UDC 004.4
DOI https://doi.org/10.32782/2663-5941/2025.4.2/21

Ilin M.O.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

Oleshchenko L.M.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

SOFTWARE-DRIVEN DATA ANALYSIS OF PROPAGANDA NARRATIVES AND SOCIAL MEDIA ENGAGEMENT

The article introduces a software framework for analyzing war propaganda distributed through Telegram during the full-scale invasion of Ukraine. The dataset comprises 5.3 million messages, 21.5 million text entities (hashtags, mentions, URLs), and 22.8 million user reactions (e.g., emojis, likes) from 116 public channels, the research approaches the problem from a structural and behavioral angle rather than through full-scale classification. The dataset for this research was extracted from a custom PostgreSQL database designed to handle high-volume, complex Telegram data related to Russian propaganda during the first two years of the full-scale invasion of Ukraine. The primary focus has been placed on building a processing pipeline capable of handling large and structurally heterogeneous data streams, most of which have been multilingual, emotionally charged, and inconsistently formatted. Exploratory analysis methods have been combined with basic feature engineering, which has made it possible to identify recurring patterns in hashtags, message structures, and audience reactions. For example, emoji usage has been found to be frequent but unevenly distributed, indicating the use of emotional framing strategies that vary across message types and channels.

Rather than presenting complete modeling results, the work has established a reproducible and modular baseline for future experimentation. Interpretable features, such as text length, hashtag count, emoji diversity, and posting times, have been extracted and organized for straightforward downstream analysis, including narrative classification, engagement prediction, or network modeling. Co-occurrence analyses of text entities and emoji-based response profiling have revealed how emotionally framed narratives have been shaped and reinforced. A few working hypotheses have been proposed, including the idea that emotionally framed content tends to receive more concentrated and homogeneous reactions.

The research has demonstrated that with a carefully designed data pipeline and relatively modest resources, it is possible to extract meaningful structure from adversarial, large-scale communication environments. The goal has not been to solve the propaganda problem in full, but rather to clarify what groundwork has been necessary for studying Telegram-based influence operations at scale. As disinformation ecosystems continue to evolve, a flexible software toolkit and foundational dataset have been offered to support future technical and analytical research on hostile media environments.

Key words: *software methods of data analysis, messaging and social media platforms, narratives spreading, Python, narrative and engagement analysis, Big Data, processing heterogeneous streaming data.*

Formulation of the problem. The use of propaganda in wartime is not new, but the way it spreads has transformed significantly in the digital age. Messaging platforms like Telegram have become vital tools in modern information warfare. What sets Telegram apart is its hybrid nature: it functions both as a broadcasting platform and a private messaging app. This structure, with public channels that can reach hundreds of thousands of subscribers, is particularly useful for spreading consistent narratives with minimal external interference.

Since the beginning of Russia's full-scale invasion of Ukraine in 2022, Telegram has played a central role in the online propaganda ecosystem. Public channels linked to Russian-aligned messaging have pushed out content at high volumes, targeting both domestic and international audiences. These messages range from traditional patriotic appeals to more

subtle or misleading information that aims to distort facts and undermine trust.

While there has been public awareness of these campaigns, the academic and technical communities still face challenges in scaling analysis to the volume and variety of content distributed across platforms like Telegram. Much of the research has focused on more mainstream social media, leaving this relatively unregulated space underexamined [1, 2]. The multilingual, emotional, and highly contextual nature of the content complicates efforts to systematically analyze it.

This research addresses that gap. By focusing on Telegram-based propaganda channels, we present a framework capable of handling large-scale, heterogeneous datasets. Our aim is not just to describe what these channels are doing, but to build software tools and methods that allow others to analyze, monitor, and potentially intervene in digital information conflicts more effectively.

Analysis of recent research and publications. Over the past few years, researchers have intensely studied how propaganda and disinformation spread on social media. The issue gained urgency after events like Brexit and the 2016 U.S. elections, and especially during the COVID-19 "infodemic," which highlighted the scale of online misinformation. In response, numerous works from 2020–2025 have focused on detecting deceptive or influential content in online platforms. Propaganda is typically defined as strategically crafted messaging aimed at influencing public opinion for a specific agenda. It often exploits rhetorical and emotional techniques – for example, using loaded language, patriotic appeals, or false dilemmas – to sway an audience. Notably, propaganda is related to but distinct from disinformation, which is deliberately false information spread with the intent to cause harm. Propaganda can be based on truths or falsehoods and may not always be overtly malicious, whereas disinformation by definition involves falsehood and harmful intent [3].

To clarify such concepts, scholars have proposed various misinformation taxonomies. Modern studies have expanded these categorizations to dozens of specific tactics. Likewise, researchers distinguish misinformation (unintentionally incorrect claims) from disinformation (intentional falsehoods) [4].

A variety of computational techniques have been applied to identify propaganda and influence campaigns in social media. One prominent approach is stance detection – determining whether a text agrees, disagrees, or is neutral toward a given claim or narrative. Stance detection plays a key role in fact-checking and rumor verification systems. Beyond stance, scholars are increasingly interested in narrative analysis of disinformation. Rather than treating each message in isolation, narrative-focused studies examine the storylines and themes that link many pieces of content. Another crucial method in propaganda detection is sentiment and emotion analysis. Propagandists frequently use emotional appeals – outrage, fear, national pride, anger – to make their message resonate. Several studies have integrated sentiment detection into misinformation classifiers.

Researchers now leverage stance detection, narrative analysis, sentiment analysis, and taxonomies to categorize different forms of propaganda and disinformation. This rich toolkit, developed in studies of Twitter, Facebook and other platforms, sets the stage for examining newer outlets. In particular, these methods inform our analysis of Russian propaganda on Telegram [5].

Telegram is a hybrid messaging and social media platform whose unique affordances have made it a hotbed for propaganda dissemination in recent years. Telegram enables users to create public channels that can broadcast messages to unlimited subscribers, all while allowing a high degree of anonymity. This anonymity, combined with Telegram's lax moderation, provides a fertile ground for disinformation actors. Telegram's minimal content moderation stands in stark contrast to the policies of Facebook, Twitter, or YouTube. In response to bans on mainstream platforms, many Russian state-affiliated actors migrated to Telegram. Telegram's channel-based architecture leads to decentralized information flows that often form echo chambers. A recent study analyzed over 1,700 Telegram channels and identified "bridge nodes" that link disparate groups by sharing messages across ideological lines [6].

The content of these propaganda posts often revolved around key narratives: justifications for the invasion, denials of atrocities, and conspiracy theories.

Compared to Twitter or Facebook, Telegram also presents hurdles for researchers and policymakers trying to monitor propaganda. Telegram has no official public API for channel messages. Researchers must manually discover relevant Telegram channels and scrape messages, which can lead to sampling biases. Telegram's openness allowed both genuine citizen reporting and false propaganda to thrive side by side. Analysts have noted that during 2022–2023, the Kremlin's information operatives treated Telegram as a digital frontline.

Recent advances in NLP, particularly the development of large pre-trained language models, have

enhanced our ability to process and interpret text at scale. A cornerstone of modern text analysis is the BERT family of models. For Russian-language text, researchers commonly use RuBERT or multilingual models like XLM-R. In propaganda detection tasks, fine-tuned BERT models have achieved state-of-the-art results. Alongside transformers, more lightweight text representations still play a role. Word embeddings such as those produced by fastText can efficiently encode words into vectors and have been used in many multilingual text applications.

To extract themes and narratives from text corpora, topic modeling techniques such as LDA and BERTopic are widely used. BERTopic combines transformer embeddings with clustering to generate more coherent topics.

Named Entity Recognition (NER) tools can automatically tag mentions of people, organizations, and places. NER also enables the construction of networks of co-occurring entities. In multilingual settings, NER models or dictionaries need to be adapted for each language [7].

Despite the arsenal of NLP tools available, analysts face several challenges when mining large-scale text data for propaganda. Firstly, there is a scarcity of labeled training data for supervised learning. Creating high-quality annotations for what counts as propaganda vs. normal content is time-consuming. Irony and sarcasm can invert the literal meaning of text, confounding simple sentiment or stance detectors. The evolving nature of language and tactics means any static model can become outdated. Propagandists continually adapt their vocabulary to evade filters. Our methodology will address these challenges by applying tools that balance performance with adaptability [8].

Task statement. This research is structured around a set of guiding questions that emerged during the process of working with large-scale Telegram data in the context of Russian wartime propaganda. Rather than focusing on predictive classification, we concentrated on questions that could be addressed through structural analysis, feature extraction, and descriptive modeling.

The core research questions are as follows.

1. What kinds of structural and content-based patterns can be observed in Telegram propaganda messages, and how are they distributed across channels and over time? This includes recurring hashtags, variations in message length and complexity, and patterns in emoji usage as a proxy for emotional tone and engagement.

2. How can a scalable, relational data model support the systematic analysis of such content across millions of records? Here, the aim is to define a reproducible schema that allows for both flexibility and performance when working with Telegram data in relational form.

3. Which features – textual, temporal, and behavioral can be extracted efficiently and interpreted meaningfully to support future analytical tasks such as narrative detection or engagement modeling?

To address these questions, we developed a modular pipeline for ingesting, exploring, and engineering features from Telegram propaganda content. The goal was to produce a reusable analytical framework that can serve both as a descriptive lens and as a foundation for future modeling efforts.

Outline of the main material of the research. The dataset used in this research was extracted from a custom-built PostgreSQL database, designed to accommodate the scale and complexity of data collected from Telegram channels disseminating Russian propaganda during the first two years of the full-scale invasion of Ukraine. The dataset comprises information from 116 public Telegram channels, identified as part of a coordinated pro-Kremlin media ecosystem. These channels were selected based on prior research, open-source intelligence, and manual vetting of content history.

In total, the database contains:

- 5,309,851 messages;
- 21,511,942 text entities (e.g., hashtags, mentions, URLs);
- 22,794,701 reactions (emojis, likes, other engagement markers).

The database is structured across four interlinked tables:

- *channels*: contains metadata about each Telegram channel, including its Telegram ID, username, and classification by type (e.g., news, commentary, satire);
- *messages*: stores individual posts along with their metadata–timestamps, author identifiers, message types, edited versions, media attachments (photo, video), and raw text content;
- *text_entities*: extracted elements from each message, such as hashtags, hyperlinks, mentions, and formatted text (bold, italic, etc.), stored with type labels and positional order;
- *reactions*: records emoji-based reactions tied to specific messages, including the emoji itself and the count of users who reacted with it.

This schema allows us to analyze messages not only as standalone units but also in terms of how

they are embedded within broader discursive and engagement contexts. Messages are linked to both their originating channels and associated interactions (entities, reactions), forming a relational structure suitable for large-scale querying, statistical analysis, and modeling.

The temporal attributes stored in both human-readable (timestamp) and Unix (unixtime) formats support granular chronological analysis, enabling us to capture patterns in propaganda bursts, narrative pivots, and audience response cycles over time. The data used in this research was collected through a combination of automated and supervised methods. Public Telegram channels were identified based on prior open-source investigations, civil society monitoring projects, and manual verification of content consistency with known patterns of Russian-aligned propaganda. The collection process relied on Telegram's open API endpoints and automated scraping tools, ensuring that all acquired data remained within the bounds of publicly accessible content. No private groups or encrypted messages were accessed. Only public channels were included, and only data visible without user authentication was collected. This guarantees that the data acquisition process adhered to Telegram's terms of service and respected the platform's intended public-private content boundaries.

The resulting database was designed to support both relational analysis and scalable processing. Data was stored in normalized tables, preserving structural integrity and making it possible to trace any record back to its channel, timestamp, or message ID. Fields such as user identifiers were retained only when they referred to public or institutional accounts (e.g., media outlets or known figures); any personal user data was either excluded or anonymized. Given the sensitive nature of the material – propaganda produced during an active war – the research followed ethical guidelines for handling war-related and potentially harmful content. The research does not reproduce or promote any of the propagandist material. Instead, all analysis is aimed at deconstructing techniques and patterns of influence for academic and counter-disinformation purposes. All results are presented in aggregate, and no attempts are made to deanonymize individuals or trace personal interactions. Furthermore, wherever relevant, disclaimers are applied to visualizations and quoted examples to contextualize their nature as subjects of critical analysis rather than information endorsement.

To process the Telegram data at scale, we developed a modular analytical pipeline consisting of four main stages: data ingestion and profiling, exploratory statistics, systematic cleansing, and feature engineering. The design emphasizes reproducibility and computational efficiency, recognizing the challenges of working with large volumes of data collected from propaganda-linked Telegram channels.

The dataset is stored in a PostgreSQL database and structured across four normalized tables: channels, messages, text_entities, and reactions. Initial profiling queries confirmed structural integrity, absence of missing values, and alignment between foreign key references (e.g., message_id links across tables). All core fields (text_content, channel_id, message_date) were present across records, enabling immediate downstream analysis.

Sampling and summary queries revealed early insights into message types and engagement patterns. A random sample of messages showed content ranging from daily greetings and calls to action to battlefield reports and emotional appeals.

Many posts featured strong emotive or rhetorical devices, often paired with emoji reactions – most frequently:

- 👍 (over 4 billion reactions);
- 😁, 🙏, ♡ and 🔥 – each with several hundred million reactions.

These reactions offer a behavioral proxy for audience engagement and emotion-triggering content. Textual entities, particularly hashtags, revealed patterns of narrative segmentation and branding. The most frequent hashtags included campaign-specific tags alongside broader geopolitical markers. This structure enables both time-series tracking and topical clustering in later analysis.

Despite the dataset's scale, quality checks showed no missing values in critical fields. Messages were filtered for completeness and deduplicated where necessary. Repeated content, truncated posts, or bot-generated noise were flagged using length and lexical diversity heuristics.

To support downstream modeling, features were engineered at multiple levels:

- *textual*: message length, punctuation density, stopword ratios, named entity counts, and TF-IDF vectors;
- *behavioral*: reaction counts by emoji, average engagement per channel, emoji entropy;
- *semantic*: hashtags, sentiment polarity (to be derived), rhetorical signal tokens.

The architecture supports stepwise iteration and selective caching of derived variables to ensure scalability across millions of records.

Fig. 1 presents proposed multi-stage pipeline for processing raw Telegram data (channels, messages,

text entities, and reactions) to enable downstream analytical tasks. The process begins with Data Ingestion & Profiling, where the structure is validated, samples are summarized, and foreign key integrity is checked. This is followed by Exploratory Analysis, which focuses on frequency statistics, emoji patterns, and hashtag surfacing.

Data Cleansing is performed to remove duplicates, normalize formats, and identify noisy or incomplete data. These cleaned and profiled data then feed into Feature Engineering, where NLP features, reaction metrics, and temporal features are extracted. The final outputs are used for various Downstream Tasks.

The initial exploratory phase aimed to surface broad patterns across message volume, content structure, and user reactions. This step provided the empirical grounding for refining data cleansing procedures and for selecting relevant features for later modeling. The initial exploratory phase utilized Python tools such as Pandas for data manipulation, Matplotlib and Seaborn for visual analysis, and NumPy for efficient numerical operations. These libraries enabled the identification of broad patterns in message volume, content structure, and user reactions, providing an empirical foundation for refining data cleansing procedures and guiding feature selection for subsequent modeling.

Across the 5.3 million messages, activity remained sustained throughout the two-year window, with noticeable peaks aligning with battlefield events or significant political developments. Due to hardware constraints, full time-series plotting was deferred; timestamp sampling showed regular posting, with specific spikes around symbolic dates and coordinated media pushes.

Sampled messages revealed considerable variability in content structure–from brief emoji-laced posts to lengthy reports or transcripts. Posts often began with rhetorical intensifiers (“!!”, “ ⚡”) and included hashtags or campaign-specific branding. Some posts lacked any meaningful text, while others spanned several paragraphs.

User reactions, recorded as over 22 million individual counts, show a clear tendency toward emotionally charged content. The top emojis – 👍, 😁, 🙏, ♡, and 🔥 – are indicative of binary approval (👍), outrage, or emotional resonance (🙏, ♡).

Particularly noteworthy is the dominance of 👍, which accounted for over 4 billion cumulative reactions, suggesting either:

- the use of 👍 as a default button on many channels,
- or bot-driven amplification, a known tactic in propaganda ecosystems.

Analysis of 21.5 million text entities shows extensive use of hashtags as narrative anchors. To under-

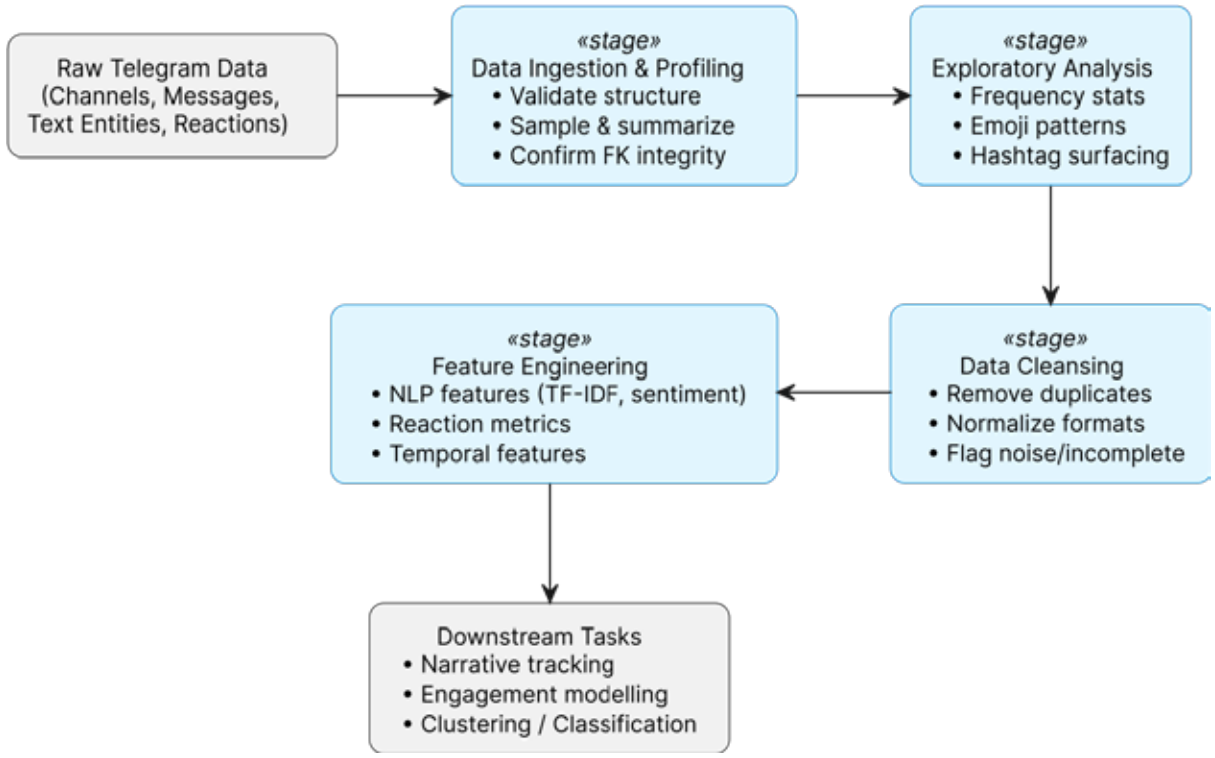

Fig. 1. Proposed overall methodological pipeline

stand how propaganda dissemination varies across actors, we conducted a comparative analysis of activity levels and engagement metrics for individual Telegram channels. Using aggregated data from the relational schema, we computed per-channel counts for total messages, word volumes, and emoji-based reactions. This revealed distinct behavioral signatures and operational scales among the channels. Emoji reactions offer insight into audience resonance. While absolute counts correlate with message volume, certain channels consistently outperform others in reaction density.

Textual entities – particularly hashtags – play a central role in shaping and signaling propaganda narratives. They function not only as indexing devices but also as rhetorical cues that frame a message's tone, urgency, or political alignment. In our dataset, over 21.5 million such entities were extracted, with hashtags representing a major subset.

To visualize broader relationships, we constructed a network graph where nodes represent hashtags and edges represent co-appearance frequency. Thicker edges denote stronger co-occurrence ties.

This analysis demonstrates that hashtags are not randomly or independently deployed. Rather, they form deliberate semantic bundles that support narrative framing, audience targeting, and automated repetition. To enable downstream analytical tasks such as clustering, classification, or narrative tracking, we engineered a set of interpretable features from the raw Telegram messages. Given the high volume of data and the heterogeneity of message formats, we prioritized lightweight, expressive features that capture structural, semantic, and behavioral characteristics of each message.

Building on this infrastructure, we outline several hypothesis for our research:

H1: messages containing emotionally charged hashtags receive significantly higher reaction counts than neutral ones;

H2: channels using more repetitive, short-format messaging (high frequency, low word count) achieve higher cumulative engagement per unit time;

H3: reaction entropy is inversely correlated with message polarity – highly polarized messages attract fewer distinct emoji types.

In follow-up work, we intend to:

- train supervised classifiers to predict reaction volume or narrative type based on engineered features and embeddings;
- integrate transformer-based sentence representations (e.g., RuBERT, XLM-R) for improved text modeling;
- apply graph-based methods to track cross-channel influence and narrative flow;
- expand the dataset to include multilingual propaganda networks, especially channels targeting foreign audiences in English or German.

To support downstream modeling tasks and exploratory narrative analysis, features were engineered at multiple levels – textual, behavioral, temporal, and semantic (Table 1). These features prioritize interpretability and computational efficiency across millions of records.

Table 1

Engineered features

Feature Name	Description	Type
text_length	Character count of the message text	Numeric
word_count	Number of words in the message	Numeric
punctuation_density	Ratio of punctuation marks to total characters	Numeric
stopword_ratio	Fraction of stopwords among all tokens (language-specific)	Numeric
named_entity_count	Count of named entities per message (if processed)	Numeric
tfidf_vector	Sparse vector representation of text content	Vector
hashtag_count	Number of hashtags in message	Numeric
url_count	Number of URLs present	Numeric
reaction_total	Sum of all reactions to the message	Numeric
reaction_entropy	Diversity of emoji reactions (Shannon entropy)	Numeric
emoji_count	Count of emojis in message body	Numeric
avg_channel_engagement	Channel-level average reactions per post	Numeric
sentiment_polarity	To be derived via sentiment analysis model	Numeric
rhetorical_tokens	Binary indicators of known propaganda keywords or formats	Categorical
hour_of_day	Posting hour (0–23)	Categorical
day_of_week	Weekday of message (0=Monday, …, 6=Sunday)	Categorical
channel_id	Foreign key linking message to its source channel	Categorical

Conclusions. This research presented a structured approach to analyzing large-scale propaganda content disseminated through Telegram during the first two years of Russia's war against Ukraine. Rather than attempting to solve classification or detection tasks directly, we focused on building a reproducible and efficient pipeline for processing, exploring, and organizing Telegram data at scale.

By designing a relational schema and extracting interpretable features from over five million messages, we were able to highlight patterns in messaging structure, hashtag use, and audience reaction. Co-occurrence analyses of text entities and emoji-based response profiling provided early insights into how narratives are framed, repeated, and emotionally charged across different channels. The exploratory phase used Python libraries like Pandas, NumPy, Matplotlib, and Seaborn to identify key patterns in message volume, structure, and reactions. These insights informed data cleaning and feature selection for later analysis.

Our primary contribution lies in creating a foundation for future work – both technical and analytical. The cleaned, structured dataset and modular feature engineering workflow are intended to support downstream tasks such as narrative classification, engagement prediction, or social graph modeling. The hypotheses outlined in this research may serve as starting points for researchers interested in tracing influence strategies or testing propaganda effectiveness over time.

The research demonstrates that with relatively modest resources, it is possible to extract meaningful structure from complex, adversarial communication environments – provided the data pipeline is thoughtfully designed. As disinformation ecosystems continue to evolve, we hope this work offers a flexible and transparent basis for analyzing them with greater clarity and scale.

Bibliography:

1. Starbird K. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media.* 2017. Vol. 11(1). P. 230–239. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14878 (date of access: 20.04.2025).

2. Zannettou S., Caulfield T., Blackburn J., Stringhini G. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the Web. *Companion Proceedings of the Web Conference 2020*. P. 218–226. DOI: https://doi.org/10.1145/3308560.3316495.

3. Cinelli M., Quattrociocchi W., Galeazzi A., et al. The COVID-19 social media infodemic. *Scientific Reports*. 2020. Vol. 10. Article 16598. DOI: https://doi.org/10.1038/s41598-020-73510-5.

4. Hardalov M., Nakov P., Koychev I. A survey on stance detection for mis- and disinformation identification. *Information Processing & Management*. 2022. Vol. 59(3). Article 102899. DOI: 10.18653/v1/2022.findings-naacl.94.

5. Tschirky M., Makhortykh M. #Azovsteel: comparing qualitative and quantitative approaches for studying framing of the siege of Mariupol on Twitter. *Media, War & Conflict*. 2023. Vol. 17 (2). P. 163–178. DOI: 10.1177/17506352231184163.

6. Li Q., Liu Q., Liu S., Di X., Chen S., Zhang H. Influence of social bots in information warfare: a case study on @UAWeapons Twitter account in the context of the Russia–Ukraine conflict. *Communication and the Public.* 2023. Vol. 8(2). P. 54–80. DOI: 10.1177/20570473231166157.

7. Lin Z., Xie J., Li Q. Multi-modal news event detection with external knowledge. *Information Processing & Management*. 2024. Vol. 61(3), Article 103697. DOI: 10.1016/j.ipm.2024.103697.

8. Javed D., Jhanjhi N.Z., Khan N.A., Ray S.K., Al Mazroa A.A., Ashfaq F., Das S.R. Towards the future of bot detection: a comprehensive taxonomical review and challenges on Twitter/X. *Computer Networks*. 2024. Vol. 254. Article 110808. DOI: 10.1016/j.comnet.2024.110808.

Ільїн М.О., Олещенко Л.М. ПРОГРАМНО-КЕРОВАНИЙ АНАЛІЗ ДАНИХ НАРАТИВІВ ПРОПАГАНДИ ТА ЗАЛУЧЕННЯ В СОЦІАЛЬНИХ МЕРЕЖАХ

У статті представлене програмне рішення для аналізу російської пропаганди, що поширюється через Telegram під час повномасштабного вторгнення в Україну. Набір даних для дослідження містить 5,3 мільйони повідомлень, 21,5 мільйони текстових сутностей (хештеги, згадки, URL) та 22,8 мільйони реакцій користувачів (емодзі, вподобання тощо) зі 116 публічних каналів. Дослідження здійснено з урахуванням структурного та поведінкового аналізу, а не повної класифікації. Було використано дані з бази PostgreSQL, спеціально розробленої для обробки великих обсягів складної інформації з Telegram, пов'язаної з російською пропагандою протягом перших двох років повномасштабного вторгнення. Основну увагу приділено побудові обробного конвеєра, здатного ефективно працювати з великим і неоднорідним потоком даних, більшість яких є багатомовними, емоційно забарвленими та нестан-

дартно структурованими. Було поєднано методи розвідкового аналізу з базовою інженерією ознак, що дозволило виявити повторювані шаблони в хештегах, структурі повідомлень і реакціях аудиторії. Зокрема, було виявлено, що використання емодзі є частим, але розподіляється нерівномірно, що вказує на застосування стратегій емоційного фреймінгу, які варіюються залежно від типу повідомлень і каналів. Замість представлення повних результатів моделювання у роботі було закладено відтворювану та модульну основу для подальших експериментів. Було виокремлено інтерпретовані ознаки, такі як довжина тексту, кількість хештегів, різноманітність емодзі та час публікацій і систематизовано їх для подальшого аналізу, зокрема, для класифікації наративів, прогнозування залучення чи моделювання мереж. Був проведений аналіз співзустрічності текстових сутностей та профілювання реакцій на основі емодзі, що дозволило виявити, як емоційно забарвлені наративи формуються та підкріплюються. Було сформульовано кілька робочих гіпотез, зокрема, припущення, що емоційно оформлений контент, як правило, отримує більш концентровані та однорідні реакції.

Було продемонстровано, що за умови використання ретельно спроєктованого конвеєра обробки даних і відносно помірних ресурсів можливо виокремити значущі структурні закономірності з великих обсягів ворожого комунікаційного контенту. Метою дослідження було формування основ, необхідних для масштабного дослідження впливових операцій у Telegram. В умовах постійної еволюції дезінформаційних екосистем було запропоновано гнучкий програмний інструментарій і базовий набір даних, покликаний підтримати майбутні технічні та аналітичні дослідження у сфері ворожих інформаційних середовищ.

Ключові слова: *програмні методи аналізу даних, платформи обміну повідомленнями, соціальні мережі, поширення наративів, Python, аналіз наративів і залучення, великі дані, обробка неоднорідних потокових даних.*

Дата надходження статті: 07.07.2025
Дата прийняття статті: 18.07.2025
Опубліковано: 27.10.2025