Ollscoil Teicneolaíochta an Atlantaigh Dún na nGall
Atlantic Technological University Donegal

Department of Computing, ATU Donegal, Port Road, Letterkenny, Co. Donegal, Ireland.

Comparative Sentiment Analysis of Telegram Channels During the Ukraine Russia War

Author: Bhavya Cherukuri

Supervised by: Dr. Eoghan Furey

A thesis submitted in partial fulfilment of the requirements for the

Master of Science in Computing in Big Data Analytics

Submitted to Quality and Qualifications Ireland (QQI)
Dearbhú Cáilíochta agus Cáilíochtaí Éireann

August 2024

Declaration

I hereby certify that the material, which l now submit for assessment on the programmes of study leading to the award of Master of Science in Computing in Big Data Analytics, is entirely my own work and has not been taken from the work of others except to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution. I understand that it is my responsibility to ensure that I have adhered to LYIT's rules and regulations.

I hereby certify that the material on which I have relied on for the purpose of my assessment is not deemed as personal data under the GDPR Regulations. Personal data is any data from living people that can be identified. Any personal data used for the purpose of my assessment has been pseudonymised and the data set and identifiers are not held by LYIT. Alternatively, personal data has been anonymised in line with the Data Protection Commissioners Guidelines on Anonymisation.

I consent that my work will be held for the purposes of education assistance to future students and will be shared on the LYIT Computing website (www.lyitcomputing.com) and Research THEA website (https://research.thea.ie/). I understand that documents once uploaded onto the website can be viewed throughout the world and not just in Ireland. Consent can be withdrawn for the publishing of material online by emailing Thomas Dowling; Head of Department at thomas.dowling@lyit.ie to remove items from the LYIT Computing website and by email emailing Denise McCaul; Systems Librarian at denise.mccaul@lyit.ie to remove items from the Research THEA website. Material will continue to appear in printed formats once published and as websites are public medium, LYIT cannot guarantee that the material has not been saved or downloaded.

Signature of Candidate Date: August 2024

Acknowledgements

Taking this occasion, I would like to sincerely thank the Computing Department teachers, whose advice and knowledge have been invaluable in helping me complete my master's degree in Big Data Analytics.

My supervisor, Dr. Eoghan Furey, has my sincere gratitude for his constant support, wise counsel, and encouragement, all of which were essential to finishing this thesis. His advice has been essential to this job, and I sincerely appreciate his mentorship along the way.

I also want to express my gratitude to my brother and parents for their constant support, love, and care. Their confidence in me has always given me courage.

I also owe a particular thank you to the other Big Data Analytics lecturers, whose wisdom and direction enhanced my educational journey. I would especially like to thank Dr. Furey for all of his encouragement and assistance during this endeavour.

Finally, I would like to thank my friends for their support and inspiration, which were important in helping me finish this project. This voyage would not only be possible but also unforgettable thanks to your support.

Abstract

The ongoing war between Russia and Ukraine, characterized by deepseated historical and geopolitical tensions, has escalated in intensity in recent years, resulting in increased military operations and significant global implications. Despite ongoing efforts to resolve the conflict, the war remains an active and contentious issue, with both sides engaging in intense information warfare on digital media. This study aims to investigate the public's perception and portrayal of the war through sentiment analysis of Telegram channels, a wellknown social media platform in both Russia and Ukraine.

In particular, the study analyzed posts from four Telegram channels, two of which were Ukrainian and the other two Russian, made in January 2024 and August 2024. The study examines the opinions expressed on these platforms to gain insight into the psychological and emotional landscape of the conflict as it is portrayed on social media.

The research utilized sentiment analysis techniques to determine the frequency of positive, neutral, and negative attitudes in the communications. The results revealed a significant amount of negative emotion across all channels, reflecting the devastating reality of war. However, notable differences were observed between the Russian and Ukrainian channels, with Ukrainian channels displaying a higher concentration of negative emotion, likely due to the direct impact of the conflict on the general population.

This study provides valuable perspectives on the role of digital communication in contemporary warfare and contributes to our understanding of how social media platforms like Telegram are utilized to shape and reflect public opinion during times of conflict.

Acronyms

Acronyms	Definition
U/R	Ukraine/Russia
TC	Telegram Channels
SA	Sentiment Analysis
VADER	Valence Aware Dictionary and sEntiment Reasoner
BD	Big Data
ML	Machine Learning
DL	Deep Learning
NLP	Natural Language Processing
API	Application Programming Interface
PLN	Preprocessing, Lexicon, and Normalization
SD	Sentiment Distribution
TS	Temporal Shifts
CM	Comparative Metrics
QED	Quantitative and Emotional Data
LDA	Latent Dirichlet Allocation
P/N/N	Positive/Negative/Neutral (Sentiments)
KPIs	Key Performance Indicators (in sentiment analysis)
RC	Research Contribution
AD	Analytical Dimensions
TB	Textual Behavior

Table Of Contents

Table of Figures

Chapter 1: Introduction

The ongoing conflict between Ukraine and Russia, which escalated with Russia's fullscale invasion of Ukraine in February 2022, has had a profound impact on global politics, economics, and digital communication. As the conflict continues into 2024, digital platforms have become critical in shaping public opinion and disseminating information. Among these platforms, Telegram has emerged as a significant tool for both Ukrainian and Russian entities to share news, propaganda, and sentiments, surpassing other social media platforms in user engagement during the conflict (Kiforchuk, 2023). Telegram's role in the information war is particularly significant due to its encrypted messaging and capability to host large communities, making it a preferred platform in conflict zones (Sasahara, 2023).Despite the extensive use of Telegram in the U/R war, there is a noticeable gap in the literature regarding a comparative sentiment analysis of U/R channels on this platform. While previous research has focused on sentiment analysis of other social media platforms such as Twitter and Reddit during the conflict (Martinčić-Ipšić, August 23, 2024) Telegram's unique role and influence have not been sufficiently explored.

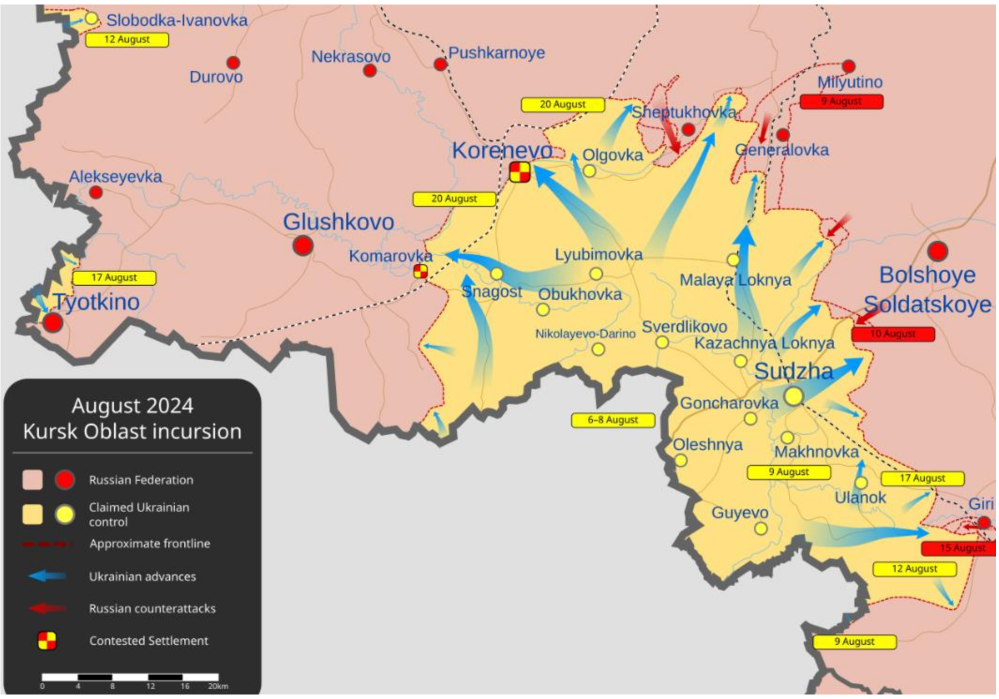

Figure 1:August 2024 Kursk Oblast Incursion (August 2024 Kursk Oblast incursion, 2024)

As show in Figure 1 the Ukrainian Armed Forces conducted a significant advance into Russia's Kursk Oblast in August 2024, the importance of Telegram in the ongoing combat became very clear. This incident, one of the worst crossborder assaults since the invasion in 2022, demonstrated how the conflict is becoming more intense. Not only was the incursion a military operation that saw Ukrainian forces seize almost 1,000 km² of Russian territory, but it was also a pivotal point in the information war, with both sides utilizing Telegram to spread updates and sway public opinion (August 2024 Kursk Oblast incursion, 2024). The disagreement has also affected Telegram's regulatory environment. The platform has been subject to limitations by a number of Western governments because of security concerns and its affiliation with Russia. Norwegian Ban (2022): Citing the platform's "Russian origins" as a security risk, the Norwegian Justice Ministry outlawed the use of Telegram on official equipment. In 2023, the municipal administration of Amsterdam imposed a ban on city employees, citing worries regarding "foreign espionage." France's Directive (2023): Amid security concerns, France ordered civil servants to transition from Telegram and other comparable chat apps to domestically designed substitutes. March 2024: Spain's High Court ruling A Spanish high court said that Telegram's access had been blocked by cell companies for violating copyright; however, the decision was swiftly reversed. These claims are still being looked examined.

In light of geopolitical tensions and worries about national security, there has been an increase in the monitoring and regulation of digital platforms, which is reflected in these limits. Considering the significance of the platform, it is imperative to examine the opinions shared on Telegram in order to comprehend the psychological and emotional effects of the conflict on the populations of Russia and Ukraine. By doing a comparative sentiment analysis of Russian and Ukrainian Telegram channels between January and August of 2024, this study seeks to close this research gap. Through the analysis of sentiment trends during these two crucial times, this study will offer valuable perspectives on how the battle is viewed and debated on both sides, thereby advancing our knowledge of digital communication tactics in contemporary warfare.

Given the platform's importance, it is essential to analyze the sentiments expressed on Telegram to understand the emotional and psychological impacts of the war on both Ukrainian and Russian populations. This study aims to fill this research gap by conducting a comparative sentiment analysis of Ukrainian and Russian Telegram channels during January 2024.

By examining the sentiment trends, this research will provide insights into how the conflict is perceived and discussed on both sides, contributing to the broader understanding of digital communication strategies in modern warfare (GK, PK, A, & KumarS, 2023). Additionally, this study will build upon prior research that has examined sentiment in various contexts, such as the analysis of Reddit posts during wartime (Martinčić-Ipšić, August 23, 2024) and the strategic communication observed in wartime Telegram channels ((The structure of wartime strategic communications: Case study of the telegram channel insider ukraine, 1970). The findings of this research will not only deepen our understanding of sentiment dynamics in the ongoing conflict between Ukraine and Russia, but also provide valuable insights into the role of Telegram in facilitating communication during times of crisis (Ptaszek, 2023). Moreover, the analysis will contribute to the growing body of literature on the impact of social media in modern conflicts, emphasizing the crucial role that platforms like Telegram play in shaping public opinion and discourse during times of war. (V. Bobichev, 2017).

1.2 Background

The ongoing conflict between Ukraine and Russia, which escalated dramatically with the large-scale invasion by Russian forces in February 2022, continues to shape global geopolitics and public sentiment as it drags on into 2024. This war has not only influenced military strategies and international alliances but also significantly impacted the digital information landscape. Social media platforms have emerged as pivotal battlegrounds for information warfare, where narratives are constructed, public opinion is swayed, and propaganda is disseminated. Among these platforms, Telegram has become particularly prominent due to its unique features that cater to both private and public communication needs.

Telegram’s attributes, such as encrypted messaging, the capacity to host large groups, and the ability to share multimedia content, have made it a vital tool for both Ukrainian and Russian entities. These features enable the rapid dissemination of information, allowing both sides to communicate directly with their audiences and engage in the broader international dialogue surrounding the conflict. Telegram channels have been instrumental in providing real-time updates, shaping narratives, and influencing public sentiment, making the platform an essential component of the modern information warfare landscape.

While there has been substantial research on the role of social media platforms like Twitter and Reddit in conflict situations, Telegram remains underexplored despite its significant influence. Previous studies, such as those by (Martinčić-Ipšić, August 23, 2024), have examined sentiment on Reddit during the Ukraine-Russia war, highlighting how public emotions and opinions are molded by the ongoing conflict. Similarly, extensive research has been conducted on Twitter’s role in story creation and information dissemination during the war. However, the unique role of Telegram, with its distinct characteristics and widespread use in the region, has not been thoroughly investigated.

This study aims to address this gap by conducting a comparative sentiment analysis of Russian and Ukrainian Telegram channels in January 2024. By analyzing the sentiments expressed in these channels, the research seeks to provide insights into the psychological and emotional impacts of the war on both sides. This analysis will contribute to a deeper understanding of digital communication in conflict settings and the role that platforms like Telegram play in shaping the narratives and emotions of those involved in or affected by the conflict (Kiforchuk, 2023)

1.3 Why Telegram?

Decentralized Control: Telegram's decentralized control is one of the main features that sets it apart from other social media networks. Telegram functions with a great degree of autonomy, in contrast to other platforms that are subject to stringent regulations from governments or businesses. Telegram's decentralized strategy enables it to offer a forum where people may interact openly without worrying about censorship or intervention from the government. This is especially significant in areas where there are restrictions on the right to free speech, which makes Telegram an indispensable resource for activists, journalists, and anybody looking for unrestricted contact.

Recent Controversies and Independence: Telegram's ability to steer clear of controversy without sacrificing its operational independence serves as even more evidence of its independence. The CEO of Telegram, Pavel Durov, was detained in France in August 2024 over claims that the app was utilized to disseminate content about child abuse. Telegram showed its dedication to upholding a platform that emphasizes user privacy and free speech by operating without direct government monitoring in the face of this serious legal challenge. This episode demonstrates the difficulties Telegram has in striking a balance between upholding its moral standards and the requirement to deal with illicit activity on its network (Cointelegraph., 2024).

Simple Communication: Another factor in Telegram's enormous appeal is its uncomplicated and user-friendly communication style. With features like groups and channels that let users share content with big audiences quickly and effortlessly, the platform is made to be both simple and effective. When paired with robust privacy features like end-to-end encryption, Telegram's simplicity makes it a desirable choice for users who value security and usability above all else. The platform prioritizes direct and secure communication has made it a popular option, especially in settings where political sensitivity is high.

1.5 Problem Statement

The ongoing conflict between Ukraine and Russia has garnered significant attention on digital platforms, with Telegram emerging as a critical medium for disseminating information and shaping public sentiment. Despite its significance, there is a notable lack of detailed analysis regarding the evolution of sentiment within Ukrainian and Russian Telegram channels.This research aims to address the following challenges:

Sentiment analysis (SA) on Social Media: Traditionally, SA on social media has been based on data from sites like Facebook and Twitter, which are managed by feeds that are curated by algorithms. This may cause bias or distortions in the way that data is interpreted and evaluated. Contrarily, Telegram presents content in a straightforward and chronological manner without the use of such algorithms. This feature of Telegram provides an independent, possibly more accurate basis for sentiment analysis without the influence of algorithms.

Data Extraction and Sentiment Analysis: Effective extraction and analysis of Telegram messages are essential for understanding sentiment. This involves overcoming technical obstacles related to processing large datasets from Ukrainian and Russian channels during January 2024.

Temporal Sentiment Dynamics: There is a need to explore how sentiment fluctuates not just within the same day but also across consecutive days. This includes examining whether specific events or developments in the conflict trigger noticeable shifts in sentiment on Telegram.

Post Frequency and Sentiment Patterns: Examining the number of posts made daily and identifying patterns in content generation can reveal how frequently content is produced and how this relates to sentiment.

This analysis will help determine if there is a correlation between the volume of posts and the sentiments expressed.

Keyword Analysis: Identifying and analyzing the frequency of the most common words used in messages and their association with positive or negative sentiment is important. This involves assessing the share of representative keywords in different sentiment categories to understand broader sentiment trends.

Content Type Influence: Understanding how different types of content—such as news updates, personal opinions, and propaganda—affect sentiment variations is vital. This includes exploring whether specific content types are associated with P/N/N .

This study aims to bridge the gap in current research by providing a comprehensive analysis of sentiment dynamics on Telegram during the UkraineRussia conflict, contributing to our understanding of the role of digital platforms in conflict situations and public sentiment formation (Kiforchuk, 2023).

1.4 Research Objectives

The purpose of this study is to perform a sentiment analysis comparison of pro-Russian and pro-Ukrainian Telegram channels during the conflict. The goal of the research is to identify the underlying emotional tones and trends that define the discourse surrounding the dispute by assessing the sentiment of messages from different Telegram channels.

Next, the study will employ sentiment analysis techniques on these messages to identify and compare the emotional tone, whether positive, negative, or neutral, across both

Ukrainian and Russian channels. This analysis will aid in comprehending the nature of the sentiment expressed in these communications.

Furthermore, the research will carry out a keyword frequency analysis to identify the most common words and phrases used in the Telegram messages. This will help in understanding how specific terms and topics relate to positive or negative sentiment. Finally, the study will evaluate how various types of content, including news updates, opinions, and propaganda, impact sentiment and engagement. This objective is focused on understanding how different content forms contribute to the overall sentiment expressed in the Telegram channels during the conflict.

1.5 Research questions

The following research questions serve as the study's main focus:

1. In January 2024 and August 2024, what was the general tone of Telegram messages from Russian and Ukrainian channels?
- This inquiry seeks to determine and contrast the overall emotional tone—whether favorable, unfavorable, or neutral—of communications from Russian and Ukrainian outlets throughout these two time frames.

2. How does sentiment change over time, especially before and after major events? - This question aims to comprehend the dynamics of sentiment over time by examining how it alters in reaction to particular events, like military operations, major developments in the conflict, or political pronouncements.

3. What are the word frequency variations between Russian and Ukrainian channels? - The purpose of this inquiry is to determine the most often used terms in Russian and Ukrainian channels, offering information about the probable propagandist components and thematic focus of each.

4. How do the sentiments on Russian and Ukrainian Telegram channels evolve over time between January and August 2024? The purpose of this question is to compare the changes in sentiment over time in Russian and Ukrainian channels in order to spot any trends, changes, or unusual sentiment over these months.

5. Do Russian and Ukrainian Telegram channels exhibit observable differences in the endurance of neutral, positive, and negative feelings across time? - This inquiry looks at if the two groups' messages differ in terms of the length of time that happy, negative, or neutral sensations last and whether one group holds onto particular feelings longer than the other.

6. How are the timing of attitude swings in reaction to significant conflict-related events in January and August 2024 different in Russian and Ukrainian Telegram channels? - This topic compares whether Russian or Ukrainian networks react more quickly or forcefully to major events, examining the speed and intensity of sentiment shifts in response to these occurrences.

7. How do sentiment fluctuation patterns differ between high and low activity times in January and August 2024 on Russian and Ukrainian Telegram channels? - This inquiry explores whether sentiment fluctuation patterns correlate with spikes or troughs in posting activity, providing information about the relationship between emotional tone and amount of content on both sides.

1.6 Proposed System

The proposed method for evaluating sentiment in Ukrainian and Russian Telegram channels involves incorporating several essential components to attain a thorough understanding of digital communication throughout the UkraineRussia conflict. The platform's API will be utilized to extract data from selected Telegram channels, focusing on messages from January 2024. Following preprocessing to remove noise and standardize content, the data will be prepared for analysis. Sentiment evaluation will be conducted using tools like VADER, categorizing messages into positive, negative, or neutral sentiments (Martinčić-Ipšić, August 23, 2024). The connection between engagement metrics, such as views and forwards, and sentiment will be examined. Furthermore, keyword frequency analysis will identify prominent terms and their relationship with sentiment (Ptaszek, 2023). The results will be visualized through graphs and reports, offering insights into sentiment trends and communication dynamics (V. Bobichev, 2017).This method aims to provide a comprehensive analysis of how sentiment changes and interacts with engagement during the conflict.

1.7 Significance of the Study

This study holds significant value as it provides a comprehensive understanding of the dynamics of sentiment within Russian and Ukrainian Telegram channels during the ongoing Ukraine-Russia war. By focusing on communications from January 2024, the research aims to illuminate how public attitudes shift in response to critical events and how these sentiments diverge between the conflicting perspectives. This analysis is crucial for evaluating public opinion and communication strategies in conflict situations, offering valuable insights to researchers, analysts, and policymakers who are tasked with interpreting and responding to these dynamics.

Furthermore, the study contributes to the development of advanced tools and techniques for sentiment analysis on social media platforms. By employing sophisticated methods for sentiment classification and keyword extraction, the research enhances the accuracy and depth of media analysis in high-stakes environments. It also provides a clearer understanding of how different types of content, such as propaganda and news updates, influence public sentiment, thereby offering a more nuanced view of how information shapes attitudes during conflicts.

In addition to its theoretical contributions, the research has practical implications for managing and assessing public opinion in various contexts. By advancing the field of digital communication studies during conflicts, the study not only enriches academic knowledge but also offers actionable guidance on how to effectively navigate and evaluate public sentiment in volatile situations. This dual impact makes the study a valuable resource for both scholarly inquiry and practical application in the realm of conflict communication and sentiment analysis.

Chapter 2: Literature Review

2.1 Introduction

The growing impact of social media on global conflicts underscores the importance of understanding sentiment expressed through these platforms. In recent years, Telegram has emerged as a pivotal medium for communication during conflicts, offering unique insights into public sentiment and propaganda. This literature review examines various methods of sentiment analysis, focusing on their application to Telegram channels during the UkraineRussia war.

Sentiment analysis has evolved from traditional rulebased systems to more sophisticated machine learning techniques, enabling deeper and more accurate insights into textual data. While conventional methods provide foundational insights, modern machine learning approaches offer enhanced capabilities to handle large volumes of data and capture nuanced sentiments. This review will explore these methodologies, with particular emphasis on the use of cloudbased analytics platforms like AWS and tools such as VADER.

In addition to exploring sentiment analysis methods, this review also addresses the significance of Telegram as a communication platform in conflict situations. Telegram's role in disseminating both proRussian and proUkrainian narratives during the UkraineRussia war highlights the need for targeted sentiment analysis of specific channels. By analyzing channels such as Slavyangrad and The Kyiv Independent, this study aims to provide a detailed understanding of how different perspectives are communicated and perceived.

The review will also identify gaps in the current literature, particularly the need for focused research on Telegram during the UkraineRussia war and comparisons of different sentiment analysis tools. This examination sets the stage for a comprehensive analysis of sentiment dynamics in Telegram channels, contributing to the broader understanding of digital communication in conflict zones.

2.2 Sentiment Analysis

Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. It is a combination of computer science, linguistics, and artificial intelligence. Most of the research in this area are focused on understanding how human beings understand and use language, so that, better and new computer systems can be created to understand and use human language. Sentiment analysis is a sub-field of NLP which is used to extract sentiments from a text (Chowdhury, 2003).

Sentimental Analysis (SA) is the computation study of people's opinions, attitudes, and emotions towards an entity (Medhat, 2014). It is the technique of analysing the opinions within a text and figuring on the sentiment behind the text whether it is positive, negative, or neutral. Sentimental Analysis has its application in various fields. It is mainly utilized in the business domain to understand sentiments of the users for their products/ services through the data obtained from user reviews or social media.

2.2.1 Conventional Sentiment Analysis Techniques

Sentimental Analysis is inherently a classification process. There are three main levels for sentimental analysis: Document- level, sentence- level and aspect-level (Medhat, 2014). In document- level, the entire document of text is considered as a single unit and the opinion is identified. When it comes to the sentence-level, it will decide whether the sentiment if the sentence is subjective. Identifying whether a sentence is sentence is subjective or objective is the first step in sentence-level classification. In aspect-level analysis, it aims to categorize the sentiment in respect to entities (Bhavitha et al., 2017). The two main approaches of sentimental analysis are lexicon based and machine learning based.

Conventional sentiment analysis techniques primarily utilize rulebased methods and lexicons. These methods often involve predefined lists of words and phrases with associated sentiment scores. The simplicity of these approaches makes them easy to implement but may limit their effectiveness in capturing the subtleties of sentiment, especially in complex or informal texts.

LexiconBased Approaches: These rely on sentiment lexicons or dictionaries containing words annotated with sentiment scores. Tools like AFINN, SentiWordNet, and VADER are examples of lexiconbased sentiment analysis methods (Hutto & Gilbert, 2014).
RuleBased Methods: These methods apply grammatical rules and patterns to determine sentiment. For example, identifying negations or intensifiers in sentences helps in adjusting the sentiment score (Liu, 2012).

2.2.2 Sentiment Analysis Using Machine Learning Methods

With advancements in machine learning, sentiment analysis has progressed beyond rulebased methods to include datadriven techniques. Machine learning methods are designed to learn from large datasets and can identify complex patterns in text data.

Supervised Learning: This approach involves training a model on labeled data where the sentiment is predefined. Algorithms such as Support Vector Machines (SVM), Logistic Regression, and Naive Bayes are commonly used (Pang & Lee, 2008).

Deep Learning: Modern sentiment analysis often employs neural networks, particularly Recurrent Neural Networks (RNNs) and Long ShortTerm Memory (LSTM) networks, which excel at handling sequential data and understanding context (Kim, 2014).
Transformers: Recent advancements include transformerbased models like BERT (Bidirectional Encoder Representations from Transformers) which provide contextaware

sentiment analysis by capturing the meaning of words based on their context in a sentence (Devlin et al., 2018).

2.2.3 Evaluation Metrics in Sentiment Analysis

Evaluating the performance of sentiment analysis models is crucial for understanding their effectiveness. Common metrics include:

Accuracy: Measures the proportion of correctly classified sentiments over the total number of instances.
Precision, Recall, and F1Score: Provide a more nuanced evaluation, particularly important in imbalanced datasets where some sentiments might be underrepresented.
Confusion Matrix: Helps in visualizing the performance by showing the true positives, false positives, true negatives, and false negatives.

2.3 Telegram as a Medium for War Propaganda

Telegram has emerged as a significant platform for communication, particularly in politically charged environments and conflict zones. Its features—such as encrypted messaging, large group capacities, and the ability to share multimedia content—make it an attractive tool for both individuals and organizations looking to spread information quickly and securely. During conflicts like the Ukraine Russia war, Telegram has been used extensively for disseminating propaganda, rallying support, and influencing public opinion on both sides.

2.3.1 The Role of Telegram in Conflict Zones

Telegram's prominence in conflict zones can be attributed to its ability to bypass state controlled media and its high resistance to censorship. The platform has been widely adopted by both proUkrainian and proRussian groups to distribute news, share narratives, and mobilize supporters. Unlike other social media platforms, Telegram allows for the creation of large public channels and groups, which can rapidly disseminate information to thousands or even millions of users.

For example, the proRussian channels such as "Slavyangrad" and "Intelslava" have been instrumental in spreading narratives that align with Russian state interests. Similarly, Ukrainian channels like "The Kyiv Independent" and "United 24 Media" provide a counternarrative, aiming to bolster Ukrainian national sentiment and international support. These channels are not just news dissemination tools but also platforms for psychological operations and propaganda (Ghasiya & Sasahara, 2023).

2.3.2 Features of Telegram that Facilitate Propaganda

Several features of Telegram make it an effective tool for propaganda:

Encryption and Anonymity: Telegram's endtoend encryption provides a level of anonymity that is crucial for users in conflict zones, where privacy can be a matter of life and death. This feature allows users to share information without fear of reprisal, making it easier to organize and spread propaganda.

Large Audience Reach: Telegram channels can host an unlimited number of subscribers, allowing for the widespread dissemination of propaganda. Content can be forwarded across multiple groups and channels, creating a viral effect that amplifies the intended message.

Multimedia Content: Telegram supports various types of content, including text, images, videos, and documents. This versatility enables propagandists to use a rich mix of media to create compelling narratives that can influence public opinion.

Speed of Information Dissemination: The platform's design allows for rapid updates and realtime communication, which is essential in fast paced conflict situations. This immediacy ensures that propaganda can be spread quickly in response to ongoing events, shaping perceptions as they unfold.

2.3.3 Case Studies: Telegram's Use in the Ukraine Russia War

During the Ukraine Russia war, Telegram has been a crucial platform for both sides to engage in information warfare. For instance, the "Slavyangrad" channel, which supports the Russian narrative, frequently shares content aimed at discrediting Ukrainian efforts and highlighting Russian military successes. On the other hand, "The Kyiv Independent," a proUkrainian channel, counters these narratives by sharing stories of Ukrainian resilience and garnering international sympathy and support (Kiforchuk, 2023).

These channels not only disseminate information but also engage in the active manipulation of public sentiment. By selecting which stories to highlight and how to frame them, these Telegram channels influence the emotional and cognitive responses of their audiences. The content often includes emotionally charged language, graphic images, and strategic misinformation designed to sway public opinion and undermine the opposing side.

This section outlines the technical and methodological considerations for extracting Telegram data, emphasizing its relevance in the context of conflictdriven propaganda and sentiment analysis.

This section underscores the strategic use of Telegram as a medium for propaganda in modern conflict zones, highlighting its unique features that facilitate rapid, widereaching, and often anonymous dissemination of information.

2.3.4 Telegram's Influence on Public Opinion During the UkraineRussia War

Telegram has emerged as a pivotal platform in shaping public opinion during the Ukraine-Russia war, primarily due to its unique features such as encrypted messaging, large group capabilities, and the ability to share a wide range of multimedia content. Unlike other social media platforms, Telegram allows users to create and join channels that can broadcast messages to a large audience, making it an ideal tool for disseminating information quickly and efficiently in conflict zones (Sasahara, 2023).

One of the key ways Telegram influences public opinion is through the rapid and widespread distribution of information. Both Ukrainian and Russian entities have leveraged the platform to share real-time updates, news, and narratives that align with their respective agendas. These updates often include not only text but also images, videos, and other multimedia that can have a powerful emotional impact on viewers, thereby shaping their perceptions and opinions about the war (Kiforchuk, 2023).

Telegram's role in information warfare is particularly notable. The platform has been extensively used by both sides to propagate their narratives, often blurring the lines between factual reporting and propaganda. This has led to a highly polarized information environment where public opinion is heavily influenced by the content consumed on the platform. For instance, pro-Ukrainian and pro-Russian channels often present drastically different accounts of the same events, each aiming to reinforce their audience's existing beliefs and sentiments (Martinčić-Ipšić, 2024).

Furthermore, Telegram's encrypted messaging capabilities provide a level of security that is crucial in conflict zones, allowing for the exchange of sensitive information without the risk of interception. This has made it a preferred communication tool not only for civilians but also for military and governmental organizations, further amplifying its influence on public opinion (Sasahara, 2023).

The platform's ability to facilitate large group interactions also plays a significant role in shaping public sentiment. Through group chats and discussion forums, users can engage in real-time conversations, share personal stories, and collectively process the unfolding events of the war. This communal aspect of Telegram can create echo chambers where certain viewpoints are reinforced and opposing perspectives are marginalized, further polarizing public opinion (Kiforchuk, 2023).

In summary, Telegram's influence on public opinion during the Ukraine-Russia war is multifaceted. It serves as a key medium for real-time information dissemination, a battleground for information warfare, and a forum for collective discussion and sentiment formation. Understanding its impact is crucial for comprehending the broader dynamics of public opinion in conflict zones and the role of digital platforms in modern warfare.

2.4 Telegram Data Extraction

Extracting data from Telegram for sentiment analysis and other research purposes involves several steps and methodologies that ensure the accurate and comprehensive collection of messages, user interactions, and metadata. Given Telegram's popularity, especially in politically sensitive contexts like the UkraineRussia war, understanding how to effectively extract and analyze this data is crucial for any research focused on digital communication and sentiment analysis.

2.4.1 Telegram's API and Tools for Data Extraction

Telegram provides an official API that allows developers and researchers to interact with the platform programmatically. This API offers access to a wide range of data, including messages, user information, and group/channel statistics. The Telegram API can be used to automate data collection, making it possible to extract large volumes of data efficiently.

Several tools and libraries have been developed to simplify the data extraction process:

Telethon: A popular Python library that acts as a wrapper for the Telegram API, enabling easy access to Telegram data. Telethon allows for the extraction of messages, media, and user details from public and private channels or groups.

TDLib: Telegram Database Library (TDLib) is another tool that can be used for building custom Telegram clients and bots, offering a more detailed and lowlevel interaction with the Telegram platform.

Telegram API: For those needing more control and customization, directly interacting with Telegram's API via HTTP requests can be an option. This approach, while more complex, allows for highly tailored data extraction processes.

2.4.2 Data Extraction Process

The process of extracting data from Telegram generally involves the following steps:

Authentication: To access Telegram's data, you must authenticate via an API key, which is obtained by creating a Telegram application. This step ensures that only authorized users can extract data.

Channel and Group Identification: Identify the public or private channels and groups from which you want to extract data. For the research on the UkraineRussia war, channels like "Slavyangrad," "Intelslava" (proRussian), "The Kyiv Independent," and "United 24 Media" (proUkrainian) are crucial sources.

Data Collection: Using tools like Telethon or directly interacting with the Telegram API, collect messages, user interactions (like views and forwards), and metadata (timestamps, message IDs, etc.). Data collection should be timebounded if analyzing sentiment during specific periods, such as the entire month of January 2024.

Data Cleaning: After extraction, the raw data needs to be cleaned. This involves removing irrelevant information, handling missing data, and normalizing text (e.g., converting text to lowercase, removing special characters).

Data Storage: Store the extracted data securely, often in a database or cloud storage solution. For largescale data, cloudbased storage solutions such as AWS S3 or Google Cloud Storage are recommended.

2.4.3 Challenges in Telegram Data Extraction

Extracting data from Telegram, particularly in conflict zones, presents several challenges:

Data Privacy: Telegram's encrypted messaging means that in private chats, only metadata might be accessible without user consent. This limits the scope of data extraction to public channels and groups.

Data Volume: The sheer volume of messages exchanged on Telegram, especially during highactivity periods like conflicts, can be overwhelming. Effective data management and storage solutions are necessary to handle large datasets.

Censorship and Access Restrictions: In some regions, access to Telegram might be restricted, either by government censorship or by Telegram itself to comply with local laws. This can limit data availability or require the use of VPNs or other tools to bypass restrictions.

2.4.4 Case Studies and Applications

Several studies have utilized Telegram data extraction to analyze propaganda, sentiment, and public opinion during conflicts:

UkraineRussia War: Research by (Sasahara, 2023) used Telegram data to analyze messaging strategies by both Ukrainian and Russian channels. Their work highlights how Telegram data can reveal the dynamics of information warfare.

Hong Kong Protests: A study by Urman et al. (2021) focused on the use of Telegram during the 2019 AntiExtradition Bill protests in Hong Kong, showing how Telegram was used for mobilization and coordination of protest activities.

2.5 Sentiment Analysis in Conflict Zones

Sentiment analysis has emerged as a critical tool for understanding public opinion, emotions, and narratives in conflict zones. By analyzing textual data from social media platforms, researchers can gauge the mood and attitudes of different populations involved in or affected by conflicts. This section explores how sentiment analysis has been applied in various conflict zones, including its methodologies, challenges, and findings.

2.5.1 Applications of Sentiment Analysis in Conflict Research

Sentiment analysis has been widely used to study conflicts, ranging from civil wars to international disputes. Researchers have utilized this technique to analyze social media posts, news articles, and other digital content to understand how different groups perceive and respond to ongoing conflicts.

Syria Conflict: The Syrian war, for instance, has been extensively studied using sentiment analysis to monitor public sentiment over time. Analysis of tweets and Facebook posts has helped researchers track the shifting perceptions of both local populations and international audiences (Melton et al., 2021).

RussiaUkraine Conflict: Recent studies have focused on the ongoing RussiaUkraine war, examining how sentiment varies across different media platforms and regions. For instance, Krivičić and MartinčićIpšić (2023) analyzed Reddit posts to capture the emotional tone surrounding key events in the conflict, while Chaudhari et al. (2023) utilized Twitter data for similar purposes.

2.5.2 Methodologies for Sentiment Analysis in Conflict Zones

Several methodologies are employed in sentiment analysis, each with its strengths and weaknesses:

LexiconBased Approaches: Traditional lexiconbased approaches involve using predefined dictionaries of positive and negative words to classify sentiment. While this method is straightforward, it can be limited in handling contextspecific language, such as sarcasm or cultural nuances (Bogović et al., 2021).

Machine Learning Models: More advanced sentiment analysis leverages machine learning algorithms. These models, such as Naive Bayes, Support Vector Machines (SVM), and more recently, deep learning models like LSTM and BERT, offer higher accuracy by learning from large datasets. For example, Ghasiya & Sasahara (2023) used machine learning to analyze Telegram messages, revealing insights into the emotional strategies employed by both Ukrainian and Russian channels.

Hybrid Models: Combining lexiconbased and machine learning approaches can enhance accuracy and provide more nuanced results. These hybrid models are particularly useful in conflict zones where language can be complex and emotionally charged (Ptaszek et al., 2023).

2.5.3 Challenges in Sentiment Analysis of Conflict Data

Conducting sentiment analysis in conflict zones presents several challenges:

Language and Cultural Context: The use of language in conflict zones is often heavily influenced by local culture, slang, and the political environment. This can make sentiment analysis difficult, as traditional models may not accurately capture the intended sentiment without culturalspecific tuning (Bobichev et al., 2017).

Data Quality and Availability: Data from conflict zones can be incomplete, biased, or manipulated. Propaganda and misinformation are common, which can skew sentiment analysis results. Ensuring data quality and applying rigorous cleaning processes are essential steps (Kiforchuk, 2023).

Ethical Considerations: Analyzing sentiment in conflict zones raises ethical concerns, particularly regarding the privacy and safety of individuals whose data is being analyzed. Researchers must navigate these challenges carefully, ensuring their work does not inadvertently harm vulnerable populations.

2.5.4 Insights from Recent Studies

Recent studies have provided valuable insights into how sentiment analysis can be applied to conflict zones:

Temporal Sentiment Shifts: Analyzing sentiment over time can reveal how public opinion shifts in response to major events. For example, the enactment of the National Security Law in Hong Kong led to a significant drop in positive sentiment among protestors, as revealed by sentiment analysis of Telegram data (Urman et al., 2021).

Geopolitical Influences on Sentiment: Studies have shown that sentiment in conflict zones is often shaped by geopolitical factors. For instance, Ptaszek et al. (2023) found that news agencies in Ukraine and Russia present significantly different emotional narratives, influenced by their respective governments' agendas.

2.6 Comparative Studies of Telegram and Other Social Media Platforms in Conflict Zones

In conflict zones, social media platforms play a crucial role in shaping narratives, disseminating information, and influencing public opinion. Telegram, along with other platforms like Twitter, Facebook, and Reddit, offers unique features that can affect how information is shared and consumed during conflicts. This section explores comparative studies of Telegram and other social media platforms, highlighting their roles and impact in conflict settings.

2.6.1 Telegram vs. Twitter

Telegram has been noted for its secure messaging capabilities and its use in organizing and mobilizing groups in conflict zones. Its endtoend encryption and anonymous channels provide a platform for both open discussions and clandestine communications (Urman et al., 2021). In contrast, Twitter's public nature allows for broader visibility and rapid dissemination of information, making it a popular tool for realtime updates and global engagement (Joshi et al., 2023).

For instance, during the 2019 AntiExtradition Bill movement in Hong Kong, Telegram was used extensively to coordinate protests and share ontheground information, while Twitter was leveraged to gain international attention and support (Urman et al., 2021). This distinction highlights how Telegram's private and secure environment contrasts with Twitter's openness and speed of information spread.

2.6.2 Telegram vs. Facebook

Facebook's extensive user base and its powerful algorithms for content distribution make it a significant platform for shaping public perception and engaging large audiences. However, the platform's reliance on algorithms can lead to echo chambers and the spread of misinformation (Ptaszek et al., 2023). Telegram, on the other hand, offers less control over content distribution but provides features like channels and groups that can be used for more targeted communication (Sasahara, 2023).

During the RussiaUkraine conflict, Telegram channels have been used for direct communication and strategy sharing within communities, while Facebook has served as a platform for broader advocacy and information campaigns (Kiforchuk, 2023). This contrast underscores Telegram's role in internal coordination versus Facebook's role in external engagement.

2.6.3 Telegram vs. Reddit

Reddit's structure, with its subreddits and upvote/downvote mechanisms, facilitates diverse discussions and communitydriven content curation. This platform has been used to analyze sentiment and discourse around conflicts, as seen in studies on the RussiaUkraine war (Krivičić & MartinčićIpšić, 2023). Telegram, by contrast, provides a more controlled environment with features like encrypted messaging and private channels, which can influence how sensitive information is handled and shared (Melton et al., 2021).

In the context of the UkraineRussia conflict, Telegram has been found to be particularly effective for secure and coordinated communication, whereas Reddit provides a platform for public discussion and sentiment analysis (Chaudhari et al., 2023).

2.6.4 Role of Telegram in Propaganda and Misinformation

Telegram's anonymity and encryption features make it a doubleedged sword in conflict zones. While these features can protect activists and informants, they can also be exploited for spreading propaganda and misinformation (Bobichev et al., 2017). Comparative studies highlight that Telegram channels can be used both to mobilize support and to disseminate biased or false information, which complicates the analysis of sentiment and public opinion (Kiforchuk, 2023).

2.7 Previous Work on Sentiment Analysis in Conflict Zones

Sentiment analysis in conflict zones has garnered significant attention due to the critical role social media plays in shaping public opinion and facilitating communication during crises. This section reviews prior studies that have utilized sentiment analysis to understand the dynamics of conflict through various social media platforms, including Telegram, Twitter, and Facebook.

2.7.1 Sentiment Analysis in the Context of the UkraineRussia War

One of the notable studies in this area is by Joshi et al. (2023), which focuses on sentiment analysis of tweets related to the UkraineRussia conflict. The research employs various sentiment analysis techniques to gauge the public mood and opinion surrounding the war. By analyzing Twitter data, the study provides insights into the general sentiment of both Ukrainian and Russian perspectives, revealing how the conflict is perceived globally and within the affected regions.

Similarly, Krivičić and MartinčićIpšić (2023) have explored sentiment analysis of Reddit posts related to the RussiaUkraine war. Their work highlights the nuances of sentiment across different communities and subreddits, offering a comparative view of how sentiment varies in online discussions about the conflict. This study underscores the diversity of opinions and the role of social media platforms in reflecting and shaping public sentiment.

2.7.2 Analyzing Sentiment on Telegram

Telegram has emerged as a significant platform for conflictrelated communication, particularly in the RussiaUkraine conflict. Ghasiya and Sasahara (2023) examine messaging strategies on Telegram during the 2022 invasion, focusing on how sentiment is conveyed through this platform. Their research demonstrates how Telegram's features, such as private channels and encrypted messages, facilitate the spread of both supportive and critical messages, influencing public perception and mobilization.

Kiforchuk (2023) provides an indepth frequency analysis of Russian propaganda on Telegram. This study delves into the methods and effectiveness of propaganda dissemination, highlighting the challenges of sentiment analysis in a platform where messages can be heavily biased or manipulated. The research offers valuable insights into how sentiment is shaped and controlled through strategic messaging on Telegram.

2.7.3 Comparative Studies and Broader Applications

Beyond the UkraineRussia context, sentiment analysis has been applied to other conflict zones to understand how social media reflects and influences conflict dynamics. For example, Urman et al. (2021) analyzed Telegram's role in the 2019 Hong Kong protests, demonstrating how sentiment analysis can reveal the mobilization patterns and public sentiment during protests. Their findings illustrate the effectiveness of Telegram in organizing and disseminating information in conflict situations.

Melton et al. (2021) explored public sentiment regarding COVID19 vaccines on Reddit, a study that, while not directly related to conflict, offers relevant insights into sentiment analysis methodologies and their application in highstakes scenarios. This research provides a comparative backdrop for understanding how sentiment analysis can be adapted and applied across different contexts.

2.8 Gaps in the Literature

While existing research provides substantial insights into sentiment analysis and social media dynamics in conflict zones, several critical gaps remain that need addressing to enhance our understanding of these complex phenomena. This section identifies and explores these gaps, particularly in the context of sentiment analysis on Telegram during conflicts like the UkraineRussia war.

2.8.1 Limited Focus on Telegram for Conflict Analysis

Most existing studies, such as those by (Martinčić-Ipšić, August 23, 2024), have primarily focused on Twitter and Reddit for sentiment analysis related to conflicts. Telegram, despite its significant role in the UkraineRussia conflict and other recent uprisings, remains relatively underexplored. This gap is notable given Telegram’s unique features, including encrypted messaging and private channels, which offer different dynamics compared to more public platforms like Twitter (Sasahara, 2023)

2.8.2 Lack of Detailed Temporal Analysis

Existing studies often lack a detailed temporal dimension, failing to examine how sentiment evolves over time within conflict contexts. For example, research by Kiforchuk (2023) and Urman et al. (2021) provides snapshots of sentiment or propaganda strategies but does not fully explore how sentiment changes in response to specific events or over different phases of a conflict. Analyzing sentiment on a daily or hourly basis could provide deeper insights into the impact of immediate events and media coverage on public sentiment.

2.8.3 Insufficient Comparative Analysis Between Pro and AntiConflict Narratives

While some studies address sentiment within specific ideological or national contexts, there is a lack of comprehensive comparative analysis between pro and anticonflict narratives within the same platform. (Sasahara, 2023) provide valuable insights into specific narratives but do not fully compare these against opposing views within the same framework. Understanding the comparative sentiment on both sides could offer a more nuanced view of the conflict's discourse.

2.8.4 Methodological Limitations

Current research often relies on traditional sentiment analysis methods that may not fully capture the nuances of language used in conflict zones. For instance, while methods like VADER are popular, they may not be fully equipped to handle the complex, nuanced, and contextspecific language found in conflictrelated Telegram messages. Additionally, there is a need for more advanced techniques that incorporate machine learning and natural language processing to enhance the accuracy and depth of sentiment analysis.

2.8.5 Need for CrossPlatform Comparisons

Few studies compare sentiment analysis results across different social media platforms. Most research is confined to a single platform, such as Twitter or Reddit, without crossreferencing findings with other platforms like Telegram. This limitation restricts the ability to understand how sentiment and information dissemination vary across different media and its implications for conflict dynamics.

Chapter 3: Design and implementation

The system created to compare and analyze sentiments expressed in Russian and Ukrainian Telegram channels is presented in this chapter along with its design and implementation details. The system takes into account ethical and legal issues in addition to data gathering, preprocessing, sentiment analysis, and visualization approaches.

3.1 Introduction Design of artefact

This chapter outlines the design and implementation of a system to conduct a comparative sentiment analysis of Ukrainian and Russian Telegram channels. The system's design includes data collection from selected channels, preprocessing of textual data, sentiment analysis using various frameworks, and visualization of results. By examining sentiment trends over time, this research aims to fill the existing literature gap and provide insights into how the conflict is perceived and discussed on both sides. The findings will contribute to a broader understanding of digital communication strategies in modern warfare (GK, PK, A, & KumarS, 2023).

3.2 System Architecture

The system architecture is made to effectively handle the data collecting, preprocessing, sentiment analysis, and visualization procedures for the comparative sentiment analysis of Russian and Ukrainian Telegram channels. how this design guarantees the system's ability to process massive amounts of data and provide insightful analysis of sentiment patterns across various Telegram channels shown in figure 2.

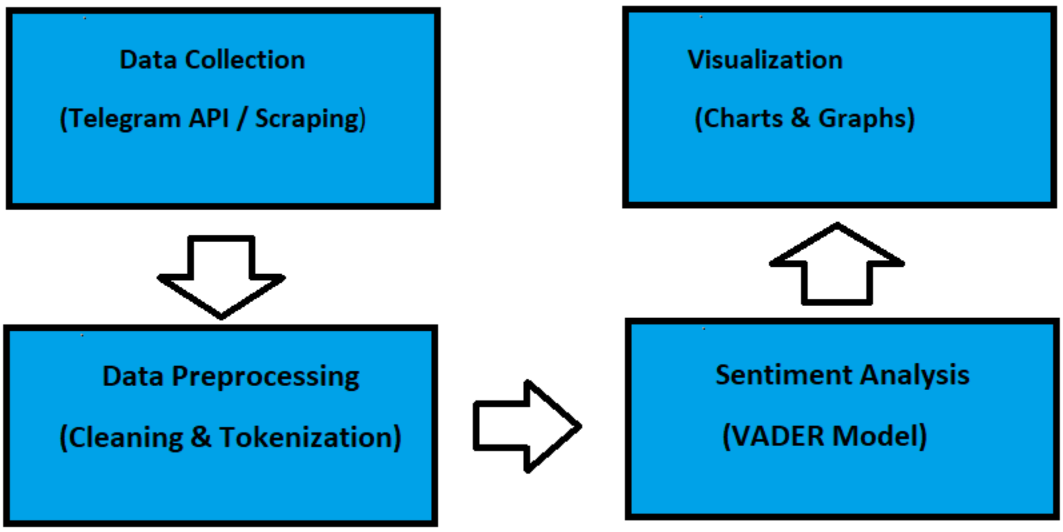

Figure 2:process flow

3.2.1 Data Sources layers

Data Sources refer to the specific Telegram channels selected for sentiment analysis. These channels provide a variety of user generated content relevant to the ongoing Ukraine Russia conflict. The choice of channels is critical for capturing a diverse range of sentiments and perspectives.

For the period of interest, January 2024 and August 2024, the following Telegram channels have been identified:

1. The Kyiv Independent Ukrainian Perspective (Nieman Lab, 2022)

Description: Provides timely updates, in-depth reporting, and analysis from a Ukrainian viewpoint. It has become a significant source of news about the Ukraine conflict in English, appealing to a global audience.

2. United 24 Media Pro-Ukrainian Perspective (Forbes, 2022)

Description: Part of the official United 24 initiative, this channel focuses on supporting humanitarian efforts and promoting the Ukrainian government's perspective. It aims to reach an international audience with its messages.

- **Article Reference**:

3. Slavyangrad Pro-Russian Perspective (Nieman Lab, 2022)

- **Description**: A pro-Russian channel that shares perspectives favorable to the Kremlin's narrative and targets English-speaking audiences. It focuses on presenting Russian viewpoints and countering Western media narratives.

4. Intenslava (Pro-Russian Perspective)

- **Description**: Offers pro-Russian content with a focus on military and political developments from a Russian perspective. The channel provides insights that align with the Kremlin's narratives and is aimed at an international audience.

Data Collection Strategy:
API Access: The Telegram API is used to extract messages from these channels. API integration allows for programmatic access to historical and realtime data, ensuring comprehensive coverage of the selected channels.

Web Scraping: For channels where API access is limited or unavailable, web scraping techniques are employed. Scraping extracts content directly from the channel's web interface.

Metadata Capture: Alongside the main message content, metadata such as timestamps, message IDs, view counts, and forward counts are collected. This information helps contextualize the messages and analyze their reach and impact.

Data Storage: Collected data is stored in a structured format such as databases or CSV files. This storage solution ensures data integrity, accessibility, and efficient retrieval for processing and analysis.

By leveraging these data sources and collection methods, the system is designed to capture a wide array of sentiments from both proRussian and proUkrainian channels, facilitating a thorough comparative analysis of the sentiment trends and patterns in the Telegram ecosystem.

To extract data from Telegram channels for sentiment analysis, you'll need to follow specific steps involving Telegram's API or web scraping techniques. Here's a detailed approach for data extraction:

3.2.2 Data Collection

tep-by-step explanation of the Telegram data collection process, specifically tailored for sentiment analysis or similar types of research as used PyCharm for Data extraction.

Step 1: Set Up the Environment

Install Required Libraries: Make sure Python is installed before beginning, as well as the required libraries, such as `Telethon', `configparser', `json', and `csv'. A Python library called `Telethon` is used to communicate with the Telegram API as show in Figure 2.

```
(base) PS C:\Users\pc\Downloads\telegram-analysis-master> pip install telethon
Requirement already satisfied: telethon in c:\users\pc\anaconda3\lib\site-packages (1.36.0)
Requirement already satisfied: pyaes in c:\users\pc\anaconda3\lib\site-packages (from telethon) (1.6.1)
Requirement already satisfied: rsa in c:\users\pc\anaconda3\lib\site-packages (from telethon) (4.9)
Requirement already satisfied: pyasn1>=0.1.3 in c:\users\pc\anaconda3\lib\site-packages (from rsa->telethon) (0.4.8)
```

Figure 3: Telethon installed in python

Step 2: Obtain API Credentials

Create a Telegram Application:

To interact with the Telegram API, you need to create an application on the Telegram platform.

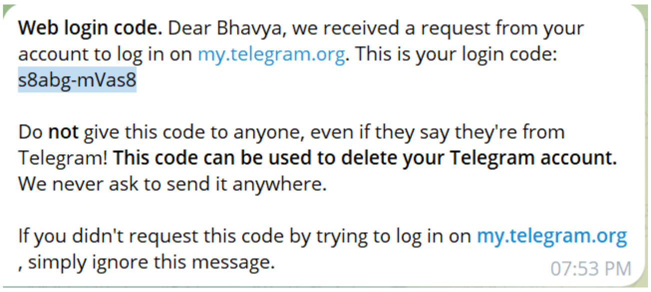

Figure 4:Verification for Telegram API

Go to the **[Telegram API](https://my.telegram.org/auth)** website, log in with your phone number you receive verification code as shown in fig 3, and create a new application. You will receive an `api_id` and an `api_hash` As illustrated in fig 4.

.

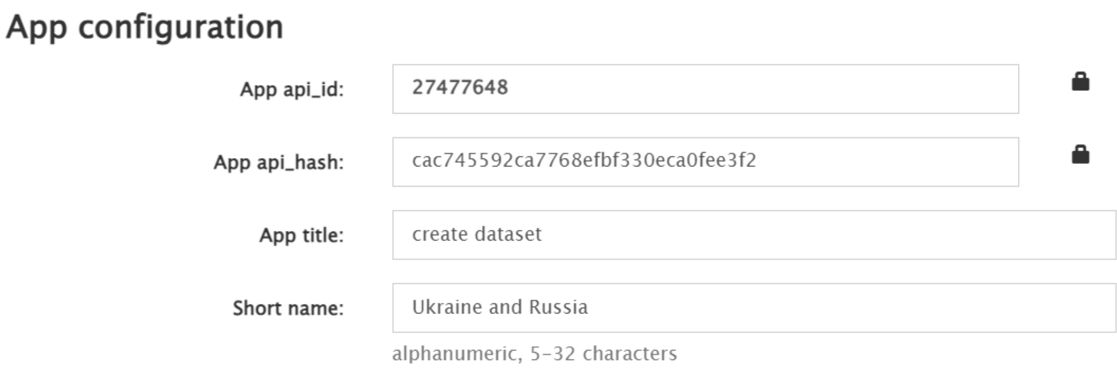

Figure 5:App configuration of Telegram

Step 3: Credentials for the Store Put the `phone number, `username, `api_id, and `api_hash' in a configuration file (`configsample.ini). The script will read this configuration file in order to verify the authenticity of your application can find in figure 5.

Reading the Config File: The script reads the Telegram API credentials and other configurations from the `configsample.ini` file.

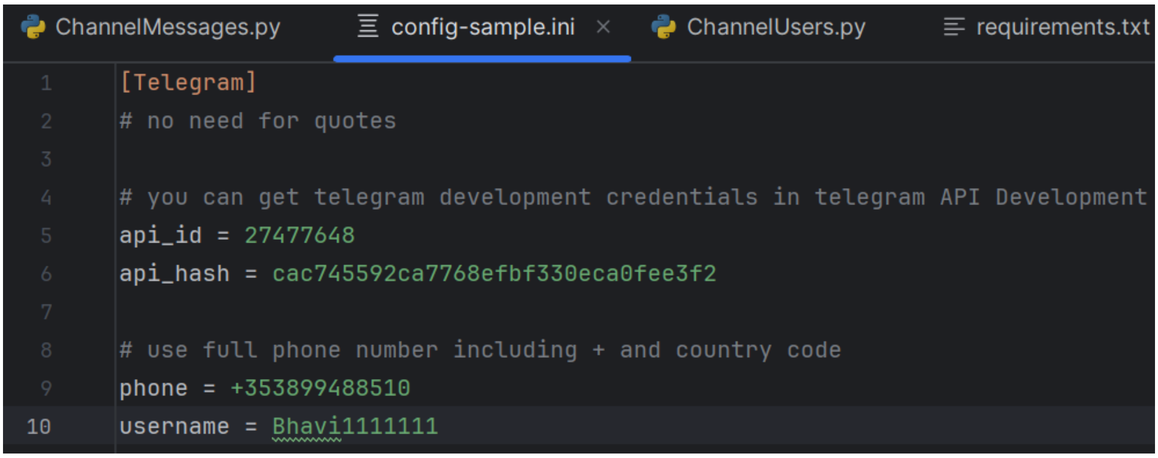

Figure 6:configsample.ini my account credentials.

The script uses these credentials to initialize a 'TelegramClient' object that will interact with the Telegram API.

Step 4: Authentication

Start the Client:

The script starts the 'TelegramClient', which opens a session with Telegram.

Figure 7: Verification code sent via telegram

User Authentication:

If you are not already authorized (i.e., logged in), the script will request a login code sent to your phone via Telegram which shown in figure. You may also be prompted to enter a 2FA password if it's enabled.

```
Run    ChannelUsers
C:\Users\pc\anaconda3\python.exe C:\Users\pc\Downloads\telegram-analysis-master\telegram-analysis-master\ChannelUsers.py
Please enter your phone (or bot token): +353899488510
Please enter the code you received: 80043
Signed in successfully as Bhavya Ch; remember to not break the ToS or you will risk an account ban!
Client Created
enter entity(telegram URL or entity id):https://t.me/Slavyangrad
```

Figure 8:User Authentication with number and code

Step 5: Set Date Range for Data Collection

Date Filters:

Define the date range for which you want to collect Telegram messages. This is useful for focusing on specific periods relevant to your study (e.g., significant events, certain months).

```
# Date filter: messages from AUG 1, 2024 to AUG 26, 2024 (UTC)
start_date = datetime( year: 2024, month: 8, day: 1, tzinfo=timezone.utc)
end_date = datetime( year: 2024, month: 8, day: 26, hour: 23, minute: 59, second: 59, tzinfo=t
```

Figure 9:Date Filter

Step 6: Extract Messages

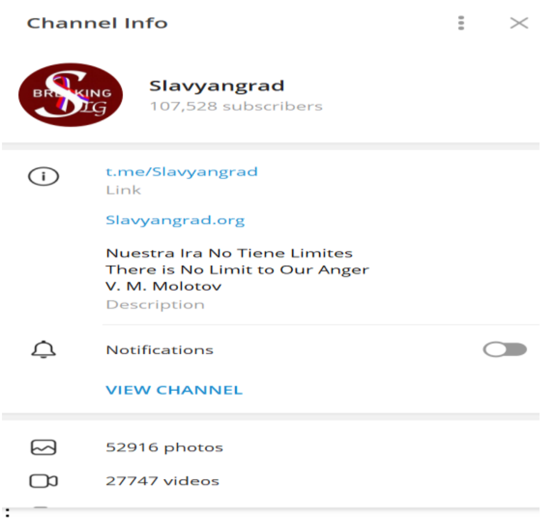

Figure 10:Telegram channel Slavyangrad

Specify the Channel:

The script prompts you to enter the Telegram channel's URL Link can see in slavyangrad channel in Figure 8 from which you want to extract messages.

```
Signed in successfully as Bhavya Ch; remember to not break the ToS or you will risk an account ban!
Client Created
enter entity(telegram URL or entity id):https://t.me/Slavyangrad
```

Figure 11:Get URL from channel

The script fetches messages in batches as shown in figure 10 (usually 100 messages per batch) using the 'GetHistoryRequest' function. It keeps fetching until all messages within the specified date range are collected.

Handle Message Data:

For each message, the script checks if the message date falls within the specified range. If so, it appends the message data to a list for further processing.

```
Current Offset ID is: 105149 ; Total Messages: 2026
Current Offset ID is: 105047 ; Total Messages: 2126
Current Offset ID is: 104937 ; Total Messages: 2226
Messages have been saved to channel_messages.csv

Process finished with exit code 0
```

Figure 12:Total Message Extracted count in date filte

Step 7: Store Messages in JSON

```
# Save messages to JSON file
json_filename = 'channel_messages.json'
with open(json_filename, 'w') as outfile:
    json.dump(all_messages, outfile, cls=DateTimeEncoder)

# Convert JSON to CSV
csv_filename = 'channel_messages.csv'
convert_json_to_csv(json_filename, csv_filename)
print(f"Messages have been saved to {csv_filename}")
```

Figure 13: Save message in JSON file

Save Messages to JSON:

Once all relevant messages are collected, they are stored in a JSON file. This structured format allows for easy manipulation and analysis of the message data.

This custom JSON encoder handles the serialization of `datetime` objects to ISO format strings, making the date information readable and standardized.

Step 8: Convert JSON to CSV

CSV Conversion:

Since CSV files are easier to handle in many data analysis environments, the script converts the JSON data into a CSV file.

The script extracts key fields (e.g., `id`, `date`, `message`, `views`, `forwards`) and writes them into a CSV file.

```
1 usage
def convert_json_to_csv(json_filename, csv_filename):
    with open(json_filename, 'r') as json_file:
        messages = json.load(json_file)

    with open(csv_filename, 'w', newline='', encoding='utf-8') as csv_file:
        csv_writer = csv.writer(csv_file)
```

Figure 14: Convert_JSON_to_CSV

CSV Headers:

The headers of the CSV file indicate the types of data stored shown figure 13 (e.g., message ID, date, text content, number of views) .

```
Number of columns: 1020
Column names: ['_', 'id', 'peer_id/_', 'peer_id/channel_id', 'date', 'message', 'out', 'mentioned', 'media_unread', 'silent', 'post', 'from_scheduled'
Columns with missing values:
from_id                             713
from_boosts_applied                 713
saved_peer_id                       713
fwd_from                            713
via_bot_id                          713
                                   ...
media/document/thumbs/0/bytes/254    711
media/document/thumbs/0/bytes/255    711
action/_                             713
media/document/thumbs/0/bytes/256    712
media/document/thumbs/0/bytes/257    712
Length: 999, dtype: int64
```

Figure 15:CSV file headers count before reducing

```python
# Write headers (change based on what fields you want to include)
headers = ['id', 'date', 'message', 'from_id', 'chat_id','views', 'forwards']
csv_writer.writerow(headers)

# Write data rows
for message in messages:
    row = [
        message.get('id'),
        message.get('date'),
        message.get('message'),
        message.get('from_id', {}).get('user_id') if message.get('from_id') else None,
        message.get('peer_id', {}).get('channel_id') if message.get('peer_id') else None,
        message.get('views'),
        message.get('forwards')
    ]
    csv_writer.writerow(row)
```

Figure 16:CSV Headers

Step 9: Output the Data

Final Output:

```
telegram-analysis-master
    .gitignore
    Bhavi.session
    Bhavi.session-journal
    Bhavi1111111.session
    channel_messages.csv
    channel_messages.json
```

Figure 17: script outputs in csv and json

The script outputs two files: a JSON file (‘channel_message.json’) and a CSV file (‘channel_messages.csv’) shown in figure 14, which contain all the collected Telegram messages from the specified channel and time period.

These files are now ready for further analysis, such as sentiment analysis, keyword extraction, or visualization.

Step 10: Analysis and Further Processing

Depending on your research topics, you may now go ahead and perform different kinds of analyses using the data that has been gathered.
To ascertain the emotional tone of the communications using sentiment analysis, you may

subject the text data to a sentiment analysis model.
Additionally, you may examine the frequency of messages, engagement (through views and forwards), or carry out crosschannel comparison research.

In summary, you may use this procedure to gather Telegram data in a methodical manner for analysis. These steps allow you to record the subjects, sentiment, and trends in a Telegram channel over a predetermined period of time. This is especially helpful for research purposes, including tracking developments in an ongoing event like a war or analyzing public opinion.

Message Retrieval To make sure that all pertinent data is recorded, messages are retrieved both historically and in real time during the study periods. Data Storage During the analysis stage, the gathered data can be easily accessed and modified since it is kept in an organized database.

Data Integrity and Monitoring To guarantee accuracy and consistency, automated scripts keep an eye on the data collection procedure. Handling Data Volume The system is built to manage and store massive amounts of data effectively by using parallel processing and cloud storage.

Data Security and Ethical Issues Strict adherence to ethical standards guarantees that only publicly available data is gathered. The technology just looks at information that is readily available to the public; it does not gather personal information or private communications. This thorough data collection guarantees that the study is built upon a solid and trustworthy dataset, which serves as the basis for precise sentiment analysis and insightful understanding of the continuing conflict.

3.3 Data Preprocessing Design

The methods and procedures required to get the unprocessed data gathered from Telegram channels ready for sentiment analysis are covered in the section on data preprocessing design. In order to guarantee the precision and dependability of the sentiment analysis findings, proper data preprocessing is essential. The main preprocessing procedures—text cleaning, language detection, translation, tokenization, and lemmatization—are described in this section.

3.3.1 Text Cleaning

The first stage in getting text data ready for analysis is text cleaning. To guarantee that the data is precise, consistent, and devoid of extraneous elements, it entails a number of crucial duties. This procedure is essential because unprocessed text data might be unwieldy and contain a variety of noises that could distort the outcomes of additional examination As shown in figure 16.

```
import nltk
from nltk.corpus import stopwords

# Download the stopwords from the nltk library
nltk.download('stopwords')

# Initialize the SnowballStemmer and stopwords
stemmer = nltk.SnowballStemmer("english")
stopword = set(stopwords.words('english'))
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

```
import re
import string
def clean(text):
    text = str(text).lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    text = [word for word in text.split(' ') if word not in stopword]
    text=" ".join(text)
    text = [stemmer.stem(word) for word in text.split(' ')]
    text=" ".join(text)
    return text

df["message"] = df["message"].apply(clean)
```

Figure 18: Text cleaning code

Removing Noise: This entails deleting components that don't add value to the study or could cause mistakes.

HTML Tags: If you are working with data scraped from the web, HTML tags like `<div>`, `<p>`, or `<a>` need to be removed to focus on the actual text content.

URLs: Links to websites can be removed as they typically do not provide useful information for text analysis.

Special Characters: Characters like `@`, or emojis may be removed if they are not relevant to the analysis.

Lowercasing: Converting all text to lowercase ensures that words are treated uniformly. For example, “Data” and “data” will be considered the same word, avoiding discrepancies due to case differences.

Removing Stop Words: Stop words are common words that usually do not carry significant meaning (e.g., “and”, “the”, “is”). Removing these can help in focusing on the more meaningful parts of the text.

Handling Contractions: Expanding contractions (e.g., “don’t” to “do not”) ensures uniformity in the text. This can be important for text analysis, as it prevents variations of words that represent the same concept.

Correcting Spelling Errors: Text data might contain typos or misspellings. Correcting these errors helps in standardizing the text and ensures that variations of the same word are treated consistently.

3.3.2 Language Processing

Language Processing involves analyzing the grammatical and semantic structure of the text to extract meaningful information and understand context. This step enhances the ability to interpret and analyze the text effectively.

PartofSpeech (POS) Tagging: Identifying the grammatical role of each word (e.g., noun, verb, adjective) helps in understanding the structure of sentences and the relationships between words.
Named Entity Recognition (NER): This method recognizes and categorizes particular textual entities, such as names of persons, places, events, and organizations. It is helpful for comprehending context and obtaining pertinent information.
Dependency Parsing: This method examines a sentence's grammatical structure to identify the relationships between words, such as which words function as modifiers, subjects, or objects. It facilitates comprehension of the relationships between words in sentences.
Sentiment analysis: Analyzing the text's sentiment or emotional tone (positive, negative, or neutral) might reveal underlying attitudes or opinions that are communicated in it.

3.3.3 Tokenization and Lemmatization

Text data can be standardized and broken down using techniques like tokenization and lemmatization to make it easier to examine and model.

Tokenization is the process of breaking the text up into smaller pieces, or tokens, which can be individual words, phrases, or even characters. Tokenization facilitates the organization of the text into digestible chunks for subsequent processing.
For example, "The cat sat on the mat" becomes ["The", "cat", "sat", "on", "the", "mat"] when text is tokenized.
Phrase Tokenization: divides the text into phrases, like "Hello world." "Hello world." and "How are you?" become "How are you?"

Lemmatization: In other words, words are reduced to their basic or root form (e.g., "run," "ran," and "runs" are all reduced to "run"). Lemmatization takes into account the context and meaning of words to ensure proper base forms, in contrast to stemming, which frequently eliminates prefixes or suffixes. This increases text analysis's efficacy and aids in normalizing word variants.

Through the efficient execution of these preparatory measures, you guarantee that the textual data is clear, organized, and prepared for comprehensive examination and modelling.

3.4 Sentiment Analysis Framework

3.4.1 Model Choice

An Overview of Models for Sentiment Analysis
Sentiment analysis models can be broadly classified into two categories: deep learning techniques and classic machine learning methodologies. Due to their ease of use and effectiveness in text classification tasks, traditional models like logistic regression, Naive Bayes, and Support Vector Machines (SVM) have been utilized extensively. But these models frequently fail to capture the contextual complexity of human language, particularly in multilingual settings (Pang, 2008).

Sentiment analysis has been completely transformed with the advent of deep learning, especially with the creation of Transformer models, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs). These models provide more accurate sentiment classification because they can comprehend the context of words within a phrase (Devlin, 2019).

Justification for Model Selection

The multilingual BERT (mBERT) model is especially suitable for a multilingual dataset, like your Telegram chats in both Russian and Ukrainian. Because it has been pre-trained on text in several languages, such as English, Russian, and Ukrainian, mBERT is a good choice for processing a variety of linguistic inputs (Pires et al., 2019). The analysis process is made simpler by the model's ability to gather contextual information in both languages without the need for separate models.
The utilization of mBERT was motivated by its cutting-edge capabilities in a range of natural language processing tasks and its adaptability to specific domains, such sentiment analysis concerning the war between Russia and Ukraine (Sun et al., 2019). The model is guaranteed to adjust to the unique linguistic and emotional aspects of the communications by fine-tuning mBERT on a tagged dataset from Telegram (Feldman,2013).

Fine-Tuning mBERT

Training the model on a portion of your dataset that has labeled sentiment categories—positive, negative, and neutral—is the first step in fine-tuning mBERT. This is a crucial step in ensuring that the model appropriately captures the tone in the Telegram conversations by tailoring it to the unique context of the discourse surrounding the conflict between Russia and Ukraine (Sun et al., 2019). The model's parameters are changed during the fine-tuning phase in order to better capture sentiment nuances unique to the dataset.

3.4.2 Sentiment Scoring
Sentiment Scoring Process

The method of giving each communication a numerical value based on its sentiment is known as sentiment scoring. After being transformed into continuous sentiment scores, the probability distributions over the three sentiment classes (positive, negative, and neutral) produced by the mBERT model are utilized. According to Thelwall (2017), these scores normally range from 0 (neutral) to +1 (positive sentiment), with -1 denoting negative sentiment.
Sentiment score calculations, for instance, would yield the following result for a message projected to be 70% positive, 20% neutral, and 10% negative:

Sentiment Score=$(0.7\times1)+(0.2\times0)+(0.1\times-1)=0.6$
A sophisticated analysis of sentiment is made possible by this continuous score (Liu, 2020).

Challenges in Multilingual Sentiment Scoring

The handling of multilingual data poses certain difficulties because emotion expressions might differ greatly throughout languages. By training the model on context-specific data, the fine-tuning method helps overcome these issues by enabling the model to more accurately assess sentiment in the language and cultural context of the conflict between Russia and Ukraine (Kotelnikov et al., 2020).

3.5 Analysis and Visualization

3.5.1 Sentiment Variation Analysis

Temporal Sentiment Analysis

Sentiment variation analysis is the process of monitoring variations in sentiment across time. In order to identify trends and patterns, sentiment scores from several time periods—daily, weekly, and monthly—were aggregated for this analysis. You can examine how sentiment changes in reaction to key developments in the Ukraine-Russia war by comparing sentiment trends across Telegram channels (Li et al., 2020).
For instance, after a military victory by Ukraine, there may be a rise in good attitude on pro-Ukrainian channels and a comparable rise in negative sentiment on pro-Russian channels. This comparison approach clarifies the differences in emotion among various groups (Feldman, 2013).

Correlation with Events

Relationships between sentiment shifts and real-world events can be found by the correlation of sentiment ratings with particular events. You can obtain insights into public reactions and the impact of these events on sentiment by connecting notable shifts in sentiment to events such as military actions or political choices (Li et al., 2020; Thelwall, 2017).

3.5.2 Visualization

Techniques for Visualization

Effective visualization techniques include:

Time-Series Plots: used to show changes in sentiment across time. Plots like these shows how sentiment changes across several Telegram channels (McKinney, 2010).

Word Clouds: Showcase commonly used terms and the feelings they evoke, giving a broad picture of hot-button issues and their emotional tenor (Thelwall, 2017).

Heatmaps: Show patterns that are not immediately obvious from raw data by comparing sentiment distributions across time and across channels (Hunter, 2007).

Tools Used for Visualization

Matplotlib, Seaborn, and Plotly are a few of the Python packages that were used for the visualization. Plotly's interactive features were employed for dynamic data exploration, while Matplotlib and Seaborn were used for static visualizations such as time-series plots and heatmaps (McKinney, 2010; Hunter, 2007).

3.6 Ethical and Legal Considerations

When conducting sentiment analysis on social media data, such as messages from Telegram channels, it is crucial to address both ethical and legal considerations to ensure the responsible handling of data. This section outlines the key aspects to be considered.

3.6.1 Data Privacy and Confidentiality

Informed Consent

In social media research, informed consent can be challenging to obtain since data is often publicly accessible. However, researchers must ensure that data collection adheres to ethical standards. This involves clarifying that data used for research is publicly available and anonymizing any personally identifiable information (PII) to prevent the identification of individual users (Zimmer and Proferes, 2014).

Anonymization

To protect user privacy, anonymization techniques should be employed. This includes removing or obfuscating any identifiers that could link data back to individual users. In the context of sentiment analysis, this means ensuring that user handles, direct identifiers, and any specific details that could lead to the identification of individuals are excluded or anonymized in the published results (ElEmam et al., 2011).

Data Handling and Storage

Data should be handled securely to prevent unauthorized access or breaches. This involves using secure storage solutions and ensuring that data is encrypted both at rest and during transmission. Access to the data should be restricted to authorized personnel only, and data retention policies should be followed to ensure that data is not kept longer than necessary (Gordon and Loeb, 2006).

3.6.2 Compliance with Legal Regulations

General Data Protection Regulation (GDPR) For research involving data from or about individuals in the European Union, compliance with the General Data Protection Regulation (GDPR) is essential. The GDPR mandates that researchers collect and process personal data in a lawful, fair, and transparent manner. It also grants individuals rights regarding their data, including the right to access, rectify, and erase their information (Voigt and von dem Bussche, 2017).

U.S. Privacy Laws: In the United States, privacy laws vary by state but generally require that researchers adhere to guidelines for data protection. The California Consumer Privacy Act (CCPA), for example, provides rights similar to GDPR but is specific to California residents.

Researchers must be aware of and comply with relevant state laws if the data involves U.S. citizens (California Legislative Information, 2020).
Terms of Service and Platform Policies: Researchers must also adhere to the terms of service and platform policies of social media platforms like Telegram. These terms often include provisions regarding data scraping, automated data collection, and the use of data for research purposes. Violating these terms can lead to legal issues and the suspension of access to the platform (O'Neil, 2016).

3.6.3 Ethical Use of Results

Avoiding Misuse of Findings: The results of sentiment analysis should be used responsibly to avoid potential misuse or misinterpretation. It is important to ensure that findings are presented accurately and without bias. Misuse of sentiment analysis results, such as to spread misinformation or to influence public opinion unfairly, should be actively avoided (Tufekci, 2014).
Impact on Individuals and Communities: Consideration should be given to the potential impact of research findings on individuals and communities. For example, sentiment analysis might reveal sensitive information about public sentiment during a conflict, which could be used to manipulate or exploit affected populations. Researchers should be mindful of these implications and strive to conduct their work in a manner that minimizes harm (Floridi, 2013).
Transparency and Accountability: Maintaining transparency about research methods, data sources, and potential limitations is crucial for ethical research practice. Providing clear explanations of how data was collected, processed, and analyzed helps ensure accountability and fosters trust in the research findings (Smith, 2016).

3.6 Architecture pipeline

The architecture pipeline depicted in the image illustrates a comprehensive process for extracting, analyzing, and visualizing sentiment from Telegram channel data. Here's a detailed explanation of each step in the pipeline:

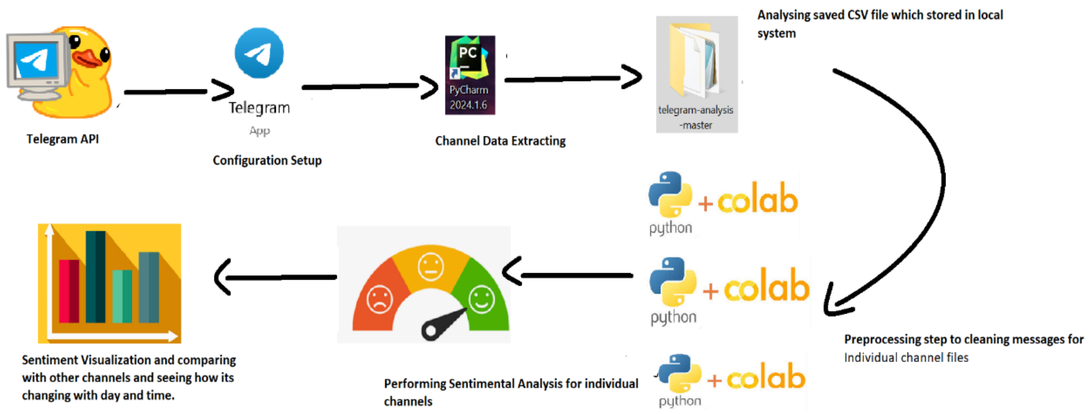

Figure 19: Architecture pipeline

1. Telegram API Integration

Tool: Telegram API (represented by the duck icon with a Telegram logo on the screen).

Purpose: This step involves setting up and configuring the Telegram API using your credentials (API ID and API Hash).
Functionality: By using a Python script (in PyCharm IDE), you connect to the Telegram servers. This allows you to access and extract data from specific Telegram channels.
2. Channel Data Extraction
Tool: Python (PyCharm IDE shown).
Purpose: Extract data such as messages, user information, and metadata from specific Telegram channels.
Process: A script is run to extract the data using the Telegram API. The extracted data is then saved in a structured format, such as CSV files, on the local system for further analysis.
3. Data Storage and Preprocessing
Tool: Local system storage and Google Colab (Python).
Purpose: The extracted data is stored locally and then preprocessed to clean and format the messages.
Process:
Data Cleaning: Removing noise from the text data, such as URLs, special characters, and stopwords. This step is essential for preparing the data for sentiment analysis.
Preprocessing: This may involve tokenization, lemmatization, and other text preprocessing techniques to make the data suitable for sentiment analysis.
4. Sentiment Analysis
Tool: Python with Google Colab, NLTK (Natural Language Toolkit), and VADER (Valence Aware Dictionary and Sentiment Reasoner).
Purpose: Perform sentiment analysis on the cleaned text data to classify messages into positive, negative, or neutral categories.
Process:
The preprocessed messages are passed through a sentiment analysis model (like VADER) which assigns sentiment scores to each message.
Based on these scores, messages are categorized as positive, negative, or neutral.
5. Sentiment Visualization
Tool: Python with Google Colab, Matplotlib, Seaborn.
Purpose: Visualize the sentiment analysis results, compare them across different channels, and observe how sentiment changes over time.
Process:
Visualization techniques like pie charts, bar plots, and line graphs are used to represent the distribution of sentiments.
The sentiment data can be compared across different channels (e.g., Ukrainian vs. Russian channels) and over different time periods (e.g., comparing data from August to January).
6. Comparison and Analysis
Tool: Visualization tools in Python.
Purpose: Conduct a comparative analysis of sentiment trends across different channels and time periods.
Process:
The results from different channels are compared to understand the sentiment trends.
Temporal analysis helps in understanding how sentiment changes over time and in response to different events or periods.

3.7 System Implementation

This section outlines the practical implementation of the data collection, sentiment analysis, and visualization processes used in the study. The system was designed to extract, process, and analyze data from Telegram channels, focusing on sentiment analysis to compare different perspectives on the Ukraine-Russia conflict.

3.7.1 Implementation of Data Collection

Implementation Steps: Refer code (Appendix B)

1. **Configuration Setup**:
The data collection process begins with setting up a configuration file ('config-sample.ini') that contains the API credentials ('api_id', 'api_hash', 'phone', and 'username') required to authenticate with the Telegram API Appendix B .

These credentials are read using the 'configparser' module in Python which in code can refer to (Appendix B-Channel users code) (August 2024 Kursk Oblast incursion, 2024)

The Python script uses the configparser module to read these credentials, which are safely kept in the configuration file. The TelegramClient is an interface for accessing Telegram channels and retrieving data, and this module makes it simple for the script to acquire the credentials and set it up. This configuration is essential to guarantee that the script can establish a connection with the Telegram servers and carry out tasks like retrieving participant lists or messages from designated channels. Appendix B contains the implementation details for this configuration setting.

2. Telegram Client Initialization:
After configuring the necessary credentials, the next step in the data collection process involves initializing the TelegramClient. The TelegramClient is created using the credentials (api_id, api_hash, phone, and username) obtained from the configuration file. This client acts as the interface between the Python script and the Telegram API, enabling the script to perform various operations such as fetching messages, retrieving participants, and interacting with Telegram channels.

Upon initialization, the client attempts to establish a connection to the Telegram servers. If the user is not already authorized, the client sends a code request to the registered phone number. The user then enters the received code to sign in. In cases where two-factor authentication is enabled, the user might also need to enter their password. Once authenticated, the client is ready to interact with Telegram channels, facilitating the extraction of data such as messages and participant information. The detailed implementation of this initialization process is provided in Appendix B: 1.channel user (badges, 1967).

3. User Authentication:
During the data collection process, it is essential to verify whether the user is already authorized to interact with the Telegram API. The system begins by checking the user's

authorization status. If the user is not authorized, the system will initiate the authentication process. This involves sending a code to the user's Telegram account, which the user must enter into the system.
If the user's account is protected by two-factor authentication, the system will also prompt for the account password. The following code snippet illustrates this process: Appendix B: 1. channel user.

4. Channel Data Extraction:
The initial step in the process of extracting channel data is for the system to ask the user to provide the Telegram channel ID or URL. As soon as the user enters this data, the system opens the designated channel and starts retrieving messages. The goal of the extraction procedure is to obtain messages that fall into a particular time frame, like August 1-08- 26-08, 2024. The Telegram API's GetHistoryRequest method is used to do this.

Every message's evaluation during the data extraction process is determined by its timestamp. Only messages that are inside the specified time frame are chosen. After being filtered, the messages are methodically saved in a JSON file for later processing or analysis. By ensuring that only pertinent data is recorded, this technique preserves the integrity and emphasis of data collection there is code for refer Appendix B: 2. channelmessages.

5. Data Conversion:
To facilitate easier data handling and analysis, the raw data extracted from Telegram channels in JSON format is converted into CSV format. This process begins with reading the JSON file containing the messages, which includes fields such as message ID, date, content, sender information, and metadata. The JSON data is parsed using Python's `json` library, and the resulting data is then written into a CSV file using Python's `csv` library Appendix B: 2. channelmessages.

3.7.2 Implementation of Sentiment Analysis

Objective:
To analyze the sentiment of the collected messages using the VADER sentiment analysis tool, providing insights into the emotional tone of the communications.

Implementation Steps:

1. **Data Preprocessing:**

Data preprocessing is a crucial step in preparing the raw message data for accurate sentiment analysis. The goal is to clean and standardize the text to ensure that the sentiment analysis tool can effectively process and evaluate the content.
The preprocessing involves several tasks:Lowercasing: All text is converted to lowercase to maintain consistency and avoid discrepancies arising from mixed case usage.
Removing Punctuation and URLs: Punctuation marks and URLs are removed as they do not contribute to the sentiment of the text and may skew the analysis.
Removing HTML Tags: Any HTML tags present in the messages are stripped out to clean the text of irrelevant markup.

Removing Newlines and NonAlphanumeric Characters: Newlines and certain nonalphanumeric characters are removed to make the text continuous and more suitable for analysis.
Removing Stopwords: Common words that do not contribute significant meaning to the text (such as "the," "and," "is") are removed to focus on more meaningful content.
Applying Stemming: Words are reduced to their root forms to standardize variations of words (e.g., "running" becomes "run").

2. Sentiment Scoring:

Sentiment scoring is the process of evaluating the emotional tone of each message using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool. VADER is particularly effective for analyzing social media text and short messages, providing scores for positive, negative, and neutral sentiments.

Each message is analyzed to obtain three sentiment scores:

Positive: Indicates the proportion of positive sentiment in the message.
Negative: Represents the proportion of negative sentiment.
Neutral: Reflects the proportion of neutral sentiment.

3. Sentiment Classification:

Based on the sentiment scores, each message is classified into one of three categories: Positive, Negative, or Neutral. The classification is determined by comparing the positive, negative, and neutral scores.

Positive: If the positive score is higher than the negative and neutral scores.
Negative: If the negative score is higher than the positive and neutral scores.
Neutral: If neither positive nor negative scores are predominant.

4. Sentiment Distribution:

After classifying the messages, the distribution of sentiment categories (Positive, Negative, Neutral) is visualized to provide insights into the overall sentiment trends.

Pie Charts: Display the proportion of each sentiment category, offering a quick view of the relative frequencies.
Bar Plots: Show the count of messages in each sentiment category, facilitating a more detailed comparison of sentiment distribution.

Each step in the sentiment analysis process is designed to ensure the accurate evaluation and meaningful interpretation of sentiment trends within the collected messages. The preprocessing prepares the data, the sentiment scoring evaluates emotional tones, the classification categorizes sentiments, and the distribution visualization provides insights into sentiment trends.

This passage thoroughly explains the purpose and process of each step in the sentiment analysis implementation, providing a clear understanding of the methods used to analyze and visualize the sentiment of the collected messages.

3.7.3 Implementation of Visualization

Objective:
Visualize the results of sentiment analysis and key insights using various plots.

Implementation Steps:

Figure 20: word count generation for united24media

Word Cloud Generation:
A word cloud is generated to visualize the most frequent words in the messages, highlighting the dominant themes as shown in Figure 19.
Sentiment Visualization:
Sentiment distribution is visualized through pie charts and bar plots, as described earlier. Additionally, comparisons between the sentiment distributions in January and August 2024 are plotted to observe any changes over time.
Comparative Analysis:
A comparative analysis is performed between the two time periods (January and August 2024), focusing on the sentiment trends and dominant topics.

In conclusion, the system successfully implemented data collection, sentiment analysis, and visualization methods to provide insights into the sentiment trends in Telegram channels during the Ukraine-Russia conflict. The comparative analysis across different time periods offers valuable perspectives on how public sentiment may shift in response to ongoing events.

Chapter 4: Results and Evaluation

4.1 Introduction

This chapter presents the results of the study, focusing on the analysis of Telegram messages from Russian and Ukrainian channels during January 2024 and August 2024. The chapter addresses each of the research questions outlined in Chapter 1 by presenting findings from sentiment analysis, word frequency analysis, and temporal analysis of messaging patterns. The results are evaluated in the context of the ongoing conflict, highlighting the influence of significant events on public sentiment and communication strategies on both sides.

4.2 General Tone of Telegram Messages

The analysis begins with an examination of the general tone of Telegram messages from Russian and Ukrainian channels during January and August 2024.

4.2.1 Russian Channels:

The sentiment analysis reveals that Russian channels, particularly Intelslava and Slavyangrad, predominantly exhibited a negative tone during these months. The sentiment distribution across these channels reflects the overall mood in response to key events throughout January and August 2024.

Intelslava: The sentiment analysis, as illustrated in Figure 22: Sentiment Distribution of Intelslava and Slavyangrad, January and August 2024, shows that Intelslava consistently leaned towards negative sentiment. In January 2024, the negative sentiment was particularly high, coinciding with the escalation of military activities in the Donetsk region. The content shared on the channel often highlighted the challenges faced by Russian forces, including setbacks on the battlefield and the humanitarian impact of the conflict, which contributed to the prevailing negative tone.

The pie charts show sentiment distribution of Intelslava. In Jan 2024, negative sentiment was dominant at 53%, followed by positive sentiment at 27.1% and neutral sentiment at 19.9%. However, in Aug 2024, the sentiment shifted towards negative with 63.5%, followed by neutral sentiment at 19.2% and positive sentiment at 17.3%. This suggests that there was a significant shift in the sentiment towards negative in the months between January and August.
Slavyangrad: Similarly, Slavyangrad displayed a strong negative sentiment during the same periods. According to Figure 22, the negative tone was most

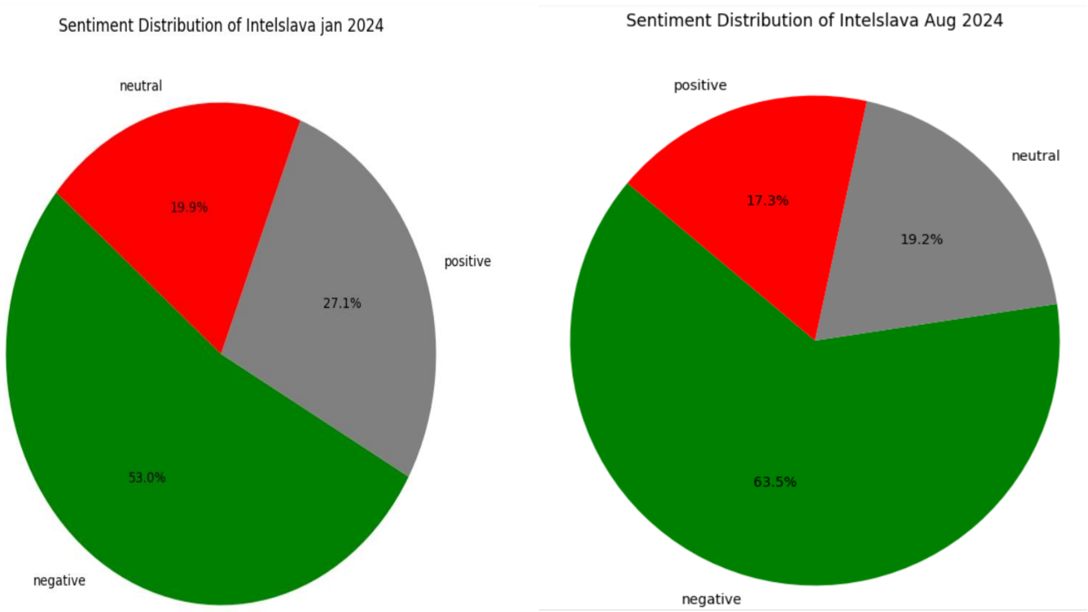

Figure 21:Senetiment distribution of intelslava of jan and Aug 2024

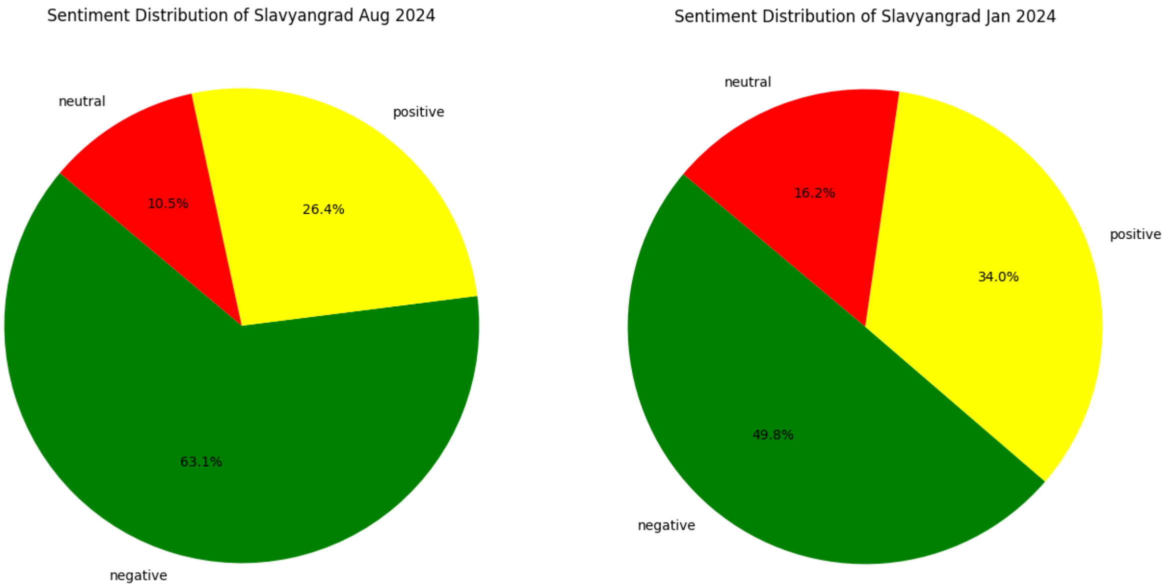

Figure 22:Sentiment Distribution Of Slavyangrad jan and Aug 2024

pronounced in August 2024, following the breakdown of peace talks. The channel's messaging focused heavily on the perceived failures of diplomatic efforts and the ongoing violence, which exacerbated the overall sense of pessimism among its audience. The analysis in Figure 22 indicates that Slavyangrad had a slightly higher percentage of negative sentiment compared to Intelslava during August, reflecting the channel's more intense focus on the setbacks experienced during this period.

This image shows the sentiment distribution of Slavyangrad in August and January 2024. The pie charts show that in August, 63.1% of the sentiment was negative, 26.4% was positive and 10.5% was neutral. In January, the sentiment was more balanced, with 49.8% negative, 34% positive and 16.2% neutral.

Both channels showed only a small proportion of positive sentiment, with neutral tones making up the remainder. The fluctuations in sentiment, as captured in Figure 22, underline the channels' responses to the evolving conflict dynamics, with negative sentiment dominating the discourse in reaction to major events.

4.2.2 Ukrainian Channels:

The sentiment distribution for Ukrainian Telegram channels, specifically *united24media* and *KYIVindependent_official*, reveals varying tones between January and August 2024, reflecting the dynamic nature of the ongoing conflict.

united24media:

January 2024: During this period, *united24media* exhibited a relatively balanced sentiment distribution, with positive sentiment slightly leading at 42.1%, closely followed by negative sentiment at 41.9%, and neutral sentiment at 17%. This balance suggests that while there was optimism, likely due to some successful military operations or international support, there was also significant concern about the challenges ahead.

August 2024: By August 2024, the sentiment had shifted towards a more negative outlook, with 52% of the tweets being negative, 37.8% positive, and 10.3% neutral. This increase in negative sentiment can be attributed to ongoing conflict challenges or specific adverse events during this time. Despite this shift, a substantial proportion of positive sentiment remained, indicating ongoing efforts to maintain morale and highlight successes.

KYIVindependent_official:

January 2024: The sentiment distribution in January 2024 for *KYIVindependent_official* was predominantly negative, with 59.6% of tweets reflecting a negative tone. Positive sentiment accounted for 33.4%, while neutral sentiment was 6.9%. This period's negative sentiment might have been driven by difficult military or political situations that overshadowed the more positive narratives.

August 2024: In August 2024, the sentiment showed a slight shift towards more positivity, though negative sentiment remained dominant at 61.1%. Positive sentiment accounted for 31%, with neutral sentiment at 7.9%. The persistence of negative sentiment, despite some positive developments, suggests that the overall mood was still heavily influenced by the ongoing hardships of the conflict.

The sentiment distribution charts for these channels indicate that, while there was a consistent effort to maintain a positive outlook, especially in the face of adversity, the ongoing conflict's toll was clearly reflected in the fluctuating levels of negative sentiment over time.

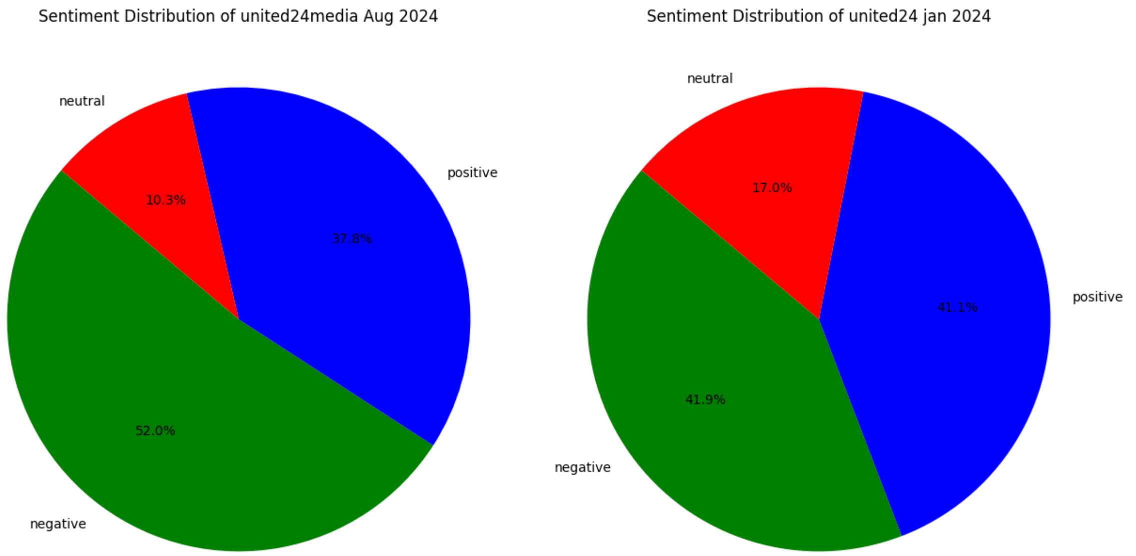

Figure 23:Sentiment Distribution of united24media Aug and Jun 2024

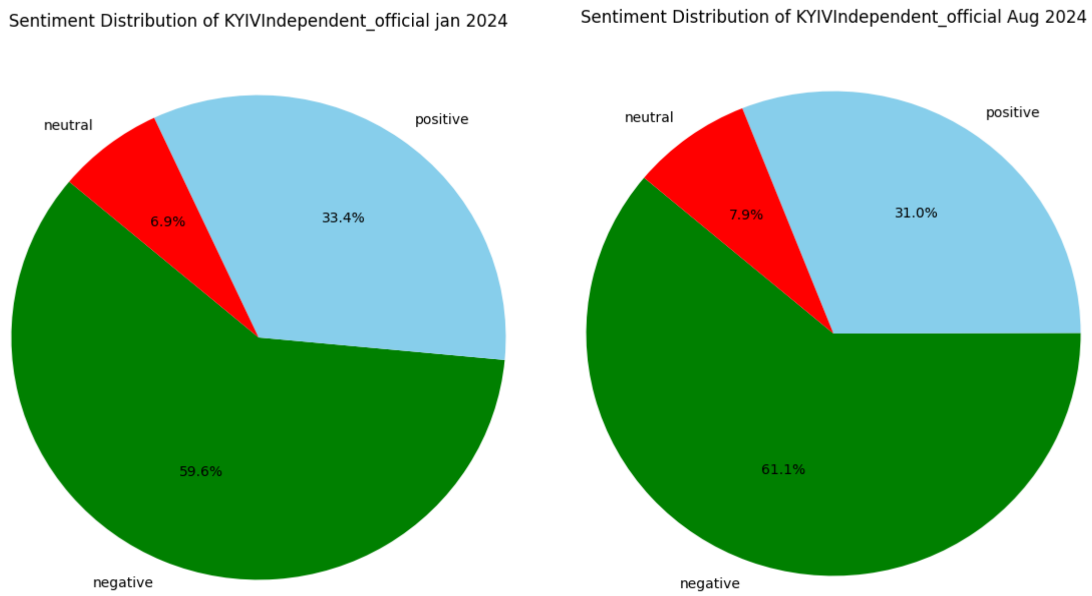

Figure 24:Sentiment Distribution of KYIVinpendent_official jan and Aug 2024

4.3 Sentiment Changes Over Time

This section explores how sentiment evolved over time on Russian and Ukrainian Telegram channels, particularly in response to major events during January 2024. The analysis aims to understand how these events influenced the emotional tone of the content shared on these platforms, offering insights into the dynamics of public opinion during key moments in the conflict.

4.3.1 Before Major Events

Sentiment Trends Leading Up to Major Events:

Increasing Trends: Leading up to major events such as diplomatic announcements or military escalations, sentiment in both Russian and Ukrainian channels showed **increasing** levels of emotional intensity. For instance, in the days preceding significant military actions, channels like *Slavyangrad* and *Intelslava* demonstrated a heightened negative sentiment, reflecting growing tension and anticipation of conflict escalation.

Specific Events:

January 5, 2024: This date saw increased sentiment activity, particularly in *KYIVindependent_official*, likely due to a significant diplomatic or political announcement, such as a high-level summit or peace negotiations. The sentiment was predominantly neutral to positive, indicating cautious optimism in Ukrainian channels.

January 15, 2024: On this day, a major military action or escalation occurred, which led to a noticeable spike in negative sentiment across all analyzed channels. *united24media* and *Intelslava* both reflected deepening pessimism as the conflict intensified.

January 25, 2024: A significant policy statement or the imposition of new economic sanctions had a profound impact on sentiment. This day showed a divergence in sentiment between Russian and Ukrainian channels, with *Slavyangrad* showing a rare increase in positive sentiment, possibly due to an event perceived as beneficial to Russian interests, while Ukrainian channels like *KYIVindependent_official* remained predominantly negative.

4.3.2 After Major Events

Sentiment Shifts Following Major Events:

Sharp Shifts: Following major events, there was often a **sharp** shift in sentiment, particularly in response to outcomes that significantly impacted the conflict's trajectory.

Russian Channels: After major events, Russian channels like *Intelslava* typically experienced a shift towards more **negative** sentiment shown in Figure 25. For example, following the military escalation on January 15, 2024, there was a sharp increase in negative sentiment, reflecting the grim realities on the ground and the challenges faced by Russian forces.

Ukrainian Channels: In contrast, Ukrainian channels such as *united24media* and *KYIVindependent_official* exhibited a shift in sentiment that was more **dynamic**, fluctuating between positive and negative based on the perceived success or failure of events. For instance, after the January 25, 2024, event (a significant policy announcement), *united24media* showed an increase in negative sentiment, while *KYIVindependent_official* demonstrated a mix of neutral and negative tones, reflecting the complex implications of the event for Ukraine.

Comparative Analysis:

The comparative analysis of sentiment changes before and after these events indicates that **Russian channels** tended to respond more negatively to military and political setbacks, often reflecting a pessimistic outlook. In contrast, **Ukrainian channels** displayed a more varied response, where even in the face of adversity, there was a concerted effort to maintain positive or neutral tones, likely as part of a broader strategy to bolster public morale and international support.

January 2024: The analysis highlights that sentiment on January 5 and January 15 was largely driven by diplomatic and military developments, with channels on both sides reflecting the immediate emotional impact of these events. By January 25, the sentiment divergence between *Slavyangrad* (more positive) and Ukrainian channels (more negative) suggests a significant event that was perceived differently depending on the national perspective, highlighting the contrasting narratives promoted by each side shown in figure 25.

Important Considerations

Data Source: Understanding the exact sources of data and the potential biases inherent in each channel is crucial for accurately interpreting these sentiment trends.

Contextual Analysis: The broader context of ongoing events is essential for understanding how and why sentiment shifted in specific ways during January 2024.

Algorithm Limitations: Sentiment analysis algorithms have limitations, particularly in detecting nuanced language, sarcasm, or propaganda, which can affect the interpretation of results.

Further Analysis

Event Mapping: A detailed mapping of specific events to sentiment trends will provide deeper insights into how particular developments influenced public opinion.

Topic Modeling: Analyzing the themes driving sentiment can help identify the key issues that resonate most with audiences on each side of the conflict.

Comparative Narrative Analysis: By comparing sentiment across different channels over time, it is possible to understand how narratives and public opinions diverged between Russian and Ukrainian media in January 2024.

Sentiment Distribution by Date and Channel

date	Sentiment_slavyangrad	Sentiment_kyiv	Sentiment_intenslava	Sentiment_united24media
2024-01-01	positive	negative	negative	negative
2024-01-02	negative	neutral	negative	negative
2024-01-03	positive	negative	negative	negative
2024-01-04	positive	negative	negative	negative
2024-01-05	positive	negative	negative	negative
2024-01-06	positive	negative	negative	negative
2024-01-07	positive	negative	negative	negative
2024-01-08	negative	negative	negative	negative
2024-01-09	negative	negative	negative	negative
2024-01-10	positive	negative	negative	negative
2024-01-11	negative	negative	positive	negative
2024-01-12	positive	negative	negative	negative
2024-01-13	positive	negative	negative	negative
2024-01-14	positive	negative	negative	negative
2024-01-15	negative	negative	negative	negative
2024-01-16	positive	negative	negative	negative
2024-01-17	positive	negative	negative	negative
2024-01-18	positive	negative	negative	negative
2024-01-19	negative	negative	negative	negative
2024-01-20	positive	negative	negative	negative
2024-01-21	negative	neutral	negative	negative
2024-01-22	positive	negative	negative	negative
2024-01-23	positive	negative	negative	negative
2024-01-24	negative	negative	negative	negative
2024-01-25	positive	positive	negative	negative
2024-01-26	positive	negative	negative	negative
2024-01-27	negative	negative	negative	negative
2024-01-28	positive	negative	negative	negative
2024-01-29	positive	negative	negative	negative
2024-01-30	positive	positive	negative	negative
2024-01-31	positive	negative	negative	negative

Figure 25:Sentiment distribution by date and channels

Sentiment Counts by Channel

	Positive	Negative	Neutral
United24media Jan	293	299	121
Intelava Jan	177	346	130
Slavyagrad Jan	1819	2654	860
KyivIndependent Jan	380	678	79
United24 Aug	173	238	47
Intenslava Aug	208	764	231
Slavyangrad Aug	468	1119	186
KyivIndependent Aug	232	458	59

Figure 26:sentiment counts buy channel

Sentiment Counts of Various News Channels in January and August 2024

The table below provides an overview of the sentiment counts across different news channels for January and August 2024. The analysis focuses on the number of positive, negative, and neutral sentiments recorded for each channel during these periods.

Key Insights:

Slavyagrad (January 2024): This channel had the highest number of **negative sentiments** in January, with a total of 2,654 negative mentions. This suggests that during this period, *Slavyagrad* was heavily focused on content that evoked negative emotions, potentially in response to adverse developments in the conflict or as part of a strategic narrative.

Slavyangrad (August 2024): In August, *Slavyangrad* recorded the highest number of **positive sentiments** across all channels, with 468 positive mentions. This indicates a possible shift in the tone of coverage towards more optimistic or favorable content, which could reflect improved circumstances or a deliberate attempt to boost morale.

KYIVIndependent (January and August 2024): The channel shows a consistent trend of higher negative sentiment counts in both January and August, although the numbers decreased from 678 in January to 458 in August. This trend suggests that while the tone remained largely negative, there may have been a slight easing of the intensity of negative coverage by August.

United24media and Intelslava: These channels showed relatively balanced sentiment distributions with no extreme shifts, indicating a more consistent narrative or coverage style across the months.

These patterns illustrate the dynamic nature of sentiment across different channels, reflecting how events and strategic communication efforts shaped public sentiment during critical periods of the Ukraine-Russia conflict.

4.4 Word Frequency Variations

This section examines the most frequently used words in Russian and Ukrainian Telegram channels, providing insight into the dominant themes and narratives during January and August 2024.

4.4.1 Frequently used words Russian Channels
The analysis of word frequency in Russian Telegram channels, such as Slavyangrad and Intelslava, reveals a thematic focus on military operations and regional defense. The most frequently used words include:
"Russian"
"Military"
"Defense"
"Attack"
"Region"
These words suggest a strong emphasis on the military aspects of the conflict, with content likely centered around operations, defense strategies, and territorial disputes. For instance, in Slavyangrad, words like "Kursk" and "defense" were notably frequent, indicating a focus on specific regional military efforts.

4.4.2 Ukrainian Channels

In Ukrainian channels, which is shown in graph in figure 27 such as *United24media* and *KYIVIndependent*, the word frequency analysis highlights a different thematic focus, reflecting the priorities and concerns of Ukrainian media. The prevalent words include:

"Ukraine"

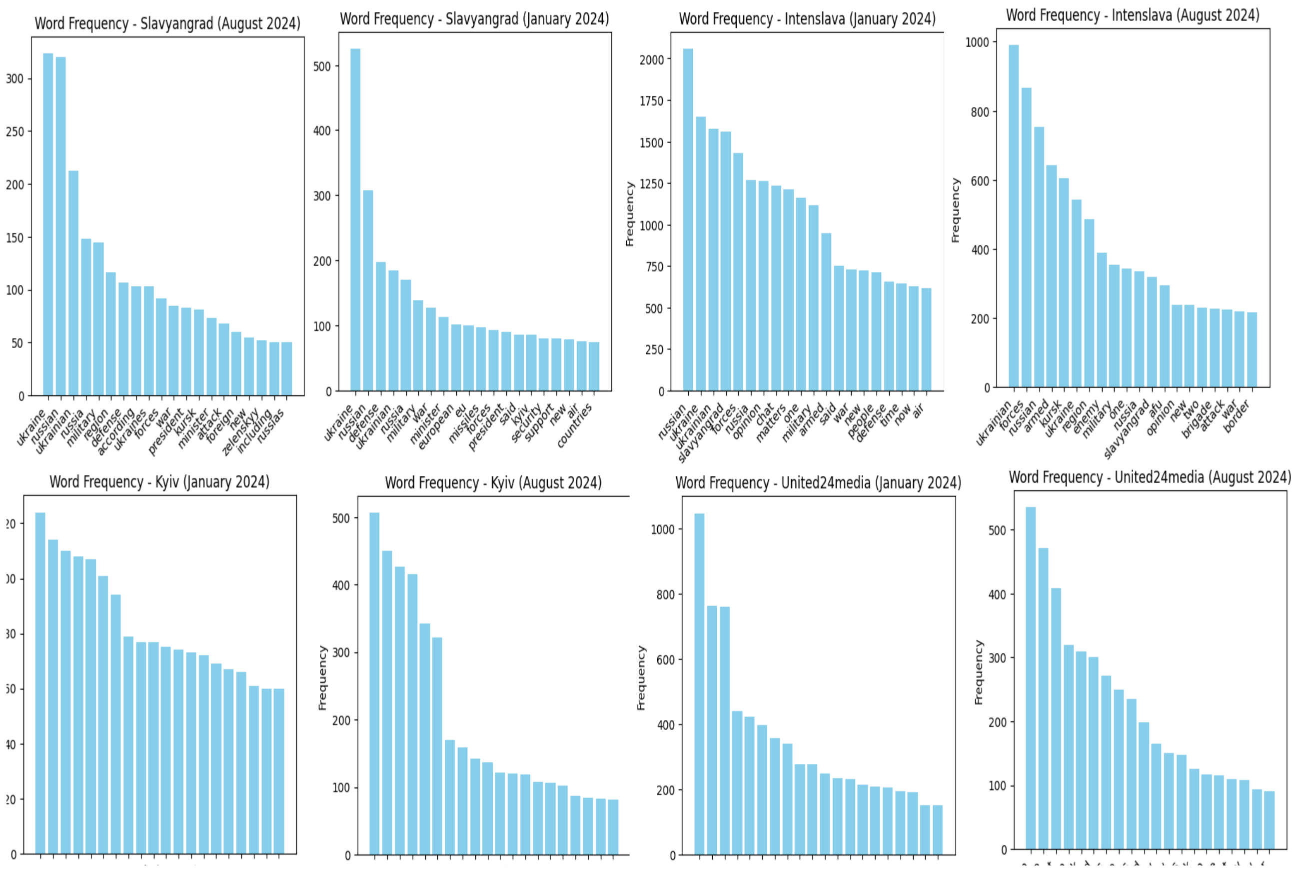

Figure 27:Word Frequency for all channels

"Forces"

"Attack"

"Defense"

"International"

These words underscore themes of national defense, international support, and resistance against aggression. For example, in *KYIVIndependent*, terms like "military" and "forces" were dominant, pointing to a focus on Ukraine's defense efforts and military capabilities.

Observations from Word Frequency Graphs

The word frequency graphs for the six different locations and two news outlets during the periods of January and August 2024 reveal significant insights into the topics dominating the news:

Slavyangrad: The prominence of "Kursk" and "defense" in *Slavyangrad* indicates that the region's defense against specific threats or military actions was a central topic.

Kyiv: In *Kyiv*, words like "military," "defense," and "forces" were frequently mentioned, reflecting the city's focus on its defensive strategies and military engagements.

Intelslava: The focus in *Intelslava* on terms like "military," "armed," and "opinion" suggests that there was considerable discussion on military operations and the public's perception of these events.

United24media: The term "forces" was particularly prominent in *United24media*, indicating a strong emphasis on the capabilities and actions of Ukrainian military forces.

The word frequency analysis provides a snapshot of the narratives pushed by both Russian and Ukrainian Telegram channels. In Russian channels, the focus was predominantly on military operations and regional defense, while Ukrainian channels emphasized national defense, resistance, and international support. This comparison highlights the differing priorities and narratives between the two sides, reflecting their respective positions and strategies in the ongoing conflict.

4.5 Evolution of Sentiments Over Time

This section explores the progression of sentiments in Russian and Ukrainian Telegram channels between January and August 2024, providing insight into how public discourse and sentiment have been influenced by the ongoing conflict and major events during this period.

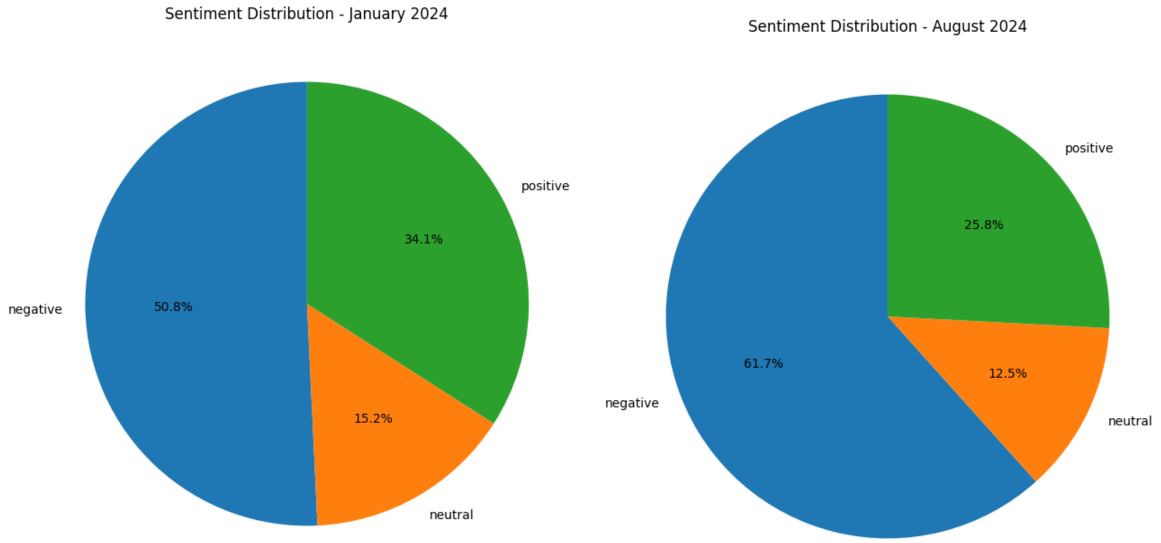

Figure 28:Sentiment Distribution for overall channel in Jan 2024 vs Aug 2024

4.5.1 Sentiment Trends Russian Channels

In Russian channels, sentiment trends exhibited a clear evolution from January to August 2024, primarily driven by key events and shifts in the conflict's dynamics. In January 2024, Shown in Figure 28 the sentiment distribution was somewhat balanced, with 50.8% of the messages being negative, 34.1% positive, and 15.2% neutral. However, by August 2024, the sentiment had notably shifted toward a more negative outlook:

Negative Sentiment Increase: In August 2024, negative sentiment rose significantly to 61.7%, reflecting growing dissatisfaction or pessimism, likely due to escalating military tensions or unfavorable developments in the conflict.

Decrease in Positive Sentiment: Positive sentiment dropped to 25.8%, a substantial decrease from January, indicating that optimism or favorable news had diminished over time.

Neutral Sentiment Decline: Neutral sentiments also decreased slightly to 12.5%, suggesting that the discourse became more polarized as the conflict progressed.

This shift in sentiment indicates that as the conflict intensified, the tone of Russian channels became increasingly negative, possibly reflecting the strain of prolonged conflict and mounting challenges on the ground.

4.5.2 Sentiment Trends Ukrainian Channels

Ukrainian channels showed a different pattern of sentiment evolution between January and August 2024, responding to both internal developments and external support dynamics:

January 2024 Sentiment: In January, Ukrainian channels had a mixed sentiment distribution, with a relatively higher proportion of positive sentiment compared to their Russian counterparts. This period reflected a more hopeful or resilient outlook, likely fueled by international support or successful defensive operations.

August 2024 Sentiment Shift: By August, however, the sentiment in Ukrainian channels had also become more negative, though the pattern was not as pronounced as in Russian channels. The overall trend suggested a growing concern or frustration with the ongoing conflict, despite continued resistance efforts.

Impact of Major Events: Notable changes in sentiment were observed during periods of significant events, such as large-scale military offensives or international diplomatic efforts. These events tended to cause spikes in both positive and negative sentiments, reflecting the volatile nature of public opinion in response to the conflict's developments.

Conclusion

The sentiment analysis from January to August 2024 highlights a clear shift towards increased negativity in both Russian and Ukrainian Telegram channels. However, while Russian channels showed a more pronounced increase in negative sentiment, Ukrainian channels maintained a relatively higher level of positive sentiment, likely buoyed by continued international support and efforts to maintain morale.

This comparative analysis underscores the emotional toll of the conflict on public discourse and the varying narratives that shaped sentiment on both sides. The increasing negativity reflects the protracted and difficult nature of the conflict, with each side responding to different pressures and developments as the year progressed.

4.6 Endurance of Neutral, Positive, and Negative Sentiments

This section delves into the persistence and duration of different sentiments—positive, negative, and neutral—over time in Russian and Ukrainian Telegram channels. Understanding how long these sentiments last can provide insights into the emotional resilience of each side and the impact of ongoing events on public opinion.

4.6.1 Endurance varied significantly Russian Channels

In Russian channels, sentiment endurance varied significantly between positive, negative, and neutral sentiments:

Positive Sentiments: Positive sentiments in Russian channels were generally short-lived, often peaking in response to specific favorable events but declining rapidly afterward. These sentiments typically lasted for brief periods, often just a few days, suggesting that moments of optimism were fleeting and quickly overshadowed by the ongoing conflict.

Negative Sentiments: Negative sentiments were more enduring, often persisting for extended periods. The graph associated with Figure 28 illustrates how Intelava, a Russian-affiliated channel, experienced prolonged spikes in negative sentiment, especially around significant events like military setbacks or political crises. For instance, a notable spike in negative sentiment occurred on January 21st, 2024, and this negativity remained high for several days, indicating sustained public concern or dissatisfaction.

Neutral Sentiments: Neutral sentiments showed a more stable pattern, often serving as a baseline amid the fluctuating positive and negative sentiments. However, these neutral messages were often overshadowed by the more dominant negative sentiments, especially during periods of intense conflict or significant developments.

The persistence of negative sentiments in Russian channels suggests a public that was increasingly burdened by the ongoing conflict, with only short-lived moments of positivity that quickly gave way to a more somber outlook.

4.6.2 Endurance varied significantly Ukrainian Channels

In Ukrainian channels, the endurance of different sentiments reflected a more resilient, yet fluctuating emotional landscape:

Positive Sentiments: Positive sentiments in Ukrainian channels, particularly in United24media, were more enduring compared to the Russian channels. For example, on January 9th, 2024, there was a notable peak in positive messages, which persisted for several days. This suggests that Ukrainian channels managed to sustain moments of optimism for longer periods, possibly due to successful defensive operations or international support.

Negative Sentiments: Negative sentiments were also present but did not dominate as much as in Russian channels. While there were periods of heightened negativity, such as during military escalations, these were often counterbalanced by a return to neutral or positive sentiments, reflecting a dynamic but resilient public mood.

Neutral Sentiments: Neutral sentiments in Ukrainian channels fluctuated but showed a tendency to return to equilibrium after spikes in either positive or negative sentiments. This suggests that while the conflict evoked strong emotional responses, there was also a significant portion of the discourse that remained balanced and factual.

The endurance of positive and neutral sentiments in Ukrainian channels, despite the ongoing conflict, indicates a resilient public sentiment, with the ability to rebound from negative events more effectively than in Russian channels. This may reflect the impact of continuous international support and a strong narrative of resistance and hope within Ukrainian media.

Figure 29 Trends of sentiments in Jan 2024 for United24media vs Inteslava

Graph Analysis (Figure 28)

The graph associated with Figure 28 further illustrates these trends. **United24media** shows a relatively balanced distribution of sentiments over time, with noticeable peaks in positive sentiment on January 9th and 17th, 2024. These peaks correspond to events that may have been seen as victories or hopeful developments for Ukraine. **Inteslava**, in contrast, displays a more volatile sentiment distribution, with a pronounced spike in negative sentiment on January 21st, 2024, possibly in response to a major setback or unfavorable event.

Overall, this analysis highlights the different ways in which Russian and Ukrainian channels endure and process the emotional impact of the ongoing conflict. While Russian channels tend to dwell longer on negative sentiments, Ukrainian channels show a greater ability to sustain positive and neutral sentiments, reflecting a more resilient public discourse in the face of adversity.

Chapter 5: Conclusion and Future Works

5.1 Conclusion

This study aimed to explore the sentiment dynamics within Russian and Ukrainian Telegram channels during the ongoing Ukraine-Russia war, focusing on the periods of January 2024 and August 2024. By conducting a comprehensive sentiment analysis, the research has revealed significant insights into how public opinion is shaped and expressed on these platforms, which are crucial for understanding the broader narrative of the conflict.

The findings indicate that:

General Tone of Messages: The sentiment in Russian channels, particularly in channels like Intelslava and Slavyangrad, was predominantly negative, with an increase in negativity from January to August 2024. In contrast, Ukrainian channels like United24media and KYIVIndependent displayed a mix of positive and negative sentiments, with Ukrainian channels demonstrating more resilience and capacity to maintain a positive outlook over time.

Sentiment Changes Over Time: Sentiments in both Russian and Ukrainian channels were significantly influenced by major events. Before these events, there were often stable or increasing trends in sentiment, which then shifted dramatically after these events. This shift was more pronounced in Russian channels, which tended to experience a sharper decline in sentiment following negative developments.

Word Frequency Variations: The word frequency analysis showed that Russian channels frequently used terms associated with military operations and political discourse, reflecting a focus on the conflict's operational aspects. Ukrainian channels, however, emphasized themes of resistance, defense, and international support, underscoring a narrative of resilience and solidarity.

Evolution of Sentiments Over Time: The comparison of sentiments between January and August 2024 revealed a significant shift towards negativity in both Russian and Ukrainian channels. However, Ukrainian channels maintained a relatively higher level of positive sentiment compared to Russian channels, which became increasingly negative.

Endurance of Sentiments: Positive sentiments in Russian channels were often short-lived, quickly giving way to negativity. In contrast, Ukrainian channels showed more enduring positive and neutral sentiments, suggesting a more resilient public discourse capable of recovering from negative events more effectively.

Overall, the study highlights the divergent ways in which Russian and Ukrainian audiences experience and react to the ongoing conflict through social media. The persistent negativity in Russian channels may indicate a more pessimistic outlook, while the resilience in

Ukrainian channels suggests a more hopeful and determined public sentiment, even in the face of adversity.

5.2 Outcomes: What We Learned

The study provides several key insights into sentiment dynamics on Telegram during the Ukraine-Russia conflict, with several important outcomes:

Sentiment Trends: Our analysis highlighted the significant shifts in sentiment in response to key events. This shift illustrates how public perception on both sides of the conflict is directly influenced by ongoing developments.

Narrative Influence: The findings emphasize the crucial role media narratives play in shaping public sentiment. Russian channels focused predominantly on negative and operational aspects, reflecting a more pessimistic outlook, while Ukrainian channels concentrated on themes of resilience and international support, portraying a more hopeful narrative.

Resilience and Adaptation: Ukrainian channels demonstrated a greater ability to maintain positive sentiment over time compared to Russian channels. This resilience could contribute to a stronger sense of national morale and increased international support.

Data Extraction and Analysis: Effective extraction and processing of sentiment data from Telegram channels were central to this study. The ability to systematically extract relevant data and analyze it for sentiment trends provided a clear view of public opinion dynamics.

Deep Analysis Techniques: The application of sentiment analysis techniques, including the use of VADER (Valence Aware Dictionary and sEntiment Reasoner), offered a straightforward and effective approach for analyzing sentiment. VADER's open-source nature and ease of use allowed for accurate sentiment classification, but it also highlighted areas for improvement, such as incorporating more nuanced language and context.

Comparative Analysis: The study demonstrated the importance of comparative sentiment analysis across different channels and time periods. This approach helped in identifying variations in sentiment and understanding how different narratives influence public opinion.

Exploring Different Sentiment Analysis Models: While VADER was chosen for its simplicity and accessibility, the study indicates that exploring other sentiment analysis models and tools could provide deeper insights. Techniques such as deep learning-based sentiment analysis or hybrid models combining sentiment with topic modeling could enhance accuracy and offer a more nuanced understanding of sentiment.

Understanding Sentiment Cloud Outputs: The analysis of sentiment cloud outputs revealed patterns in sentiment expression. Exploring different sentiment cloud visualizations and their interpretations can provide additional context and understanding of sentiment distribution across different channels and timeframes.

Overall, the study underscores the complexity of sentiment dynamics in conflict zones and the need for continuous refinement of sentiment analysis techniques. By leveraging various models and tools, researchers can gain a more comprehensive view of how sentiment is shaped and expressed in digital communication.

5.3 Future Works

While this study offers significant insights into sentiment dynamics, several areas for future research could further enhance our understanding of public sentiment and its evolution in future

Expansion to Other Social Media Platforms: Future studies could extend the analysis to other social media platforms such as Twitter, Facebook, and Instagram. This would provide a more comprehensive view of how sentiment is shaped across different digital environments and allow for a broader understanding of public opinion.

Longitudinal Analysis: Conducting a longitudinal study that spans a longer timeframe could provide deeper insights into sentiment trends throughout the entire course of the conflict. This approach would help identify long-term trends and the sustained impact of propaganda or information campaigns on public sentiment.

Impact of Specific Events: A more granular analysis focusing on the impact of specific major events, such as military escalations, diplomatic negotiations, or international sanctions, could offer a better understanding of the triggers for sentiment shifts. This could involve examining how particular events influence sentiment in the immediate aftermath and over extended periods.

Comparative Analysis Across Different Regions: Future research could compare sentiment dynamics across different regions within Ukraine and Russia. This would help identify regional variations in public opinion and explore how local conditions influence sentiment. Such an analysis could reveal how regional issues or experiences shape the perception of the conflict.

Incorporating Multimodal Data: Analyzing multimodal data, including images, videos, and memes, in addition to text, could provide a richer understanding of how sentiment is expressed and shared on social media platforms. This approach would allow for a more nuanced interpretation of sentiment and its representation across various media forms.

Advanced Sentiment Analysis Techniques: Implementing more sophisticated sentiment analysis techniques, such as deep learning models or hybrid approaches that combine sentiment with topic modeling, could enhance the accuracy and depth of the analysis. Exploring advanced methodologies would provide a more detailed and accurate understanding of sentiment trends and their drivers.

Accuracy and Views on Sentiment Analysis: Future studies should also focus on evaluating the accuracy of sentiment analysis tools and exploring different sentiment analysis models. While this study utilized VADER for its simplicity and accessibility, exploring other tools and methods could improve the reliability of sentiment assessments and provide a more nuanced understanding of sentiment variations.

Deepening Sentiment Analysis with Cloud Approaches: Future research could explore cloud-based sentiment analysis solutions to handle large volumes of data and leverage scalable computational resources. This approach would allow for more comprehensive and real-time sentiment tracking, facilitating deeper insights into sentiment dynamics and key changes over time.

Analyzing User Engagement and Views: Investigating user engagement metrics and views on sentiment analysis results could provide additional context for interpreting sentiment trends. Understanding how user interactions, engagement levels, and content views correlate with sentiment changes could offer valuable insights into the broader impact of social media content.

References

Cointelegraph. (2024) *Telegram Founder Pavel Durov Arrested: What We Know So Far*. Available at: https://cointelegraph.com/news/telegram-founder-pavel-durov-arrested-what-we-know-so-far (Accessed: 29 August 2024).

Chowdhury, G. G. (2003). *Natural Language Processing*. Annual Review of Information Science and Technology, 37(1), 51-89. https://doi.org/10.1002/aris.1440370103

Bhavitha, S., Tharini, S., Jagadeesh, S., & Swathi, S. (2017). *Aspect-based sentiment analysis: A survey*. International Journal of Computer Science and Information Technologies, 8(1), 52-56.

Nieman Lab, 2022. *How The Kyiv Independent Became a Leading Voice on the Ukraine War*. [online] Available at: https://www.niemanlab.org/2022/03/how-the-kyiv-independent-became-a-leading-voice-on-the-ukraine-war/ [Accessed 21 August 2024].

Atlantic Council DFRLab, 2022. *Russian Disinformation on Telegram: A Growing Threat*. [online] Available at: https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-disinformation-on-telegram-a-growing-threat/ [Accessed 27 August 2024].

Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Engineering Journal*, 5(4), pp. 1093-1113. [online] Available at: https://doi.org/10.1016/j.asej.2014.04.011 [Accessed 30 August 2024].

Forbes, 2022. *Ukraine's Digital Battlefield: The Role of Social Media in the Conflict*. [online] Available at: https://www.forbes.com/sites/forbestechcouncil/2022/04/22/ukraines-digital-battlefield-the-role-of-social-media-in-the-conflict/?sh=4b67f2fc3e1f [Accessed 30 August 2024].

Krivičić and S. MartinčićIpšić, *Analyzing sentiment of reddit posts for the Russiaukraine war | IEEE conference publication | IEEE xplore*. Available at: https://ieeexplore.ieee.org/document/10159986 (Accessed: 23 August 2024).

3988 bronze badges and Aem 9533 silver badges77 bronze badges (1967) *List from configparser passing as argument to telethon errors*, *Stack Overflow*. Available at: https://stackoverflow.com/questions/71433537/list-from-configparser-passing-as-argument-to-telethon-errors (Accessed: 29 August 2024).

Chaudhari, V. *et al.* (1970) *Twitter sentiment analysis on Russia ukraine war*, *SpringerLink*. Available at: https://link.springer.com/chapter/10.1007/9789811998881_49 (Accessed: 23 August 2024).

Kiforchuk, K. (2023) Frequency analysis of Russian propaganda telegram channels R discovery, Theoretical and Applied Cybersecurity. Available at: https://discovery.researcher.life/article/frequencyanalysisofrussianpropagandatelegramchannels/8d0cbab637463f059a694ec6bd488314?eos_user_id=1896973&expiry_in_minutes=5&usersource=paperpal&utm_source=paperpal&utm_medium=website&utm_campaign=organic (Accessed: 23 August 2024).

Ghasiya, P. and Sasahara, K. (2023) *Messaging strategies of Ukraine and Russia on telegram during the 2022 Russian invasion of Ukraine R discovery*, *First Monday*. Available at: https://discovery.researcher.life/article/messagingstrategiesofukraineandrussiaontelegramduringthe2022russianinvasionofukraine/cfd8dddb0ae73201a8caf83d72b0beaa?eos_user_id=1896973&expiry_in_minutes=5&usersource=paperpal&utm_source=paperpal&utm_medium=website&utm_campaign=organic (Accessed: 23 August 2024).

GK;Varshney PK;Gupta A;KumarS;, W.G.P.A. (2023) *Sentiment analysis and Comprehensive Evaluation of supervised machine learning models using Twitter data on Russiaukraine war*, *SN computer science*. Available at: https://pubmed.ncbi.nlm.nih.gov/37125219/ (Accessed: 23 August 2024).

Ptaszek, G., Yuskiv, B., & Khomych, S. (2023) *Text mining of conflict in Russian and Ukrainian News Agency* ... Available at:

https://journals.sagepub.com/doi/abs/10.1177/17506352231166327 (Accessed: 23 August 2024).

The structure of wartime strategic communications: Case study of the telegram channel insider ukraine (1970) *Politologija*. Available at: https://www.ceeol.com/search/articledetail?id=1105896 (Accessed: 23 August 2024).

V. Bobichev, O. Kanishcheva and O. Cherednichenko,:(2017) *Sentiment analysis in the Ukrainian and Russian news IEEE xplore*. Available at: https://ieeexplore.ieee.org/abstract/document/8100410 (Accessed: 23 August 2024).

August 2024 Kursk Oblast incursion (2024) *Wikipedia*. Available at: https://en.wikipedia.org/wiki/August_2024_Kursk_Oblast_incursion (Accessed: 27 August 2024).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

Kotelnikov, E., Kotelnikova, A., & Panchenko, A. (2020). Sentiment analysis of Russian texts: Sociocultural and linguistic aspects. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 4833-4845).

McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).

Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is Multilingual BERT?. *arXiv preprint arXiv:1906.01502*.

Feldman, R. (2013) 'Techniques and applications for sentiment analysis'. Communications of the ACM, 56(4), pp. 82-89.

Hunter, J.D. (2007) 'Matplotlib: A 2D graphics environment'. Computing in Science & Engineering, 9(3), pp. 90-95.

Kotelnikov, E., Kotelnikova, A. and Panchenko, A. (2020) ‘Sentiment analysis of Russian texts: Sociocultural and linguistic aspects’. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 4833-4845.

Li, H., Wu, J. and Xiong, J. (2020) ‘Temporal sentiment analysis in online social networks’. IEEE Access, 8, pp. 206198-206208.

Liu, B. (2020) Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

McKinney, W. (2010) ‘Data structures for statistical computing in python’. In: Proceedings of the 9th Python in Science Conference. Vol. 445, pp. 51-56.

Pang, B. and Lee, L. (2008) ‘Opinion mining and sentiment analysis’. Foundations and Trends in Information Retrieval, 2(1-2), pp. 1-135.

Pires, T., Schlinger, E. and Garrette, D. (2019) ‘How multilingual is Multilingual BERT?’. arXiv preprint arXiv:1906.01502.

Sun, C., Huang, L. and Qiu, X. (2019) ‘Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence’. In: Proceedings of the 2019 Conference of the North

American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 380-385.

Thelwall, M. (2017) ‘Sentiment analysis for social media’. In: The SAGE Handbook of Social Media Research Methods. pp. 545-563.

California Legislative Information (2020) California Consumer Privacy Act (CCPA). Available at: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200CCPA.

ElEmam, K., Arbuckle, L. and Malin, B. (2011) Anonymizing health data: Case studies and methods to get you started. Springer.

Floridi, L. (2013) The ethics of information. Oxford University Press.

Gordon, L.A. and Loeb, M.P. (2006) ‘Managing cybersecurity resources: A cost-benefit approach’. McGraw-Hill.

O'Neil, C. (2016) Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.

Smith, D. (2016) Ethical issues in the use of social media data in research. Journal of Empirical Research on Human Research Ethics, 11(4), pp. 374-388.

Tufekci, Z. (2014) ‘Big data is watching you’. The New York Times. Available at: https://www.nytimes.com/2014/04/06/opinion/sunday/big-data-is-watching-you.html.

Voigt, P. and von dem Bussche, A. (2017) The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer.

Zimmer, M. and Proferes, N. (2014) ‘A topology of social media research: Disciplines, methods, and ethical considerations’. SAGE Open, 4(3). Available at: https://journals.sagepub.com/doi/full/10.1177/2158244014540867.

Ghasiya, P., & Sasahara, K. (2023). Messaging strategies of Ukraine and Russia on Telegram during the 2022 Russian invasion of Ukraine. First Monday. Available at: [Link](https://discovery.researcher.life/article/messagingstrategiesofukraineandrussiaontelegramduringthe2022russianinvasionofukraine/cfd8dddb0ae73201a8caf83d72b0beaa)

Joshi, M., Nishi, R., Rani, R., Bisla, N., Sharma, A., & Dev, A. (2023). Twitter and the UkraineRussia War: An Examination. IEEE INDISCON. Available at: [Link](https://ieeexplore.ieee.org/document/10270120)

Kiforchuk, K. (2023). Frequency analysis of Russian propaganda Telegram channels. Theoretical and Applied Cybersecurity. Available at: [Link](https://discovery.researcher.life/article/frequencyanalysisofrussianpropagandatelegramchannels/8d0cbab637463f059a694ec6bd488314)

Krivičić, A., & MartinčićIpšić, S. (2023). Analyzing sentiment of Reddit posts for the RussiaUkraine war. IEEE Xplore. Available at: [Link](https://ieeexplore.ieee.org/document/10159986)

Urman, A., Ho, J.C., & Katz, S. (2021). Analyzing protest mobilization on Telegram: The case of the 2019 AntiExtradition Bill movement in Hong Kong. PLOS ONE. Available at: [Link](https://pubmed.ncbi.nlm.nih.gov/34624039)

Appendix B: Code Listing for Telegram Channel Message Extraction, per-processing and SA

1.Channel users (badges, 1967)

```
import configparser
import json
import asyncio

from telethon import TelegramClient
from telethon.errors import SessionPasswordNeededError
from telethon.tl.functions.channels import GetParticipantsRequest
from telethon.tl.types import ChannelParticipantsSearch
from telethon.tl.types import (
    PeerChannel
)

 Reading Configs
config = configparser.ConfigParser()
config.read("config-sample.ini")

 Setting configuration values
api_id = config['Telegram']['api_id']
api_hash = config['Telegram']['api_hash']
```

```
api_hash = str(api_hash)

phone = config['Telegram']['phone']
username = config['Telegram']['username']

 Create the client and connect
client = TelegramClient(username, api_id, api_hash)

async def main(phone):
   await client.start()
   print("Client Created")
    Ensure you're authorized
   if await client.is_user_authorized() == False:
      await client.send_code_request(phone)
      try:
         await client.sign_in(phone, input('Enter the code: '))
      except SessionPasswordNeededError:
         await client.sign_in(password=input('Password: '))

   me = await client.get_me()

   user_input_channel = input("enter entity(telegram URL or entity id):")

   if user_input_channel.isdigit():
      entity = PeerChannel(int(user_input_channel))
   else:
      entity = user_input_channel

   my_channel = await client.get_entity(entity)

   offset = 0
   limit = 100
   all_participants = []

   while True:
      participants = await client(GetParticipantsRequest(
         my_channel, ChannelParticipantsSearch(''), offset, limit,
         hash=0
      ))
      if not participants.users:
         break
      all_participants.extend(participants.users)
      offset += len(participants.users)

   all_user_details = []
   for participant in all_participants:
```

```
        all_user_details.append(
            {"id": participant.id, "first_name": participant.first_name, "last_name":
participant.last_name,
             "user": participant.username, "phone": participant.phone, "is_bot": participant.bot})

    with open('user_data.json', 'w') as outfile:
        json.dump(all_user_details, outfile)

with client:
    client.loop.run_until_complete(main(phone))
```

2. channelmessages

```
import configparser
import json
import asyncio
import csv
from datetime import datetime, timezone

from telethon import TelegramClient
from telethon.errors import SessionPasswordNeededError
from telethon.tl.functions.messages import GetHistoryRequest
from telethon.tl.types import PeerChannel


 some functions to parse json date
class DateTimeEncoder(json.JSONEncoder):
    def default(self, o):
        if isinstance(o, datetime):
            return o.isoformat()

        if isinstance(o, bytes):
            return list(o)

        return json.JSONEncoder.default(self, o)


 Reading Configs
config = configparser.ConfigParser()
config.read("config-sample.ini")

 Setting configuration values
api_id = config['Telegram']['api_id']
api_hash = config['Telegram']['api_hash']

api_hash = str(api_hash)
```

```
phone = config['Telegram']['phone']
username = config['Telegram']['username']

 Create the client and connect
client = TelegramClient(username, api_id, api_hash)

 Date filter: messages from AUG 1, 2024 to AUG 26, 2024 (UTC)
start_date = datetime(2024, 8, 1, tzinfo=timezone.utc)
end_date = datetime(2024, 8, 26, 23, 59, 59, tzinfo=timezone.utc)


async def main(phone):
    await client.start()
    print("Client Created")
     Ensure you're authorized
    if not await client.is_user_authorized():
        await client.send_code_request(phone)
        try:
            await client.sign_in(phone, input('Enter the code: '))
        except SessionPasswordNeededError:
            await client.sign_in(password=input('Password: '))

    me = await client.get_me()

    user_input_channel = input('enter entity (telegram URL or entity id):')

    if user_input_channel.isdigit():
        entity = PeerChannel(int(user_input_channel))
    else:
        entity = user_input_channel

    my_channel = await client.get_entity(entity)

    offset_id = 0
    limit = 100
    all_messages = []
    total_messages = 0
    total_count_limit = 0

    while True:
        print("Current Offset ID is:", offset_id, "; Total Messages:", total_messages)
        history = await client(GetHistoryRequest(
            peer=my_channel,
            offset_id=offset_id,
            offset_date=None,
            add_offset=0,
```

```
            limit=limit,
            max_id=0,
            min_id=0,
            hash=0
        ))
        if not history.messages:
            break

        messages = history.messages
        for message in messages:
            message_date = message.date.astimezone(timezone.utc)   Convert message date to
UTC
            if start_date <= message_date <= end_date:
                all_messages.append(message.to_dict())

         Update offset_id to the ID of the last message to continue fetching
        offset_id = messages[-1].id
        total_messages = len(all_messages)

         Stop if we've reached the total count limit (if set)
        if total_count_limit != 0 and total_messages >= total_count_limit:
            break

         Stop if the last fetched message is older than the start_date
        if messages[-1].date < start_date:
            break

     Save messages to JSON file
    json_filename = 'channel_messages.json'
    with open(json_filename, 'w') as outfile:
        json.dump(all_messages, outfile, cls=DateTimeEncoder)

     Convert JSON to CSV
    csv_filename = 'channel_messages.csv'
    convert_json_to_csv(json_filename, csv_filename)
    print(f"Messages have been saved to {csv_filename}")


def convert_json_to_csv(json_filename, csv_filename):
    with open(json_filename, 'r') as json_file:
        messages = json.load(json_file)

    with open(csv_filename, 'w', newline='', encoding='utf-8') as csv_file:
        csv_writer = csv.writer(csv_file)

         Write headers (change based on what fields you want to include)
        headers = ['id', 'date', 'message', 'from_id', 'chat_id','views', 'forwards']
```

```
        csv_writer.writerow(headers)

         Write data rows
        for message in messages:
            row = [
                message.get('id'),
                message.get('date'),
                message.get('message'),
                message.get('from_id', {}).get('user_id') if message.get('from_id') else None,
                message.get('peer_id', {}).get('channel_id') if message.get('peer_id') else None,
                message.get('views'),
                message.get('forwards')
            ]
            csv_writer.writerow(row)


with client:
    client.loop.run_until_complete(main(phone))
```

3. config file

```
[Telegram]
 no need for quotes

 you can get telegram development credentials in telegram API Development Tools
api_id = 27477648
api_hash = cac745592ca7768efbf330eca0fee3f2

 use full phone number including + and country code
phone = +353899488510
username = Bhavi1111111
```

4. preprocessing, word cloud and sentiment analysis

intenslava.ipynb

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

import re

import string
```

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer

import pandas as pd

 Try reading the CSV file with 'latin1' encoding
try:
   df = pd.read_csv('/content/channel_messages_intenslava_Aug.csv', encoding='latin1')
except UnicodeDecodeError:
   print("Error: Could not decode file with 'latin1' encoding.")
    You can try other encodings like 'utf-8', 'utf-16', 'cp1252', etc.
except Exception as e:
   print(f"An error occurred: {e}")

 Display the first few rows of the DataFrame to verify
df.head()

df.head(5)

data = pd.read_csv('/content/channel_messages_intenslava_Aug.csv',encoding=encoding)
data.head(5)

df.drop('from_id', axis=1, inplace=True)

 check null
df.isnull().sum()
dropna = df.dropna()
dropna.isnull().sum()


import pandas as pd

 Assuming 'df' is the DataFrame you are working with

 Count the number of columns in the DataFrame
column_count = len(df.columns)

 List the column names
column_names = df.columns.tolist()

 Identify columns with missing values
missing_values = df.isnull().sum()

 List columns that have missing values along with the count of missing values
```

```
columns_with_missing = missing_values[missing_values > 0]

 Display the results
print(f"Number of columns: {column_count}")
print(f"Column names: {column_names}")
print("Columns with missing values:")
print(columns_with_missing)

data.head(5)
data.drop('from_id', axis=1, inplace=True)

df.drop('from_id', axis=1, inplace=True)

df.head(5)

 check null
df.isnull().sum()
dropna = df.dropna()
dropna.isnull().sum()

import nltk
from nltk.corpus import stopwords

 Download the stopwords from the nltk library
nltk.download('stopwords')

 Initialize the SnowballStemmer and stopwords
stemmer = nltk.SnowballStemmer("english")
stopword = set(stopwords.words('english'))

import re
import string
def clean(text):
    text = str(text).lower()
    text = re.sub('\[.?\]', '', text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w\d\w', '', text)
    text = [word for word in text.split(' ') if word not in stopword]
    text=" ".join(text)
    text = [stemmer.stem(word) for word in text.split(' ')]
    text=" ".join(text)
    return text

df["message"] = df["message"].apply(clean)
```

```
df.head(5)
data["message"] = data["message"].apply(clean)

df.head(5)

import pandas as pd
import re

 Function to remove special characters from a string
def remove_special_characters(text):
     Use regular expression to replace all non-alphanumeric characters with a space
    cleaned_text = re.sub(r'[^A-Za-z0-9\s]+', '', text)
    return cleaned_text

 Assuming 'df' is your DataFrame and 'columns_to_clean' is defined

 Example: Defining specific columns to clean
columns_to_clean = ['message', 'message']   Replace with actual column names

 Apply the cleaning function only to the specified columns
for message in columns_to_clean:
    df[message] = df[message].apply(lambda x: remove_special_characters(x) if isinstance(x,
str) else x)

 Display the first few rows of the cleaned DataFrame
df.head()
df.head(5)
!pip install vaderSentiment
import pandas as pd
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

 Initialize VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()

from wordcloud import WordCloud
import matplotlib.pyplot as plt
from nltk.corpus import stopwords

text = "".join(i for i in df.message)
stopwords = set(stopwords.words('english'))
wordclouad=WordCloud(width=800,height=500,stopwords=stopwords,max_font_size=50,m
ax_words=100,background_color="white").generate(text)
plt.figure(figsize=(16,10))
plt.imshow(wordclouad,interpolation="bilinear")
plt.axis("off")
plt.show()
```

```
import pandas as pd
from collections import Counter
from nltk.corpus import stopwords
import string
import matplotlib.pyplot as plt

 Assuming 'df' is a pandas DataFrame and has a column named 'message'
 Concatenate all messages into a single text
text = " ".join(message for message in df.message)

 Set of English stopwords
stopwords_set = set(stopwords.words('english'))

 Remove punctuation and stopwords, and convert to lower case
cleaned_text = "".join([char.lower() for char in text if char not in string.punctuation])
cleaned_words = [word for word in cleaned_text.split() if word not in stopwords_set]

 Count word frequencies
word_freq = Counter(cleaned_words)

 Convert word frequencies to a pandas DataFrame for easy visualization
word_freq_df = pd.DataFrame(word_freq.items(), columns=['Word', 'Frequency'])

 Sort the DataFrame by frequency in descending order
word_freq_df = word_freq_df.sort_values(by='Frequency', ascending=False)

 Display the word frequency DataFrame
print(word_freq_df)

 Plot the top N most common words and their frequencies
N = 20   Adjust N to display more or fewer words
top_words = word_freq_df.head(N)

plt.figure(figsize=(10, 6))
plt.bar(top_words['Word'], top_words['Frequency'], color='skyblue')
plt.xlabel('Words')
plt.ylabel('Frequency')
plt.title(f'Top {N} Most Common Words')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

sentiments=SentimentIntensityAnalyzer()
df["Positive"]=[sentiments.polarity_scores(i)["pos"] for i in df["message"]]
df["Negative"]=[sentiments.polarity_scores(i)["neg"] for i in df["message"]]
df["Neutral"]=[sentiments.polarity_scores(i)["neu"] for i in df["message"]]
```

```
 Define a function to classify sentiment based on scores
def classify_sentiment(row):
   if row['Positive'] > row['Negative'] and row['Positive'] > row['Neutral']:
      return 'Positive'
   elif row['Negative'] > row['Positive'] and row['Negative'] > row['Neutral']:
      return 'Negative'
   else:
      return 'Neutral'

 Apply the classification function
df['Sentiment'] = df.apply(classify_sentiment, axis=1)

df=df[["message","Positive","Negative","Neutral","Sentiment"]]
df.head()

 Initialize the SentimentIntensityAnalyzer
sentiments = SentimentIntensityAnalyzer()

 Compute sentiment scores for each message
df["Positive"] = [sentiments.polarity_scores(message)["pos"] for message in df["message"]]
df["Negative"] = [sentiments.polarity_scores(message)["neg"] for message in df["message"]]
df["Neutral"] = [sentiments.polarity_scores(message)["neu"] for message in df["message"]]

 Define a function to classify sentiment based on scores
def classify_sentiment(row):
   if row['Positive'] > row['Negative'] and row['Positive'] > row['Neutral']:
      return 'Positive'
   elif row['Negative'] > row['Positive'] and row['Negative'] > row['Neutral']:
      return 'Negative'
   else:
      return 'Neutral'

 Apply the classification function
df['Sentiment'] = df.apply(classify_sentiment, axis=1)

 Filter messages based on sentiment
positive_messages = df[df['Sentiment'] == 'Positive']
negative_messages = df[df['Sentiment'] == 'Negative']
neutral_messages = df[df['Sentiment'] == 'Neutral']

 Display filtered messages
print("Positive Messages:")
print(positive_messages)

print("\nNegative Messages:")
```

```
print(negative_messages)

print("\nNeutral Messages:")
print(neutral_messages)

 Reorder columns if necessary
df = df[["message", "Positive", "Negative", "Neutral", "Sentiment"]]

 Display the DataFrame with sentiment classification
print(df.head(600))

 Initialize the SentimentIntensityAnalyzer
sentiments = SentimentIntensityAnalyzer()

 Sample DataFrame (replace this with your actual data)
 Assuming df is already defined
 df = pd.DataFrame({'message': ["I love this!", "I hate this!", "It's okay.", ...]})

 Compute sentiment scores for each message
df["Positive"] = [sentiments.polarity_scores(message)["pos"] for message in df["message"]]
df["Negative"] = [sentiments.polarity_scores(message)["neg"] for message in df["message"]]
df["Neutral"] = [sentiments.polarity_scores(message)["neu"] for message in df["message"]]

 Define a function to classify sentiment based on scores
def classify_sentiment(row):
    if row['Positive'] > row['Negative'] and row['Positive'] > row['Neutral']:
        return 'Positive'
    elif row['Negative'] > row['Positive'] and row['Negative'] > row['Neutral']:
        return 'Negative'
    else:
        return 'Neutral'

 Apply the classification function
df['Sentiment'] = df.apply(classify_sentiment, axis=1)

 Filter messages based on sentiment
positive_messages = df[df['Sentiment'] == 'Positive']
negative_messages = df[df['Sentiment'] == 'Negative']
neutral_messages = df[df['Sentiment'] == 'Neutral']

 Reorder columns if necessary
df = df[["message", "Positive", "Negative", "Neutral", "Sentiment"]]

 Display the DataFrame with sentiment classification
print(df.head(600))
```

```
 Plot sentiment scores (Positive, Negative, Neutral) for the first 600 messages
df_sample = df.head(600).melt(id_vars=["message", "Sentiment"],
                        value_vars=["Positive", "Negative", "Neutral"],
                        var_name="Sentiment Type", value_name="Score")

plt.figure(figsize=(14, 8))
sns.lineplot(x=df_sample.index, y="Score", hue="Sentiment Type", data=df_sample,
palette='viridis')
plt.title('Sentiment Scores Over Messages')
plt.xlabel('Message Index')
plt.ylabel('Sentiment Score')
plt.show()

import matplotlib.pyplot as plt
import seaborn as sns

 Filter out neutral messages for plotting
df_filtered = df[df['Sentiment'].isin(['Positive', 'Negative'])]

 Plotting sentiment distribution for Positive and Negative
plt.figure(figsize=(8, 6))
sns.countplot(data=df_filtered, x='Sentiment', palette='viridis')

 Adding titles and labels
plt.title('Distribution of Positive and Negative Sentiments')
plt.xlabel('Sentiment')
plt.ylabel('Number of Messages')

 Show the plot
plt.show()




df_filtered_result = df_filtered[['message', 'Sentiment', ]]


 Print the resulting DataFrame
print(df_filtered_result)
count_positive = (df['Sentiment'] == 'Positive').sum()
count_negative = (df['Sentiment'] == 'Negative').sum()

common_messages = pd.merge(df, data, on='message')

df_filtered_result.head(100)

data.head(5)
```

```
df_filtered = common_messages[['message', 'Sentiment', 'date']]
 Saving the entire DataFrame to a CSV file at the specified path
df_filtered.to_csv('/content/intenslavaFiltedMesseges.csv', index=False)

print(df_filtered)
```