Socio-Emotional Computational Analysis of Propaganda Campaigns on Social Media Users in the Middle East

Conference Paper · April 2023

DOI: 10.1145/3543873.3587677

Zain Halloush
Augusta University

Craig Albert
Augusta University

Socio-Emotional Computational Analysis of Propaganda Campaigns on Social Media Users in the Middle East

Zain Halloush
zhaloush@augusta.edu
Augusta University
Augusta, Georgia, USA

Ahmed Aleroud
aaleroud@augusta.edu
Augusta University
Augusta, Georgia, USA

Craig Albert
calbert@augusta.edu
Augusta University
Augusta, Georgia, USA

ABSTRACT

Society has been significantly impacted by social media platforms in almost every aspect of their life. This impact has been effectively formulating people's global mindsets and opinions on political, economic, and social events. Such waves of opinion formation are referred to as propagandas and misinformation. Online propaganda influences the emotional and psychological orientation of people. The remarkable leaps in Machine Learning models and Natural Language Processing have helped in analyzing the emotional and psychological effects of cyber social threats such as propaganda campaigns on different nations, specifically in the Middle East, where rates of disputes have risen after the Arab Spring and the ongoing crises. In this paper, we present an approach to detect propagandas and the associated emotional and psychological aspects from social media news headlines that contain such a contextualized cyber social attack. We created a new dataset of headlines containing propaganda tweets and another dataset of potential emotions that the audience might endure when being exposed to such propaganda headlines. We believe that this is the first research to address the detection of emotional reactions linked to propaganda types on social media in the Middle East.

CCS CONCEPTS

• **Human-centered computing** → **Social networks**; **Social media**.

KEYWORDS

propaganda, misinformation, social media analysis, deep learning, language transformers

ACM Reference Format:
Zain Halloush, Ahmed Aleroud, and Craig Albert. 2023. Socio-Emotional Computational Analysis of Propaganda Campaigns on Social Media Users in the Middle East. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30–May 04, 2023, Austin, TX, USA*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3543873.3587677

ACM ISBN 978-1-4503-9419-2/23/04…$15.00
https://doi.org/10.1145/3543873.3587677

1 INTRODUCTION

Recently, contemporary society has witnessed waves of posted information through different platforms on social media. People now formulate their political, social, religious, and economic views based on the course of events around the world [21, 38, 39]. This extensive amount of publicly available resources and ease-of-access of information opened the doors for many malicious targeted campaigns that aim to create a state of unrest and disputes in different countries, locally and/or regionally [16, 19].

The last decade witnessed rising conflicts and disputes across the world in general, and in the Middle East specifically. As a result, targeted media has gained ubiquitous channels by utilizing social media platforms, and managed to create monolithic and polylithic campaigns in different communities [45]. These opinion maneuver campaigns built emotional and psychological crises for users and communities who receive this information, and might affect people's mindsets towards local and global events, and how they react to them [7, 19].

Malicious campaigns include propaganda's intentional and non-intentional attacks. Propaganda is a biased information used to promote ideological or political assumptions in a misleading nature to influence the attitudes and behavior of groups of people in favor of certain agendas and governments [26, 30]. Social media platforms and news websites reflect the influence created by this propaganda and how people react towards this mass information. People in some regions are highly interactive and effected emotionally and psychologically, which plays an important role in spreading these types of targeted campaigns.

Propaganda campaigns are generally related to Socio-political aspects of information cyber operations, which encompass three dimensions 1) psychological operations (renamed to military information support operations), which focus on spreading propaganda and dis/misinformation to change human behavior, 2) Electronic warfare operations: which disrupt adversaries' information networks, and 3) Military deception operations, which seek to mislead adversaries on military matters.

With all such ongoing cyber social attacks, social cyber security becomes a new direction in the science that focuses on the first types of operations [11]. It involves humans using technology for hacking other humans and societies. The main limitation of the current effort is their focus on English as the only language to create such attacks. Most of the existing research on information warfare focuses mainly on

- Russian efforts to influence western democracy [6, 33].
- China's use of fake social media accounts to manipulate political discourse [18, 37].

- The efforts of Iran to utilize Twitter, Facebook, and other platforms to shape conversation about American foreign policy [17].
- Russia's use of social media manipulation with other unconventional tactics in its direct war with Ukraine[20].
- The use of social media by the Islamic State of Iraq and Syria (ISIS) to spread misinformation, to recruit and to amplify the effects of extremism [10].

Motivation and Contributions: Research on information warfare ignores other regions of the world such the Middle East. According to the authors in [45] proxy-communications ignited and hijacked the events in the Arab world and how individuals worked to shape history. Recent examples of information operations in the region are related to Russian efforts to achieve political or diplomatic ends and influencing leadership and public opinion in the region. In [41] the authors refers to several examples on Russian propaganda efforts in the Middle East, and how they used the word "Ukraine" in Arabic and the loaded language to redirect users to irrelevant content on the United States. This also includes several older examples such as in 2016, when the United States (at the request of the Iraqi military) targeted a hospital being used as an ISIS operations center in the battle to retake Mosul from ISIS control. Russians then used their media to spread words such as "The so-called humanity of Obama in Iraq as he bombed a hospital in Mosul with internationally banned chemical weapons. This is how the children of Mosul were buried."

The era of Machine Learning (ML) and Deep Learning (DL) models has helped the research community in studying and addressing those state-linked attempts and the interaction between people in the Middle East through analysis obtained using Natural Language Processing (NLP) [4, 23, 29]. In addition, the emergence NLP and their capabilities in detecting salient and latent features has helped in building frameworks to detect people's opinions and feelings towards events in the Middle east. Due to data scarcity in low resources languages such as Arabic, recent language transformers are now trained on large datasets in various languages such as French, Chinese, and Arabic, which makes it easier to analyze the socio-emotional aspects of social media content [25, 40]. However, research on Arabic language to detect emotional impacts due to cyber social threats remains sparse.

The Arabic language is a widely spoken language with more than 360 million native speakers in different countries. Arabic is known for its rich morphological nature and several special traits related to the written formation of words and the existence of eight special symbols called diacritics that are capable of changing the meaning of the word when altered [22, 43]. In addition, the latent understanding of certain Arabic words is subject to context. The detection of emotional implications in Arabic is also a difficult task due to these reasons. These specifications limit the potential analytical aspects that can be studied for different NLP tasks.

In this paper, we introduce a framework for detecting propaganda headlines posted on twitter social media pages of well-known Arabic news networks and identify the associated potential emotional reactions on those headlines. We used a recently released unlabeled dataset that contains tweeted news headlines and articles' bodies on different topics related to recent conflicts in the Arab World. The dataset was a collection of headlines with no propaganda labels. Therefore, we conducted a robust labeling approach to detect whether specific headline targets creating a political propaganda. Later, the positive headlines (i.e., labeled with a propaganda) was also labeled by psychologists to identify the type of emotions it may target. We then studied the socio-emotional effect on how these headlines may have on Arab audiences. We used different language transformers in our classification tasks and a typical baseline models for propaganda labeling. The novelty of this ongoing research is originated from studying people emotional reactions and attitudes towards the propaganda agendas in the region. Additionally, this research introduces new datasets that can be used in future studies on social cyber threats in the Middle East.

The rest of the paper is organized as follows: Section 2 reviews important related works in the field. Section 3 gives a glance over the datasets we used and labeled, while Section 4 explains the methodology of work. In Section 5 we present the results of our experiments. Section 6 discusses the socio-emotional analysis in propaganda headlines, and Section 7 concludes the work.

2 RELATED WORK

Propaganda campaigns usually operate covertly, exploiting the anonymity and ease of creating fake online accounts to disseminate persuasive messages to a target audience [34]. Detecting coordinated propaganda campaigns is crucial in thwarting propaganda dissemination including those conducted by troll armies, internet water armies [9], sockpuppets [31], seminar users [13], and botnets. The study in [13] developed a framework for detecting seminar users engaged in political propaganda dissemination in the Arab World. The authors analyzed seminar users' behaviors, interactions, linguistic traits, and their use of aggressive terms in political discussions on Twitter in different Arab countries such as the United Arab Emirates, Saudi Arabia, and Yemen. The authors found that seminar users published more tweets than average users and had a significant impact on mainstreaming political discourse using hashtag campaigns. Their framework was able to successfully detect these seminar users with a precision of 84% by studying the behavior of 150 users who were tagged as seminar users using SVM classifier. Their proposed framework is not language specific and can be used to detect propaganda groups in other languages.

Studying the mechanisms of propaganda dissemination is crucial, as it involves a range of tactics from paralogism to emotional pressure. Paralogism involves the use of red herring, which distracts readers by presenting irrelevant materials, while another tactic is the straw man approach, where someone substitutes a proposition with fallacious ones and begins refuting it. Other tactic is called Whataboutism which is very common where someone charges another with hypocrisy to discredit their proposition, rather than directly opposing them. The black-and-white fallacy tactic is also used to force audiences to choose between specific options, when there are many other options available. Additionally, there is the appeal to authority approach which is used to lend credibility to a claim by relying solely on the authority of the person making the claim, without any supporting evidence. Emotional pressure is another common tactic and involves flag-waving, which promotes an idea by playing on strong national feelings. Loaded language, on the other hand, is used to influence the public either positively or negatively using words with strong emotional connotations. Thought terminating clichés, as well as slogans are also frequently employed in propaganda dissemination which rely on critical thinking to distract the audiences' minds from the truth.

These meanings of the different types of propaganda were employed and investigated in [12, 36]. The authors in [12] built a large manually annotated corpus with the help of six domain experts to detect the types of propaganda and build a multi-granularity neural network and compare it with BERT-based model as a baseline. Their work was expert driven to mark specific text spans with eighteen techniques mentioned earlier. Their work was built with 13 propagandistic news outlets and 36 non-propagandistic headlines in a sequence labeling fashion. They built their work on fragment level and sentence level propaganda detection.

In the same context the authors in [8] explored the effectiveness of several large language models, including AraBERT, ARBERT and MARBERT for detecting multi-label propaganda utilizing 21 propaganda techniques. The challenge addressed in their work was the contextual formation of the headlines of fake news as in many situations they are constructed based on real events which can easily deceive the audience, and how these headlines are written by experts to make them sound acceptable to the audience. The dataset was highly skewed in terms of labels and the distribution was also biased in terms of the propaganda types with high number of loaded language type. The authors explored several models' performance solely and in an ensemble fashion including DeHateBERT. Ensemble structure scored the highest with 51.7% micro-F1 and a similar score of micro-F1 was obtained by using ARBERT as a standalone model.

Scholars have also addressed studying the spread of propaganda by political extremists and strategies to counter and identify such activities, such as in [28], where the authors investigated Twitter accounts that are affiliated with Mijahideen groups and their supporters that disseminate terrorist-related propaganda on social media. The authors presented and trained AdaBoost classifier in binary classification fashion to detect these types of propaganda disseminators using data independent features such as time features, emotion words, and data dependent features which rely on certain keywords with high frequencies used by such accounts. They experimented with two datasets in English and in Arabic. Their accuracy scores varied when using data dependent features and when using data independent features.

Another work was conducted by the authors in [27] who constructed a deep learning model to recognize extremist and terrorist groups (specifically Sunni extremists) that disseminate jihadist propaganda and engage in recruiting other extremists. The authors built an expert-driven corpus by employing 40 experts, English and Arabic native speakers, to collect terrorist and non-terrorist content in English and Arabic languages which is a challenging task due to the contextual differences between these languages. To overcome this problem, the content was converted into ASCII characters, binarized, and vectorized for ease of training and analysis. The outcomes revealed a recall rate of 95.3%, a precision rate of 95.5%, an F1 score of 93% and an accuracy level of 93.2%.

3 DATASET

The scarcity of Arabic datasets and the complex process associated with collecting and labeling Arabic datasets due to the language's special traits and complex grammatical, morphological, and synthetic nature makes it hard to label a large volume of data elements. The datasets we used in this research is a recent twitter API dataset that contains non-labeled social media news headlines, articles, and tweets in response to those articles [44]. We didn't use the labeled response tweets to the articles since they were mainly collected to study cyber bullying attacks. The headlines on the other hand, were collected from different Twitter accounts that follows Aljazeera.net [2]. Those headlines were what we investigated in this research to identify signs of propagandas and potential emotions. Aljazeera news website is a well known hub for real-time news streaming in the Arab world. The network is widely approached by users in the Middle East; therefore, it is considered a good source for collecting such data. The dataset was preprocessed to remove redundancy and irrelevant titles and was fined into a dataset of 9530 headlines.

3.1 Binary Classification

A binary classification was considered as a Span Identification (SI) task to separate titles with potential propaganda with the labels (0) and (1) to identify and extract contiguous spans of text that corresponds to at least one propaganda technique(s) [3]. The label (0) indicates no observed aspect of propaganda in the title, and (1) indicates a potential observation of at least one propaganda category. We didn't conduct a multi-label classification in this work, but we are currently extending our labeling process to label headlines using a multi-label approach.

The labeling process was performed over a period of four weeks by four domain experts on media and political sciences. The annotators hold graduate degrees in each of the fields with about five years of work experience in Jordan media institute [24]. The annotators were given some examples on each of propaganda techniques based on many publicly available examples on the web. The average disagreement in the labeling processes using the Kappa index was 16%, which consists of cases where there is a disagreement between annotators. In those cases, the annotators conducted a second round of labeling to avoid labeling unintentional errors, we then used the majority voting rule to produce the final labeled dataset. The overall contextual formation of the headline defines whether this headline contains a perceived perspective of propaganda. The annotators were given several examples of news headlines that contain propaganda.

As shown in Figure 1, the higher number of headlines that contain potential propaganda can be attributed to the use of controversial phrases to increase the level of people engagements. Note that the dataset includes accounts that retweeted those headlines and are therefore potential candidates who belong to propaganda campaigns.

The conflicts in the Middle East create waves of different aspects of how social media news providers formulate the headline to build chaotic points of views of agreements and disagreements and obtain broader reach and interactions on certain events. Conflicts in Yemen, Syria, and West Bank are some of the most controversial and critical events in the Middle East, especially when relating these conflicts to foreign agendas and countries such as the United States, Russia, Israel, and Iran.

Table 1: Example of the labeled headlines data

English Translation of Arabic Headline	Classification
America, Russia and Iran are accused of collusion against the Syrians	Propaganda
Obama and Netanyahu.. No disagreement is higher than support for Israel	Propaganda
Russia: America refuses to exchange information about ISIS in Syria	Propaganda
Jordan and Lebanon affirm the political solution in Syria	No Propaganda

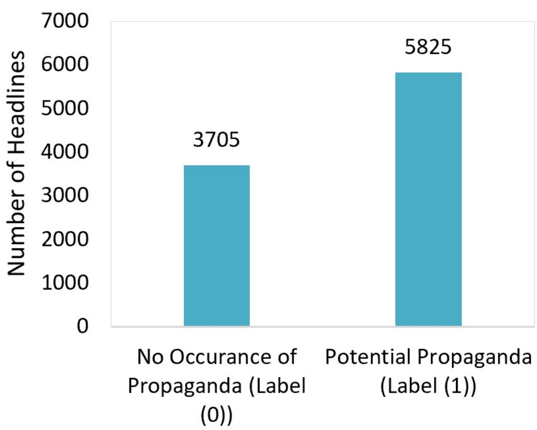

Figure 1: Data labeling results

The annotators focused on the contextual formation of these headlines and excluded sentences that do not contain potential propaganda or are irrelevant. Our labeling process was conducted with consideration of ethical aspects related to cultural, linguistic, and contextual factors without bias, harm or pre-judgments of any group, country, or individuals. Examples of the headlines in our dataset and their labels are shown in Table 1.

3.2 Multi-label Classification of Emotions towards Propaganda

The second stage of the data labeling process was with respect to the expected socio-emotional reactions on the audience seeing the headlines that contained the propaganda. For this purpose, we consulted two Psychology and Sociology Arabic domain experts to provide the potential labels for each headline. It is important to mention that annotators have taken into consideration the different perspectives expected to be present as a reflection on the audience's feelings and reactions. The labeling process for the emotions was based on studying the headline with the potential propaganda with corresponding emotions from the Wheel of emotions [42]. These emotions include joy, sadness, acceptance, disgust, fear, anger, surprise, anticipation, intolerance, and confusion. It is worth mentioning that certain headlines reflect a combination of expected feelings and mixed reactions, therefore, we relied on the multi-labeling approach.

The domain experts were Arabic speakers who are aware of the current conflicts in the Middle East such as the Israeli-Palestinian conflict which ignites anger, anticipation and in some sadness altogether. Another example includes the issues faced by refugees from Syria and Yemen due to the civil wars in both countries, which create waves of sadness and anger. The labeling process of reactions took almost two weeks. Sample of the identified emotions is shown in Table 2.

4 DATA CLASSIFICATION METHODOLOGY

This section demonstrates the approach used to create our data classification framework. The dataset was divided using the 80:20 rule, where 80% of the data was used for training and 20% was used for evaluation respectively. A stratified sampling method was used to represent each of the classes.

4.1 Propaganda Classification

4.1.1 MARBERT Model. Arabic has special traits, thus, working with the Arabic language requires different steps of data wrangling to be read and processed correctly. These steps included removing diacritics and special characters and symbols such as (@, #, $, &, { }, ...etc.), along with normalizing certain letters in Arabic which take different shapes and turning them into one shape. In addition, we removed different types of emojis. The preprocessing was performed using the "pyarabic.araby" library in Tensorflow to strip the diacritics. Furthermore, we utilized the Arabic stop words dictonary to remove stop words from the dataset. User mentions were also removed along with removing the redundant headlines.

The classification stage was to detect whether a headline contains a potential propaganda. To perform this task, we used MARBERT model as our baseline transformer-based classifier [1]. The MARBERT model is based on the original BERT(see Figure 2), which is a transformer based neural network architecture that has achieved a state-of-the-art performance on various natural language processing tasks, including text classification [15].

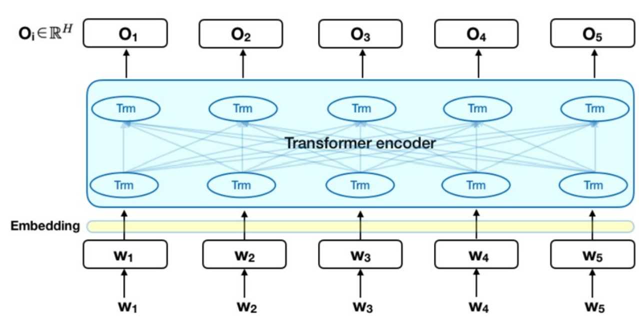

Figure 2: Illustration of the structure of the BERT model

Table 2: Emotional multi-label classification of propaganda headlines

English Translation	Emotion Classification
Diseases and malnutrition threaten the children of Yemen	Anger, Sadness
Accusations of the occupation liquidating the martyrs of Hebron in cold blood	Anger, Disgust, Fear, Intolerance
An agreement between Iran and the major powers after a decade of crisis	Anticipation, Confusion
Why does Europe support Obama with regard to Iran?	Anger, Confusion, Surprise

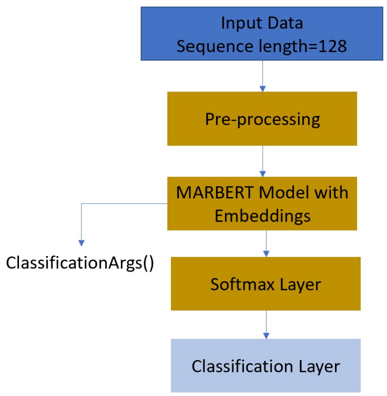

Figure 3: Architecture of the proposed MARBERT-based classification model

We fine-tuned the MARBERT model on our propaganda detection task using binary cross-entropy function and AdamW as the optimization function with a learning rate of 3e-5 and a training batch size of 16 headlines. In this experiment, we configure the MARBERT model for the classification task as demonstrated in the HuggingFace Transformers library [14]. The structure of the model is shown in Figure 3. All of our code and data is available on github[1].

4.1.2 AraBERT Model. Antoun et al. in [5] presented AraBERT as a fine-tuned model for the Arabic language. The model reached a state-of-the-art performance compared to the baseline Multilingual BERT on Arabic and was trained to obtain remarkable results on Sentiment Analysis , Question and Answering, and Name Entity Recognition datasets. This fine-tuning and language specification framework has outperformed previous multilingual frameworks due to the large set of pre-training data used (more than 24GB of data). The model is built using a stacked encoder and a classifier to decide the correct sense of the word. The main aspect of AraBERT model is the use of the attention mechanism and transformers to map the relations between the words in a sentence and determine the importance and calculating the relevance of one word to another. Once the transformer determines the importance of the word; an adjusted weight is given to that word and therefore a newly generated embedding is given to that word. The transformers generate the textual representations by recurrently coupling the input textual data with the layer of the attention mechanism, and therefore can extract features more efficiently. Thus, the model can be described as two parts: the first is the generalized form of features' learning, and the second is the task specific learning which occurs by only the available data (dense layer). An illustration is provided in Figure 4.

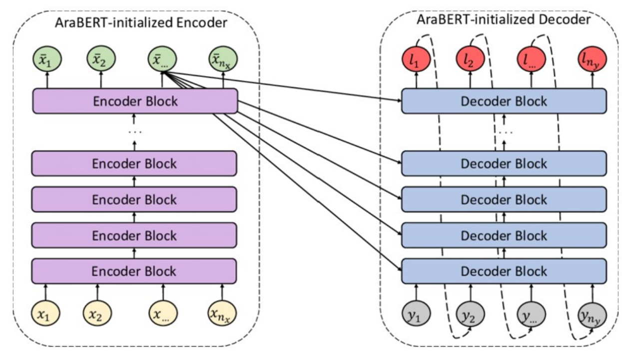

Figure 4: Illustration of the structure of the AraBERT model

We fine-tuned the AraBERT v02 model on our propaganda detection task using the same settings of the previous experiment as shown in Figure 5.

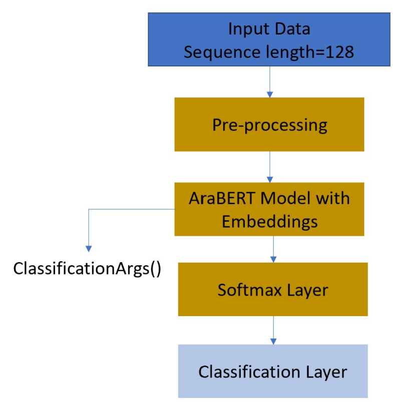

Figure 5: Architecture of the proposed AraBERT-based classification model

5 RESULTS

5.1 Results of Transformer Models

Figures 6 and 7 illustrate the results obtained using our classification methods. The results obtained are promising in terms of the

[1]Code and data are available at https://github.com/ahmed-aleroud/Arabic-Propaganda.

accuracy and other evaluation metrics such as F1-score, recall, and precision. While our dataset is unbalanced and has higher number of potential propaganda compared to non propaganda, it is less skewed compared to the existing datasets. In addition one objective of our initial labeling is to further conduct span detection of other propaganda categories and study the impact of emotions per each type of propaganda.

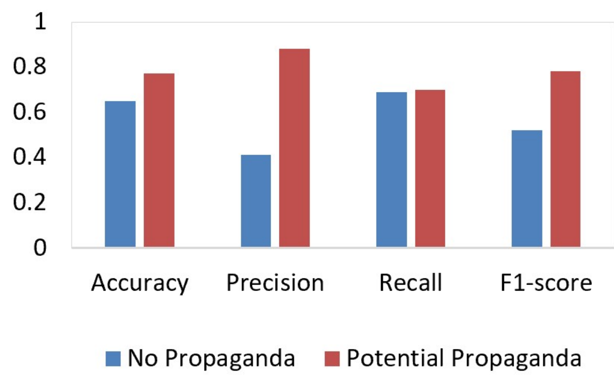

Figure 6: MARBERT classification results

AraBERT model outperformed the MARBERT model with a slight difference in the binary classification accuracy. The accuracy obtained using MARBERT model was 70%, while the accuracy obtained using AraBERT was 71.41%. This is not a significant difference, however, we believe that for multi-label task we need to fine-tune more language models.

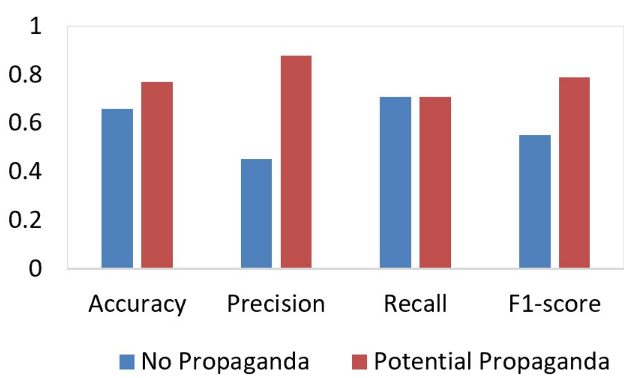

Figure 7: AraBERT classification results

5.2 Results of Baseline Models

We compared our results with two well known text classification models, a supervised CBOW model, which is a vector space model that is used for information retrieval. For this classification task, each headline was represented as a vector [35]. The vector components represent the weights or importance of each word in the headline. As shown in Figure 8, most of our headlines has 6-8 words in their text. As such using this model is relevant since we don't have headlines with varied lengths.

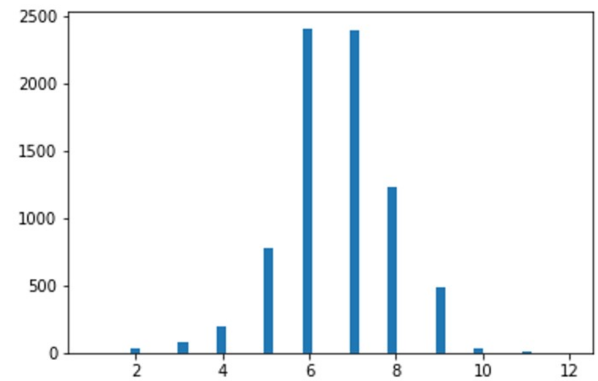

Figure 8: Distribution of the number of tokens in the dataset, x-axis: the number of tokens, y-axis: the number of headlines

We also compared with another baseline model. We built a Bi-LSTM model for the propaganda classification step [32]. For this task we used the Long-short Term Memory (LSTM), which is a type of Recurrent Neural Networks (RNNs). LSTMs have the ability to remember previous inputs for a long period of time which allows them to capture patterns in the textual data that are not visible in a single input.

LSTM units include a memory cell that can maintain information in memory for long periods of time. The LSTM cell has three gates namely input gate, forget gate, and output gate, which controls the information flow through the cell and the neural network. At time t, the input is the headline x_t, the hidden layer output is h_t and its former output is h_{t-1}. The input gate state is denoted by i_t, the output gate state is denoted by o_t, and the cell input state is denoted by $\widetilde{c_t}$. The structure of the LSTM cell indicates that both c_t and h_t are transmitted to the next neural network in RNN. To calculate c_t and h_t, the following equations are used to also calculate the state in each of the three gates' and the cell input state.

Input gate:

$$i_t = \sigma\left(W_i x_t + U_i h_t - 1 + b_i\right) \tag{1}$$

Forget gate:

$$f_t = \sigma\left(W_f x_t + U_f h_t - 1 + b_f\right) \tag{2}$$

Output gate:

$$o_t = \sigma\left(W_o x_t + U_o h_t - 1 + b_o\right) \tag{3}$$

Cell input state:

$$\widetilde{c_t} = tanh\left(W_c x_t + U_c h_t - 1 + b_c\right) \tag{4}$$

where W_i, W_f, W_o and, W_c are the weight matrices connecting x_t to the three gates and the cell input state. U_i, U_f, U_o, and U_c are the weight matrices connecting previous cell output state $h_t - 1$ to the three gates and the cell input state. b_i, b_f, b_o, b_c are the bias vectors of the three gates and the cell input state, σ represents the gate activation function and tanh represents the hyperbolic tangent function. The following formula calculates cell output state:

$$C_t = i_t \cdot \widetilde{c_t} + f_t \cdot C_{t-1} \tag{5}$$

Finally, the hidden layer output can be calculated as

$$h_t = o_t * tanh(C_t) \tag{6}$$

Where $i_t, f_t, \widetilde{c_t}$, and C_t have the same dimension.

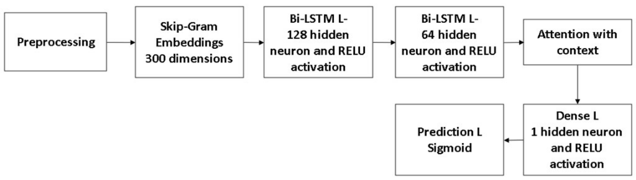

Figure 9: Bi-LSTM model summary

A summary of the LSTM model used for classification is shown in Figure 9. We first initialized the Skip-gram embeddings to extract the context features as the first layer with the input of 300 input dimensions embeddings. We used two Bi-LSTM layers stacked on top of the word embeddings' layer, with 128 and 64 units, respectively. The attention mechanism was used on the top of the Bi-LSTM. The attention with context layer was augmented by two fully connected layers with 1 unit to predict propaganda. The model was built using Keras API with SGD optimizer with 0.01 learning rate, coupled with a batch size of 32.

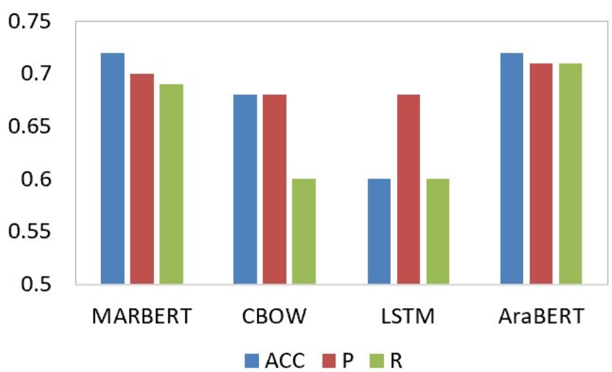

Figure 10: Comparison of different classification models

The results of all four classification models are shown in Figure 10. Both transformer-based language models yield relatively better results in terms of Accuracy, Precision, and Recall.

6 SOCIO-EMOTIONAL ANALYSIS IN PROPAGANDA HEADLINES

6.1 Emotion Labels

The conflicts in the Middle East and Especially in the the Arab World has been having significant impacts on the psychological and mental aspects of the Arab social media audience. This is due to the solidarity that unite these nations and the feeling of guilt and desire to help in some hard circumstances. In this research we labeled 5825 headlines to identify the potential emotions that the audience might endure when being exposed to those headlines. The preliminary statistics in Figure 11, show that the Arab audience will most likely experience a wide range of *anticipation* followed by *anger* and *surprise* emotions. This is explained based on the contextual formation of the headlines which always contain obscurity and vagueness as the crises in the Middle East are not always controlled by the local authorities. The majority of feelings are concentrated under *anger*, *anticipation*, and even *confusion*. The *confusion* can be justified by the rapid pace of changes in those countries.

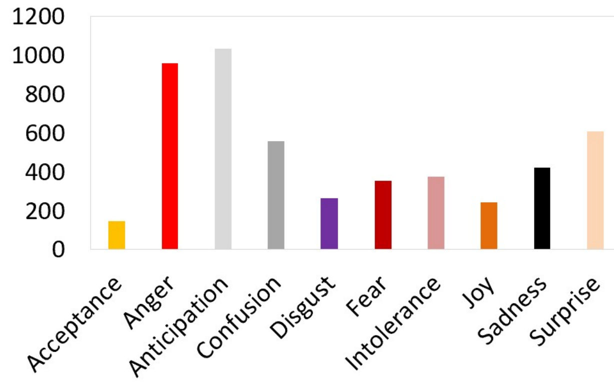

Figure 11: Frequency of occurrence of emotions

6.2 Summary of Findings

The following findings

- It is observed that out of all emotions *acceptance* and *joy* were at minimum, which is understandable considering the current crises in the region and the opinion polarities on the political issues.
- It is also observed from analyzing the dataset, that there are certain keywords that are associated with certain emotions based on the occurrence. For example, words such as *initiatives*, *attempts*, *rehabilitation* are mainly associated with *acceptance* and *anticipation*. While on the other hand, words such as *refugees*, *dead*, *killing*, *bombing* are associated with *anger*, *sadness*, and *intolerance*.
- An interesting finding, on the other hand, was with occurrence of words such as *martyrs* in contexts related to attacks on Israeli people, which mainly reflect a *joy* among the investigated emotions. Words such as *Aqsa* resulted in *anger*, *fear*, *disgust*, and *sadness*.
- Most of the headlines that included the United States, Russia, or Iran also included reactions such as *anticipation*, *confusion* and in *anger* in some cases.

6.3 Ongoing Work

We are currently working on creating classification models to detect emotions in response to tweets when different propaganda categories are detected [3]. Specifically, we will conduct a multi-label classification of propaganda and the response tweets. We are mainly interested in labeling the headlines into the following categories, then conducting a correlation study between emotions and each of the those categories:

- Loaded Language: influencing an audience by using words or phrases that have strong emotional connotations (either positive or negative).
- Name Calling / Labeling: Labeling the campaign's target with certain names or adjectives to agitate the audience either positive or negative.
- Repetition: Repeating the same context/message multiple times so that it is normalized between people.
- Exaggeration or Minimization: making certain events or decisions appear to be larger or smaller than they originally were.
- Doubt: questioning the capability of someone/party and questioning the credibility of certain decisions.
- Appeal to fear/Prejudice: convincing people with a certain idea/option in favor of other options/ideas.
- Flag Waving: adjuration of national beliefs and country belonging.
- Slogans: emotional appeals with strictly brief and effective phrases.
- Whataboutism: Neglecting opposite ideas and thoughts by accusing their holders with bad qualities such as hypocrisy.
- Red Herring: diverting the audience's attention from the main topic being discussed with irrelevant details or surrounding events.

7 CONCLUSIONS AND FUTURE WORK

In this research, we proposed an approach for studying the socio-emotional effects on the audience in the Middle East considering their exposure to headlines with potential propaganda. The study, to the furthest we know, is one of the first attempts in this area in general, and on the Arabic language in specific. We relied on MARBERT and AraBERT models to classify data that was labeled by domain experts. The second stage was an expert-driven multi-labeling process which was conducted by experts with social and psychosocial expertise to detect the type of emotions that might be shown when being exposed to such types of headlines. In addition to our ongoing work on multi-label classification of propaganda, we plan to study targeted propaganda attacks, which are attacks against specific countries, groups, or entities then compared different language models on general versus targeted propaganda.

ACKNOWLEDGMENTS

This material is based upon research supported by the U.S. Office of Naval Research under award number N000142212549.

REFERENCES

[1] Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2020. ARBERT & MARBERT: deep bidirectional transformers for Arabic. *arXiv preprint arXiv:2101.01785* (2020).

[2] Mohamed Abuagla. 2023. *Aljazeera Network.* Retrieved March 12, 2023 from https://www.aljazeera.net/

[3] Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 Shared Task on Propaganda Detection in Arabic. *arXiv preprint arXiv:2211.10057* (2022).

[4] Ibrahim Aljarah, Maria Habib, Neveen Hijazi, Hossam Faris, Raneem Qaddoura, Bassam Hammo, Mohammad Abushariah, and Mohammad Alfawareh. 2021. Intelligent detection of hate speech in Arabic social network: A machine learning approach. *Journal of Information Science* 47, 4 (2021), 483–501.

[5] Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104* (2020).

[6] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. 2019. Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining* 9 (2019), 1–11.

[7] Michael Chan. 2016. Social network sites and political engagement: Exploring the impact of Facebook connections and uses on political protest and participation. *Mass communication and society* 19, 4 (2016), 430–451.

[8] Tanmay Chavan and Aditya Kane. 2022. Large Language Models for Multi-label Propaganda Detection. *arXiv preprint arXiv:2210.08209* (2022).

[9] Long Chen, Jianguo Chen, and Chunhe Xia. 2022. Social network behavior and public opinion manipulation. *Journal of Information Security and Applications* 64 (2022), 103060.

[10] Innocent E Chiluwa. 2019. Deception in online terrorist propaganda: A study of ISIS and Boko Haram. In *Handbook of research on deception, fake news, and misinformation online.* IGI Global, 520–537.

[11] Blane R Clark. 2010. Information operations as a deterrent to armed conflict. *Science* 90, 3 (2010), 97–104.

[12] Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeno, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP).* 5636–5646.

[13] Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, and Yelena Mejova. 2017. Seminar users in the Arabic Twitter sphere. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9.* Springer, 91–108.

[14] Clement Delangue and Julien Chaumond. 2023. *HuggingFace.* Retrieved March 12, 2023 from https://huggingface.co/

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[16] Daniela V Dimitrova and Jörg Matthes. 2018. Social media in political campaigning around the world: Theoretical and methodological challenges. , 333–342 pages.

[17] Douglas Farah and Alexa Tavarez. 2021. *Iran in Latin America: Malign Alliances," super Spreaders," and Alternative Narratives.* National Defense University Press.

[18] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. 2020. Characterizing social media manipulation in the 2020 US presidential election. *First Monday* (2020).

[19] R Kelly Garrett, Brian E Weeks, and Rachel L Neo. 2016. Driving a wedge between evidence and beliefs: How online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication* 21, 5 (2016), 331–348.

[20] Yevgeniy Golovchenko. 2022. Fighting propaganda with censorship: A study of the Ukrainian ban on Russian social media. *The Journal of Politics* 84, 2 (2022), 639–654.

[21] Rebecca Goolsby. 2010. Social media as crisis platform: The future of community maps/crisis maps. *ACM Transactions on Intelligent Systems and Technology (TIST)* 1, 1 (2010), 1–11.

[22] Imane Guellil, Ahsan Adeel, Faical Azouaou, Fodil Benali, Ala-eddine Hachani, and Amir Hussain. 2018. Arabizi sentiment analysis based on transliteration and automatic corpus annotation.. In *WASSA@ EMNLP.* 335–341.

[23] Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, and Damien Nouvel. 2021. Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences* 33, 5 (2021), 497–507.

[24] Mustafa Hamarneh. 2023. *Jordan Media Institute.* Retrieved March 12, 2023 from https://jmi.edu.jo/

[25] Michael A Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. 2021. A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 2545–2568.

[26] Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, and Preslav Nakov. 2022. The spread of propaganda by coordinated communities on social media. In *14th ACM Web Science Conference 2022.* 191–201.

[27] Andrew H Johnston and Gary M Weiss. 2017. Identifying Sunni extremist propaganda with deep learning. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI).* IEEE, 1–6.

[28] Lisa Kaati, Enghin Omer, Nico Prucha, and Amendra Shrestha. 2015. Detecting multipliers of jihadism on twitter. In *2015 IEEE international conference on data mining workshop (ICDMW).* IEEE, 954–960.

[29] Tarek Kanan, Odai Sadaqa, Amal Aldajeh, Hanadi Alshwabka, Shadi AlZu'bi, Mohammed Elbes, Bilal Hawashin, Mohammad A Alia, et al. 2019. A review of natural language processing and machine learning tools used to analyze arabic social media. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).* IEEE, 622–628.

[30] Gary King, Benjamin Schneer, and Ariel White. 2017. How the news media activate public expression and influence national agendas. *Science* 358, 6364 (2017), 776–780.

[31] Srijan Kumar, Justin Cheng, Jure Leskovec, and VS Subrahmanian. 2017. An army of me: Sockpuppets in online discussion communities. In *Proceedings of the 26th international conference on world wide web*. 857–866.

[32] Chenbin Li, Guohua Zhan, and Zhihua Li. 2018. News text classification based on improved Bi-LSTM-CNN. In *2018 9th International conference on information technology in medicine and education (ITME)*. IEEE, 890–893.

[33] Darren L Linvill, Brandon C Boatwright, Will J Grant, and Patrick L Warren. 2019. "The Russian are hacking my brain!" investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior* 99 (2019), 292–300.

[34] Darren L Linvill and Patrick L Warren. 2020. Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication* 37, 4 (2020), 447–467.

[35] Bing Liu. 2020. Text sentiment analysis based on CBOW model and deep learning in big data environment. *Journal of ambient intelligence and humanized computing* 11 (2020), 451–458.

[36] Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. *arXiv preprint arXiv:2007.08024* (2020).

[37] Michael J Mazarr, Abigail Casey, Alyssa Demus, Scott W Harold, Luke J Matthews, Nathan Beauchamp-Mustafaga, and James Sladden. 2019. Hostile social manipulation present realities and emerging trends. (2019).

[38] Shannon C McGregor. 2019. Social media as public opinion: How journalists use social media to represent public opinion. *Journalism* 20, 8 (2019), 1070–1086.

[39] Joe Murphy, Michael W Link, Jennifer Hunter Childs, Casey Langer Tesfaye, Elizabeth Dean, Michael Stern, Josh Pasek, Jon Cohen, Mario Callegaro, and Paul Harwood. 2014. Social media in public opinion research: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly* 78, 4 (2014), 788–794.

[40] Shuteng Niu, Yongxin Liu, Jian Wang, and Houbing Song. 2020. A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence* 1, 2 (2020), 151–166.

[41] Nadia Oweidat. 2023. *The Russian Propaganda in Arabic Hidden from the West*. Retrieved March 12, 2023 from https://www.washingtoninstitute.org/policy-analysis/russian-propaganda-arabic-hidden-west

[42] Robert Plutchik. 1982. A psychoevolutionary theory of emotions. *Social sciences information* 21, 4-5 (1982), 529–553.

[43] Mohammad Salameh, Houda Bouamor, and Nizar Habash. 2018. Fine-grained arabic dialect identification. In *27th International Conference on Computational Linguistics, COLING 2018*. Association for Computational Linguistics (ACL), 1332–1344.

[44] Yasir Shaker. 2012. *Aravic Cyber Bullying*. Retrieved March 27, 2013 from https://www.kaggle.com/datasets/yasserhessein/arabic-cyberbullying

[45] Shabnoor Siddiqui, Tajinder Singh, et al. 2016. Social media its impact with positive and negative aspects. *International journal of computer applications technology and research* 5, 2 (2016), 71–75.