text ensures that BERT is able to capture not only the sentiments of text, but do so in such a manner that the model is able to detect certain nuances, such as sarcasm or irony [16].

BERT's architecture is built on transformers, an architecture of neural networks that uses a combination of recurrent and convolutional networks [21]. BERT is primarily pre-trained in two phases: first, using a large dataset of unlabeled data, then second; a smaller set of labeled data, usually for fine-tuning the BERT model according to some NLP task. One such NLP task is Sentiment Analysis [21]. In pretraining, BERT operates on two main objectives. The first is the Masked Language Model (MLM henceforth). A random sample of tokens in the input sequence is selected and replaced with a special mask token [MASK]. The objective is then for the BERT model to be able to predict what these masked tokens are. Next is Next Sentence Prediction (NSP), a binary classification task. The goal is for the model to be able to predict whether two text segments follow each other. Both positive examples (consecutive sentences from the training set) and negative examples (pairs of segments from different documents) are provided and are sampled with equal probability.

In 2019, the Facebook AI research team found that BERT was "significantly undertrained", and thus proposed RoBERTa, short for "[A] Robustly Optimized BERT Pretraining Approach". The team sought to improve the training process of the BERT model by (1) training the model over a longer period of time, and with bigger batches of data, (2) removing the NSP objective in pretraining, and (3) dynamically applying the masking pattern applied to the training data. The results of this optimized pretraining process do show, indeed, that RoBERTa is able to either match or exceed the performance of BERT in NLP tasks, the former scoring higher than the latter in multiple NLP model evaluation tests such as GLUE, SQuAD, and RACE [26].

As BERT and RoBERTa have seen usage in analysing large datasets of text, it is able to aid in research on social media. Social media is considered a rapidly evolving form of text widely different from more traditional text formats such as novels mainly due to the widespread usage of informal language, abbreviations, and emojis, among other elements, which can be challenging to understand without the proper context.

For one, Kumar and Sadanandam [30] were able to use BERT and RoBERTa to classify a large dataset of some 8,225 posts related to the Coronavirus into three general sentiments: positive, neutral, and negative. Both BERT and RoBERTa were able to perform sentiment analysis across the entire dataset, achieving high accuracies (at least 88%), precision (at least 0.88), recall (at least

0.74 but can go as high as 0.91), and F1-score (at least 0.78 but can go as high as 0.90). Prasanthi et al. [31] were also able to accomplish a similar feat using both BERT and RoBERTa, performing sentiment analysis on large social media datasets with extremely high accuracies, these accuracies only improving with each succeeding epoch. BERT was able to achieve a base accuracy of 95.10% on the first epoch, which only increased to 99.16% on the tenth epoch. Similarly, RoBERTa was able to achieve a base accuracy of 99.53%, with a final accuracy of 99.70% on the tenth epoch [31].

As mentioned earlier, BERT and RoBERTa are able to capture nuances such as sarcasm and irony in texts. Detecting sarcasm, in particular, has proven to be a highly challenging NLP task as a sarcastic statement implies a negative sentiment whilst seemingly conveying a positive one surface-level. Nevertheless, Dong et al. [16] were able to train RoBERTa on a dataset of posts from Reddit and X to give it the ability to detect sarcasm in a given text, with an F1 score of 80.2.

These studies illustrate the importance of RoBERTa and BERT in the context of sentiment analysis on highly informal and nuance-laden texts such as social media posts.

Since RoBERTa's release, there have been updated revisions of the model that further increase RoBERTa's performance on certain NLP tasks. Conneau et al. [10] released a version of RoBERTa that is also able to handle texts in languages other than English but has stronger gains than the previous models on classification, sequence labeling and question answering. XLM-RoBERTa, short for Cross-lingual Model RoBERTa (referred to as XLM-R henceforth), is trained with some 2.5 terabytes of data using texts on 100 languages, and trained using the MLM objective [10].

The team then evaluated the model, alongside other versions of BERT, in different tests. Firstly, the XLNI, short for the Cross-lingual Natural Language Interface, is a dataset containing training and testing datasets in 15 languages. The goal of the dataset is to evaluate a model's cross-lingual transfer from English to another language. XLM-R was able to outperform mBERT (multilingual BERT) in this test, with an advantage of around 10% in terms of per-test accuracy.

Next is the Named Entity Recognition test, which evaluates not only cross-lingual transfer, but also the model's performance per-language as well as multilingual learning ability. Also included are the Cross-lingual Question Answering test and finally the GLUE benchmark. Again, XLM-R outperforms mBERT on all tests, and was also shown to be able to effectively model low-resource languages such as Swahili and Urdu [10].

Other studies have been published since the release of XLM-R showing its prowess in processing multilingual texts, and outperforming other models in the same tasks. Azadi et al. [6], for example, trained both XLM-R and GPT 3.5 to classify a bilingual dataset of posts (composed of English and Spanish posts) through two tasks. The first task was to determine whether or not the post has sexist content, and the second was to determine what type of textual content the post has with regards to sexism (if the post contained purposively sexist content, was reporting a sexist situation, or was condemning certain sexist behaviors). XLM-R was shown to outperform GPT 3.5 on both tasks: in Task 1, XLM-R and GPT scored an over F1 score of 0.78 and 0.71, respectively; while in Task 2, XLM-R scored 0.48 as opposed to GPT's 0.43.

XLM-R may also be improved further by incorporating Zero-Shot Learning. With the proposal of chain of thought prompting (CoT) instead of the conventional question-and-answer format, Zero-shot LLMs in text classification can utilize pre-trained models to predict both seen and unseen classes, making them powerful text classifiers in NLP studies due to their representational capacity and scalability [36, 37].

## 2.2 Simulating Through Agent-Based Modeling

Analyzing social phenomena does not stop at merely analyzing the sentiments behind social media posts – with sufficient data, machine learning models can be used to create models of certain social actors and simulate the different interactions that take place between them. For this, Agent-Based Modeling is necessary. Agent-Based Modeling (ABM henceforth) is a computational model that simulates the actions and interactions of autonomous individuals or "agents" [27]. Macal and North [27] define an agent as an autonomous decision-making entity, each of which is self-contained and possessing a set of characteristics as well as behaviors that the agent can perform. Agents are required to interact with each other through an environment, as interactions create changes to their respective states, imitating an iterative learning process [27]. Thus, such a method of modeling helps analyze social events at a large scale.

### 2.2.1 Linking ABM with Social Networks and Social Media

Recent studies have shown that utilizing ABM can help learn about the spread of information on social media [17, 9, 15, 12]. As social media consists of a user interacting with others, it forms a "network" of interactions between them. This network becomes the basis of how ABM can be used to study user behavior as well as the flow of information within a social media platform in a specific time period.

Past studies have delved into studying online social media phenomena using ABM. Chen [9], for example, created an ABM to investigate the process of information diffusion in online social networks. Chen [9] defines information diffusion as the spread of information; and efficient diffusion means widespread awareness of certain events.

A similar study was used to simulate information diffusion in social media platforms, taking into account that users might have different sociopolitical views from one another. Coscia and Rossi [12] used ABM to study the phenomenon of polarization on social media by defining an environment meant to mirror social media, as well as agents who operate in said environment. Agents assumed the form of one of two categories: either Users with personal accounts, or News Sources with official accounts. Users are able to choose from the following actions: either reshare or flag a news item post depending on the polarity between the user and post, change their polarity, and “unfriend” and/or “unfollow” other social media users. Through a series of simulations performed through ABM, Coscia and Rossi [12] were able to glean valuable insights about how social media users act with regards to emotionally-charged situations such as political posts, such as how sharing such posts can lead to polarization among online spaces, and how minimizing conflicts on social media might actually lead to increased polarization online.

### 2.2.2 Linking LLMs with ABM

Large Language Models (LLMs) with their human-like reasoning and decision-making have opened new avenues to enhance ABM's simulations of social phenomena. Gao et. al [18] discussed a promising foundation for simulations in ABM, especially simulations that require sophisticated cognitive abilities, such as heterogeneity and personalization of agents' internal states and behaviors, and they have adaptive learning and evolution. However, ABM poses challenges such as computational expenses, robustness, and ethical risks. It was used across multiple domains, especially in the social domain, where there are studies that used a combination of LLMs and ABM to study social networks, especially in simulating emotions, attitudes, and information propagation in online social networks when discussing current issues that cause extreme emotions [18].

In recent years, there have been studies addressing the combination of LLM and ABM. They use decoder models like GPT to generate interactive behaviors in the simulated social network environment, which in turn updates the agents and environment [19, 22, 29]. This methodology was applied in tackling political biases and detecting stances. The combination of two machine learning models has helped Piao et al. [29] design and implement a social experiment that involves Bandura's social learning theory to mitigate political biases on LLM-generated social bots. Additionally, in a different study [22], it helps them develop a multi-dimensional text analysis and reasoning-enhanced debating to determine the stance of a post, where the agents play roles through the lenses of linguistics, domain knowledge, and social media expression, and the two sides of a debate. Therefore, based on these sources, developing a similar methodology is feasible, though certain limitations, such as the intensity of computing resources, should still be considered.

In the context of XLM-RoBERTa, there have been no existing studies on its interaction with ABM. Yet, it has the potential to become a unique advantage in tackling multilingual social dynamics and being an encoder transformer model that classifies texts. Therefore, this research gap is what this study aims to tap into regarding how being an encoder model with cross-lingual properties can help develop agents to create a simulation of Philippine social media during an election season.

## 2.3 Social Media Use in the Elections

To properly ground our study and simulation in real-world phenomena, it is important to note the different ways in which social media has been used throughout past election periods and how it has been taken advantage of by politicians.

Over the years, social media has become the first reference of the voters when it comes to perceiving political content and eyeing political information, political groups, and political parties [8]. Tapsell's [34] analysis brings attention to an interview with Nic Gabunada, Duterte's 'campaign friend', where he says Facebook became vital to their campaign after realizing that 45% of Filipinos are on Facebook, via mobile phone, and that their goal for the campaign was to maximize awareness. Contractor et al. [11] also found that, alongside holding discourse and reactions, social media provides information ahead of traditional news sources by a few hours.

### 2.3.1 Public Opinion

Tapsell's [34] paper reveals how social media is used to encourage citizens to be more active online in supporting a candidate.

As observed by Sinpeng et al. [33], despite Duterte's unprofessional online presence, his supporters are committed and constantly rallied to his defense against the criticism of other candidates. Tapsell's [34] interview with Jonji Gonzales, a PR professional from Visayas, reveals that any form of discourse that engages with content about a political figure is a form of support, especially in cases where there's information to correct or clarify. According to the interview, this is how the average Filipino becomes a 'keyboard warrior' for the candidate.

### 2.3.2 Voter Preference

Student voters give importance to political attributes that reflect competencies such as profession, experience, accomplishments, and priorities [3]. To see and verify the qualities they are looking for, social media helps voters evaluate a leader's current and future service to their community and people by making (political) information easily accessible [8].

Almarez and Malawani's [2] study found that 44% of their respondents indicated that their presidential preferences were influenced by social media and 75% of their respondents indicated that social media is a determining factor in the process of presidential campaign as a channel of campaign information. Having open access to political information at any time allows for change in political opinions and preferences in voters.

Other factors that guide voter behavior are "social identity, family voting, gender differences, ideology and emotions" [32].

# CHAPTER III

# METHODOLOGY

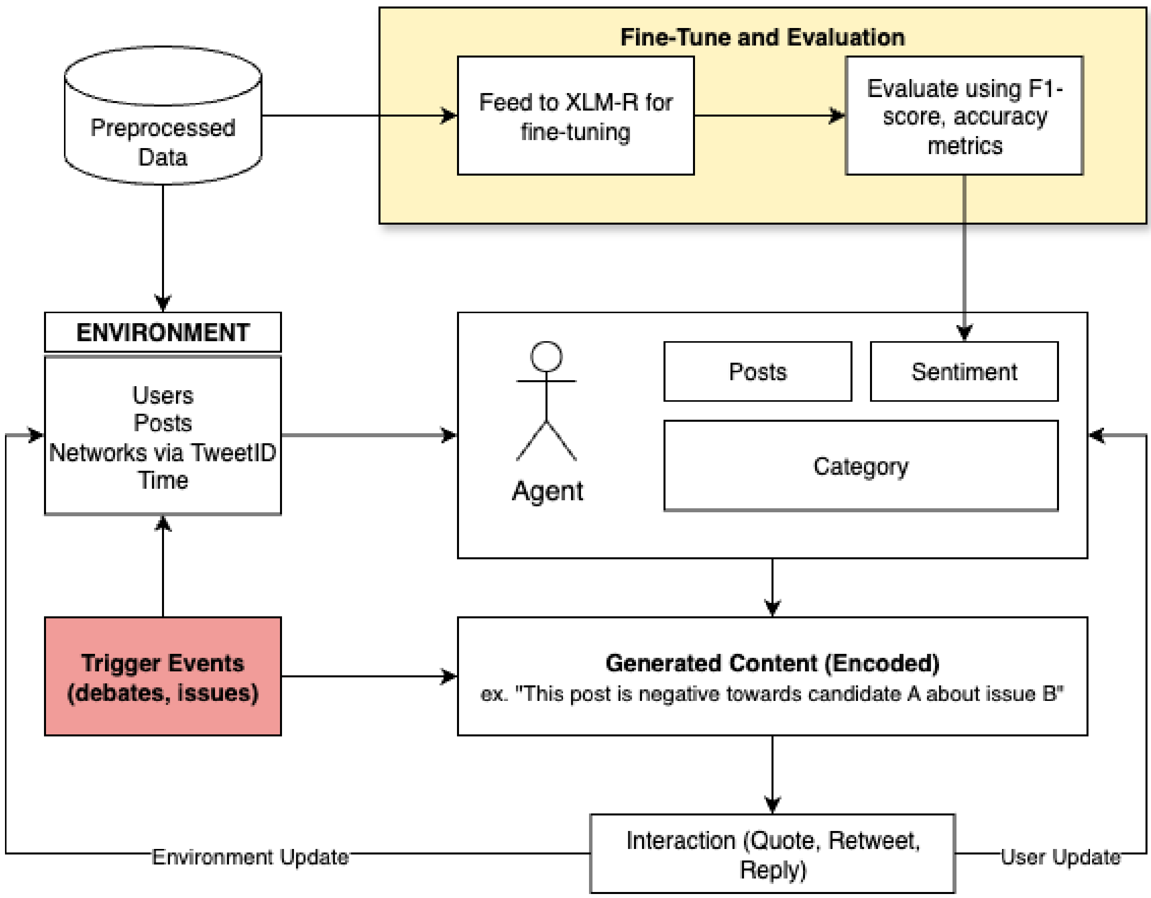


Figure 3.1: Methodology Flowchart

Figure 3.1 shows an overview of the methodology for this study. Posts collected will first be annotated according to their sentiment, and then used to fine-tune the model. Posts will then be used as a basis of the environment of our simulation, whilst the sentiments extracted by XLM-R will be used to define our agents. The agent-based model will then simulate interactions made regarding electoral candidates, which may or may not trigger a change in the user's views and attitudes.

## 3.1 Data Collection

Due to the discontinuation of the official free Academic API from X, the researchers will have to use a third-party Application Programming Interface (API) called Twikit for scraping posts about the 2022 Philippine Presidential elections. In collecting posts, keywords and dates will be utilized to perform an advanced search to get the posts needed for the analysis.

Table 3.1: Table for Philippine Dataset Ranges

| Election Year | Start Date (Deadline for Filing of CoC) | End Date (Election Day Proper) |
|---|---|---|
| 2019 Midterm | October 17, 2018 | May 12, 2019 |
| 2022 Presidential | October 8, 2021 | May 8, 2022 |
| 2025 Midterm | October 8, 2024 | May 11, 2025 |

The rationale behind the inclusive start-end dates is the high likelihood of a spike in sentiments from the public in two major events: the finalization of official candidates and election day proper.

Posts scraped per query will be saved in a CSV file with relevant information about each post, such as the account that posted it, engagement metrics (e.g. likes, replies, reposts), and whether or not the post is a standalone post, a reply to a prior existing post, or a quote repost of another post.

## 3.2 Data Preprocessing

Before a group of datasets is fed into a tokenizer, they will undergo text processing. Stop words such as “the,” “a,” “is”, etc., will not be removed as they might be considered for the full context of a sentence when performing byte-tokenization. The following steps to preprocess the data will be as follows:

- Omitting a post from the dataset if it is not in English, Tagalog, or a mix of them. Since Twikit takes care of labeling languages during the collection, they can be filtered using the Python library Pandas.

- Removing punctuation marks that have no significance for sentiment analysis.

- Replacing emojis with special tags.

- Removing unnecessary emojis or replacing emojis with special tags describing them if it is necessary for sentiment analysis.

- Lowercasing the text.
- Handling links and email addresses by replacing them with a placeholder.
- Removing whitespaces and replacing multiple spaces with a single space.
- Adding paddings to equalize the length of sentences.
- Anonymize the user mentions by replacing them with @user except for candidates' user handles, if this is present in the post.

The preprocessed dataset will be placed in a new CSV file. It is expected that the dataset, after they are preprocessed, will be 20,000 posts per election.

## 3.3 Text Annotation

### 3.3.1 XLM-RoBERTa

To detect the sentiment of election posts, they will be manually annotated as positive, negative, or neutral. Each has a value of 0, 1, and 2, respectively, once it is converted for analysis. These labels will help build the model dedicated to this study by adding a classification layer.

Pre-existing XLM-R models that are already capable of determining sentiments of posts on X will be used. For this study, Anke et al.'s [7] XLM-T-Base model will be employed, with additional fine-tuning to accurately detect

Table 3.2: Sample Sentiment Annotation

| | |
|---|---|
| **Positive** | "Showing my support to a good friend and Bagumbayan candidate for Senator Rafael Acuman" |
| **Neutral** | "A total of 129 party-list groups have filed their Certificate of Nomination and Acceptance CONA so far COMELEC expects a hundred more last-minute filers today." |
| **Negative** | "Mayoral candidate Isko Moreno blames incumbent Mayor Joseph Estrada for the worsening state of Manila adding that he will never build a political dynasty if he wins" |

Filipino and English election sentiments while keeping the general sentiments. The 2019 and 2022 election datasets will be used to train sentiments, while the 2025 dataset will serve as an evaluation dataset. The metrics to be used will be:

- Accuracy for overall correctness
- Precision, Recall, and F1-score, especially when the dataset is imbalanced
- A Confusion Matrix to determine the misclassification during the evaluation process

A dataset of posts with hashtags #Eleksyon[Year] and #Halalan[Year] will be used for fine-tuning.

The proportion of the three sentiments is significantly uneven. To prevent the model from dominantly inferring neutral sentiment, it will use Focal Loss

Table 3.3: Number of Tweets to Feed for Fine-Tuning

| **Year** | **Sentiments** | | |
|---|---|---|---|
| | **Positive** | **Neutral** | **Negative** |
| 2019 | 118 | 2498 | 120 |
| 2022 | 126 | 3642 | 245 |
| 2025 | 149 | 1155 | 178 |
| **TOTAL** | **393** | 7295 | **543** |

instead of Cross Entropy Loss (CEL) to address the class imbalances, which uses a modified version of CEL with additional parameters such as $\gamma$ (focusing factor) and optionally $\alpha$ (class weight) to put more emphasis on the classes that are difficult to classify [23].

### 3.3.2 Preparation for Fine-Tuning

Table 3.4 shows the hyperparameters that will be used to fine-tune the model. First, three epochs are used to train and evaluate the model. Then, the learning rate is set at 2e-5, which differs from some models that use a rate of 5e-5, as this model is being fine-tuned from a pre-trained version. Then, batch sizes for training and evaluation differ to accelerate the evaluation process. Lastly, the weight decay rate is set at 0.01 and the warmup ratio is 0.06.

Table 3.4: Hyperparameters for Fine-Tuning Anke et al.'s [7] Twitter-XLM-RoBERTa-Base using Election Tweets

| **Hyperparameter** | **Value** |
|---|---|
| Epochs | 3 |
| Train Batch Size per Device | 16 |
| Evaluation Batch Size per Device | 32 |
| Learning Rate | 2e-5 |
| Weight Decay Rate | 0.01 |
| Warm-up Steps | 0.06 |

## 3.4 Agent-Based Modeling

### 3.4.1 Creating the baseline simulation: 2019 Midterm and 2022 Presidential Elections

The output of the fine-tuned XLM-R model serves as a foundation for the creation of an agent. The following are the categories an agent belongs to: candidates, news outlets mentioning the candidates, influencers, and the general public. Influencer agents are neither candidates nor news outlets, yet they have more average interactions in their tweets than the general public. Therefore, they can serve as a "seed" in the simulation process.

### 3.4.2 Changing State

To change state, agents must interact with other agents and the relevant issues occurring during a specific time period. Therefore, the following events in Table 3.5 during the election have the potential to cause spikes in sentiments.

Table 3.5: Table of Debates

| Election Year | Key Dates | Events |
|---|---|---|
| 2019 Midterm | February 9, 2019 | GMA Network's Senatorial Face-Off |
| | February 18, 2019 | ABS-CBN's Harapan 2019 (1st Senatorial Town Hall Debate) |
| | March 1, 2019 | ABS-CBN's Harapan 2019 (2nd Senatorial Town Hall Debate) |
| | April 26, 2019 | CNN Philippines Senatorial Debate at UST |
| 2022 Presidential | February 15, 2022 | SMNI Presidential Debate |
| | February 26, 2022 | CNN Philippines Vice Presidential Debate |
| | March 19, 2022 | COMELEC PiliPinas Presidential Debate (1st) |
| | March 20, 2022 | COMELEC PiliPinas Vice Presidential Debate |
| 2025 Midterm | No COMELEC-hosted debates in this election year [4]. | |

To mitigate the lack of debates during the 2025 elections, key issues and surveys during the last three elections can factor into how the agents will react towards them. These terms are picked to ensure impartiality.

- Surveys (such as Pulse Asia)

- Issues: corruption, infrastructure, inflation, disinformation, fake news

### 3.4.3 Simulating the 2025 Midterm Elections

Using the calibrated 2019 and 2022 elections simulation, the succeeding 2025 midterm election can now be simulated. Posts will be gathered about the 2025 midterms elections for the top 20 winning candidates. Sentiments from those posts will be extracted, and then fed as input into the simulation. The simulation will then determine election results. To verify the accuracy of these simulated results, t-testing will be done to determine if there is any statistically significant differences between simulated voter proportions and actual voter proportions during the 2025 midterm elections.

### 3.4.4 Predicting Future Presidential Elections

Once the simulation has been verified to have accurately predicted the 2025 elections, the study will go beyond elections that have already passed to try and predict preliminary results for upcoming elections. With the agent-based model, a possible question of interest would be: “If the presidential elections were to happen at this moment, who would be the winners?”.

To try and answer this question, recent tweets will be gathered on prospecting presidential candidates based on existing poll and/or survey data in real-time. Using a similar method as in simulating the 2025 midterm elections, the simulation will now try to predict potential winners if the presidential elections were to happen this year.

# CHAPTER IV

# RESULTS AND DISCUSSION

## 4.1 Fine-Tuning the Model

Table 4.1: Per-epoch Results of Fine-Tuning Twitter-XLM-RoBERTa-Base using Manually Annotated Election Tweets

| Epoch | Training Loss | Validation Loss | Accuracy | F1 |
|---|---|---|---|---|
| 1 | 0.352100 | 0.960999 | 0.790290 | 0.794193 |
| 2 | 0.263000 | 0.648351 | 0.763318 | 0.783774 |
| 3 | 0.127700 | 0.631237 | 0.794336 | 0.807629 |

| Epoch | Precision | Macro F1 | Confusion Matrix |
|---|---|---|---|
| 1 | 0.834747 | 0.683953 | [[107, 34, 8], [80, 944, 131], [14, 38, 127]] |
| 2 | 0.833354 | 0.657547 | [[98, 32, 19], [81, 891, 183], [10, 26, 143]] |
| 3 | 0.804361 | 0.616454 | [[57, 61, 31], [48, 998, 109], [10, 52, 117]] |

Table 4.2: Overall Evaluation Results from Fine-Tuning Anke et al.'s [7] Twitter-XLM-RoBERTa-Base using Manually Annotated Election Tweets

| **Loss** | **Accuracy** | **F1** | **Precision** | **Macro F1** | **Confusion Matrix** |
|---|---|---|---|---|---|
| 0.6312 | 0.7943 | 0.8076 | 0.8347 | 0.6840 | [107, 34, 8], [80, 944, 131], [14, 38, 127 ] |

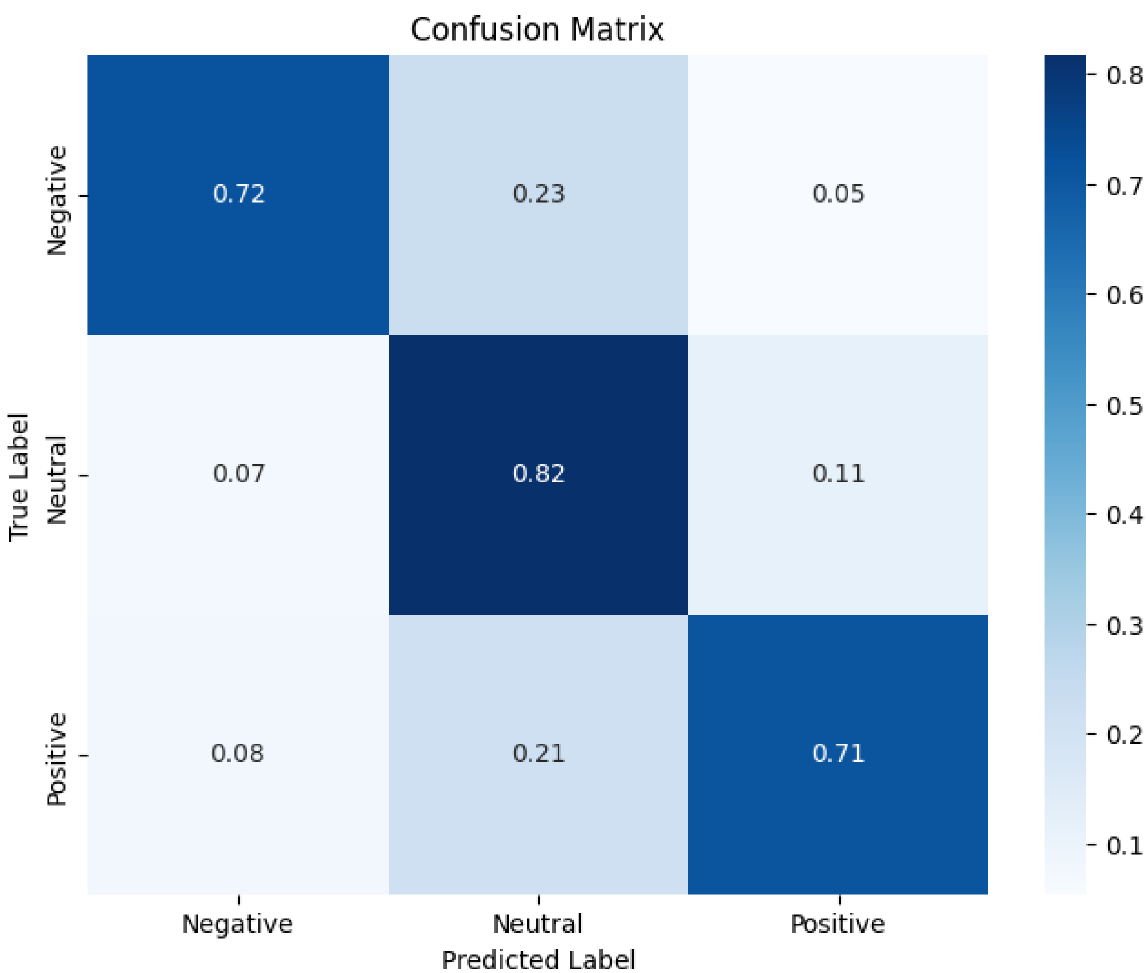


Figure 4.1: Confusion Matrix

Shown in tables 4.1 and 4.2 are the initial results of fine-tuning Twitter-XLM-R-Base using datasets gathered on the 2019 midterm, 2022 presidential, and 2025 midterm elections. The model was trained over three epochs using, as aforementioned, datasets on the 2019 and 2022 elections as training data, and the dataset on the 2025 elections as validation data. The tables show a steady increase in accuracy with each epoch, with an overall accuracy of fine-tuned Twitter-XLM-R-Base of about 79.02% and an overall F1 score of 0.7474. Finally, figure 4.1 shows the overall confusion matrix of our finetuned model.

# BIBLIOGRAPHY


[1] ABS-CBN. Halalan results 2022. `https://halalanresults.abs-cbn.com/`, 2022.

[2] Almarez, D., and Malawani, A. The social media as a transformative agent in the electoral process. *Journal of Government and Politics 7* (08 2016).

[3] Anabo, F. D. What matters most to student voters?: A conjoint analysis of political candidate attributes. *Philippine Social Science Journal 4*, 2 (June 2021), 18–29.

[4] Aning, J. No more comelec debates, but private events pushed. *INQUIRER.NET* (25, Jan 2025).

[5] Aquino, J. A., Liew, D. J., and Chang, Y.-C. Graph-aware pre-trained language model for political sentiment analysis in filipino social media. *Engineering Applications of Artificial Intelligence 146* (2025), 110317.

[6] Azadi, A., Ansari, B., Zamani, S., and Eetemadi, S. Bilingual sexism classification: Fine-tuned xlm-roberta and gpt-3.5 few-shot learning, 2025.

[7] BARBIERI, F., ANKE, L. E., AND CAMACHO-COLLADOS, J. Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond, 2022.

[8] CAMPANERO, C. A. D., DEOFERIO, E. J. A., LLENA, M. J. A., MATUNOG III, N. L., AND SANTOS, M. L. A. *The Influence of Social Media to the Voting Preference as Perceived by Selected Voters: A Case Study*. PhD thesis, Rizal Technological University College of Education, June 2021.

[9] CHEN, Z. An agent-based model for information diffusion over online social networks. *Papers in Applied Geography 5*, 1-2 (2019), 77–97.

[10] CONNEAU, A., KHANDELWAL, K., GOYAL, N., CHAUDHARY, V., WENZEK, G., GUZMÁN, F., GRAVE, E., OTT, M., ZETTLEMOYER, L., AND STOYANOV, V. Unsupervised cross-lingual representation learning at scale, 2020.

[11] CONTRACTOR, D., CHAWDA, B., MEHTA, S., SUBRAMANIAM, L. V., AND FARUQUIE, T. A. Tracking political elections on social media: Applications and experience. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (2015), IJCAI'15, AAAI Press, p. 2320–2326.

[12] COSCIA, M., AND ROSSI, L. How minimizing conflicts could lead to polarization on social media: An agent-based model investigation. *PLOS ONE 17*, 1 (Jan 2022), 1–23.

[13] Cruz, L. C., dela Cruz, J. N., Maglangit, S. F., Magtira, M., Imperial, J. M., and Rodriguez, R. Is twitter an echo chamber? connecting online public sentiments to actual results from the 2019 philippine midterm elections. In *2022 International Conference on Asian Language Processing (IALP)* (2022), IEEE, pp. 57–62.

[14] Demillo, R. E., Solano, G., and Oco, N. Philippine national elections 2022: Voter preferences and topics of discussion on twitter. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (2023), pp. 724–729.

[15] DiCarlo, M. F., and Zechman Berglund, E. Connected communities improve hazard response: An agent-based model of social media behaviors during hurricanes. *Sustainable Cities and Society 69* (2021), 102836.

[16] Dong, X., Li, C., and Choi, J. D. Transformer-based context-aware sarcasm detection in conversation threads from social media, 2020.

[17] Fan, R., Xu, K., and Zhao, J. An agent-based model for emotion contagion and competition in online social media. *Physica A: Statistical Mechanics and its Applications 495* (2018), 245–259.

[18] Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., and Li, Y. Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Humanities and Social Sciences Communications 11*, 1259 (27, Sep 2024).

[19] Gao, C., Lan, X., Lu, Z., Mao, J., Piao, J., Wang, H., Jin, D., and Li, Y. $S^3$: Social-network simulation system with large language model-empowered agents, 2025.

[20] Johnson, H., and Head, J. Leni robredo: The woman leading the philippines' 'pink revolution'. *BBC* (7 May 2022).

[21] Koroteev, M. V. Bert: A review of applications in natural language processing and understanding, 2021.

[22] Lan, X., Gao, C., Jin, D., and Li, Y. Stance detection with collaborative role-infused llm-based agents, 2024.

[23] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection, 2018.

[24] LIU, B. *Sentiment analysis and opinion mining*. SPRINGER INTERNATIONAL PU, 2012.

[25] Liu, B., Zhang, W., Deng, Y., Pan, S. J., and Bing, L. Sentiment analysis in the era of large language models: A reality check, 2023.

[26] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. Roberta: A robustly optimized bert pretraining approach, 2019.

[27] Macal, C., and North, M. Agent-based modeling and simulation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)* (12 2009), pp. 86–98.

[28] Macrohon, J. J. E., Villavicencio, C. N., Inbaraj, X. A., and Jeng, J.-H. A semi-supervised approach to sentiment analysis of tweets during the 2022 philippine presidential election. *Information 13*, 10 (2022).

[29] Piao, J., Lu, Z., Gao, C., and Li, Y. Social bots meet large language model: Political bias and social learning inspired mitigation strategies. In *Proceedings of the ACM on Web Conference 2025* (New York, NY, USA, 2025), WWW '25, Association for Computing Machinery, p. 5202–5211.

[30] Pranay Kumar, B., and Sadanandam, M. A fusion architecture of bert and roberta for enhanced performance of sentiment analysis of social media platforms. *Manchala, A Fusion Architecture of BERT and RoBERTa for En-*

*hanced Performance of Sentiment Analysis of Social Media Platforms* (27, May 2023).

[31] Prasanthi, N., Madhavi, R., Sabarinadh, D., and Sravani, B. A novel approach for sentiment analysis on social media using bert and roberta transformer-based models. In *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)* (04 2023), pp. 1–6.

[32] Qorri, F. The psychology behind voting behavior in kosovo, July 2018.

[33] Sinpeng, A., Gueorguiev, D., and Arugay, A. A. Strong fans, weak campaigns: Social media and duterte in the 2016 philippine election. *Journal of East Asian Studies 20*, 3 (2020), 353–374.

[34] Tapsell, R. Social media and elections in southeast asia: The emergence of subversive, underground campaigning. *Asian Studies Review 45* (09 Nov 2020), 117–134.

[35] Timoneda, J. C., and Vera, S. V. Bert, roberta, or deberta? comparing performance across transformers models in political science text. *The Journal of Politics 87*, 1 (2025), 347–364.

[36] Wang, Z., Pang, Y., and Lin, Y. Large language models are zero-shot text classifiers, 2023.

[37] Zhang, K., Zhang, Q., Wang, C.-C., and Roger Jang, J.-S. Comprehensive study on zero-shot text classification using category mapping. *IEEE Access 13* (2025), 23526–23546.