Bachelor's Thesis

Characterizing spreaders of Russian propaganda

Institute of Artificial Intelligence (AI) in Management
Ludwig-Maximilians-Universität München

Mihnea Andrei Stefanescu

Prof. Dr. Stefan Feuerriegel
Academic Advisor: Dominik Bär

Munich, August 26th, 2024

Abstract

This work examines the spread and characteristics of Russian propaganda on TikTok, in the context of Ukrainian invasion. The goal of the research is to provide a detailed description of propaganda videos and demonstrate the difference between them and non-propaganda content on the platform. By combining the content analysis, sentiment analysis, and machine-learning methods, the research probes into how these videos employ the particular hashtags and narratives to sync with the Russian side. Among the important findings was the use of targeted hashtags like #istandwithrussia and #putin besides brief descriptions, to force TikTok's algorithm to elevate the video. Rather, non-propagandistic materials supply more elaborated information. Sentiment analysis showed a mostly unfavorable mood in both video categories, which reflects the emotional tension of conflict-related content. Nonetheless, the work also points out a considerable number of limitations in the detection of sentiments and classification of propaganda in the small and informal text that is usually found in TikTok. These findings indicate the need to develop more advanced tools for analytical purposes to cope with the complexities of misinformation in social media platforms. Propaganda's spreading on digital platforms is the main issue explored in this research and note that the algorithm plays an important role in content visibility. Future research should focus on improving sentiment analysis models for social media and studying the influence of algorithms on content dissemination.

Table of Contents

List of Figures

List of Tables

1 Introduction

Propaganda is the spreading of ideas, information, or rumor for the purpose of helping or injuring an institution, a cause, or a person (Dictionary.com, 2024). Social media platforms play a critical role in propaganda spreading as they make it very easy for propaganda spreaders to reach big masses of people. Nowadays, platforms such as TikTok play a critical role in war efforts (e.g. Russia-Ukraine war), as they provide a very efficient way of pushing certain narratives or changing publics opinions.

On February 24, 2022 Russia launched the invasion of Ukraine. Soon after, all social media platforms were flooded with propaganda. Given the reach that social media has, it represents a very important tool in propaganda efforts.

Due to it being a relatively new platform, TikTok is still understudied. Not much has been studied about the russian propaganda on Tiktok regarding the war. The primary goal of this thesis the characterization of russian propaganda spreaders and how ti differs from normal users on the platform. Given that TikTok focuses on video content and not on text content (e.g. X or Facebook), it can prove to be quite difficult to detect propaganda with great accuracy only using NLP (Natural Language Processing) techniques. Despite extensive research on propaganda in traditional media and other social media platforms, there is a gap in understanding how propaganda tactics are employed on TikTok and how they can be detected.

In this paper we characterize spreaders of russian propaganda by understanding common techniques and trends in propaganda videos. It employes a mixed-method approach by taking into consideration multiple criterias such as descriptions, hashtags, language used, region codes, topics of videos and more.

1.1 Research Questions

To address the problem, the following research questions guide this thesis:

1. **RQ1**: How can propaganda be detected on TikTok by only using text-format informations?

2. **RQ2**: What are the characteristics of russian propaganda spreaders regarding the invasion of Ukraine?

3. **RQ3**: How does it differ from normal content about the war?

The first research question focuses on propaganda classification methods by only using text data such as descriptions, hashtags, comments, transcriptions etc. This research analyzes whether using a zero-shot classifier is reliable, considering the limited available data about TikTok videos. What will be done in the case where there isn't enough data available for a TikTok video (e.g. very short descriptions)? It checks whether more context for a video is needed and if using a zero-shot classifier to detect propaganda is sufficient enough in order to satisy the goal of the thesis.

The second question aims to characterize spreaders of russian propaganda and create a profile for the propaganda spreaders. What similarities are there between all the propaganda videos? Do the propaganda videos use a specific tone in their posts? Do they focus

on certain topics? What are the naratives pushed forwards by the russian propaganda? We answer this question by analyzing in more detail the used language and what naratives are being pushed forwards the most.

The third question focuses on comparing the propaganda videos to the non-propaganda videos. It does so by comparing all the available data such as descriptions, reach, transcriptions, region codes etc. What is different for the propaganda videos compared to the other content on the platform? How can we make the difference between propaganda and non-propaganda content regarding the war in Ukraine?

In order to answer the three questions, this research aims to classify a dataset of TikTok videos and analyze it. The final goal is to make a characterization of the russian propaganda spreaders regarding the war in Ukraine.

2 Related Work

For a better understanding of the the work done in this thesis, explaining what TikTok is and what methods were used in order to detect and characterize russia propaganda is needed.

2.1 TikTok

TikTok is a short-form video sharing platform owned by the Chinese internet company ByteDance and was launched in September 2016. It has become one of the most popular and the fastest growing social media platform in a very short period of time, reaching 2 billion mobile downloads by April 2020. Most of TikTok users are young people and according to Cloudflare, TikTok was the most searched web-domain in 2021, above Google and Facebook (Cloudflare, 2021). According to the parent comapny ByteDance, TikTok has reached 1 billion monthly users in September 2021 and has continued to grow since then. To put these numbers into perspective, almost 20% of the internet users are also monthly active TikTok users. The following image shows the growth in monthly active users over time:

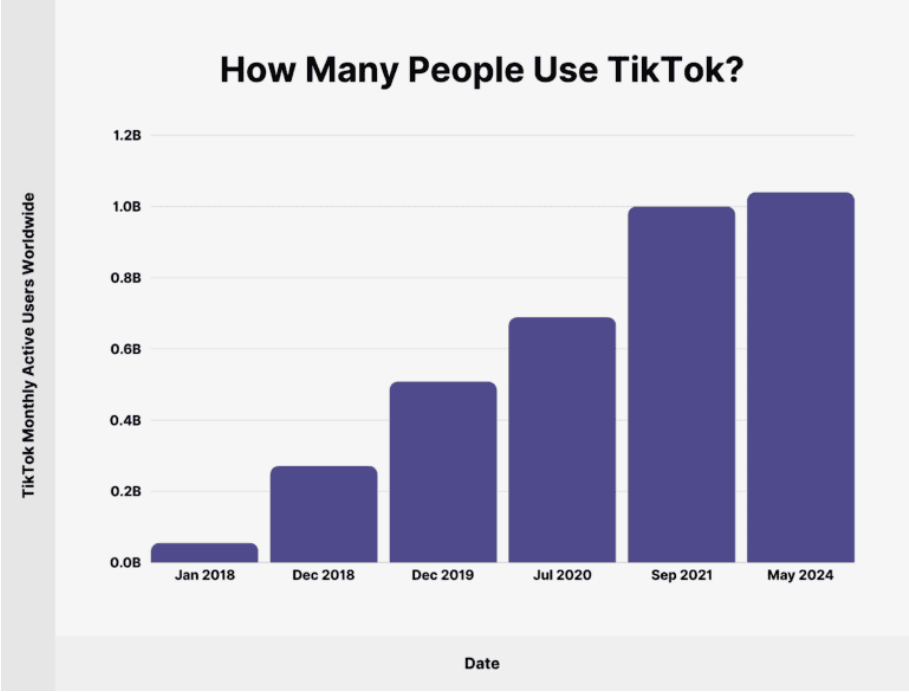

Figure 1: How Many People Use TikTok? Source: Backlinko (n.d.).

Another very important aspect is the age demographic of the platform. According to Oberlo (Oberlo, 2024) 70% of the platforms users are between the ages of 18 and 34 making it extremly popular among young people. The following image shows the users distribution by age category:

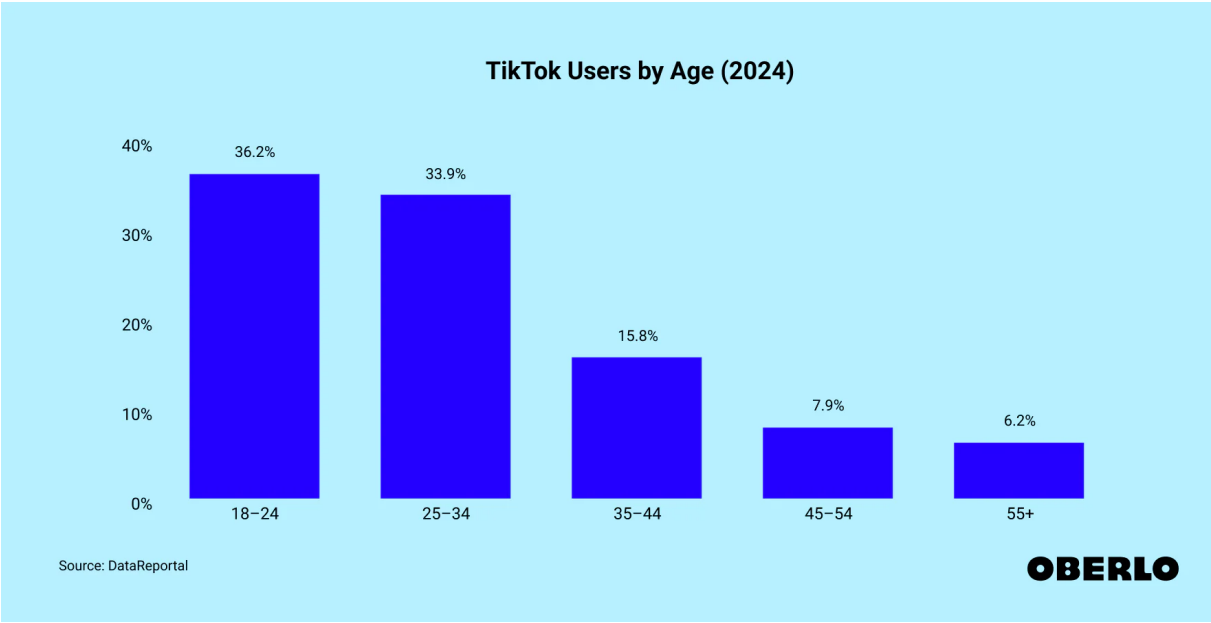

Figure 2: Monthly users growth

2.2 Literature Review

In the following we review already existing literature regarding the characteristics of russian propaganda spreaders and how it differs from non-propaganda content on social media.

Propaganda has been a part of human history for a long time and has been used mostly in ideological or religious wars. The first records of propaganda use can be traced back to ancient Greece, the Roman Empire and early Christians. (Garth S. Jowett, 2012). We often have the impression that propaganda is something new that was invented in the 21st century and we often associate it with the digital age. Although nowadays spreading propaganda is easier and more efficient than it ever was, the first records of mass spreading of propaganda can be traced back to the invention of the printing press. The printing press was quickly adopted by Marthin Luther, a german priest, theologian, author, hymnwriter and professor, in order to fight against the Catholic Church (Garth S. Jowett, 2012). According to Garth S. Jowett and Victoria O'Donnell there are 10 steps in the identification of propaganda: " identification of ideology and purpose, identification of context, identification of the propagandist, investigation of the structure of the propaganda organization, identification of the target audience, understanding of media utilization techniques, analysis of special techniques to maximize effect, analysis of audience reaction, identification and analysis of counterpropaganda, and completion of an assessment and evaluation".

Nowadays most of the propaganda spreaders put a big effort into flooding the social media networks with propaganda. A very good example are the propaganda efforts of the russian gouvernment following the invasion in Ukraine. Although it's sometimes hard to differentiate between a russian propaganda video on TikTok and a persons opinion regarding the war, there are studies who try to characterize russian propaganda spreaders.

Such a paper is (Klim Kireev, 2024) which aims to characterize propaganda accounts on Telegram. According to the paper, a propaganda account usually has a lifespan of less than a day, after which they become inactive. Another characteristic of propaganda is the language used in the messages. Usually normal users tend to reply or comment with shorter texts (e.g: "Yes", "No", "Why?") while propaganda accounts replies are longer.

In the beginning of the war hashtags supporting the war went viral on all social media platforms. This could be a very good indicator of propaganda content, as a lot of pro-russian hashtags have been used. For example: "#istandwithrussia","#standwithrussia", "#istandwithputin", "#standwithputin" were one of the most used hashtags by the accounts sharing russian propaganda. (Dominique Geissler, 2023).

According to another study, one of the most used weapons for spreading russian propaganda are URL's which are included in messages/descriptions and point mostly towards the propaganda spreaders own content. Repeating the same messages and naratives on mutiple social media platforms aims at altering the users perception of reality and to make them not distinguish between reality and fiction (Larissa Doroschenko, 2021).

Due to TikTok being a very new platform and after researching all the publicly available literature, we find the following gap in the literature. Most of the work done so far regarding russian propaganda profiling focuses on social media platforms such as Facebook, X (formerly Twitter), Telegram. This thesis aims to fill that gap by analyzing what strategies are employed by russian propaganda spreaders on TikTok. Due to TikTok being a video-focused social media platform, using traditional Natural Language Processing techniques can prove to be more challenging as users rely on video content to share their opinions and messages.

3 Methods

As mentioned in the beginning of the thesis, the aim is to characterize the spreaders of russian propaganda. We do so by answering the three research questions in Section 1.1. The following section explains how the data has been extracted from TikTok and what methods have been used to characterize russian propaganda spreaders. In order to better understand and analyze the fetched data, a database with three tables has been created, containing three different datasets. The database has the following structure:

- *"data"* table consisting of $\tilde{2}$0.000 TikTok posts with relevant data.
- *"comments"* table consisting of $\tilde{1}$8.000 comments from propaganda and non propaganda videos.
- *"transcriptions"* table consisting of 1000 transcript records of TikTok videos.

3.1 Dataset construction

The data used for this study was collected from the social media platform TikTok using the TikTok Researcher API (TikTok, n.d.) and saved in a database. As mentioned above, the database consists of three tables: *data, comments* and *transcriptions* which can be better visualised in the following diagramm:

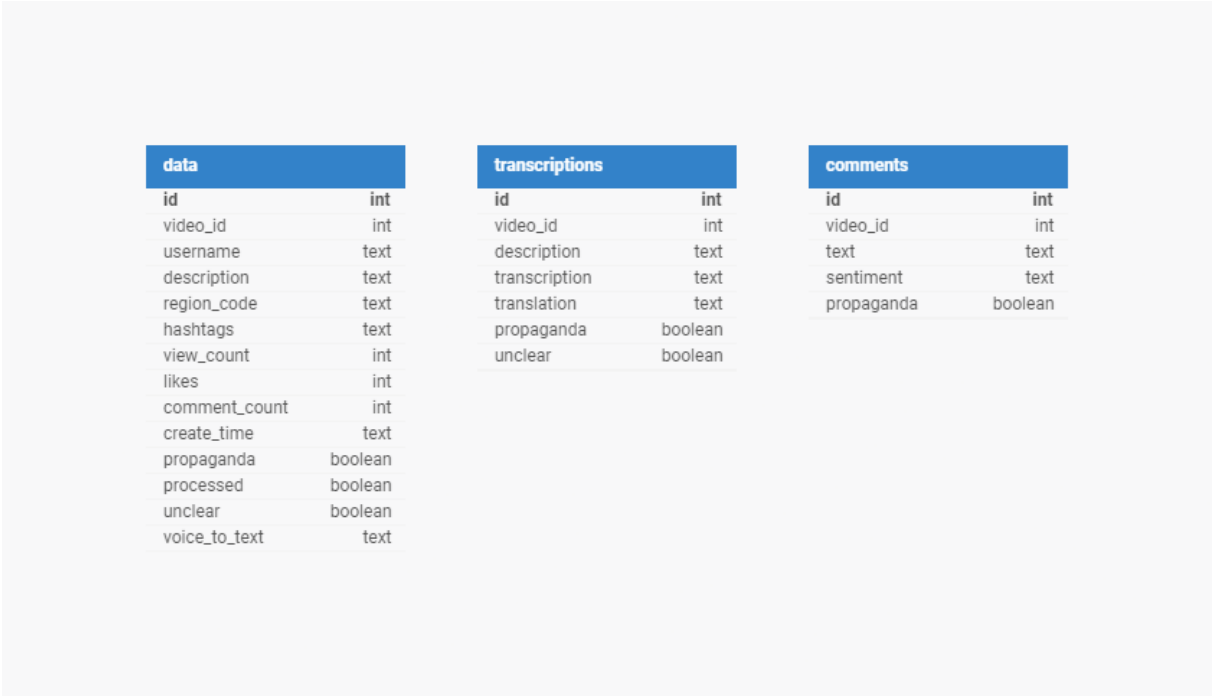

Figure 3: Dataset Tables

3.2 Data Preprocessing

In order to analyze only relevant data for our study, the preprocessing step has been done before storing the data in the database by filtering it. It is important to note that this technique only saves a chunck of the propaganda videos, not all of them. The purpose of this study is to analyze propaganda videos and identify common characteristics and techniques of propaganda spreaders.

The first step in the data preprocessing was to create a table called *"data"* where all of the fetched video informations would be saved, such as: video id, description, username, posting date, hashtags etc. (see graphic above). Considering that the russian invasion started on 24th of Februrary 2022, only TikToks posted after 20th of February have been queried. After that, the focus was on quering TikToks only containing relevant hashtags and keywords. The first hashtag/keyword queried was "istandwithrussia" which was a trending hashtag in the beginning of the war (Dominique Geissler, 2023). Afterwards a subsample of 500 random videos were analyzed and additional hashtags were manually chosen. Thus the following hashtags were identified to be pro russian and could hint towards other propaganda videos: "russia", "putin", "vladimirputin", "iloverussia", "wagner", "istandwithputin", "istandwithrussia", "loverussia", "prayforrussia", "donbas", "bakhmut", "slavarussia", "naziukraine", "ukrainecorrupt". Further the language of the descriptions was detected and only the ones with an english description were saved. The final step in the creation and preprocessing of the "data" table was refining the already saved videos in the database. This implied converting the Unix epoch timestamp used by TikTok API to a human readable format, converting all the region codes to capital letters (EX.: us to US, de to DE) and also removing the duplicate video ids. No further preprocessing such as removing emojis or special characters from the descriptions was needed, as the model used for classifying the videos is capable of reading and interpreting special characters such as emojis.

Down below we can see an illustration of how the data preprocessing works (for the data dataset):

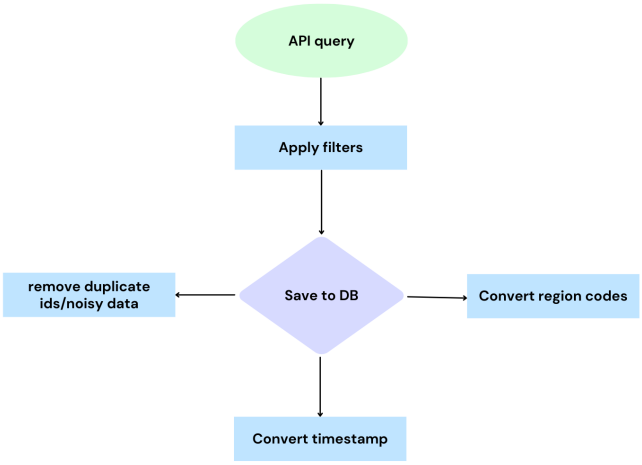

Figure 4: Data Preprocessing Pipeline

Analyzing videos only based on the descriptions or hashtags can sometimes be quite challenging as TikTok is a video-sharing platform. It focuses mainly on video content and a lot of the videos have very short descriptions or very general hashtags, which can be dificult to identify as pro/anti -russian. In order to have better accuracy when classifying the videos, all the entries in the database were checked for transcriptions. After going through all of the 20.000 saved videos, approximately 1000 of them also had transcriptions. A new table named "transcriptions" was created in which all of the transcriptions were saved. Further preprocessing of the table was needed as a lot of the transcriptions were not in english. For this a popular translation model was used (**?**).

The third and final step was to extract comments from the candidates in the transcriptions table in order to analyze them. This has been done by querying a random sample of $\tilde{1}00$ propaganda and $\tilde{1}00$ non-propaganda video ids and fetching all of their comments. After removing all the empty comments we were left with 7600 comments made on propaganda videos and 5300 made on non propaganda videos.

3.3 Models

In order to properly interpret and analyze the collected data, various models and techniques for machine learning tasks were used. The scope of using these models and techniques is to find out more about the propaganda spreaders and to properly characterize them.

The following techniques were used:

- Zero-Shot Classifier: MoritzLaurer/deberta-v3-large-zeroshot-v2.0, (MoritzLaurer, 2024)
- Sentiment Analysis: cardiffnlp/twitter-roberta-base-sentiment-latest, (CardiffNLP, 2024)
- Topic Extraction: Latent Dirichlet Allocation (LDA)

3.3.1 Zero-Shot Classification

Given that there aren't any open source models that are trained to detect russian propaganda regarding the invasion in Ukraine, but also because fine-tunning a model for this task would require lot of computational power and ressources, a zero-shot classifier was used. A zero-shot classifier is a machine learning technique in Natural Language Processing where a model classifies data into labels/categories it has not been trained on yet. Unlike traditional classification models which require a lot of labeled data, a zero-shot classifier uses pre-existing knowledge to classify data.

The model used in this thesis, MoritzLaurer/deberta-v3-large-zeroshot-v2.0, is a fine-tunned version of the transformers model DeBERTa. This speific model has been used due to DeBERTa's capability of innovative use of disentangled attention, which helps the model better capture the relationships between words in a sentence, and its enhanced decoding process, which improves its language understanding capabilities. The model is designed to handle a wide range of zero-shot classification tasks in NLP, making it a powerful tool for text categorization and more without needing task-specific labeled data. The following image shows how deberta-v3-large-zeroshot-v2.0 performs in comparison to other popular zero-shot classifiers:

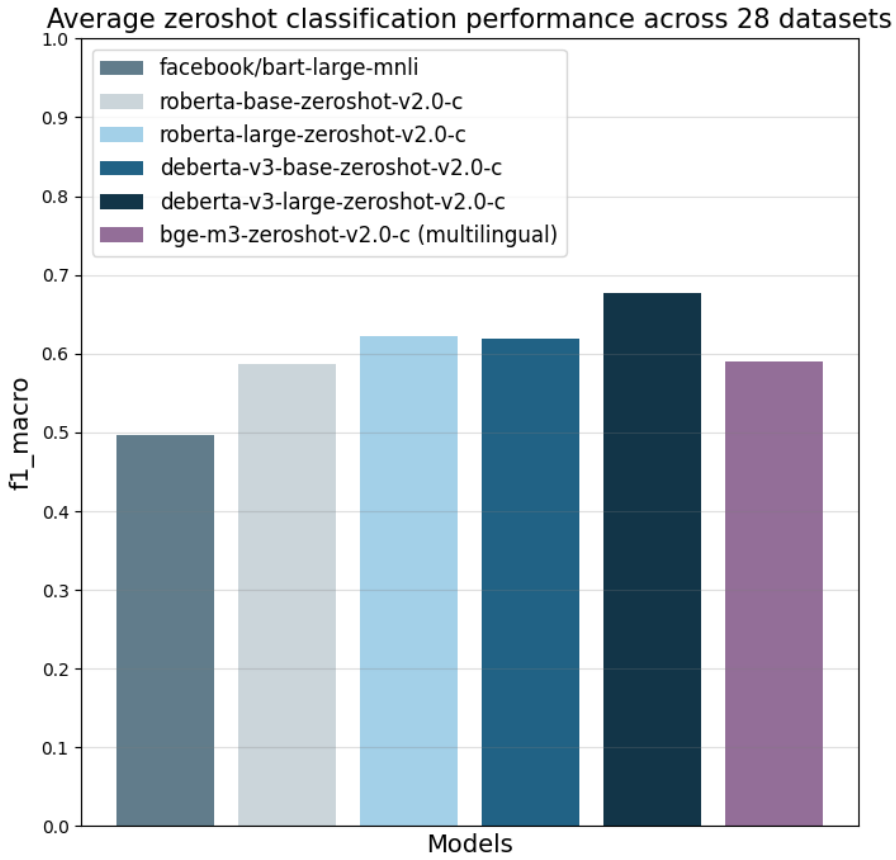

Figure 5: Models performance
(Huggingface, n.d.)

The following settings have been chosen for the classifier:

- Hypothesis Template: *This text is ... regarding the Russian invasion of Ukraine*
- Labels: *Pro Russia, Anti Russia, Unclear*

After experimenting with multiple prompts and labels, the above ones were identified as having the best results. Considering this, the zero-shot classifier was run on all of the 20.000 descriptions from the "data" table in the database.

Given that a lot of TikToks have very vague or general descriptions, the classifier has been also run on the "transcriptions" table which contains the description and transcription of the videos. Having more context/text up to 400 words makes the model perform better and return more accurate results.

3.3.2 Aspect-Based Sentiment Analysis

In order to analyze the sentiments towards specific topics, Aspect-Based Sentiment Analysis has been used. Before understanding what ABSA (Aspect-Based Sentiment Analysis) is and how it has been implemented, it's important to clarify two subtasks: Sentiment Analysis and Named Entity Recognition (NER).

Sentiment analysis is a powerful technique in Natural Language Processing (NLP) used to determine the attitude or emotional tone of a text, whether it's positive, negative, or neutral. This technique is widely used in various applications, such as analyzing social media posts, customer reviews and feedback etc. Sentiment analysis works by first preprocessing the text(Tokenization, Lowercasing, Removing Stopwords, Stemming/Lemmatization). It then proceeds to extract features such as TF-IDF and Word Embeddings and finally it classifies the text in a sentiment category.

Named Entity Recognition (NER) is a subtask of information extraction in NLP that aims to locate and classify named entities mentioned in unstructured text into predefined categories such as: people, organizations, locations etc.

Aspect-Based Sentiment Analysis is a task in Natural Language Processing that determines the sentiment of a text with respect to a specific aspect. ABSA (Aspect-Based Sentiment Analysis) is a new method that came as an improvement to the limitations of traditional sentiment analysis techniques. Unlike standard sentiment analysis, ABSA allows us to understand the sentiment expressed towards different aspects of a text. (Ezzameli & Mahersia, 2023)

In this thesis, Aspect-Based Sentiment Analysis (ABSA) has been used in order to identify sentiments towards specific relevant aspects, such as: "Russia", "Ukraine", "Nato", "Putin" and "Zelensky". The model employed for this task is deberta-v3-base-absa-v1.1 (yangheng, 2024b), which is a version of DeBERTa fine-tuned with English datasets from ABSADatasets (yangheng, 2024a).

First the extracted text is cleaned by converting all the characters to lowercase to ensure uniformity and by removing URLs and any non-alphanumeric characters except for emojis and spaces. This helps remove the noise and keep only relevant content improving the accuracy of the model. Next the ABSA model and its tokenizer are loaded using Hugging Face's transformers library. Using the python library "spacy", we perform NER on the text in order to extract relevant entities that match the predefined aspects. If no entities are detected, we check directly for mentions of aspects within the text. Once the entities are identified using NER, we filter them to match the predefined aspects ("Russia", "Ukraine", "Nato", "Putin" and "Zelensky"). This ensures that only the sentiments of relevant aspects are analyzed. With the relevant aspects identified, the ABSA model is then used to determine the sentiment expressed towards each aspect.

Aspect-Based Sentiment analysis has been chosen in this thesis in order to analyze the comments of propaganda and non-propaganda videos, to better understand what impact the propaganda videos have on the regular TikTok users and to also analyze the opinion of the people when seeing propaganda videos. ASBA has been applied on all of the 18.000 comments from the *"comments"* dataset but also on the *"transcriptions"* dataset.

3.3.3 Topic Extraction

Topic extraction is a NLP technique used to identify and extract the main topics/subjects of a text. Unlike traditional classification techniques, topic extraction works in an unsupervised manner. Latent Dirichlet Allocation (LDA) is used in this study in order to extract relevant topics from videos and see what topics are the most common among propaganda videos. LDA is a very popular statistical method which assumes that each document is a mixture of a small number of topics and each topic is a mixture of words.

LDA was applied to the "data" dataset but also to the "transcriptions" dataset as it performs better when there is more text. It performs better with more text because it has more data to infer distinct and coherent topics.

The used LDA method works by first removing stopwords like *and, the* etc. because they do not contribute to the meaning of the sentence. The stopwords corpus from NLTK was used in this study. Next lemmatization is done by reducing words to their base form (e.g., "running" and "run"). Next the extremes of the text are filtered. The most commong word appearing in more than 50% of the text are removed but also the very rare ones. This contributes to removing the noise. Next, the corpus is prepared by converting each document into a bag-of-words (BoW) format using the doc2bow method. In this format each document is represented as a list of tuples where each tuple contains a word ID and its frequency in that document. Finally LDA is applied on all of the data.

For this analysis, the LDA implementation provided by the gensim library in Python were used, with the following settings: The number of topics (num_topics) was set to 5. This parameter determines how many distinct themes the algorithm attempts to identify in the corpus. After experimenting with different values, 5 topics were found to provide the most coherent informations for the dataset "transcriptions". The number of passes ("passes") was set to 100. This parameter controls how many times the LDA model will iterate over the entire corpus during training. Given that the used dataset only contains around 1000 candidates, a large number of passes has been chosen for better accuracy. Afterwards we proceeded with Text Preprocessing (Stop Words Removal, Tokenization and Lemmatization). A dictionary was created from the processed text, mapping each unique word to a unique ID. The corpus was then created by converting the processed documents into a bag-of-words format using this dictionary.

4 Results

The aim of this thesis is to examine and characterize spreaders of russian propaganda on TikTok. For this, an in-depth analysis of the queried data wasd done. The following graphic shows a visual representation of the hashtag trends over time:

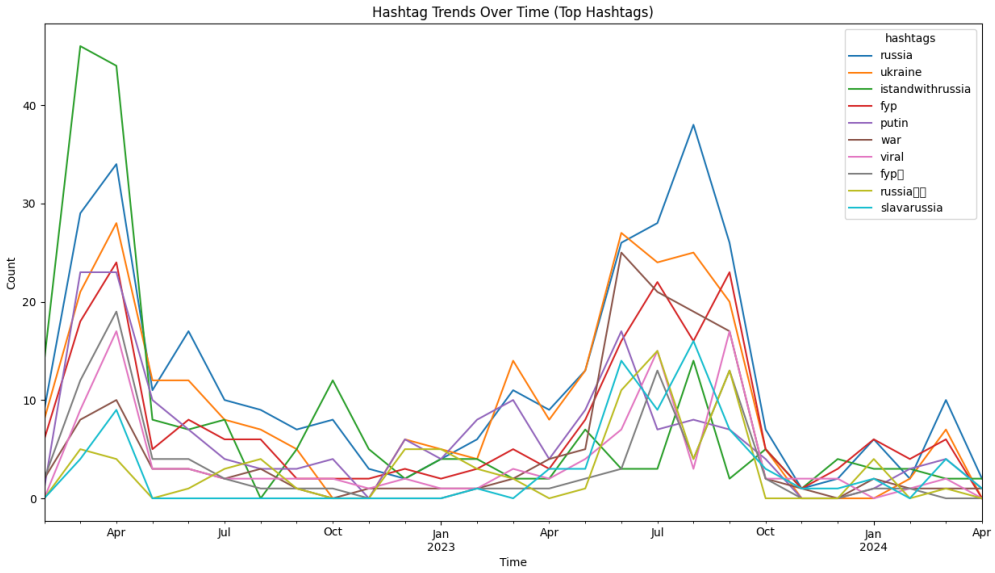

Figure 6: Hashtag Trends

As we can see, hashtags such as *"istandwithrussia", "russia", "ukraine", "war"* spiked around April 2022, which is also when the russian invasion of Ukraine started. This spike illustrates how the invasion immediately captured global attention, leading to a surge of reactions online. Considering that very specific pro-russian hashtags went viral during the beginning of the invasion, it indicates the intensification of russian propaganda campaigns aimed at justifying the invasion and spreading misinformation. Although the reason for the second spike in the graph above is not completly clear, in that period there has been an attempted massive counter-attack on the Ukraine side which could justify the intensification of russian propaganda, trying to undermine the scale of the counter-attack and to portray the russian defense in a good light (Wikipedia, n.d.). The 5 most common hashtags for propaganda videos are: istandwithrussia - 229 occurrences, russia - 197 occurrences, fyp - 138 occurrences, ukraine - 114 occurrences, putin - 100 occurrences. It shows that propaganda videos focus heavily around themes involving Russia, particularly in relation to Ukraine. The use of hashtags like #istandwithrussia and #putin suggests an effort to build a narrative that supports Russian perspectives. The use of popular hashtags like #fyp indicates a strategic effort to reach a broader audience and maximize engagement by exploting TikToks algorithm.

The spike in these specific hashtags may have been further amplified by TikTok's algorithm, which tends to promote content that generates high engagement. This raises questions about how algorithmic curation can inadvertently help the spread of propaganda.

In order to get more in-depth results, the descriptions and transcriptions of the videos have been analyzed. In the table below we can see the average descriptions and transcription length:

Shorter descriptions are used by propaganda videos in order to capture viewers attention quickly and to exploit TikTok's algorithm better, increasing the chances of the videos

	Descriptions	**Transcriptions**
Propaganda	117	3680
Not Propaganda	249	1823

Table 1: average length of descriptions and transcriptions

being shown to a broader or more targeted audience. Propaganda videos are designed to quickly engage and manipulate viewer perception through the extensive use of keywords and hashtags in descriptions. In contrast, non-propaganda videos prioritize comprehensive, informative content with longer transcriptions that provide more context, relying less on descriptions and focusing more on the content itself.

Additionally, we can further analyze the descriptions. The most frequent words from each group (propaganda = 1 and propaganda = 0) have been extracted. The tables below show a comparison between the two groups:

Word	**Frequency (Propaganda = 1)**
ukraine	112
russia	102
fyp	60
war	56
bakhmut	53
nato	43
putin	40
donbass	40
russian	36
soldier	31

Table 2: Most Frequent Words in Descriptions(Propaganda = 1)

Word	**Frequency (Propaganda = 0)**
ukraine	204
bakhmut	177
war	171
fyp	171
worldnews	159
vairal	156
soldier	154
crimea	132
nato	128
kherson	116

Table 3: Most Frequent Words in Descriptions (Propaganda = 0)

Both groups have words like "ukraine," "bakhmut," "war," "fyp," "soldier," and "nato" among the most frequent, which suggests that both propaganda and non-propaganda videos focus heavily on the conflict between Ukraine and Russia, key locations, and military elements. Propaganda videos use words like "russia," "putin," "donbass," and "russian" suggesting a focus on russian related content and figures. It also reflects a narrative centered around Russia.

On the other hand, non-propaganda videos use words like "worldnews," "crimea," and "kherson" suggesting a broader coverage of the conflict, possibly aiming to provide more comprehensive news. Terms like "worldnews" and "vairal" imply an intent to reach a wide audience, highlighting trending content or global attention rather than specific narratives.

4.1 Propaganda Classification

As mentioned in Section 3.3.1, in order to label the data as propaganda or not propaganda, the zero-shot classifier *"MoritzLaurer/deberta-v3-large-zeroshot-v2.0"* was used. Consid-

ering that a lot of the hashstags from the videos are very vague and contradictory (e.g. istandwithrussia, istandwithukraine, fyp), the model has been run on the descriptions of the videos (of both the data and transcriptions dataset) for better results.

After running the model on the "data" dataset, 500 videos have been classified as propaganda, 3600 as not propaganda and 15.000 were labeled as neutral. Almost all of the neutral videos either had a too short of a description (e.g.: *"duet with @donbasher" , "#donbass #italy"*) or were unrelated to the war in Ukraine (e.g.: *"My last day in Russia: eating caviar as last time , follow for more"*). When it comes to the reach of both groups (propaganda and non-propaganda), the non-propaganda videos have more numbers in terms of Views, Comments and Likes. In the table below we can see the reach of propaganda and non propaganda videos:

	Views	**Likes**	**Comments**
Propaganda	4.59 M	0.36 M	0.018 M
Not Propaganda	56 M	4.8 M	0.029 M

Table 4: Total reach of the videos

These numbers represent the total reach only for the collected data. To be able to properly compare the two groups, an average reach for each video is needed. The table below shows how many Views, Likes and Comments each video gets on average:

	Views	**Likes**	**Comments**
Propaganda	10.500	770	41
Not Propaganda	15.500	1300	82

Table 5: Average reach

As we can see, in terms of likes, views and comments the non-propaganda videos have a bigger reach than the propaganda ones. Although the number of views is 50% more for non-propaganda videos, we can see that there are twice as many likes and comments. This shows that the TikTok public resonated more with the non-propaganda videos and that even though the propaganda videos get simmilar number of views, people don't interact with them as much.

Additionally, another interesting measure is the percentage distribution of region codes. The following pie chart shows that the "us" region code accounts for 68% of the total propaganda content. This suggests that a significant proportion of propaganda content is either being created in or targeted towards audiences in the United States. This reveals a strategic focus to influence the population of the West, mainly the United States of America. As mentioned in Section 3.2 only videos with an english description have been saved, which could influence the results of the chart below.

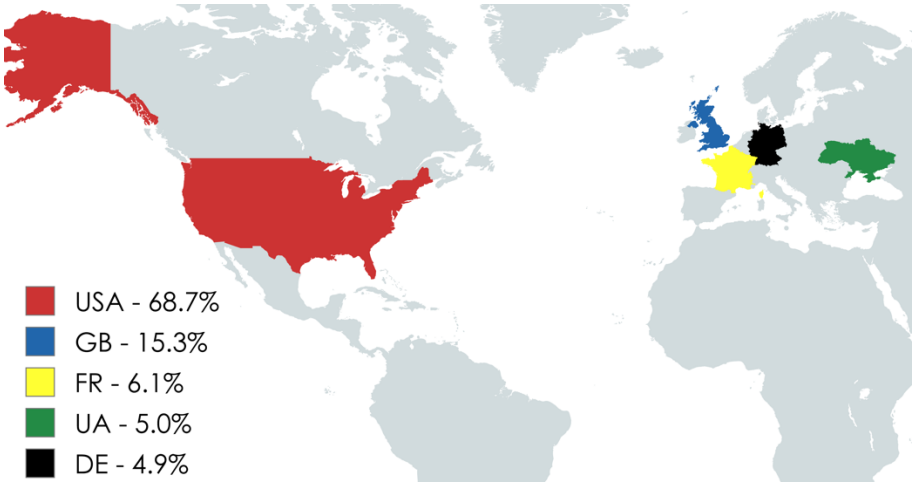

Figure 7: Region Codes Map Chart

When it comes to classifying the data from the "transcriptions" dataset, better accuracy is expected as the zero-shot classifier has more text to analyze as mentioned in Section 3.2. This serves also a robustness check for the results on the whole "data" dataset.

Out of all the 1037 videos that contained a transcription, 134 have been classified as propaganda and 211 as non-propaganda. Approximately 12% of the videos from the "transcriptions" dataset which were labeled as propaganda or not propaganda, had different results after running the classifier on the transcriptions of the videos. Most of the videos that had different results on both dataset were first labeled as "unclear" and then after having more context (transcription) they were classified as propaganda or not propaganda.

4.2 Topic Extraction

As mentioned in Section 4.2, LDA was used on the *transcriptions* dataset in order to better understand the narrative of the russian propaganda and what it is mainly aimed at. The *transcriptions* table was used because the classification model returns better results on it.

LDA (Propaganda)	
Topic 1	people, want, country, government, like, war, u, would, nato, know
Topic 2	ukrainian, force, soldier, army, troop, unit, region, brigade, military, defense
Topic 3	tank, drone, ukrainian, missile, hit, armored, another, vehicle, two, grenade
Topic 4	video, news, force, channel, armed, destroyed, one, subscribe, enemy, like
Topic 5	know, say, go, uh, war, going, get, see, let, come

Table 6: LDA Topic Modeling Results for Propaganda Videos

LDA (Non-Propaganda)	
Topic 1	people, putin, like, say, know, storm, u, shadow, missile, make
Topic 2	vehicle, drone, missile, fire, system, artillery, hit, ammunition, armored, destroyed
Topic 3	said, enemy, offensive, operation, brigade, unit, armed, direction, artillery, th
Topic 4	drone, missile, hit, target, system, destroyed, helicopter, two, boat, like
Topic 5	crimea, bridge, offensive, month, explosion, peninsula, occupied, missile, counter, territory

Table 7: LDA Topic Modeling Results for Non-Propaganda Videos

The figure above shows five extracted topics for both groups (propaganda = 1 and propaganda = 0) which we can interpret to understand more on how the russian propaganda works and how it differs from non-propaganda.

In the propaganda content, the first topic appears to focus on public sentiment and geopolitical concerns. The frequent use of words like "people," "country," and "government" suggests content aimed at shaping public opinions about national and international issues. The presence of "nato" and "war" indicates a narrative that may be addressing or criticizing Western alliances, especially NATO, and possibly framing them as a threat or discussing their involvement in the ongoing conflict. The words "want" and "know" suggest discussions around what people desire or perceive, possibly aiming to influence public opinion on what the government or the country should do, especially in the context of war and international relations.

Topic 2 is heavily centered on the Ukrainian military. The focus on words like "Ukrainian," "soldier," "army," and "troop" suggests that the content may be discussing the actions and effectiveness of Ukrainian forces. The inclusion of "region," "brigade," and "defense" indicates that the narrative might involve detailed reports on military engagements, possibly framing Ukrainian forces in a particular light, either to emphasize

their resistance or to undermine their efforts. This topic likely serves to inform or misinform the audience about the state of the conflict from a military perspective, possibly with a slant that serves the propagandistic goals.

Next, the third topic, focuses on military hardware and combat operations. The dominance of terms like "tank," "drone," "missile," and "armored" suggests a narrative centered around the technical aspects of the war. The mention of "Ukrainian" alongside terms like "hit", "dron", "grenade" indicates that the content may be showcasing or discussing the destruction of Ukrainian military assets.

The fourth topic is centered on media and communication, particularly the dissemination of propaganda through video content. Words like "subscribe" and "like" imply a call to action, encouraging viewers to engage with the content, which is a common tactic in online propaganda to increase the reach and influence of the videos. The mention of "enemy," "armed," and "destroyed" suggests that these videos or news pieces might focus on military victories or the depiction of the enemy (likely Ukrainian forces) being defeated, which serves to portray the russian army as very powerful.

The final topic appears to capture conversational and speculative discussions. The informal nature of words like "uh," "go," "get," and "come" suggests that the content might involve interviews, casual commentary, or discussions that speculate about the war or other related events. The inclusion of "war" amidst these informal terms indicates that the discussions are likely centered around the ongoing conflict, with an aim to engage the audience in a more relatable or persuasive manner.

In contrast, the non-propaganda content also touches on public sentiment but with a less manipulative tone. Words like "Putin," "storm," and "shadow" for the first topic suggest a focus on specific events or figures without the same level of ideological framing. The absence of terms like "NATO" or "war" in this topic shows a less aggressive attitude towards geopolitical issues. The lack of words like "Ukrainian" together with words with military conotation such as "drone", "strike", "hit", indicates a more straightforward reporting style, where the emphasis is on describing military actions without the ideological bias evident in the propaganda content. The absence of direct references to Ukrainian forces in the topics suggests a less aggressive, more fact-based approach, where the content likely aims to inform rather than influence public perception. Non-propaganda content does not show the same level of informal, speculative discussion. Instead, it tends to focus on specific issues like "bridge," "explosion," and "peninsula," providing fact based coverage of events.

4.3 Aspect-Based Sentiment Analysis

The first dataset on which the ABSA Algorithm has been run is the "comments" dataset which contains comments from propaganda and non-propaganda videos, as mentioned in Section 3.1. You can see the results for the 5 chosen aspects in the tables down below:

	Positive	**Negative**
Russia	121	1447
Ukraine	38	520
Nato	19	329
Putin	42	379
Zelensy	3	67

Table 8: ABSA Results for propaganda = 1

	Positive	**Negative**
Russia	31	816
Ukraine	46	296
Nato	1	127
Putin	3	240
Zelensy	2	31

Table 9: ABSA Results for propaganda = 0

Both groups (propaganda = 0 and propaganda = 1) have mostly only negative comments regarding the chosen aspects, indicating that most of the interaction on the TikTok videos is critical towards the aspects mentioned above.

The majority of comments regardless if they are from propaganda or non-propaganda videos, express negative sentiments. This trend is consistent across all the aspects analyzed (e.g. "Russia," "Ukraine," "NATO," "Putin," "Zelensky"). This could indicate the controversial nature of the topic (the russian invasion of Ukraine).

The sentiment analysis suggests that propaganda videos are likely designed to provoke and polarize, creating strong reactions. The negative sentiments might be a result of provocative content that aims to criticize or attack opposing views.

After applying the ABSA algorithm on the descriptions of the videos from the "transcriptions" dataset, we are left with the following results:

Aspect	**Positive**	**Neutral**	**Negative**
Nato	0	26	2
Putin	0	9	6
Russia	3	79	20
Ukraine	1	31	9

Table 10: ABSA Results for propaganda = 1 in Transcriptions

Aspect	Positive	Neutral	Negative
Nato	0	117	0
Putin	0	6	2
Russia	1	84	18
Ukraine	0	88	8

Table 11: ABSA Results for propaganda = 0 in Transcriptions

ABSA has been applied to the descriptions of the videos because the transcriptions of the videos are very long and also it's harder for the algorithm to understand context and what the text is reffered to. In the non-propaganda videos there is a substantial dominance of neutral sentiments for all aspects. For example "Nato" has 117 neutral mentions, "Russia" has 84, and "Ukraine" has 88, with almost no positive or negative sentiments.

This suggests that non-propaganda videos tend to provide a more balanced view, presenting information without a strong emotional or biased tone. This could indicate an intent to inform or discuss topics from a neutral perspective and not to manipulate viewers. These results are in accordance with the results from Section 4.2, which showed that non-propaganda videos focus on informing the public, not to change their opinions.

Observation: Propaganda videos (propaganda = 1) show a higher count of both positive and negative sentiments directed at specific entities, particularly "Russia" and "Ukraine." This contrasts with non-propaganda videos, where the sentiment towards these entities is predominantly neutral. Interpretation: This indicates that propaganda videos often have specific targets or focal points they aim to influence public opinion about, whether to criticize or subtly support these entities. The variability in sentiment (both positive and negative) could be a tactic to address different audience segments or to sow confusion and division by presenting conflicting messages.

The lack of positive sentiment suggests that both types of videos, even those classified as propaganda, are not heavily focused on creating a positive narrative. Instead, the focus is either on presenting information neutrally (in non-propaganda videos) or on highlighting negative aspects (in propaganda videos).

4.4 Evaluation

In order to validate the accuracy of the zero-shot classifier, human validation is needed. In that regard two methods of verification were employed. First a random subset of 100 videos (50 propaganda and 50 not propaganda) was extracted for human review. After analyzing them, 82 out of 100 were correctly classified resulting in an accuracy of 82%.

4.4.1 Survey

To further check the accuracy of the model, a survey consisting of 11 questions has been created and distributed to 13 people. The participants were asked to classify 11 descriptions from TikTok videos as either Propaganda, Not Propaganda or Unclear. The survey can be found in Section A of the thesis.

	Propaganda	**Not Propaganda**	**Unclear**
Q1	0%	100%	0%
Q2	100%	0%	0%
Q3	0%	100%	0%
Q4	0%	100%	0%
Q5	84.6%	7.7%	7.7%
Q6	53.8%	23.1%	23.1%
Q7	0%	100%	0%
Q8	0%	100%	0%
Q9	23.1%	30.8%	46.2%
Q10	7.7%	92.3%	0%
Q11	0%	100%	0%

Table 12: Survey percentage results

When comparing the results from the survey with the results of the zero-shot classifier, we can notice similarities but with slight differences. There are 3 descriptions which have been classified wrongly by the model (Questions: 1, 3 and 4). Question 1 has been classified as unclear by the model but in reality it is not propaganda. This further proves that the model is classifying a lot of the videos as unclear and further analysis of those unclear videos is needed. Questions 3 and 4 have been completely missclassified, as the model classified them as propaganda but in reality they are not propaganda. A question that had mixed responses from the survey takers is Question 9 which has been classified as unclear by the algorithm.

5 Discussion

This is an exploratory study and it examined the characteristics of russian propaganda spreaders regarding the invasion of Ukraine. This study examined the characteristics of russian propaganda spreaders on TikTok regarding the invasion of Ukraine. The aim of the thesis was to find similarities between all the russian propaganda videos and to compare them to normal content on TikTok.

As mentioned in Section 1.1, this thesis aims to answer three research questions. The first question aimed at finding a method to differentiate between russia propaganda and non-propaganda content on TikTok. In order to answer this question, a zero-shot classifier has been employed (MoritzLaurer, 2024). Following the results, most of the videos from the dataset have been classified as unclear, demonstrating the difficulty to classify propaganda videos only based on descriptions and text, given that TikTok is a video-focused social media platform. As a robustness check and to see if we can get better results, the zero-shot classifier has been employed also on the transcriptions of the videos (approx. 1000). A slightly better accuracy has been noticed especially for the videos that have previously been classified as unclear. While this strategy doesn't capture all the propaganda content on the platform, it can help us identify a decent chunck of propaganda videos which contribute at characterizing the propaganda spreaders. The second and third research questions focused on characterizing the russian propaganda spreaders and comparing it to normal content on the social media platform. When it comes down to the numbers and how propaganda videos are used in Russias favor, we clearly saw that propaganda efforts tend to skyrocket during important events, such as the beginning of the war or counter-offenses by the ukrainian side. Another trend in propaganda videos is the short description used which tends to quickly capture the users attention, followed by long transcriptions (audio content) which aims at pushing certain narratives while providing extensive narrative detail. Hashtags play a critical role in the propaganda efforts, as they are a powerful tool to increase the visibility and reach of videos. The most common hashtags used by russian propaganda are #istandwithrussia and #putin, clearly suggesting a focus on russian related narratives. Very general hashtags such as #fyp are used to increase reach and engagement by exploiting TikToks algorithmic curation. The spikes in specific hashtags around certain events (e.g. the start of the invasion and Ukrainian counter attacks) indicate that propaganda efforts are not constant but instead intensify during significant events.

The most frequent words used by propaganda spreaders are "russia", "putin", "donbass," and "russian" suggesting a focus on russian related content compared to the non-propaganda content which focuses on a more general view of the conflict, trying to inform the public rather than pushing narratives. We also saw that the propaganda videos have a better views-interaction ratio than the non-propaganda videos, which suggests that the topics portrayed in those videos spark more reactions from the users.

To further analyse the language used by propaganda spreaders, LDA was used in order to extract the most relevant topics from the transcriptions of the videos. We clearly saw that propaganda content focuses on topics such as Nato, Ukrainian military and geopolitical conerns, but it also potrayes a call to action by using terms such as "like" and "subscribe". The presence of terms like "uh", "oh", "get" and "come" suggest that

the content might involve interviews, casual commentary, or discussion that speculate about the war or other related events. It also suggests a tactic to create a more relatable or conversational tone, potentially lowering the audience's critical thinking. On the other hand, the non-propaganda content has a less manipulative tone, a less aggressive attitude towards geopolitical issues and a more fact-based approach.

When it comes to analyzing the sentiments towards specific aspects, as described in Section 4.3, the results are mostly negative. ABSA models struggle with context, especially in complex text like social media content or when the content is focused on a very specific topic. More than that, social media posts tend to be multi-faceted, where a post can contain positive and negative sentiments towards different aspects. Better accuracy can be obtained if the respective model is trained on a dataset that is more related to the topic of the Thesis.

5.1 Strengths and Limitations

As mentioned in Section 2.2, there is a gap in the research regarding russian propaganda on TikTok. To the best of my knowledge, this thesis represents the only analysis of russian propaganda regarding the war on TikTok out of the ones that are publicly available.

One of the strengths of this study is that the data has been extracted by the author himself. Extracting videos from TikTok based on the filters mentioned in Section 3.2, allowed the author to work with only relevant data for the study but also to have full control over it. This helped to fetch only data that is related to the war in Ukraine, making the analysis of russian propaganda spreaders easier.

On the other hand, this thesis also has some limitations. The first limitation is the TikTok Researcher API as it doesn't provide as much data as other social media platforms. TikTok is still considered a non-researcher friendly platform making it difficult to extract all the data needed. For example informations such as User Age, User Reposts and Account Creation Date are not available throught the Researcher API. Another difficulty encountered is that TikTok focuses on video-content, not on text-content as mentioned in Section 1. Only being able to work with hastags, video descriptions and video transcriptions (if available) made it difficult to properly analyze the videos and to classify them as propaganda or non-propaganda. There is no option to download the audio of a video and also using third-party audio downloaders ends up with TikTok blocking the computers IP. Having said so, collecting the needed data on TikTok proved to be quite challenging considering that this thesis focuses mainly on Natural Language Processing tasks.

Another limitation is the use of the zero-shot classifier mentioned in section 3.3.1. Although our analysis showed that the classifier returns decent accuracy, training a model specifically for the task of detecting russian propaganda would be much more benefical and accurate. Doing this would be very ressource consuming as it would require a lot of computational power to fine-tune a model, but it would also require a lot of human ressources in order to create a big database with russian propaganda examples. The high presence of very short descriptions in the datasets also representz a limitation. Considering that it would be impossible even for a human to classify some of the TikToks only based on the descriptions, this task can't be solved by any algorithm or Machine Learning Model.

5.2 Future Research Recommendations

Based on the findings and limitations of this study, there are aspects which can be further analysed and improved. Future research could focus on developing a more complex and accurate way of classifying the propaganda videos, taking into consideration the ambiguous nature of some descriptions and the very long transcriptions. Future studies should also take into consideration expanding the dataset with more videos in order to avoid misidentification. Another aspect which could be improved is ASBA, explained in Section 3.3.2. Given the unique linguistic features on social media and the presence of sarcasm, slangs and ambiguity, employing a context-aware model such as transformer-based architectures (e.g. BERT, GPT), could prove to be more efficient when trained on the right datasets. Future research can also be expanded to image vision, which can analyse the images from videos and identify patterns.

By pursuing these recommended research directions, scholars can enhance the understanding of russian propaganda spreaders and develop more efficient tools and strategies for it.

6 Conclusion

This thesis explored the characteristic of russian propaganda spreaders on TikTok, especially in the context of the invasion of Ukraine. The research aimed to highlight the features of propaganda videos, understand the narratives they push, understand the used language and see how they compare to regular TikTok content. By combining methods like content analysis, sentiment analysis, and machine learning, the study managed to make a good comparison between the two groups described above.

The study found that propaganda videos often use specific hashtags such as #istandwithrussia and #putin to change narratives in the favor of Russia. These videos typically have shorter descriptions, likely to catch viewers attention quickly and take advantage of TikToks algorithm by using hastags such as #fyp. On the other hand, non-propaganda videos generally aim to provide more comprehensive informations and a more general view about the war. The ABSA method revealed that both propaganda and non-propaganda videos tended to have a predominantly negative tone, reflecting the emotionally charged nature of the topic.

However this research also encountered some challenges. One major limitation was accurately identifying propaganda in short descriptions and very ambiguous texts. Another limitation of the study was also the accuracy of the ABSA algorithm, which is not accuractely understanding the context of the text.

Despite these challenges, these findings are significant for understanding what techniques and strategies the propaganda spreaders use on social media platforms such as TikTok.

Future research could contribute to this topic by improving the classification models but also the sentiment analysis models by trainig them on social media datasets. Further studies might also look into how social media algorithms impact what content gets seen and how it spreads, offering a better view of how content on social media becomes viral or has a big reach.

In conclusion, this research highlights the characteristics of russian propaganda spreaders and makes a comparison between russian propaganda and normal content. As social media becomes more relevant and important every single day, the propaganda efforts of hostile actors become bigger as well. We see more and more efforts put into propaganda on social media and this is concerning given that most of the social media users are young people. By researching how propaganda works in the digital age, we can understand what tactis are employed and we can better understand what is real and what is not on the internet.

A Appendix

As explained in Section 4.4.1, a survey has been created and distributed to people in order to humanly validate the classification results. You can find the survey down below:

Hello, I am studying characteristics of russian propaganda regarding the war in Ukraine on TikTok. In order to test the accuracy of the algorithms I used, I created a survey. Please select if the text down below is Propaganda, Not Propaganda, or Unclear. The survey should not take more than 5 minutes. Thanks!

1. 'Think about why the #Russians attacked! They did not #attack, they came to #protect the #people of #Donbass'. A Donbass #resident responded to a #Ukrainian #commentator about what is really happening and what is the #real #problem confronting the brain-washed #Ukrainians: 'You have the #Russians to #blame for everything'.
    - Propaganda
    - Not Propaganda
    - Unclear
2. Russia economy is stronger than it has ever been #CapCut #biden #putin #UkraineRussiaWar #militaryindustrialcomplex
    - Propaganda
    - Not Propaganda
    - Unclear
3. Russian forces have begun to push through Ukrainian defences along eastern Ukraine. Ukrainian President Volodymyr Zelenskiy calls this the "Battle of the Donbas.". #onenews #onenewsph #newsph #tiktoknews #tiktoknewsph #worldnews #russia #ukraine
    - Propaganda
    - Not Propaganda
    - Unclear
4. Mykhailo Podolyak is talking about the ruined Donbas, that used to be a peaceful and developing region before 2014. #donbas #ukraine #supportukraine #istandwithukraine #podolyak
    - Propaganda
    - Not Propaganda
    - Unclear
5. Brave Vodner soldier shoots down Ukrainian Mig 29 over the skies of Donetsky #russia #istandwithrussia #ukraine #war #fyp #wagnerpmc
    - Propaganda
    - Not Propaganda
    - Unclear
6. Russian atrocities know no bounds as they kidnap Ukranian children for adoption. #War #Russia #Ukraine #NATO #UN #UnitedNations #US #UK #Germany #Poland #France #Italy #Romania #Lithuania #Moldova #Bulgaria #Czechia #USA #UnitedStates #Britian #Norway #Finland #Sweden #Turkey #Spain #NATO #EU #EuropeanUnion #slovakia #Europe #News #Newsforthought #Newzforthought #SocialNews #SocialMedia #NewsTok #MillTok #icc #warcrimes #steal #kidnapping
    - Propaganda
    - Not Propaganda

- Unclear

7. Russian Ice Cream Myth! #shorts #fyp #fyp #foryou #viral #canigetviral #russian #myth #funfacts #knowledge

    - Propaganda
    - Not Propaganda
    - Unclear

References

Backlinko. (n.d.). *Number of tiktok users.* ://backlinko.com/tiktok-users.

CardiffNLP. (2024). *Sentiment analysis.* Retrieved from `cardiffnlp/twitter-roberta-base-sentiment-latest`

Cloudflare. (2021). *Most popular domains of 2021.* `https://blog.cloudflare.com/popular-domains-year-in-review-2021`. (Accessed: July 21, 2024)

Dictionary.com. (2024). *Propaganda definition.* Retrieved from `https://www.dictionary.com/browse/propaganda` (Accessed: August 01, 2024)

Dominique Geissler, N. P. S. F., Dominik Bär. (2023). *Russian propaganda on social media during the 2022 invasion of ukraine.*

Ezzameli, K., & Mahersia, H. (2023). Emotion recognition from unimodal to multimodal analysis: A review. In *Information fusion* (p. 7.1-7.2). Elsevier. Retrieved from `https://www.sciencedirect.com/topics/computer-science/aspect-based-sentiment-analysis` (Accessed: 2024-08-25)

Garth S. Jowett, V. O. (2012). *Propaganda and persuasion* (5th ed.). Sage Publications.

Huggingface. (n.d.). *Zero-shot performance.* `https://huggingface.co/MoritzLaurer/deberta-v3-large-zeroshot-v2.0`.

Klim Kireev, C. T. R. O., Yevhen Mykhno. (2024). Characterizing and detecting propaganda-spreading accounts on telegram. *arXiv preprint arXiv:2406.08084* .

Larissa Doroschenko, J. L. (2021). *Trollfare:russia's disinformation campaign during military conflict in ukraine.*

MoritzLaurer. (2024). *Zero-shot classifier.* Retrieved from `https://huggingface.co/MoritzLaurer/deberta-v3-large-zeroshot-v2.0`

Oberlo. (2024). *Tiktok age demographic.* `https://www.oberlo.com/statistics/tiktok-age-demographics`. (Accessed: July 25, 2024)

TikTok. (n.d.). *Researcher api.* `https://developers.tiktok.com/doc/research-api-specs-query-user-info/`.

Wikipedia. (n.d.). *Ukrainian counter-offensesive.* `https://en.wikipedia.org/wiki/2023_Ukrainian_counteroffensive`.

yangheng. (2024a). *Absa datasets.* Retrieved from `https://github.com/yangheng95/ABSADatasets`

yangheng. (2024b). *Aspect-base sentiment analysis.* Retrieved from `https://huggingface.co/yangheng/deberta-v3-base-absa-v1.1`