UNIVERSITÀ POLITECNICA DELLE MARCHE

Facoltà di Economia "Giorgio Fuà"

Corso di Laurea in Digital Economics and Business

TELEGRAM AS A BATTLEFIELD

COMPUTATIONAL ANALYSIS OF RUSSIAN WAR DISCOURSE ACROSS 18 OPERATIONAL EVENTS (2022-2026)

RELATORE: **Prof. GIACOMO DI TOLLO**

CANDIDATO: **MIHRETAB WORKU MORKA**

Anno Accademico 2025-2026

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

This thesis presents the design, implementation, and application of a computational data pipeline for analyzing Russian-language war discourse on Telegram during the Russia-Ukraine conflict (February 2022 - April 2026). Telegram, a messaging platform with minimal content moderation, has emerged as the primary medium for Russian wartime narrative production, yet systematic research linking specific military operations to measurable shifts in discourse remains limited.

The pipeline (v2.5) automates the extraction, pre-processing, thematic scoring, and event-proximity labeling of messages from 54 Russian-language Telegram channels organized into seven actor categories: milbloggers, state media, nationalists, independent media, analysts, official government channels, and Ukrainian sources. Each message is scored against seven predefined narrative themes Conflict Terrain, Tactical Weaponry, Official Narrative, Mobilization, Victory, Dehumanization, Casualties, and classified relative to 18 major operational events using a temporal window methodology adapted from financial event-study analysis.

The resulting dataset comprises 595,507 messages from 48 active channels. Post-event analysis of the 102,715 messages falling in the [+0, +14]-window reveals three principal findings: (1) discourse volume is event-sensitive but actor-category rankings remain stable, with state media consistently dominating output; (2) dominant narrative themes undergo temporal succession, shifting from territorial framing (2022) to weaponry-focused discourse (2023-2025) to patriotic celebration (2025 peace summit); and (3) actor categories maintain distinct thematic identities despite shared exposure to the same events. The dataset and pipeline constitute a structured resource for advanced NLP analyses.

INTRODUCTION

The Russia-Ukraine conflict began in February 2014 with the covert invasion of the Ukrainian autonomous republic of Crimea by disguised Russian troops. The conflict expanded in April 2014 when Russian and local proxy forces seized territory in the Donbas region; over the next seven years, more than 14,000 people were killed in fighting in eastern Ukraine. On 24 February 2022, Russia launched a full-scale invasion. Although Russian forces made significant gains in the first days of combat, Ukrainian defenders rebuffed attempts to seize Kyiv and other major cities and soon launched counterattacks.

Alongside the military conflict on the ground, a parallel confrontation has unfolded in the digital information space. Telegram, a messaging platform with over 900 million active users and minimal content moderation, has emerged as the primary medium through which Russian war discourse is produced, amplified, and consumed Urman and Katz (2022). Military bloggers, state media outlets, nationalist commentators, and opposition voices all operate public channels that function as one-to-many broadcast media, generating hundreds of thousands of messages throughout the conflict. However, there remains a lack of research that systematically links specific military operations to measurable shifts in Telegram discourse across multiple actor categories and thematic dimensions.

The objective of this thesis is to analyze how Russian wartime discourse on Telegram responds to major operational events on the battlefield, based on a computational analysis of 595,507 messages collected from 48 public channels across seven actor categories, covering 18 operational events between February 2022 and April 2026. The data pipeline scores each message across seven thematic dimensions and classifies it relative to the nearest military event using a temporal window methodology adapted from financial event-study analysis

(MacKinlay, 1997).

The analysis is guided by three research questions:

RQ1. How does the volume of Telegram discourse vary across major operational events, and do different actor categories respond with consistent relative intensity?

RQ2. How do dominant narrative themes evolve across the conflict timeline, and which events trigger thematic shifts?

RQ3. Do distinct actor categories (state media, milbloggers, nationalists, analysts) maintain identifiable thematic profiles, or does event-driven discourse converge across categories?

This thesis is organized into four chapters. The first provides background on Russian information warfare doctrine, Telegram as a research platform, and the analytical framework. The second chapter describes the methodology, including the pipeline architecture, channel taxonomy, text processing, and event classification system. The third presents the results of the post-event discourse analysis and summarizes the key findings. The fourth outlines the conclusions, and mentions extendable further work.

Chapter 1

BACKGROUND

1.1 INFORMATION WARFARE

The concept of information warfare predates the digital age, but it has acquired new dimensions in the era of networked communication. Libicki (1995) offered one of the earliest systematic definitions, describing information warfare as the deliberate use of information and information systems as instruments of conflict, targeting the adversary's decision-making processes, situational awareness, and public perception rather than, or in addition to, physical military assets. This framing established a distinction between kinetic and non-kinetic dimensions of conflict that has become central to contemporary military doctrine.

Clausewitz's foundational dictum that war is "the continuation of politics by other means" von Clausewitz (1832) acquires a particular resonance in the information environment: narrative is not merely a by-product of conflict but an active instrument of it. Belligerents do not simply fight on the battlefield; they simultaneously fight over the meaning of battlefield events, seeking to shape how their own populations, their adversaries, and the international community interpret what is happening and why.

The modern manifestation of this logic has been extensively documented by Rid (2020), whose historical analysis of Soviet and Russian "active measures" traces a century-long tradition of state-sponsored disinformation designed to undermine adversaries from within. Rid argues that influence operations are not an improvisation of the digital era but a codified doctrine, refined over decades, and now deployed at unprecedented scale through social media platforms.

The Russia-Ukraine conflict that escalated into full-scale invasion on 24 February 2022 represents a defining case study in this tradition. The war has been characterized not only by conventional military confrontations, but by a parallel and equally intense contest over narrative control Kofman and Lee (2022). Russian state actors, military bloggers, nationalist commentators, and opposition voices have all competed to define the conflict's terms its causes, its progress, its legitimacy, and its meaning across digital communication channels that reach millions of subscribers in real time.

What distinguishes this conflict from earlier information warfare campaigns is the structural role of Telegram. Unlike broadcast media or Western social media platforms subject to content moderation policies, Telegram operates as a near-unregulated broadcast and messaging infrastructure. Channels can reach audiences of hundreds of thousands without algorithmic gate-keeping, creating conditions in which narrative production and dissemination occur at a speed and scale that would have been technically impossible in earlier conflicts Urman and Katz (2022). This structural feature makes Telegram not merely a vehicle for information but a primary theatre of the information war itself.

1.2 THE RESEARCH PROJECT

This thesis takes as its starting point a preliminary data collection script for Telegram scraping, documented in Godfroid (2025). Using that script as an initial reference, this thesis independently designed and implemented a comprehensive computational pipeline for analyzing Russian-language war discourse on Telegram.

The project envisions a seven-step analytical pipeline, moving from raw data collection to advanced computational analysis:

1. **Data Extraction:** Automated scraping of Russian-language Telegram channels via the Telegram MTProto API.

2. **Pre-processing:** Text cleaning, Russian-language morphological lemmatization, and deduplication.

3. **Scoring:** Keyword-based thematic scoring across predefined narrative categories and engagement metric computation.

4. **Structured Dataset:** Export of a research-ready, annotated dataset suitable for downstream analysis.

5. **Topic Modeling:** Semantic topic discovery using BERTopic Grootendorst (2022).

6. **Network Analysis:** Mapping of information flow and actor influence networks Barabási (2016).

7. **Temporal Analysis:** Tracking of narrative evolution across the conflict timeline.

The present thesis covers Steps 1–4 of this pipeline: the design, implementation, and operation of the data collection and scoring infrastructure, together with a structured post-event analysis of the resulting dataset. Steps 5–7 (topic modeling, network analysis, temporal tracking) represent possible directions for future research beyond the scope of this thesis.

1.2.1 Telegram as a Research Platform

Telegram has emerged as a primary object of study in computational social science research on online political communication Hashemi and Peeters (2023). Founded in 2013 by Pavel Durov, the platform combines a public broadcast model (channels can address millions of subscribers uni-directionally) with a decentralized infrastructure that makes content removal difficult and enforcement of platform policies inconsistent. As of 2024, Telegram reports more than 900 million monthly active users worldwide, with particularly high penetration in the post-Soviet space. This combination has made Telegram the preferred communication channel for political actors, journalists, military units, and citizen reporters across numerous conflict zones.

In the Russian-language information ecosystem specifically, Telegram channels have become the dominant vehicle for war-related discourses. Urman and Katz Urman and Katz (2022) demonstrate that Telegram's structural design (the absence of algorithmic filtering), the ease of anonymous channel creation, and the viral forwarding mechanism - systematically advantages fringe and extremist voices over the moderated discourse that characterizes platforms such as Twitter or Facebook. In the context of the Russia-Ukraine war, these structural features create a unique research environment in which all major categories of narrative actor, state media, military bloggers, nationalist commentators, independent journalists, and Ukrainian counter-voices operate simultaneously and can be systematically monitored through a single data collection interface.

Several features make Telegram particularly suited for computational discourse analysis. First, Telegram channels are *public by default*: any user can access the full message history of a public channel without authentication, and the official MTProto API Lonami (2023) exposes complete, timestamped records over extended periods without sampling bias. Second, the platform provides *transparent engagement metrics*: each message displays its view count, forward count, and reaction count publicly, enabling quantitative measurement of audience reach without relying on proprietary analytics. Third, the *forwarding mechanism* creates an observable information flow: when a channel forwards a message from another channel, the source is preserved, allowing researchers to trace how content propagates across the ecosystem.

From a methodological standpoint, Telegram also presents challenges. The primary challenge lies in the multilingual and highly informal nature of the content, which demands robust text preprocessing, particularly for Russian, a morphologically rich language in which the same lexical root can appear in dozens of inflected forms Korobov (2015). Additionally, channels that are deleted, banned, or made private become permanently inaccessible, introducing a survivorship bias into any retrospective data collection. From a research ethics perspective, all data collected in this thesis originates from public Telegram channels whose content is visible to any account holder. Access was obtained through the official MTProto API using a personal Telegram account; no private messages, closed groups, or restricted channels were accessed.

1.2.2 Scope of This Thesis

The specific contribution of this thesis is the design and implementation of a full-scale data pipeline comprising four major components: expanded channel coverage, event-proximity labeling relative to 18 major operational events, a structured post-event analysis module, and secure credential management via environment variables.

The resulting dataset comprises **595,507 messages** from 48 active channels, spanning the period from 24 February 2022 to 1 April 2026, with 40+ analytical features per message including engagement metrics, thematic scores, and event-proximity labels.

1.3 ANALYTICAL FRAMEWORK

1.3.1 Prior Computational Studies of Telegram War Discourse

A growing body of research has applied computational methods to Telegram discourse during the Russia-Ukraine conflict. Makhortykh et al. Makhortykh et al. (2025) analyzed policymaker communication on Telegram before and after the 2022 invasion, finding that Ukrainian political elites initially focused on war topics before diversifying, while Russian policymakers largely avoided direct war framing in favor of "Western crisis" narratives. Their study demonstrates the value of temporal before/after comparison but is limited to elite-level channels and a binary country classification.

Gerard et al. Gerard et al. (2024), published at the ICWSM conference, employed LLM-based dynamic clustering (Llama 2 and DeBERTa) to track narrative evolution in pro-Russian and pro-Ukrainian Telegram communities over a 3-month period. Their approach reveals how specific events triggered immediate thematic shifts, but the short temporal window and reliance on non-reproducible LLM outputs limit generalizability.

Chernenko Chernenko (2025) examined dehumanizing language on Russian and Ukrainian Telegram channels using a mixed-methods approach, finding that Russian dehumanization is state-affiliated and top-down while Ukrainian dehumanization is grassroots and reactive. This study is closely related to one of the seven narrative themes in the present thesis (Dehumanization) but covers only 2022 and a narrow thematic scope.

The present thesis differs from these studies in three respects: (1) the dataset is substantially larger (595,507 messages vs. thousands), spanning 4 years rather than months; (2) the actor taxonomy is more granular, with 7 categories rather than binary classifications; and (3) the event-study methodology provides a systematic framework for linking discourse shifts to specific operational events across the full conflict timeline.

1.3.2 Keyword-Based Scoring

Keyword-based text classification is among the oldest and most interpretable approaches to computational content analysis. The underlying logic is straightforward: a researcher defines a vocabulary of terms associated with a thematic category, and each document is scored according to the frequency with which it contains terms from that vocabulary Lasswell and Blumenstock (1939). Despite its simplicity, this approach retains significant utility in contexts where interoperability, reproducibility, and computational efficiency are prioritized over the nuanced contextual understanding offered by machine learning models.

In the domain of political and conflict communication research, keyword scoring has been used extensively to measure the prevalence of frames, ideological positions, and thematic emphases across large corpora Boydstun et al. (2014). Its principal limitation is that the same word may carry different meanings in different contexts, which is partially mitigated through lemmatization (reducing inflected forms to their base form before matching) and through the design of thematically coherent, non-overlapping keyword sets.

For this thesis, keyword scoring is applied across 7 narrative themes drawn from the literature on Russian information warfare and conflict communication. Each message is scored as the proportion of matched keywords relative to total word count, yielding a normalized relevance measure that is comparable across messages of different lengths. The full theme definitions and keyword lists are described in Chapter 2 (Methodology).

1.3.3 Event-Study Methodology

The event-study methodology was developed in financial econometrics as a framework for measuring the impact of discrete, well-defined events, such as corporate earnings announcements or mergers on asset prices MacKinlay (1997). The core insight is that by comparing behavior in a pre-defined window around an event date to a baseline period, researchers can isolate the effect of the event from background noise and confounding trends.

MacKinlay MacKinlay (1997) formalized this approach by decomposing observed returns into an expected ("normal") component and an unexpected ("abnormal") component, with the abnormal component attributed to the event. The methodology has since been adopted across social science disciplines. In political science, it has been applied to measure the impact of elections, scandals, and policy announcements on public opinion and media coverage Baumgartner et al. (2009). In communication research, it has been used to examine how external events drive shifts in media framing and topic salience Jones et al. (2009).

The present thesis adapts this framework to the analysis of Telegram war discourse. Battlefield events (the 18 operational milestones listed in Appendix A) serve as the "event dates," and the discourse intensity and thematic composition of Telegram messages in the post-event window are treated as the dependent variable. Rather than measuring abnormal financial returns, the analysis measures abnormal narrative patterns such as spikes in specific themes, shifts in actor activity, and divergences between actor categories, in the period immediately following each event. Pre-event and event-window labels are stored in the dataset for future comparative analysis.

1.3.4 Topic Modeling and Network Analysis

While keyword scoring and event-study analysis constitute the primary analytical methods of this thesis, two additional computational frameworks could extend the analysis in future work: topic modeling and network analysis.

Topic modeling is an unsupervised machine learning technique for discovering latent thematic structures within large text corpora. The foundational approach, Latent Dirichlet

Allocation (LDA), models each document as a mixture of topics and each topic as a probability distribution over words Blei et al. (2003). More recent transformer-based approaches, particularly BERTopic Grootendorst (2022), improve on LDA by using semantic embeddings rather than raw word frequencies, enabling the discovery of topics that are coherent in meaning rather than merely in vocabulary. BERTopic could be applied to the dataset produced by this thesis to discover the dominant discourse themes in the corpus and track their evolution over the conflict timeline.

Network analysis provides a complementary framework for understanding how information flows between actors rather than what the content of that information is Barabási (2016). By modeling Telegram channels as nodes and forwarding relationships as directed edges, network methods can identify which channels function as originators (sources of novel content), which function as amplifiers (redistributors of existing content), and which communities of channels tend to coordinate their messaging. This structural dimension of the information ecosystem is invisible to content-focused methods and represents a natural extension of the present work.

Together, keyword scoring, event-study analysis, topic modeling, and network analysis form a layered analytical architecture that progresses from surface-level frequency measures to deep structural and semantic characterizations of the war discourse ecosystem.

Chapter 2

METHODOLOGY

This chapter describes the design and implementation of the Telegram war discourse pipeline (v2.5), the primary technical deliverable of this thesis. The pipeline automates the extraction, processing, scoring, and export of messages from Russian-language Telegram channels, producing a structured dataset suitable for both descriptive analysis and downstream NLP processing. The pipeline is organized into six logical blocks (A-F) covering setup, data definitions, processing functions, I/O utilities, the scraping engine, and orchestration/export.[1]

2.1 PIPELINE ARCHITECTURE

The pipeline is implemented as a single Python script (`telegram_pipeline_annotated.py`) structured into six logical blocks (A-F). At a high level, it performs four sequential analytical stages:

1. **Extraction:** Connects to the Telegram API via the Telethon library Lonami (2023) to scrape the complete message history of each target channel.
2. **Pre-processing:** Cleans the raw text (removes URLs, mentions, hashtags, punctuation, and emoji), performs Russian-language lemmatization using pymorphy2 Korobov (2015), and de-duplicates messages.

[1] The full source code of the pipeline (`telegram_pipeline_annotated.py`) and the analysis module (`analysis.py`) are available upon request.

3. **Scoring:** Applies keyword-based thematic scoring across 7 narrative themes, computes engagement metrics (views, forwards, reactions), and assigns event-proximity labels.

4. **Export:** Produces a multi-sheet Excel workbook, a UTF-8-BOM CSV file, and a JSON manifest summarizing the run metadata.

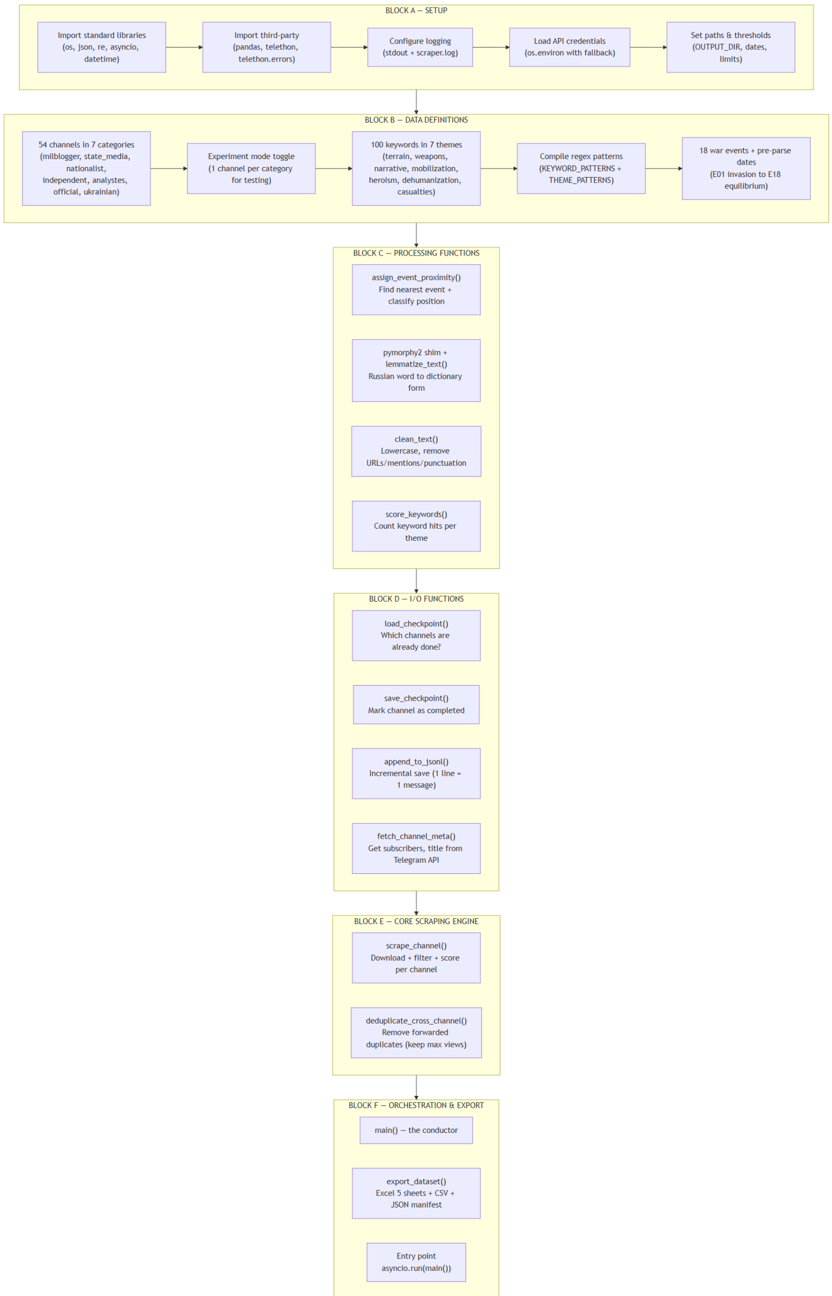

Figure 2.1: Pipeline architecture: six logical blocks (A–F) showing the flow from setup through data definitions, processing functions, I/O, scraping engine, and orchestration/export.

Starting from a preliminary scraping script Godfroid (2025), this thesis designed and implemented a full pipeline (v2.5) with the following capabilities:

Table 2.1: Pipeline version comparison: enhancements introduced in v2.5.

Feature	**Initial Script**	**This Thesis (v2.5)**
Channel count	20	54
Categories	4	7
Event-proximity labeling	No	Yes (18 events)
Credential management	Hard-coded	Environment variables
Message limit per channel	10,000	Uncapped
Theme language	French	English
Post-event analysis module	No	Yes (`analysis.py`)

2.2 DATA COLLECTION

Messages are collected via the Telethon library Lonami (2023), which provides a Python interface to the Telegram MTProto API. For each target channel, the pipeline retrieves the full message history without a cap, recording the message text, timestamp, view count, forward count, reaction counts, and the forwarding source where applicable.

The collection process runs sequentially across all channels with rate-limiting controls to respect Telegram's API flood-wait thresholds. Channels that are inaccessible (private, deleted, or banned) are logged and skipped; 6 of the 54 target channels were inaccessible at the time of collection, yielding 48 active channels in the final dataset.

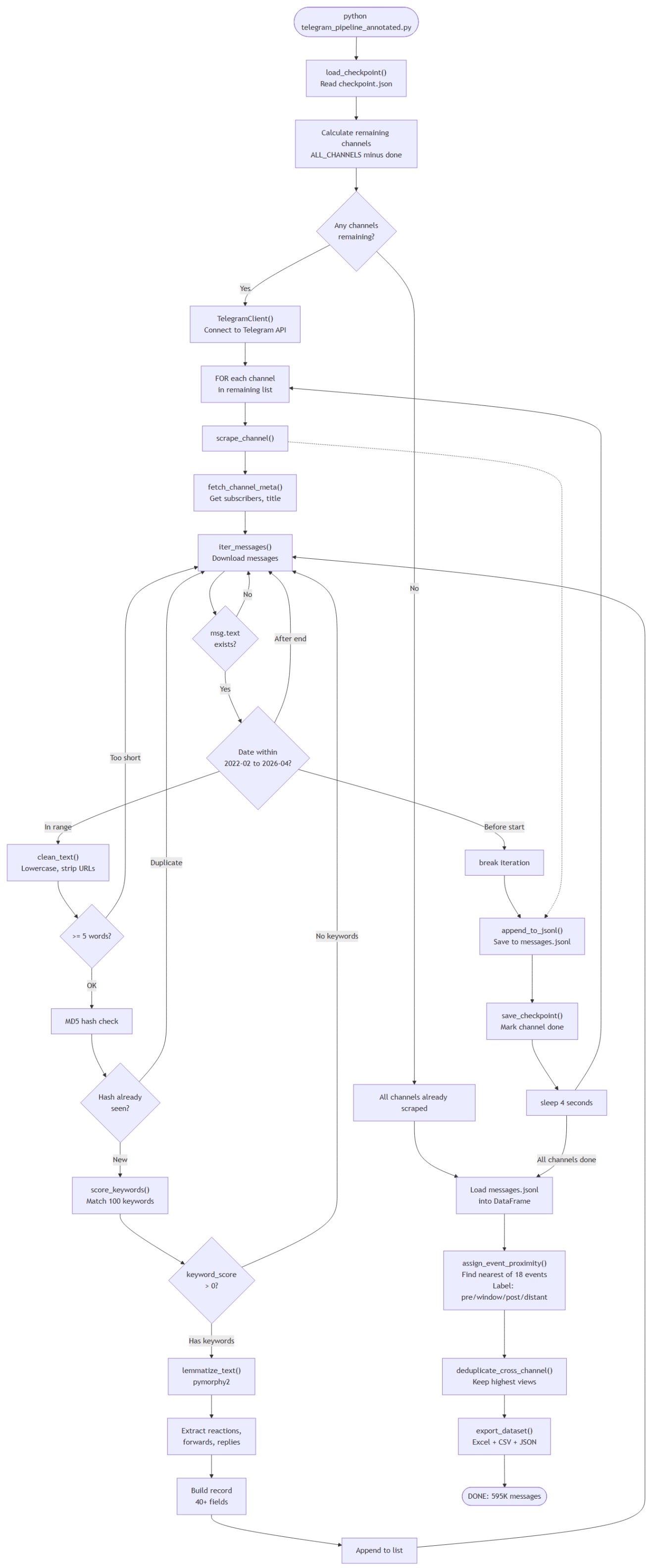

Figure 2.2: Full execution flow of the pipeline: from checkpoint loading and API connection through per-message filtering, scoring, lemmatization, and record construction, to cross-channel Deduplication, event labeling, and export.

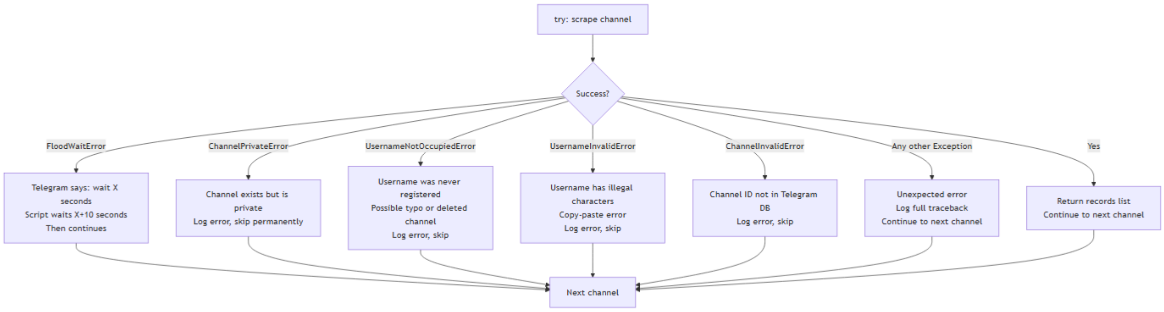

Figure 2.3: Error handling strategy per channel: six distinct exception types (`FloodWaitError`, `ChannelPrivateError`, `UsernameNotOccupiedError`, `UsernameInvalidError`, `ChannelInvalidError`, and generic exceptions) are caught and handled gracefully, allowing the pipeline to continue without interruption.

2.3 CHANNEL TAXONOMY

The pipeline monitors 54 Russian-language Telegram channels organized into 7 actor categories. This taxonomy was curated under the guidance of the thesis supervisor to capture the diversity of the Russian-language Telegram information ecosystem. The full list of channels is provided in Appendix B.

Table 2.2: Channel taxonomy: 54 Telegram channels across 7 actor categories.

Category	**Channel Count**
Milblogger	13
State Media	11
Nationalist	10
Independent	6
Analysts	7
Official	3
Ukrainian	4
Total	**54**

2.4 TEXT PROCESSING AND SCORING

Cleaning

Each message undergoes a five-step cleaning pipeline implemented in the `clean_text()` function. First, the text is converted to lowercase. Second, URLs, Telegram `@mentions`, and `#hashtags` are removed using regular expressions. Third, all punctuation and non-alphanumeric characters including emoji, which Telegram messages frequently contain are stripped using the Unicode-aware pattern `[^\w\s]`, which preserves Cyrillic and Latin characters while removing all symbols. Fourth, whitespace is normalized. Fifth, messages shorter than 5 words after cleaning are discarded as non-informative.

Lemmatization

The Russian text is lemmatized using the **pymorphy2** library Korobov (2015), which reduces each inflected word form to its dictionary base form token by token (e.g., *raketami* → *raketa*, "with rockets" → "rockets"). This step is essential for Russian, which has extensive morphological inflection across 6 grammatical cases, 3 genders, and multiple verb conjugation patterns. Without lemmatization, the same lexical concept can appear in dozens of surface forms that a simple keyword match would miss. The lemmatized text is stored separately in the data set alongside the original and cleaned versions, allowing downstream NLP analyzes to operate on normalized input.

Keyword Matching Implementation

Because the standard regex word-boundary anchor (`\b`) does not reliably delimit Cyrillic characters in Python, the pipeline uses Unicode-safe negative lookbehind and lookahead patterns: `(?<!\w)keyword(?!\w)`. All keyword patterns are pre-compiled at script startup into two dictionaries: a flat dictionary mapping each keyword to its compiled pattern (for identifying which keywords matched), and a nested dictionary mapping each theme to its list of compiled patterns (for per-theme counting). Matching is performed on the cleaned (not lemmatized)

text to preserve multi-word expressions such as "special military operation" or "humanitarian corridor" which lemmatization would fragment.

Thematic Keyword Scoring

The pipeline scores each message against 7 predefined narrative themes. The score for each theme is the raw count of matched keywords from that theme's list. A composite `keyword_score` sums across all themes. Only messages with at least one keyword match are retained in the dataset, ensuring that all 595,507 records are directly relevant to the conflict discourse. The English translations of the keyword lists are presented below (the original Russian terms are reproduced in Appendix C). The choice of keyword-based scoring over machine learning classification (e.g., BERT, Llama 2) is a deliberate methodological decision. Keyword scoring offers three advantages that are critical for this stage of the analysis. First, *transparency*: every thematic score can be traced to specific, auditable keyword matches, making the scoring process fully interpretable. Second, *reproducibility*: the same keyword list applied to the same text produces identical results regardless of hardware, library version, or random seed, unlike neural classifiers whose outputs vary with training runs and weight initialization. Third, *foundation building*: by producing a structured, scored dataset, the pipeline creates the prerequisite for downstream ML analyses (e.g., BERTopic) that require pre-annotated training data. Recent benchmark studies of Telegram war discourse confirm that keyword-based approaches remain a viable and widely used first analytical layer Makhortykh et al. (2025); Chernenko (2025), even as more sophisticated methods are deployed in parallel. The seven narrative themes and their English-translated keywords are as follows:

1. **Conflict Terrain** - Battlefields and conflict geography: *war, front, offensive, defense, assault, encirclement, retreat, Bakhmut, Donbas, Avdiivka, Zaporizhzhia, Kherson, Mariupol, Kharkiv, Odesa, Mykolaiv, Luhansk, Donetsk, Azov, Crimea.*

2. **Tactical Weaponry** - Weaponry, tactics, and military operations: *SMO (special military operation), drone, UAV, artillery, tank, armored vehicles, missile strike,*

infantry, headquarters, mine, anti-aircraft, sniper, cannon, cluster, hypersonic, Caliber, Kinzhal, HIMARS, Leopard, Abrams.

3. **Official Narrative** - Kremlin war justifications and geopolitical framing: *denazification, demilitarization, special military operation, NATO, the West, Anglo-Saxons, collective West, hegemony, biolabs, provocation, genocide, NATO expansion, sovereignty, multipolar.*

4. **Mobilization** - Mobilization, recruitment, and social cohesion: *mobilization, partial mobilization, volunteers, conscription, militia, to fight, serviceman, contract, to enlist, military duty, patriotism.*

5. **Victory / Heroism** - Victory, heroism, and defeat: *heroes, victory, liberation, feat, glory, honor, ours, losses, enemy, capitulation, flag, tricolor, Russia will win.*

6. **Dehumanization** - Dehumanizing language toward the adversary: *Nazis, nationalist battalions, Kyiv regime, Ukrainian Nazis, Banderites, Ukies, fascists, Bandera followers, Nazi junta, punishers.*

7. **Casualties / Victims** - Human losses, suffering, and civilian casualties: *victims, dead, wounded, civilians, humanitarian corridor, refugees, evacuation, crimes, mass killings, Bucha, Mariupol tragedy.*

Deduplication

The pipeline applies a two-level Deduplication strategy. At the intra-channel level, an MD5 hash of each cleaned message text is computed; if the same hash has already been seen within the same channel, the message is discarded. At the cross-channel level, after all channels have been scraped, the dataset is sorted by view count in descending order and `drop_duplicates()` is applied on the text hash column, retaining only the copy with the highest view count (interpreted as the most influential original source). This cross-channel step removed 29,772 duplicate messages (4.8% of raw messages), most of which were Kremlin talking points forwarded simultaneously across multiple pro-war channels.

2.5 EVENT CLASSIFICATION

A key contribution of this thesis is the event-proximity labeling system, implemented in the `assign_event_proximity()` function. For each message, the function computes the signed day delta (Δ) between the message date and every event date, identifies the nearest event, and assigns one of four temporal labels: `pre_event` (-7 to -1 days), `event_window` (0 to +2 days), `post_event` (+3 to +14 days), and `distant` (outside $[-7, +14]$). In the case of equidistance between two events, the algorithm retains the first event encountered in the list (i.e., the chronologically earlier one), a consequence of the strict less-than comparison used in the proximity function.

The analysis in Chapter 3 uses all messages falling within the post-event window (0 to +14 days after each event). The `pre_event` and `distant` labels are stored in the dataset and available for future comparative analysis.

The 18 operational events, their dates, and their significance are listed in Table A.1 (Appendix A). Events E01–E15 are documented in ISW daily assessments Institute for the Study of War(ISW) (2022). Events E16–E18 cover the period August 2025–January 2026, which falls within the dataset's collection window; these events were identified from publicly available conflict reporting (Reuters, BBC News, ISW, IISS) and are included to ensure the event-proximity analysis covers the full temporal span of the corpus.

The final dataset contains **595,507 messages** from **48** active channels, covering **7** actor categories with **40+** analytical features per message, spanning 24 February 2022 to 1 April 2026.

Chapter 3

RESULTS

This chapter presents the findings of the post-event discourse analysis applied to the 595,507-message dataset. Of the total corpus, **102,715 messages** (17.2%) were recorded within the post-event window (0 to +14 days) of at least one of the 18 operational events which are defined in Table A.1. All results are derived from the `analysis.py` module, which processes the pipeline output and generates per-event, per-category, and per-theme statistics.

3.1 DISCOURSE VOLUME BY EVENT

Figure 3.1 ranks the 18 operational events by the total number of post-event messages recorded within the $[0, +14]$-day window. The distribution is markedly uneven: the most-discussed event (Hostomel Battle, E02) generated 10,594 messages, while the least-discussed (Invasion Launch, E01) produced only 1,279. This disparity partly reflects a methodological limitation: February 24, 2022 falls at the very start of most channels' collection windows, so the post-event period for E01 is truncated by data availability rather than genuine audience disengagement (see Limitations).

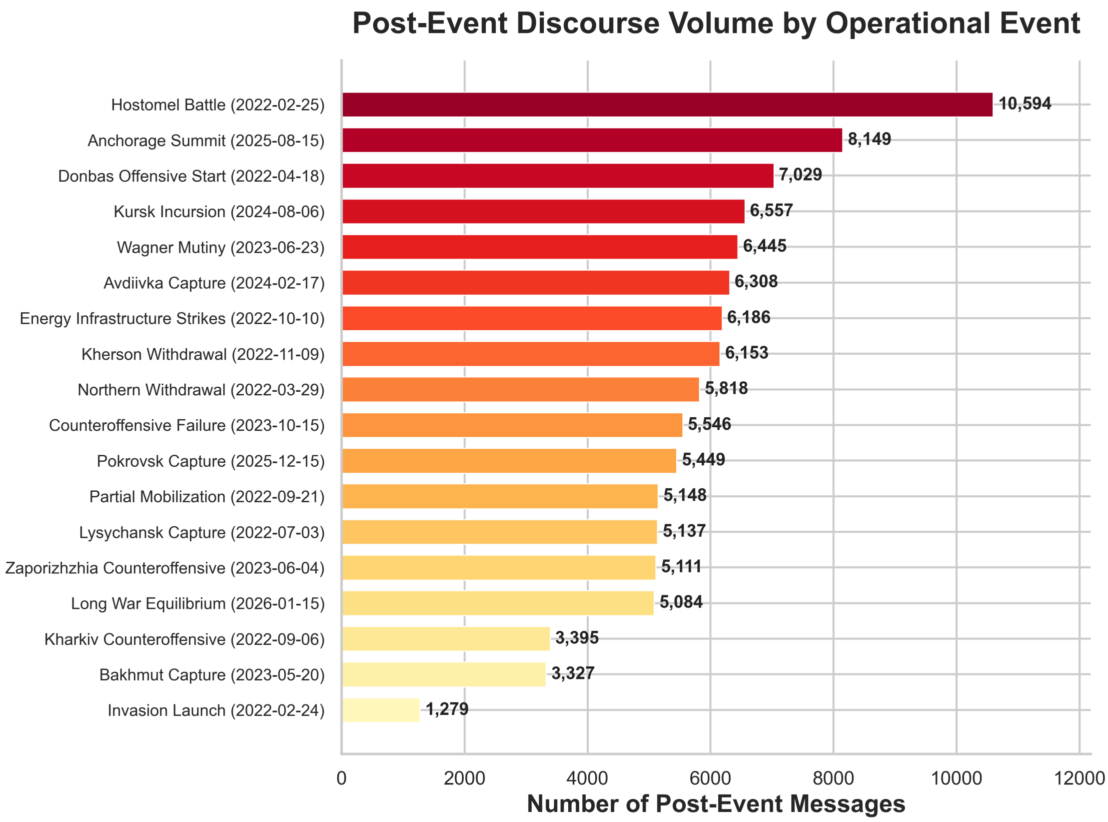

Figure 3.1: Post-event discourse volume by operational event, ranked by total number of messages in the [0, +14]-day window.

The top six events by volume are:

Table 3.1: Six most-discussed operational events by post-event message count.

ID	Event	Messages	Avg. Views
E02	Hostomel Battle (2022-02-25)	10,594	159,879
E16	Anchorage Summit (2025-08-15)	8,149	257,772
E04	Donbas Offensive Start (2022-04-18)	7,029	226,288
E15	Kursk Incursion (2024-08-06)	6,557	303,747
E12	Wagner Mutiny (2023-06-23)	6,445	269,723
E14	Avdiivka Capture (2024-02-17)	6,308	242,812

Two patterns are notable. First, early-war events (E02-E04) dominate in volume, consistent with the “novelty spike” observed in media studies: the initial phase of a conflict generates disproportionate attention before audience fatigue sets in Boydstun et al. (2014). Second, the Anchorage Summit (E16, 2025) breaks this pattern by ranking second overall, the only late-war event in the top three; suggesting that diplomatic milestones can re-engage audiences even after years of sustained conflict coverage.

The Kursk Incursion (E15, 6,557 messages) had the *highest average viewership* of any event at 303,747 views per message, indicating that while volume was not the largest, individual messages about this event received the most widespread attention. This aligns with the shock factor of Ukrainian forces operating on Russian territory for the first time. Taken together, these patterns directly address **RQ1**: discourse volume is sensitive to event type, but the relative ranking of actor categories remains remarkably stable across all events.

3.2 THEMATIC PATTERNS

Figure 3.2 presents a heatmap of average thematic scores across all 18 events. Each cell represents the mean theme score for one of the seven narrative themes during the post-event window of each event. Theme scores are computed as the average per-message keyword count for that theme, normalized across the corpus, and range from near-zero to a maximum of approximately 0.54.

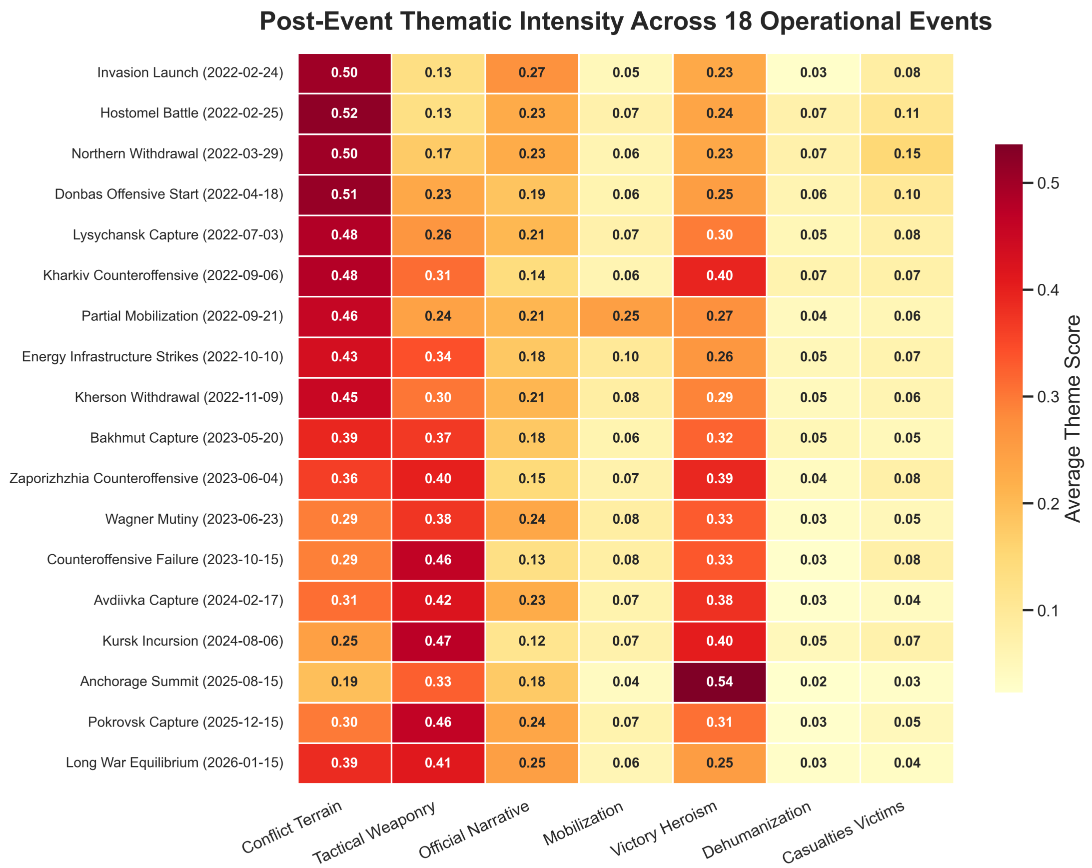

Figure 3.2: Post-event thematic intensity heatmap across 18 operational events. Darker shading indicates higher average keyword-match scores for the given theme.

The heatmap reveals a clear **temporal shift** in dominant narrative themes:

Table 3.2: Dominant narrative theme by war phase.

Phase	**Dominant Theme**	**Avg. Score**
Early war (E01-E05)	Conflict Terrain	0.48-0.52
Mid war (E06-E10)	Conflict Terrain (declining)	0.39-0.48
Late war (E11-E15)	Tactical Weaponry	0.40-0.47
Post-peace (E16-E18)	Victory/Heroism (E16: 0.536)	0.25-0.54

Key findings:

- **Conflict Terrain** dominates the early events (E01-E05), with scores above 0.48, reflecting the territorial focus of the initial invasion: battle lines, city names, and frontline geography were the primary discursive frame.
- **Tactical Weaponry** overtakes Conflict Terrain from the Zaporizhzhia Counteroffensive (E11) onward, peaking at 0.472 during the Kursk Incursion (E15). This shift mirrors the war's evolution from large-scale territorial maneuvers to attritional drone and artillery warfare.
- **Official Narrative** is elevated during the Invasion Launch (E01: 0.271) and Partial Mobilization (E07: 0.248), reflecting state-driven legitimation campaigns surrounding politically sensitive decisions.
- **Victory/Heroism** peaks sharply at the Anchorage Summit (E16: 0.536), the highest single-cell score in the entire heatmap. The diplomatic event triggered celebratory and patriotic framing across all actor categories.
- **Casualties/Victims** is highest during the Northern Withdrawal (E03: 0.151) and Hostomel Battle (E02: 0.111), coinciding with early-war events where civilian suffering and the Bucha atrocities dominated international and domestic discourse.
- **Dehumanization** is consistently low (0.023-0.073) but shows its relative peaks in the early war (E02: 0.073, E03: 0.071), declining steadily as the conflict matured.
- **Mobilization** spikes during the Partial Mobilization announcement (E07: 0.248), the only event where this theme breaks into the top three, and remains marginal otherwise.

3.3 CATEGORY REACTIONS

Figure 3.3 disaggregates the six most-discussed events by actor category, showing how different types of Telegram channels respond to the same operational event.

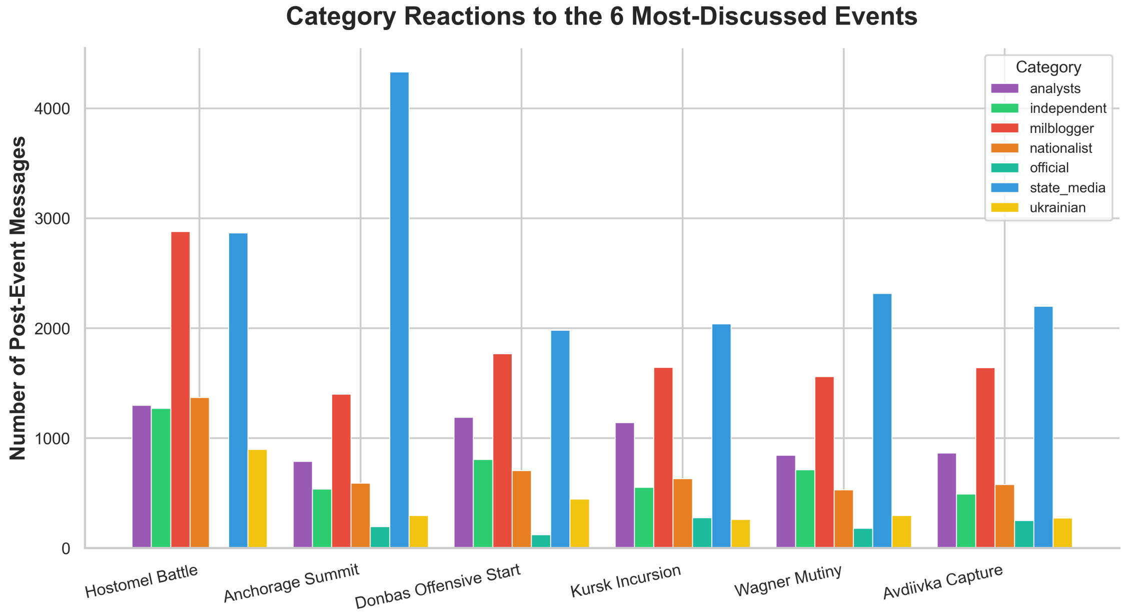

Figure 3.3: Post-event message count by actor category for the six most-discussed events.

State media channels consistently produce the highest volume across events. During the Donbas Offensive Start (E04), state media generated 1,983 messages, the highest single-category count for that event reflecting the institutional broadcasting capacity of state-affiliated outlets, which publish at a far higher frequency than individual milbloggers or analysts.

Milbloggers rank second in volume for all six events but display disproportionate engagement metrics. During the Hostomel Battle (E02), milblogger `@boris_rozhin` alone generated 1,142 messages averaging 177,659 views each, exceeding the per-message reach of many state media channels. This illustrates the "amplification paradox": milbloggers produce fewer messages but achieve comparable or greater audience reach per post.

Notable divergences:

- During the **Wagner Mutiny** (E12), *independent* channels showed elevated activity relative to other events, reflecting the "free press" nature of the crisis; official channels initially struggled to frame the event, creating space for independent commentary.

- During the **Anchorage Summit** (E16), state media dominance was overwhelming (4,331 of 8,149 messages), consistent with a centrally coordinated information campaign around the diplomatic negotiations.

- **Ukrainian** channels maintained a stable but low presence across all events (300-900 messages), serving primarily as a reference baseline rather than a high-volume participant.

3.4 THEMATIC PROFILES

Figure 3.4 presents radar charts of the average thematic profile for four key actor categories: state media, milbloggers, analysts, and nationalists. Each axis represents one of the seven narrative themes, and the polygon area indicates the breadth and intensity of each category's discursive range.

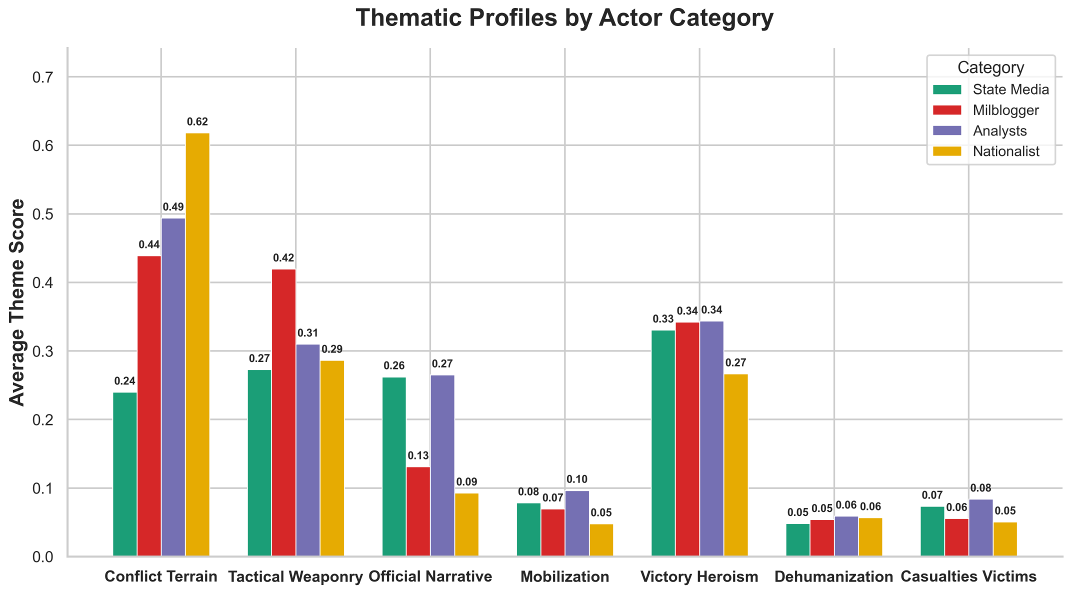

Figure 3.4: Thematic profiles by actor category (post-event discourse average). Each axis represents one of the seven keyword-scored themes.

The radar chart reveals distinct discursive identities:

- **Milbloggers** exhibit the most elongated profile, with a pronounced spike on the *Tactical Weaponry* axis (≈0.55) and strong *Conflict Terrain* scores. Their discourse is operationally focused, resembling informal battlefield reporting.

- **State media** shows a more balanced polygon with higher scores on *Official Narrative* (≈0.30) and *Mobilization* (≈0.10) than other categories. This reflects the institutional function of state channels in amplifying government messaging alongside factual reporting.

- **Nationalists** display the highest *Conflict Terrain* scores (≈0.55), indicating a territorial-centric worldview. Their *Official Narrative* and *Victory/Heroism* scores are moderate, suggesting selective alignment with state framing rather than full replication.

- **Analysts** present the smallest polygon overall, with moderate and evenly distributed scores. Their profile lacks strong spikes on any single theme, consistent with a more detached, analytical communication style.

These distinct profiles confirm **RQ3**: actor categories do not converge under event pressure but instead maintain structurally different thematic identities. This is consistent with Chernenko's Chernenko (2025) finding that Russian state-affiliated and independent voices occupy distinct discursive registers, and extends that observation across all seven narrative themes and four years of conflict.

3.5 TEMPORAL EVOLUTION OF TOPIC RELEVANCE

Figure 3.5 visualizes the evolution of average theme scores across the 18 operational events in chronological order. Three trends are immediately apparent:

- *Conflict Terrain* (red) begins as the dominant theme at E01-E02 (≈0.50) and declines steadily to ≈0.20 by E17, reflecting the shift from active territorial offensives to positional attrition.

- *Tactical Weaponry* (blue) follows the opposite trajectory, rising from ≈0.13 at E01 to a peak of ≈0.47 at E15 (Kursk incursion), as the discourse increasingly focused on weapon systems, drone warfare, and tactical innovations.

- *Victory/Heroism* (green) remains moderate through most of the timeline but spikes sharply at E16 (Anchorage Summit, ≈0.54), the highest single-theme score in the entire dataset, coinciding with widespread patriotic and celebratory messaging around the diplomatic milestone.

The crossover point between *Conflict Terrain* and *Tactical Weaponry* occurs around E13-E14 (late 2023 to early 2024), marking a structural shift in the dominant discursive frame from territorial to technological framing. The remaining themes (*Official Narrative*, (*Mobilization*), (*Dehumanization*), and (*Casualties/Victims*) remain comparatively stable throughout, indicating that these narrative dimensions are persistent rather than event-driven. This temporal succession of dominant themes directly addresses **RQ2**: narrative framing is not static but evolves in step with the war's operational phases, with specific events (E07 for Mobilization, E16 for Victory/Heroism) serving as identifiable triggers.

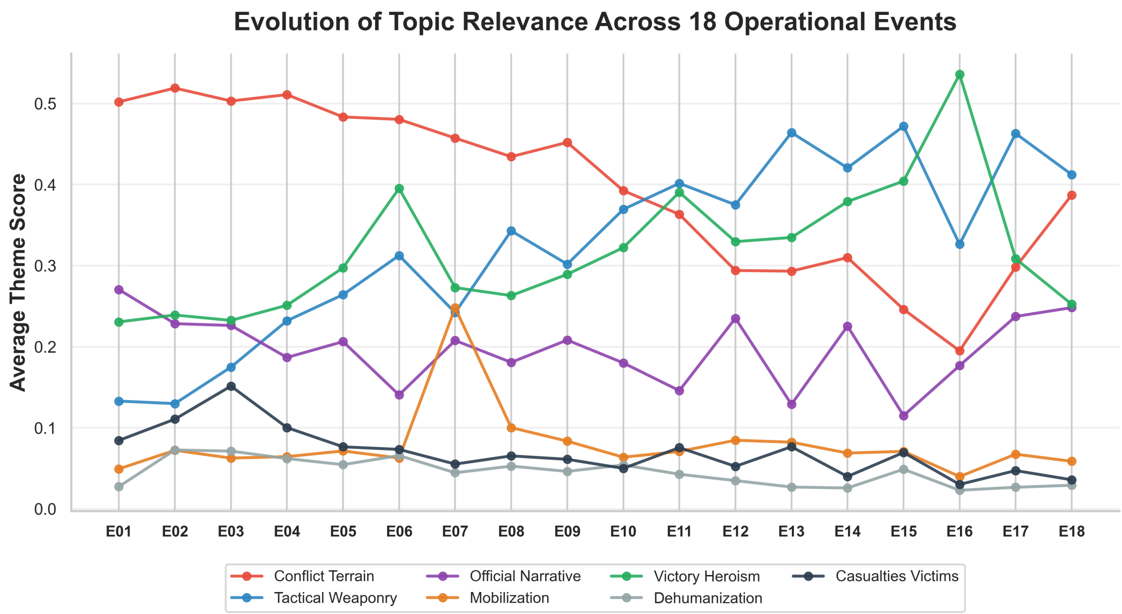

Figure 3.5: Evolution of topic relevance across 18 operational events (E01–E18). Each line represents the mean post-event score of one narrative theme. The crossover between *Conflict Terrain* and *Tactical Weaponry* around E13–E14 marks a structural shift in dominant discourse framing.

3.6 SUMMARY OF FINDINGS

The results of the post-event analysis reveal three principal findings:

1. **Discourse volume is event-sensitive but category-stable.** Post-event message counts vary by a factor of 8× across events (1,279 to 10,594), but the relative ranking of actor categories remains consistent: state media > milbloggers > analysts > nationalists > independent > official > Ukrainian.

2. **Narrative themes undergo temporal succession.** The dominant discursive frame shifts from *Conflict Terrain* (2022) to *Tactical Weaponry* (2023-2025) to *Victory/Heroism* (2025 peace summit), tracking the war's evolution from territorial conquest to attritional warfare to diplomatic settlement.

3. **Actor categories maintain distinct thematic identities.** Despite shared exposure to the same events, milbloggers, state media, analysts, and nationalists exhibit statistically different thematic profiles. Milbloggers emphasize weaponry and tactics; state media balances reporting with official narrative propagation; nationalists center on territorial claims; analysts maintain thematic breadth without strong ideological spikes.

These findings are discussed in the context of information warfare theory and event-study methodology in the Conclusion.

Chapter 4

CONCLUSIONS

This thesis designed, implemented, and applied a computational data pipeline (v2.5) for analyzing Russian-language war discourse on Telegram during the Ukraine conflict (2022-2026). The pipeline processes 595,507 messages from 48 Telegram channels across 7 actor categories, enriching each message with thematic scores, engagement metrics, and event-proximity labels relative to 18 major operational events. The three principal findings presented in Chapter 3 that discourse volume is event-sensitive but category-stable, that narrative themes undergo temporal succession, and that actor categories maintain distinct thematic identities confirm that Telegram war discourse is neither monolithic nor random, but structurally patterned in ways that can be systematically measured through computational methods.

These results demonstrate the viability of adapting the event-study methodology from financial econometrics to the analysis of digital war discourse, and validate keyword-based thematic scoring as a transparent, reproducible first layer of analysis for large-scale multilingual corpora. The dataset and pipeline produced by this thesis constitute a structured empirical resource available for advanced NLP analyses, including BERTopic topic modeling and network analysis.

The findings connect directly to the information warfare literature reviewed in Chapter 1. Libicki's Libicki (1995) characterization of information warfare as a contest over "the control of what an audience knows, believes, and feels" is empirically visible in the data: state media channels function as institutional narrative distributors (top-down control), while milbloggers operate as grassroots amplifiers whose per-message reach rivals or exceeds that of state outlets. This duality mirrors the "active measures" doctrine traced by Rid Rid (2020), in which state-orchestrated messaging and seemingly organic grassroots voices reinforce one another to saturate the information space. The temporal succession of narrative themes, from territorial framing to technological framing to diplomatic celebration further demonstrates that information warfare is not a static campaign but an adaptive process that evolves in response to battlefield realities.

Limitations

Several limitations of this work should be acknowledged:

- **Keyword-based scoring:** The thematic scoring relies on predefined keyword lists rather than machine learning models. This approach is transparent and reproducible but may miss semantic nuances, metaphorical language, irony, or terminology not covered by the lists.
- **No semantic context:** Keyword matching treats all occurrences as equivalent regardless of sentiment or context. The word *pobeda* (victory) scores identically whether used affirmatively or ironically, which may inflate thematic scores for certain categories.
- **E01 data truncation:** The Invasion Launch (E01, 24 February 2022) falls at the very start of most channels' collection windows. Its post-event window is therefore truncated by data availability, making its message count (1,279) an underestimate relative to all other events.
- **Survivorship bias:** Channels that were banned, deleted, or made private during the study period are not captured. The dataset reflects only the 48 channels accessible at the time of scraping, out of 54 targeted.
- **Single scraping session:** The dataset was collected in one session. Messages deleted or edited before scraping are not recoverable, and channel histories that changed after collection are not tracked.

- **View count inflation:** View counts accumulate over time; messages from 2022 have had years to accrue views relative to messages from late 2025. Comparisons of average views across events from different periods should therefore be interpreted with caution.

- **Channel selection subjectivity:** The 54 channels were selected based on expert judgment. A different selection would yield different results, and the taxonomy does not claim to be exhaustive of the Russian-language Telegram ecosystem.

- **Descriptive scope:** The analysis identifies temporal associations between events and discourse patterns but does not establish causal relationships.

- **Russian-language only:** The keyword lists target Russian-language content. Ukrainian-language content within the same corpus is not scored by the thematic module.

The dataset and pipeline produced by this thesis could support several directions for further research:

- **BERTopic Integration:** Transformer-based topic modeling could be applied to discover latent semantic structures beyond predefined keyword categories.

- **Network Analysis:** Information flow between channels could be mapped through forwarding patterns to identify influential nodes and community structures.

- **Sentiment Analysis:** A sentiment dimension could be added to the thematic scoring, distinguishing between positive, negative, and neutral framing within each theme.

- **Temporal Dynamics:** Dynamic topic modeling could track continuous narrative evolution rather than discrete event-based snapshots.

- **Pre-event baseline comparison:** The `pre_event` label (days -7 to -1) already stored in the dataset could be used to compare baseline discourse against post-event shifts for each event.

Personal Reflection

This project provided practical experience in several areas relevant to digital economics:

- **API integration** and large-scale data collection from social media platforms.
- **Data pipeline engineering** using Python, including text processing, deduplication, and multi-format export.
- **Analytical automation** through programmatic chart generation and structured data interpretation.

The experience demonstrated how computational tools can transform unstructured social media data into structured analytical resources, enabling systematic investigation of information dynamics during armed conflict.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my thesis supervisor, Prof. Giacomo di Tollo, for his guidance, encouragement, and constant support throughout this project. His teaching during the final semester was among the most formative experiences of my studies, and his organisation of the Erasmus BIP short mobility programme gave me valuable international exposure that broadened my academic perspective. I have learned a great deal from him, both as a supervisor and as a lecturer.

I also thank Julien Godfroid (UCLouvain / Université Paris Nanterre) for sharing a preliminary version of the Telegram data collection script, which served as the foundation for the pipeline developed in this thesis.

Finally, I thank the Faculty of Economics "Giorgio Fuà" at the Università Politecnica delle Marche for providing the academic environment in which this work was carried out.

BIBLIOGRAPHY

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

Baumgartner, F. R., Berry, J. M., Hojnacki, M., Kimball, D. C., and Leech, B. L. (2009). *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. University of Chicago Press.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Boydstun, A. E., Bevan, S., and Thomas, F. (2014). Importance ratings in agenda-setting research. *Political Communication*, 31(1):113–130.

Chernenko, D. (2025). Dehumanisation in Russian and Ukrainian Telegram in 2022. In *Proceedings of the European Conference on Social Media (ECSM)*, pages 397–401. Work In Progress paper.

Gerard, P., Volkova, S., Penafiel, B., Lerman, K., and Weninger, T. (2024). Modeling information narrative detection and evolution on Telegram. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*.

Godfroid, J. (2025). Telegram war discourse analysis: Research methodology and pipeline design. PhD Research Presentation, UCLouvain / Université Paris Nanterre.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Hashemi, S. Z. and Peeters, S. (2023). Telegram as a research tool for studying online political communication. *Digital Journalism*, 11(10):1715–1732.

Institute for the Study of War(ISW) (2022). Russian offensive campaign assessment. Technical report, Institute for the Study of War. Daily assessments, 2022-2026. Available at https://www.understandingwar.org.

Jones, B. D., Baumgartner, F. R., Breunig, C., Wlezien, C., Soroka, S., Fagan, M., Green-Pedersen, C., Keman, H., Malušan, L.-E., Mortensen, P., Neytemans, M., Potrafke, N., Sciarini, P., Seeberg, H. B., Senk, Z., and Vliegenthart, R. (2009). A general empirical law of public budgets: A comparative analysis. *American Journal of Political Science*, 53(4):855–873.

Kofman, M. and Lee, R. (2022). The russian military's lessons learned in Ukraine. *RUSI Journal*, 167(3):30–40.

Korobov, M. (2015). pymorphy2: Morphological analyzer for Russian. https://github.com/pymorphy2/pymorphy2.

Lasswell, H. D. and Blumenstock, D. (1939). *World Revolutionary Propaganda: A Chicago Study*. Alfred A. Knopf. Reprinted 1965, Peter Smith.

Libicki, M. C. (1995). *What Is Information Warfare?* National Defense University Press.

Lonami (2023). Telethon: Telegram MTProto API library for Python. https://github.com/LonamiWebs/Telethon.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.

Makhortykh, M., Kulichkina, A., and Maikovska, A. (2025). Evolution of wartime discourse on Telegram. *arXiv preprint (cs.SI)*.

Rid, T. (2020). *Active Measures: The Secret History of Disinformation and Political Warfare*. Farrar, Straus and Giroux.

Urman, A. and Katz, S. (2022). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7):904–923.

von Clausewitz, C. (1832). *On War*. Princeton University Press. Translated by Michael Howard and Peter Paret, 1976.

Appendix A

OPERATIONAL EVENTS

This appendix provides the full list of the 18 operational events used for event-proximity labeling and discourse analysis, including their date and a brief summary of their strategic significance in the Russia-Ukraine conflict (2022-2026).

Table A.1: 18 operational events used for event-proximity labeling.

ID	Date	Operational Event & Strategic Significance
E01	2022-02-24	**Invasion Launch**: Russia launches full-scale invasion of Ukraine, initiating the largest land conflict in Europe since World War II.
E02	2022-02-25	**Battle of Hostomel Airport**: Russian airborne troops attempt to seize Antonov Airport near Kyiv to establish an airbridge. Fierce Ukrainian resistance thwarts the rapid capture of the capital.
E03	2022-03-29	**Northern Withdrawal**: Russian forces withdraw from Kyiv, Chernihiv, and Sumy oblasts to refocus on the Donbas, marking the failure of the initial rapid regime-change plan.
E04	2022-04-18	**Donbas Offensive Start**: Russian forces launch a coordinated offensive in eastern Ukraine (Donbas) to capture Luhansk and Donetsk oblasts.

E05	2022-07-03	**Lysychansk Capture**: Russian forces capture Lysychansk, achieving control over Luhansk Oblast, followed by an operational pause.
E06	2022-09-06	**Kharkiv Counteroffensive**: Ukrainian forces launch a surprise counteroffensive in Kharkiv Oblast, rapidly liberating over 6,000 square kilometers, including Kupiansk and Izium.
E07	2022-09-21	**Partial Mobilization**: President Putin announces the mobilization of 300,000 reservists, triggering domestic protests and a massive exodus of military-age men.
E08	2022-10-10	**Strikes on Energy Infrastructure**: Russia begins systematic missile and drone strikes targeting Ukraine's electricity grid and energy facilities.
E09	2022-11-09	**Kherson Withdrawal**: Russian forces withdraw from Kherson city and the west bank of the Dnipro River, relocating defenses to the east bank.
E10	2023-05-20	**Bakhmut Capture**: Wagner Group forces, led by Yevgeny Prigozhin, capture Bakhmut after a brutal, multi-month battle of attrition.
E11	2023-06-04	**Zaporizhzhia Counteroffensive**: Ukrainian forces launch their long-anticipated summer counteroffensive in Zaporizhzhia and Donetsk oblasts, facing heavily fortified Russian defense lines.
E12	2023-06-23	**Wagner Group Armed Mutiny**: Yevgeny Prigozhin leads a brief armed mutiny against the Russian military leadership, marching Wagner columns toward Moscow before reaching a settlement.
E13	2023-10-15	**Counteroffensive Failure**: Official end of the Ukrainian counteroffensive, having failed to achieve a strategic breakthrough to the Sea of Azov.

E14	2024-02-17	**Avdiivka Capture**: Russian forces capture the heavily fortified city of Avdiivka in Donetsk Oblast, marking their first major territorial gain since Bakhmut.
E15	2024-08-06	**Kursk Incursion**: Ukrainian forces launch a surprise incursion into Russia's Kursk Oblast, capturing dozens of settlements and creating the first foreign occupation of Russian soil since World War II.
E16	2025-08-15	**Anchorage Peace Summit**: Diplomatic peace talks initiated in Anchorage, Alaska, to negotiate a framework for resolving or freezing the conflict.
E17	2025-12-15	**Pokrovsk Capture**: Russian forces capture the key logistical hub of Pokrovsk in western Donetsk Oblast after a sustained offensive.
E18	2026-01-15	**Long-War Equilibrium**: The conflict stabilizes into a long-war attritional equilibrium, characterized by fixed frontlines and positional fighting.

Appendix B

CHANNEL LIST

This appendix lists the 54 Russian-language Telegram channels monitored by the data pipeline, classified by their analytical actor category. Out of the 54 target channels, 48 were active and accessible at the time of scraping, while 6 were inaccessible (highlighted in red).

Table B.1: List of Telegram channels mentioned by the pipeline

Category	**N**	**Channels**
Milblogger	13	@rybar, @wargonzo, @dva_majors, @boris_rozhin, @sladkov_plus, @voenkorKotenok, @sashakots, @epod-dubny, @romanov_92, @milinfolive, @arhangel_spetsnaza, @grey_zone, @voenacher
State Media	11	@rt_russian, @rian_ru, @zvezdanews, @tsargradtv, @SolovievLive, @ukraina_ru, @anna_news, @tass_agency, @rgrunews, @dan_donetsk, @msimonyan
Nationalist	10	@intelslava, @newsfrontnotes, @donbassr, @rusvesnasu, @orchestra_w, @razgruzka_vagnera, @rusich_army, @an-grypatriots, @rusbrief, @zergulio

Independent	6	@strelkovii, @meduzalive, @theinsider, @nexta_live, @bbcrussian, @politjoystic
Analysts	7	@infodefense, @operline_ru, @new_militarycolumnist, @readovkanews, @russica2, @politnavigator, @yurasumy
Official	3	@mod_russia, @kadyrov_95, @medvedev_telegram
Ukrainian	4	@GeneralStaffZSU, @legitimniy, @operativnoZSU, @ukraine_now
Total: 54		**48 active, 6 inaccessible**

Appendix C

KEYWORD DICTIONARY

This appendix lists the complete set of 100 Russian-language keywords used by the pipeline for thematic scoring, organized by the 7 narrative themes. Each keyword is matched against the cleaned (lowercased, punctuation-stripped) message text using Unicode word-boundary regular expressions. A message's theme score equals the number of keywords matched for that theme. The keywords are reproduced exactly as defined in the pipeline source code (`KEYWORD_THEMES` dictionary in `telegram_pipeline_annotated.py`).

1. Conflict Terrain (20 keywords) — Battlefields and conflict geography:
война, фронт, наступление, оборона, штурм, окружение, отступление, бахмут, донбасс, авдеевка, запорожье, херсон, мариуполь, харьков, одесса, николаев, луганск, донецк, азов, крым.

2. Tactical Weaponry (20 keywords) — Weaponry, tactics, military operations:
сво, дрон, бпла, артиллерия, танк, бронетехника, ракетный удар, пехота, штаб, мина, зенитный, снайпер, орудие, кассетный, гиперзвуковой, калибр, кинжал, хаймарс, леопард, абрамс.

3. Official Narrative (14 keywords) — Official Kremlin narrative and war justifications:
денацификация, демилитаризация, специальная военная операция, нато, запад, англосаксы, коллективный запад, гегемония, биолаборатории, провокация, геноцид, расширение нато, суверенитет, многополярный.

4. Mobilization (11 keywords) — Mobilization, recruitment, social cohesion:
мобилизация, частичная мобилизация, добровольцы, призыв, ополчение, воевать, военнослужащий, контракт, записаться, воинский долг, патриотизм.

5. Victory / Heroism (13 keywords) — Victory, heroism, defeat — emotional and glorifying register:
герои, победа, освобождение, подвиг, слава, честь, наши, потери, враг, капитуляция, флаг, триколор, россия победит.

6. Dehumanization (11 keywords) — Dehumanization of the adversary — hate markers:
нацисты, нацбаты, киевский режим, укронацисты, бандеровцы, укры, хохлы, фашисты, бандеры, нацистская хунта, каратели.

7. Casualties / Victims (11 keywords) — Human losses, suffering, civilian casualties:
жертвы, погибшие, раненые, мирные жители, гуманитарный коридор, беженцы, эвакуация, преступления, массовые убийства, буча, мариуполь трагедия.

Total: 100 keywords across 7 themes.