Graph-Based Detection of Disinformation Narrative Diffusion between Russian and Ukrainian Telegram Channels

Yuliia Vistak
Ukrainian Catholic University
vistak.pn@ucu.edu.ua

Viktoriia Makovska
Ukrainian Catholic University
makovska.pn@ucu.edu.ua

Vera Schmitt
Technische Universität Berlin
vera.schmitt@tu-berlin.de

Veronika Solopova
Technische Universität Berlin
veronika.solopova@tu-berlin.de

Abstract

Detecting disinformation narratives on social media is challenging due to the scale of amplification, rapid evolution, and linguistic variability of online content. We propose a graph-based framework for identifying and analyzing disinformation narratives in Telegram ecosystems by combining weak supervision with propagation graph analysis. The approach aggregates semantically related claims into narrative-level clusters and models their diffusion across interconnected channels. This enables the detection of coordinated narrative amplification that is difficult to capture through post-level analysis alone. Our results demonstrate that integrating textual signals with network structure provides a scalable method for detecting disinformation narratives and offers insights into how they propagate within large-scale messaging environments.

1 Introduction

Disinformation at scale remains a persistent challenge for modern information ecosystems, with content volumes far exceeding the capacity of manual verification and fact-checking (Wardle and Derakhshan, 2017). This challenge is especially acute in conflict settings, where disinformation evolves rapidly and is repackaged across platforms and communities. That's why narrative-level representation offers a practical abstraction for analyzing high-volume environments: rather than tracking individual posts and factuality of their claims, they aggregate semantically related claims into higher-level frames that remain stable despite linguistic variation and cross-platform adaptation (Nikolaidis et al., 2025). Prior work in computational propaganda and fake news analysis shows that rhetorical framing and narrative structure capture systematic signals that generalize beyond surface text, enabling scalable trend detection across large corpora (Rashkin et al., 2017).

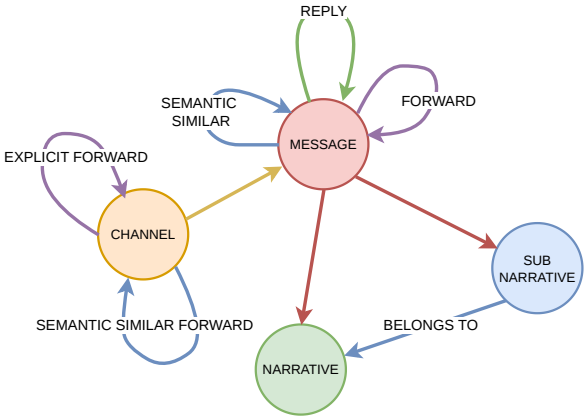

Figure 1: Schema of the graph used in our analysis, with channel, message, narrative, and sub-narrative nodes, and edges for posting, reply, explicit forwarding, semantic similarity, and narrative assignment.

Telegram is a particularly important platform for studying these dynamics, with a particularly high share of media space in Eastern Europe (Makhortykh et al., 2025). With moderation enforcement increased on mainstream social media in the early 2020s, harmful and conspiratorial communities migrated toward lower-moderation spaces such as Telegram (Kalkbrenner et al., 2025). Its channel-based broadcast architecture and native forwarding mechanism make information diffusion *structurally observable*: forwarded/forwarded-from metadata enables reconstruction of cross-channel propagation networks (La Morgia et al., 2025). Telegram is also a central venue for Kremlin-related and anti-Kremlin communications during the Russia-Ukraine conflict, operating at substantial scale (Bawa et al., 2025). Despite this, misinformation detection research remains heavily skewed toward English-language and platforms such as X (Kalkbrenner et al., 2025). Content-only classifiers are often brittle across languages and domains, while disinformation is frequently better characterized by amplifiers (e.g., botnets, troll

Proceedings of the Fifth Ukrainian Natural Language Processing Conference (UNLP 2026), pages 80–96

farms, and coordinated sockpuppet accounts) and their patterns, rather than by the lexicon of the messages. Propagation-aware methods exploit the empirical observation that deceptive and credible information spreads differently, thereby providing more language-agnostic and manipulation-resistant signals (Monti et al., 2019). Therefore, in this study, we pursue network-driven disinformation narrative analysis for the Ukrainian Telegram ecosystem, building on MisinfoTeleGraph (Kalkbrenner et al., 2025) while adapting it to a conflict-specific, narrative-centric setting. We collect posts from a manually curated set of Telegram channels and construct a directed *share graph* based on cross-channel forwarding (Figure 1). Our supervision target is *multi-class narrative assignment*: each message may activate multiple disinformation narratives. We operationalize the label space extending narrative taxonomy from the VoxCheck Propaganda Diary (database of narratives, created by fact-checking organization VoxCheck) and match messages against a fine-grained inventory of **380** pro-Russian disinformation subnarratives. [1] [2]

Methodologically, we compare three families of weak labeling approaches: semantic similarity to narrative descriptions, multilingual NLI-based zero-shot classification, and instruction-tuned LLM zero-, few-shot prompting. These methods are evaluated on a seed set of human-labeled messages, with model selection or ensembling guided by precision–coverage trade-offs. The resulting weak labels are attached to message nodes and, where needed, aggregated to the channel level. This labeled content is then analyzed in two complementary graph layers: an explicit share graph induced by Telegram forwarding metadata, and a semantic propagation graph induced by cross-channel message similarity.The full codebase, including labeling pipelines, graph construction, and evaluation scripts, is publicly available at GitHub repository. [3]

2 Background and Related Work

This section reviews the main strands of work that inform our approach: disinformation narrative analysis, weak supervision for large-scale labeling, and graph-based modeling of information spread. Together, these lines of research motivate our focus on narrative-level detection in Telegram and our combination of weak labeling with graph-based propagation analysis.

2.1 From Claims to Narratives

Narratives are recurring ideological structures that organize multiple claims into coherent worldviews (Hellman, 2024). Their effectiveness lies not in factual accuracy but in emotional resonance and identity alignment, which makes them resilient to corrective information (Dahlstrom, 2014). As a result, disinformation detection increasingly focuses on identifying narrative patterns rather than isolated falsehoods. Prior work explicitly models narratives as hierarchical or multi-level structures. PolyNarrative introduces a multilingual dataset with coarse- and fine-grained narrative labels (Nikolaidis et al., 2025), while datasets such as EUvsDisinfo provide benchmarks for detecting pro-Kremlin EU-targeting disinformation narratives in news articles (Leite et al., 2024). Domain-specific studies have also examined narrative ecosystems in contexts such as climate change denial (Upravitelev et al., 2026a,b). We build on this by adopting a hierarchical narrative representation tailored to the Ukrainian information environment.

2.2 Weak Supervision for Narrative Detection

Expert annotation remains a bottleneck for narrative-level disinformation detection. Programmatic Weak Supervision addresses this challenge by combining multiple noisy labeling sources (known as labeling functions) to generate scalable, probabilistic supervision (Ratner et al., 2017). Such sources may include heuristic rules, semantic similarity, or outputs of zero-shot classifiers.

In multilingual settings, weak labeling functions commonly rely on semantic similarity between texts and narrative descriptions, natural language inference (NLI)–based entailment scoring, or zero-shot classification with instruction-tuned language models (Yin et al., 2019; Schick and Schütze, 2021; Ratner et al., 2017). Weak supervision is well-suited for Telegram, where narrative inventories can be externalized and calibrated using small expert-labeled seed sets (Kalkbrenner et al., 2025).

2.3 Disinfo Spreading Networks Analysis via Graphs

Graph-based approaches model misinformation not only through message content, but also through the structure of its spread. Early work showed that

[1] https://russiandisinfo.voxukraine.org/
[2] https://voxukraine.org/en/voxcheck
[3] https://github.com/yuliavistak/TeleNarratives.git

credibility can be inferred in part from diffusion and interaction patterns in social media streams (Castillo et al., 2011). Large-scale evidence from Twitter demonstrated that false news spreads faster, farther, and more broadly than true news, which motivated propagation-aware approaches as an alternative to content-only classification (Vosoughi et al., 2018). Building on this, later work explicitly represented misinformation spread as graphs (Monti et al., 2019; Bian et al., 2020). Our work shifts the emphasis from message- or story-level misinformation detection to *narrative-level* propagation analysis in Telegram.

3 Methodology

Our methodology consists of four main components: (i) data collection, (ii) definition of a multi-label narrative assignment task using VoxCheck subnarratives, (iii) weak labeling via multiple few-shot and similarity-based labeling functions, and (iv) share graph construction from Telegram forwarding metadata.

3.1 Data Collection

We manually compiled a set of public Telegram channels relevant to war-related political discourse. The channel selection heavily relied on data from the fact-checking organization such as Spravdi. [4] We began with channels they identified as "unreliable" or "suspicious" and supplemented these with popular Ukrainian news channels exceeding 300,000 followers. This resulted in a total of **98** channels. Although we recognize that any curated list involves some bias, our aim was to make the final dataset as balanced as possible.

We also adopt the **VoxCheck Propaganda Diary**, which organizes pro-Russian disinformation into a structured taxonomy of **26** high-level **narratives** and **360** fine-grained **sub-narratives**, identified and curated during 2022–2023. Some of these narratives directly relate to the Russian–Ukrainian conflict (e.g. "The war in Ukraine demonstrates the supremacy of Russian weapons"), while others address broader societal allegations (e.g. "The Russian language is being suppressed in Ukraine"). This taxonomy reflects real-world monitoring practices and enables multi-class narrative assignment, capturing the fact that individual messages may activate multiple narrative elements simultaneously.

[4] Spravdi channel list.

The messages dataset includes (i) message text, (ii) timestamps, (iii) channel and message identifiers, and (iv) native forwarding provenance fields (`is_forwarded`, `fwd_from_channel_id`, `fwd_from_message_id`). This design tracks how content spreads across channels without using any private user data.

The resulting messages corpus is primarily bilingual, containing messages in both **Ukrainian** and **Russian**. Our primary objective was to analyze the discourse throughout 2025 and to include the early months of 2026, up to the date of our final data extraction. Consequently, the dataset consists of **1,352,668 messages** published between **16 December 2024** and **27 February 2026**. Figure 2 shows the temporal distribution of messages during the observation period.

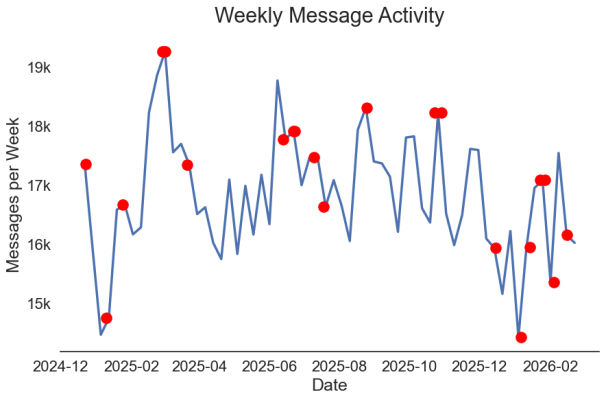

Figure 2: Weekly message activity within the curated Telegram corpus. Red markers denote significant political or conflict-related events (see Table 4 in the appendix for a comprehensive list). The peaks in message volume frequently coincide with these events.

To investigate the propagation of specific narratives and the underlying network topology between channels, we focused our analysis on the subset of messages involved in forwarding chains. This includes both **forwarded messages** and the **original posts** from which content was forwarded. From the initial corpus, this refined subset consists of **70,595** messages.

Some channels do not use the official "forward" button, copying and pasting text with only tiny changes. We call this *implicit forwarding*. To catch these cases, we looked for messages with very similar meanings. For that reason, the decision was to transform messages into embedding vectors via `baai/bge-m3`, compute the similarity score between them, and take pairs with high score

in further analysis. [5]

This step enabled the inclusion of **10,872** semantically similar pairs (representing **10,774** additional messages). By tracking both official forwards and these "copy-paste" reposts, we get a much clearer picture of how news and narratives actually spread.

The final dataset analyzed in this study consists of **81,369** messages, which were subsequently annotated for further investigation.

3.2 Multi-Class Narrative Assignment

We formulate disinformation detection as a **multi-class narrative assignment** problem. The narratives dataset is organized into a two-level hierarchy:

Narratives (26 total): Broad, general themes of disinformation (e.g., "The West controls Ukraine and uses it for its own purposes").

Sub-narratives (360 total): Specific claims and real-world examples that fit within those broader themes (e.g., "The US is using Ukraine to weaken Europe").

For manual annotation, we sampled 412 messages (*'golden set'*). To capture the corpus's temporal structure and account for activity fluctuations throughout the year, we employed **stratified random sampling** by week of the year.

Three independent annotators conducted the labeling in two phases. First, to establish inter-annotator agreement (IAA), they labeled *a shared subset* of 103 messages. Second, each annotator independently labeled an additional 103 distinct messages to expand the dataset. Annotators assigned each message to the best-matching narrative from the 26 predefined ones, or marked it as *not containing a narrative*.

During analysis, we observed that many messages contained pro-Russian narratives absent from the original taxonomy. This likely occurred because the messages in the dataset were relatively recent, while the existing set of narratives had been defined earlier and therefore did not fully capture newly emerging disinformation narratives. To address this gap, annotators identified **20 new sub-narratives**, which were subsequently harmonized and grouped into **6 new narrative categories**.

Furthermore, because single messages frequently contained multiple contextually related narratives, we grouped these interrelated narratives under broader, overarching meta-narratives.

The final dataset, **TeleNarratives**, is organized as a three-level taxonomy with **380 sub-narratives**, **32 narratives**, and **9 meta-narratives**. The complete resource, including a **Neo4j**[6] **graph database dump**, is publicly available at the project repository.

3.3 Weak Labeling

Given the scale of the corpus, full manual annotation was impractical due to limited human resources. Therefore, we adopted a **weak labeling** approach.

To implement this approach, we explored three different methods. The core idea behind these strategies was to compare each message with the sub-narratives rather than broader narratives. It allows models to perform more precise semantic comparisons between the message content and the narrative descriptions. Our goal was to evaluate the performance of these strategies and select the most effective one based on its agreement with the **golden set**.

Semantic similarity-based labeling. This approach utilizes **sentence-transformer models** to represent messages and sub-narratives as embedding vectors. Using cosine similarity, each message is assigned to the most similar sub-narrative, provided the score exceeds a predefined threshold. Messages that do not meet this criterion are classified as **not containing a narrative**.

To generate embeddings, we experimented with three multilingual models: `paraphrase-multilingual-MiniLM-L12-v2`, `BAAI/bge-m3`, `text-embedding-3-large (OpenAI)`[7] [8] [9]. These models were selected to compare different multilingual embedding approaches, including lightweight open-source models and larger, high-capacity proprietary systems.

Natural Language Inference (NLI)-based labeling. Natural Language Inference (NLI) models identify the logical relationship between two text segments as *entailment*, *contradiction*, or *neutral*. In this study, NLI models assess whether a message supports a particular sub-narrative.

Under this framework, the message acts as the *premise*, and the sub-narrative description

[5] https://huggingface.co/BAAI/bge-m3

[6] https://neo4j.com/

[7] https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

[8] https://huggingface.co/BAAI/bge-m3

[9] https://developers.openai.com/api/docs/models/text-embedding-3-large

as the *hypothesis*. The model calculates the probability of entailment for each pair; the message then receives the label of the sub-narrative with the highest score, provided it exceeds a predefined threshold. Messages that do not meet this criterion are classified as **not containing a narrative**. We experimented with `mDeBERTa-v3-base-xnli-multilingual-nli`[10] as a multilingual NLI model. We selected this model for its rare Ukrainian language support and strong community validation on Hugging Face.

LLM-based labeling with zero-shot and few-shot prompting. In this approach, the model identifies whether a message contains a specific sub-narrative and provides a confidence score. In this sense, the LLM acts as an 'additional annotator', producing labels that can be compared with human annotations.

We explored two prompting approaches: **zero-shot** and **few-shot**. In the zero-shot setting, the model receives only task instructions. In the few-shot setting, the prompt also includes several labeled examples with brief explanations. These examples show the model how to identify narratives in practice.

To evaluate the performance of this method and identify the optimal price-quality ratio for the task, the following LLMs were selected: `GPT-4.1`[11], `GPT-4o-mini`[12], `Gemini-2.5-flash`[13], `Claude-sonnet-4-20250514`[14].

3.4 Share Graph Construction

We model message diffusion using two complementary layers: (i) an explicit channel-level share graph derived from Telegram forwarding metadata, and (ii) a semantic message-level graph designed to recover repost-like diffusion that is not marked as a native forward.

Forward-based channel share graph. Let $G_{\text{share}} = (V_C, E_F, w_F)$, where V_C is the set of observed Telegram channels. We add a directed edge $(u \to v) \in E_F$ when channel v contains a message forwarded from channel u. Edge weights count observed forwarding events:

$$w_F(u, v) = \sum_{m \in M_v} \mathbf{1}[\texttt{fwd\_from\_channel}(m) = u] \,,$$

where M_v denotes the set of messages posted in channel v. In implementation, we retain a forward edge only when the referenced source message is present in the collected corpus. This restriction avoids links to unobserved sources and ensures that all edges in G_{share} are supported by directly observed data.

Semantic message graph for repost-like diffusion. Explicit forwarding provides a high-precision, platform-native signal of diffusion, but it does not capture copy-paste reposts or lightly edited message reuse. To approximate such hidden propagation, we construct a message-level semantic graph with edges of type `SIMILAR_TO` marked in the dataset.

Before embedding, we normalize message text and exclude duplicates attributable to explicit forwarding chains. We then encode the remaining Ukrainian- and Russian-language messages using `baai/bge-m3` and retrieve approximate nearest neighbors with an HNSW index (Malkov and Yashunin, 2018).

For each message, we retrieve the top-k neighbors with $k = 80$, $M = 48$, $ef_{\text{construction}} = 200$, and $ef_{\text{search}} = 200$; these settings preserved recall for subtle multilingual paraphrases while keeping index size and query latency tractable, with only marginal gains beyond them. We retain only high-confidence cross-channel pairs above a calibrated cosine threshold $\tau = 0.985$, excluding same-channel pairs and pairs already linked by explicit forwarding metadata. We selected τ using 3,215 mapped hidden-forward pairs: recall was 81.0% at 0.98, 76.39% at 0.985, and 69.58% at 0.99, while the number of retained pairs after filtering decreased from 35,092 to 24,692 and 19,095, respectively, providing a practical balance between recall and edge inflation.

We selected `baai/bge-m3`, `intfloat/multilingual-e5-large`, and `gemini-embedding-001` as pilot candidates because all three are strong multilingual embedding models suitable for cross-lingual semantic retrieval. In pilot calibration on manually inspected Ukrainian/Russian message pairs, `baai/bge-m3` showed the clearest separation among these candidates.

[10] https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7

[11] https://developers.openai.com/api/docs/models/gpt-4.1

[12] https://developers.openai.com/api/docs/models/gpt-4o-mini

[13] https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash

[14] https://platform.claude.com/docs/en/about-claude/models/overview

In a 16-pair pilot containing exact-like positives and manually identified hard negatives, the hard-negative pairs had a mean cosine similarity of 0.328 under `baai/bge-m3`, compared with 0.744 for `intfloat/multilingual-e5-large` and 0.762 for `gemini-embedding-001`. We therefore used `baai/bge-m3` for the full embedding run.

4 Experiments, Evaluation, Results

We evaluate the proposed framework in three stages. First, we measure IAA to verify the reliability of the manually labeled data. Second, we compare the weak labeling strategies from Section 3.3 and identify the most effective setup for large-scale narrative assignment. Third, using the selected labels, we analyze the diffusion graphs to study source behavior, cross-group sharing, multi-hop propagation, and narrative-level dissemination patterns. This organization allows us to move from label quality to structural findings and to assess whether combining weak supervision with graph analysis yields an informative view of the spread of disinformation narratives on Ukrainian Telegram.

4.1 Annotator Agreement

To evaluate the quality of the manual annotation, we computed **Cohen's Kappa**, **Fleiss' Kappa**, and **Krippendorff's Alpha**. Cohen's Kappa measures agreement between pairs of annotators while accounting for chance agreement. Fleiss' Kappa generalizes this measure to multiple annotators and categorical data. Krippendorff's Alpha provides a flexible measure of inter-annotator reliability that can handle multiple annotators, missing data, and different data types.

The results are presented in Table 1. Binary classification achieved near-perfect agreement ($\kappa/\alpha \approx 0.887$), while meta-narrative labeling reached substantial agreement (≈ 0.719). Fine-grained narrative annotation showed moderate agreement (≈ 0.626), suggesting that this task is at the narrative level relatively subjective and harder to annotate reliably. Notably, all three metrics (Cohen's Kappa, Fleiss' Kappa, and Krippendorff's Alpha) produce almost identical within each category, which strongly confirms the reliability and consistency of these measurements.

Metric	Narrative	Meta-narrative	Binary
Cohen's Kappa (mean)	0.626	0.719	0.887
Fleiss' Kappa	0.626	0.718	0.887
Krippendorff's Alpha	0.627	0.719	0.887

Table 1: Inter-annotator agreement results.

4.2 Weak Labeling Results

The models were executed to evaluate each labeling strategy across the taxonomy. Since each sub-narrative is linked to a specific narrative and meta-narrative, identifying a sub-narrative automatically determines its higher-level categories. This mapping allows predictions to be evaluated at all levels of the hierarchical structure.

Due to label imbalance (i.e., a majority of non-narrative messages and an uneven distribution of specific narratives), we focused on metrics robust to skewed data: **F1** for binary classification (whether a message contains a narrative), **Weighted F1** for multi-class classification, **Matthews Correlation Coefficient (MCC)** for both.

First, we evaluated the models used in the **semantic similarity**, **multilingual NLI**, and **LLM zero-shot prompting** strategies. However, their performance on *the shared subset* of 103 messages was relatively limited (see Table 4.2 for detailed results).

We also explored **ensemble approaches** that combined predictions from multiple LLMs. In particular, we tested ensembles consisting of `Claude`, `GPT-4.1`, and `Gemini`, as well as `GPT-4o-mini`, `GPT-4.1`, and `Gemini`, assigning a higher weight to Gemini due to its stronger individual performance. However, these ensemble configurations did not lead to a meaningful improvement in overall results (as Gemini consistently outperformed them across all evaluated metrics).

We found that the main difficulty is that the relationship between a message and a narrative is often **implicit**. As a result, the message may not be semantically similar to the corresponding narrative description. Instead, the narrative needs to be inferred from the common (sometimes political) context.

Because of this, the **semantic similarity** and **multilingual NLI** approaches were not suitable for our task. We therefore modified the LLM-based strategy by using **few-shot prompting**, adding several annotated examples to the prompt to illustrate the implicit relationship between messages and narratives. This approach produced significantly better results. To ensure that the method performs consistently on a larger sample, we repeated the evaluation on the full set of 412 manually annotated messages. The results remained strong ($F_1 = 0.82$,

$\text{MCC}_{\text{bin}} = 0.71$, W-$F_1 = 0.76$, $\text{MCC}_{\text{meta}} = 0.57$). The final version of the prompt used in the experiments is provided in Appendix E.

Classifier	Binary		Meta	
	F1	MCC	W-F1	MCC
Semantic Similarity				
MiniLM-L12-v2*	0.51	0.14	0.41	0.11
BGE-M3*	0.37	0.08	0.54	0.13
OpenAI Text-3*	0.41	-0.07	0.36	0.03
Multilingual NLI				
mDeBERTa-v3-base*	0.54	0.15	0.32	0.13
LLM (Zero-Shot)				
GPT-4o-mini*	0.51	0.23	0.53	0.17
GPT-4.1*	0.65	0.60	0.69	0.50
Gemini-2.5-Flash*	0.76	0.63	0.74	0.52
Claude-Sonnet-3.5*	0.60	0.55	0.67	0.40
LLM (Ensemble)				
GPT-4.1 + Claude + Gemini*	0.69	0.61	0.72	0.50
GPT-4.1 + GPT-4o-mini + Gemini*	0.71	0.61	0.70	0.44
LLM (Few-Shot)				
Gemini-2.5-Flash*	**0.85**	**0.78**	**0.80**	**0.62**
Gemini-2.5-Flash**	**0.82**	**0.71**	**0.76**	**0.57**

Table 2: LM-based approaches significantly outperform Semantic Similarity and NLI baselines, with **Gemini-2.5-Flash (Few-Shot)** achieving the highest overall performance. Results marked with (*) denote the *shared subset*, while (**) represents the full *golden set*. Meta denoted Meta-narrative.

4.3 Share Graph Results

We next analyze the share graph to understand what its structure reveals about how narratives spread across channels. Our analysis focuses on three questions: which channels are likely to occupy upstream positions, how often sharing occurs across channel groups, and how these patterns differ between the explicit-forwarding and semantic-similarity layers. Formal definitions of the metrics and additional robustness details are provided in Appendix A.

Narrative source detection. We first examine whether explicit forwarding structure can identify likely upstream sources of narrative dissemination. For this analysis, a channel is classified as *narrative-active* if it satisfies two conditions: (i) it has at least 50 labeled messages, and (ii) at least 60% of those labeled messages receive a narrative label. Under this rule, 41 of the 98 observed channels are classified as narrative-active. The remaining 57 channels are treated as non-narrative for this grouping analysis, including 25 channels that meet the minimum label-count requirement but fall below the 60% narrative threshold, and 32 channels with limited evidence.

To characterize source behavior in the explicit forwarding graph, we use three complementary channel-level metrics: total forwarding volume (*spread_events*), the number of distinct recipient channels reached (*downstream_channels*), and a normalized source tendency score (*source_share*) that distinguishes net sources from channels that mostly relay content from others.

Across the 41 narrative-active channels, we observe 3,016 explicit spread events. By forwarding volume, `rian_ru` is the dominant source with 612 forwarding events, accounting for 20.3% of all spread events in this subset. It is followed by `depzdravzo` (337), `Pukhov_M` (303), `hersonruss` (228), and `readovkanews` (209). `rian_ru` also ranks first in dissemination breadth, reaching 17 distinct downstream channels, and has a source-share score of 1.00, indicating a purely source-like position in the observed forwarding network. These results show that the explicit share graph provides a useful signal for identifying likely upstream narrative spreaders, while still reflecting source positions only within the collected network rather than absolute first creators outside it.

Explicit graph hop calibration and sensitivity. We selected the hop budget by examining the shortest-path distribution over reachable ordered pairs. The smallest value covering at least 85% of reachable ordered pairs is $h = 4$, which covers 92.28% of them. At $h = 4$, overall cross-group reachability is 0.0211, with reachability of 0.0153 from narrative-active to non-narrative channels and 0.0268 in the reverse direction. Increasing the hop budget to $h = 8$ raises reachability from non-narrative to narrative-active channels to 0.03, while reachability from narrative-active to non-narrative channels remains 0.0153.

Cross-group sharing analysis. We next examine if explicit forwarding crosses channel groups defined by channel-level narrative presence. To quantify this, we compare within-group and cross-group forwarding shares and then examine whether

cross-group contact becomes more common when short multi-hop paths are allowed.

Direct cross-group forwarding is rare: among 34,967 forwarding events, only 121 cross group boundaries, corresponding to a share of 0.0035. These events are limited in both directions, with 57 events from narrative-active to non-narrative channels and 64 in the reverse direction. Allowing short multi-hop paths does not substantially change this conclusion. Using a hop budget calibrated on the shortest-path distribution, overall cross-group reachability in the explicit forwarding graph is only 0.02, with somewhat higher reachability from non-narrative channels into the narrative-active group than in the reverse direction. Overall, most forwarding remains in-group, and bridge-like cross-group routes are uncommon.

Semantic graph flow and reachability details. In the semantic similarity graph, message-level semantic links are oriented by time and aggregated into channel-level semantic flow counts, where $\mathcal{W}^{\text{sem}}_{u \to v}$ denotes the number of semantic propagation events from channel u to channel v. Direct cross-group semantic flow is summarized using the same within-group and cross-group share logic as in the explicit forwarding graph. For indirect connectivity, we calibrate the hop budget on the shortest-path distribution over reachable ordered channel pairs. The smallest cutoff covering at least 85% of reachable ordered pairs is $h = 3$, corresponding to 85.65% coverage. At $h = 3$, overall cross-group reachability is 0.5385, with reachability of 0.5049 from narrative-active to non-narrative channels and 0.5721 in the reverse direction. A larger hop budget, $h = 8$, increases these values to 0.669 and 0.6387, respectively, with 0.6538 overall.

Cross-group sharing in the semantic similarity graph. To complement the explicit forwarding analysis, we examine cross-group diffusion in the semantic similarity graph, where message-level semantic links are time-oriented and aggregated to channel-level semantic flows. We then compare within-group and cross-group semantic flows and assess whether the two groups are connected via short multi-hop paths.

Direct cross-group semantic flow remains limited but is more common than in the explicit forwarding graph. Among 52,414 semantic flow events, 1,078 cross group boundaries, corresponding to a share of 0.02. The directional split is also asymmetric: 256 events run from narrative-active to non-narrative channels, while 822 run in the reverse direction. The main contrast appears in the multi-hop setting. At hop budget $h = 3$, overall cross-group reachability rises to 0.54, far above the corresponding value in the explicit forwarding graph. This suggests a repost-like narrative transfer frequently connects the two groups through intermediate channels even when direct cross-group links remain relatively uncommon.

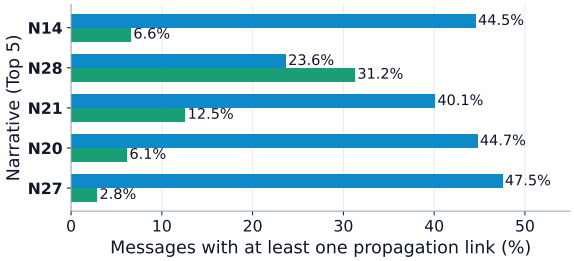

Figure 3: Coverage-based propagation comparison for the five most frequent narratives: N14 (Discrediting or ridiculing representatives of Ukrainian authorities), N28 (Russia improves life in occupied territories), N21 (Ukraine's victory is impossible), N20 (The West controls Ukraine and uses it for its own goals), and N27 (Discrediting the EU and the West). Blue bars show the share of messages with at least one explicit forward (`FORWARD_FROM`); green bars show the share of messages with at least one semantic forward (`SIMILAR_TO`).

4.4 Narrative Distribution

We next examine narrative distribution in the labeled graph scope. Of the 81,369 labeled messages, 29,649 (36%) receive a narrative assignment. Figure 3 compares propagation coverage for these five narratives across the explicit and semantic graph layers. At the macro-family level, the distribution is dominated by *Discrediting Ukraine and its institutions* (33%), followed by *Narratives about Russian welfare and "liberating" role* (17.2%) and *Discrediting the EU and the West* (14.2%) (See Appendix D). This suggests that narrative prominence in the corpus and propagation strength do not align uniformly across diffusion layers (explicit forward and semantic similar forward).

5 Discussion and Conclusion

This work introduces a graph-based framework for detecting and analyzing disinformation narratives in Telegram ecosystems by combining weak supervision with propagation graph analysis. Modeling disinformation at the narrative level enables the

system to capture semantically related claims that appear in different linguistic forms, addressing the challenge of repeated paraphrasing and reposting common in Telegram channels. The propagation graph further provides a structural perspective on information spread, revealing how narratives circulate across interconnected channels rather than appearing as isolated posts. Graph representation highlights clusters of channels that repeatedly amplify similar narratives, offering insights into how disinformation spreads through platform-specific communication patterns. Beyond improving large-scale detection, this approach provides a framework for studying narrative diffusion and monitoring emerging disinformation campaigns in rapidly evolving online environments such as Telegram, where traditional post-level analysis often fails to capture broader information dynamics.

Limitations

Our approach has several limitations. First, the framework relies on weak supervision for narrative labeling, which can introduce label noise. While this enables scalable annotation of large Telegram datasets, automatically generated labels may include ambiguous or incorrectly assigned instances, potentially affecting downstream analysis. In the study, we estimated the amount of errors which one can expect from our best-performing solution. However, on different channels and over time the performance might degrade or perform better.

Second, the narrative taxonomy used for labeling may be incomplete and subject to drift over time. As new narratives emerge or existing ones evolve, the predefined taxonomy may fail to capture all relevant claims. In our experiments, we observe indications of narrative drift when additional narratives are introduced, suggesting that narrative boundaries are dynamic and context-dependent.

Third, the dataset is limited to a selected set of Telegram channels, which may introduce sampling bias, especially as we only subsample from the messages that have forwarding connections. The analyzed network may therefore not fully represent the broader Telegram information ecosystem nor the channels where they come from.

Finally, we do not investigate variability of Ukrainian language and code-switch varieties, and how performance of narrative detection decreases on such instances, while this is expected behaviour based on prior research Shynkarov et al. (2025).

Ethical Considerations

Our analysis relies on publicly accessible Telegram channels. Although these data are publicly available, users may not anticipate their messages being analyzed in large-scale computational studies. To mitigate potential privacy risks, we focus on aggregate patterns of narrative propagation and big public channels rather than individual user behavior.

Automated systems for detecting disinformation may produce false positives or misclassify legitimate content. Such errors could contribute to unfair labeling of channels or narratives if used without appropriate human oversight. Our framework is intended as a research tool for analyzing information dynamics rather than as a standalone moderation system.

Methods for detecting narrative propagation may also inform adversarial actors about how such systems operate. Increased awareness of detection strategies could encourage actors to adapt their communication patterns to evade analysis.

Using an external narrative inventory (VoxCheck) embeds expert judgments about what constitutes a disinformation narrative. While grounded in professional fact-checking practice, such inventories reflect particular epistemic and institutional perspectives and may not capture all interpretations of contested claims.

Acknowledgements

This research was partially supported by ELEKS through a grant dedicated to the memory of Oleksiy Skrypnyk. The work on this paper is partially performed in the scope of the project “VeraXtract” (16IS24066) funded by the German Federal Ministry for Research, Technology and Aeronautics (BMFTR).

References

Apaar Bawa, Ugur Kursuncu, Dilshod Achilov, Valerie L. Shalin, Nitin Agarwal, and Esra Akbas. 2025. Telegram as a Battlefield: Kremlin-Related Communications During the Russia-Ukraine Conflict. *Proceedings of the International AAAI Conference on Web and Social Media*, 19(1):2361–2370.

Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. *Preprint*, arXiv:2001.06362.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, page 675–684, New York, NY, USA. Association for Computing Machinery.

Michael F. Dahlstrom. 2014. Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences*, 111(Supplement 4):13614–13620.

Maria Hellman. 2024. *Narrative Analysis and Framing Analysis of Disinformation*, pages 101–121. Springer Nature Switzerland, Cham.

Lu Kalkbrenner, Veronika Solopova, Steffen Zeiler, Robert Nickel, and Dorothea Kolossa. 2025. MisinfoTeleGraph: Network-driven misinformation detection for German telegram messages. In *Proceedings of the The 9th Workshop on Online Abuse and Harms (WOAH)*, pages 179–191, Vienna, Austria. Association for Computational Linguistics.

Massimo La Morgia, Alessandro Mei, and Alberto Maria Mongardini. 2025. Tgdataset: Collecting and exploring the largest telegram channels dataset. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, page 2325–2334, New York, NY, USA. Association for Computing Machinery.

João A. Leite, Olesya Razuvayevskaya, Kalina Bontcheva, and Carolina Scarton. 2024. Euvsdisinfo: A dataset for multilingual detection of pro-kremlin disinformation in news articles. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, CIKM '24, page 5380–5384, New York, NY, USA. Association for Computing Machinery.

Mykola Makhortykh, Aytalina Kulichkina, and Kateryna Maikovska. 2025. Evolution of wartime discourse on telegram: A comparative study of ukrainian and russian policymakers' communication before and after russia's full-scale invasion of ukraine. *Preprint*, arXiv:2510.11746.

Yu. A. Malkov and D. A. Yashunin. 2018. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *Preprint*, arXiv:1603.09320.

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. 2019. Fake News Detection on Social Media using Geometric Deep Learning. *Preprint*, arXiv:1902.06673.

Nikolaos Nikolaidis, Nicolas Stefanovitch, Purificação Silvano, Dimitar Iliyanov Dimitrov, Roman Yangarber, Nuno Guimarães, Elisa Sartori, Ion Androutsopoulos, Preslav Nakov, Giovanni Da San Martino, and Jakub Piskorski. 2025. PolyNarrative: A multilingual, multilabel, multi-domain dataset for narrative extraction from news articles. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 31323–31345, Vienna, Austria. Association for Computational Linguistics.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937, Copenhagen, Denmark. Association for Computational Linguistics.

Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. 2017. Snorkel: rapid training data creation with weak supervision. *Proc. VLDB Endow.*, 11(3):269–282.

Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 255–269, Online. Association for Computational Linguistics.

Yurii Shynkarov, Veronika Solopova, and Vera Schmitt. 2025. Improving sentiment analysis for Ukrainian social media code-switching data. In *Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025)*, pages 179–193, Vienna, Austria (online). Association for Computational Linguistics.

Max Upravitelev, Veronika Solopova, Charlott Jakob, Premtim Sahitaj, Sebastian Möller, and Vera Schmitt. 2026a. Retrieving climate change disinformation by narrative. *Preprint*, arXiv:2603.22015.

Max Upravitelev, Veronika Solopova, Charlott Jakob, Premtim Sahitaj, Sebastian Möller, and Vera Schmitt. 2026b. Retrieving climate change disinformation by narrative. *Preprint*, arXiv:2603.22015.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Claire Wardle and Hossein Derakhshan. 2017. Information disorder: Toward an interdisciplinary framework for research and policy making. Technical report, Council of Europe.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3914–3923, Hong Kong, China. Association for Computational Linguistics.

A Details for Share-Graph Results

Source metrics in the explicit forwarding graph. Let $W_{u\to v}$ denote the number of explicit forwarding events in which channel v forwards content originally posted by channel u. We summarize each channel's source behavior using three forwarding-based metrics:

$$\begin{aligned} \textit{spread\_events}(u) &= \sum_{v\neq u} W_{u\to v}, \\ \textit{downstream\_channels}(u) &= \left|\{v \neq u : W_{u\to v} > 0\}\right|, \\ \textit{source\_share}(u) &= \frac{\sum_{v\neq u} W_{u\to v}}{\sum_{v\neq u}(W_{u\to v} + W_{v\to u})}. \end{aligned}$$

The first metric captures total propagation volume, the second captures dissemination breadth across distinct recipient channels, and the third measures net-source tendency on a bounded $[0, 1]$ scale.

Cross-group forwarding metrics. To quantify within-group versus cross-group forwarding in the explicit share graph, we compute

$$\begin{aligned} \textit{same\_label\_share} &= \frac{\sum_{u\neq v} W_{u\to v}\,\mathbf{1}[y(u) = y(v)]}{\sum_{u\neq v} W_{u\to v}}, \\ \textit{cross\_label\_share} &= 1 - \textit{same\_label\_share}. \end{aligned}$$

where $y(u)$ denotes the group label of channel u.

Cross-group reachability definition. For ordered channel pairs (u, v), let $d(u, v)$ be the directed hop distance from u to v. We define cross-group reachability under hop budget h as

$$\begin{aligned} R_{\text{cross}}(h) &= \frac{|\{(u, v) \in C : d(u, v) \leq h\}|}{|C|}, \\ C &= \{(u, v) : y(u) \neq y(v)\}. \end{aligned}$$

B Cross-Group Shares

Table 3 summarizes the direction of cross-group events, separately for explicit forwarding and semantic similarity.

Additional examples of such messages, along with their detected narratives and sub-narratives, are presented in Table 5 for cross-group explicit-forwarding from narrative-active to non-narrative channels, Table 6 for cross-group explicit-forwarding from non-narrative to narrative-active channels, Table 7 for cross-group semantic-similarity from narrative-active to non-narrative channels, and Table 8 for cross-group semantic-similarity from non-narrative to narrative-active channels.

C Chronological List of Events

A chronological list of events corresponding to markers in Figure 2 is shown in Table 4.

From	To	Events	Share
Explicit forwarding			
Narrative-active	Non-narrative	57	0.4711
Non-narrative	Narrative-active	64	0.5289
Semantic similarity			
Narrative-active	Non-narrative	256	0.2375
Non-narrative	Narrative-active	822	0.7625

Table 3: Summary of cross-group events for explicit forwarding and semantic similarity.

D Narrative Distribution

Supplementary visualizations for the narrative distribution analysis discussed in Section 4.4. Table 4 shows narrative distribution at the meta-narrative level.

E Final LLM Prompt

This appendix provides both the original Ukrainian version (Table 9) and the English translation (Table 10) of the final prompt used in the LLM-based weak labeling approach. The prompt was executed using `Gemini-2.5-flash`.

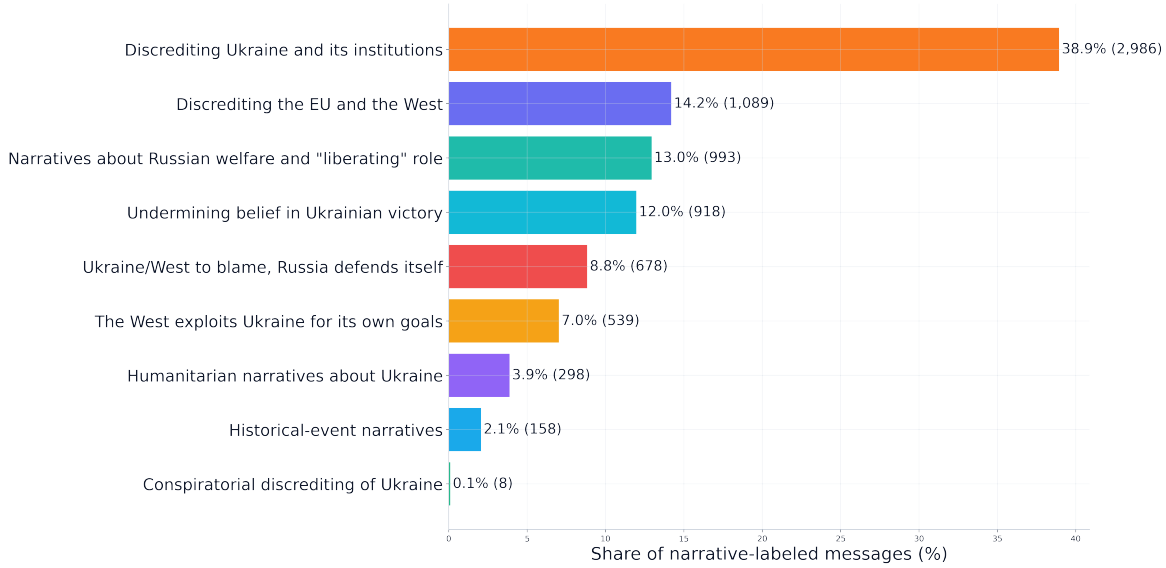

Figure 4: Narrative distribution at the meta-narrative level.

Date	Event
2024-12-19	EU agreed on a €90B loan programme for Ukraine (2026–2027).
2024-12-22	Slovak PM Fico visits Moscow; protests begin.
2025-01-10	Large-scale protests in Slovakia.
2025-01-24	New protests following "attempted coup" statements.
2025-02-28	Ukraine-US meeting at White House on strategy.
2025-03-02	London Summit on Ukraine.
2025-03-21	German Bundestag approves €3B military aid.
2025-06-13	Iran's massive missile strike on Israel.
2025-06-22	US air strikes on Iranian nuclear facilities.
2025-06-23	Iran attacks US base Al-Udeid in Qatar.
2025-07-10	EU announces €2.3B in reconstruction funding.
2025-07-18	EU adopts 18th package of sanctions.
2025-08-25	EU announces additional €4B in aid.
2025-10-23	EU adopts 19th package of sanctions.
2025-10-29	Updated EU-Ukraine trade agreement in force.
2025-12-15	European leaders propose peace plan.
2026-01-06	Declaration for multinational security force.
2026-01-14	EC proposes €90B in aid for 2026–2027.
2026-01-23	1st round of negotiations (Abu Dhabi).
2026-02-04	2nd round of negotiations (Abu Dhabi).
2026-02-15	Munich Security Conference 2026.
2026-02-23	Hungary blocks new EU sanctions/aid.

Table 4: Chronological list of events corresponding to markers in Figure 2.

Message (English translation)	Share count	Narrative	Sub-narrative
In Kremenchuk, TCC officers chased a young man and deliberately crashed into parked cars. The video shows how the TCC men deliberately ram the car, because they know that the law does not exist for them. A day before that, these same draft officers in uniform opened fire. The head of the Main Directorate of the National Police in Poltava oblast remains silent.	1	Discrediting the Ukrainian army	Mobilization in Ukraine violates all standards
The Wall Street Journal, citing sources in the American administration: the United States has removed the key restriction on Ukraine's use of Western long-range missiles, which will allow Kyiv to strike deep into Russia. This shift coincided with President Trump's attempt in early October to pressure the Kremlin to begin negotiations on ending the war.	1	The West controls Ukraine and uses it for its own goals	Ukraine is an instrument of the United States
Trump has become more detached on the issue of the Ukrainian conflict and no longer lashes out at Russia and Ukraine over the lack of progress in the negotiations, ABC reports, citing American officials.	1	Ukraine and the West refuse to start peace negotiations	Ukraine must immediately begin peace negotiations and make compromises with Russia

Table 5: Representative cross-group explicit-forwarding examples from narrative-active to non-narrative channels.

Message (English translation)	Share count	Narrative	Sub-narrative
It feels as though the wrong people were called occupiers. This is how the TCC tried to abduct a serf in front of his pregnant wife and children. People fought them off, but what moral freaks they are. Ze mobilization is the genocide of the Ukrainian people.	3	Discrediting or ridiculing representatives of Ukrainian authorities	Zelensky wants to sacrifice Ukrainians for his own interests
To call Putin an old fantasist while having 78-year-old Trump next to him is a brilliant move. I love it when this idiot buries himself.	1	Discrediting or ridiculing representatives of Ukrainian authorities	Ridiculing representatives of the Ukrainian authorities
"Ukraine is an artificially created country. When Putin takes Odesa, everything will be fine for us," just a survey on the streets of Odesa. More than 3 years of full-scale war. What is in their heads?	1	Narrative about historical events	Ukraine developed as an artificial state

Table 6: Representative cross-group explicit-forwarding examples from non-narrative to narrative-active channels.

Message (English translation)	Share count	Narrative	Sub-narrative
War has been declared today on the Russian world — Vladimir Putin.	2	Actions of Ukraine and the West forced Russia to start the war	Russia was forced to enter the war to ensure its survival
Putin supported Trump's idea of a mutual refusal by Russia and Ukraine for 30 days to strike energy infrastructure and gave such an order to the military — Kremlin.	2	Ukraine and the West refuse to start peace negotiations	Unlike Ukraine, Russia demonstrates readiness for negotiations
Poland does not want simply to be in solidarity with Ukraine; it wants to profit from it. "We will no longer help in a naive way. It will not be that Poland is solidaristic and others profit from the reconstruction of Ukraine," said Polish Prime Minister Tusk.	2	The West controls Ukraine and uses it for its own goals	Western elites are the main beneficiaries of the war

Table 7: Representative cross-group semantic-similarity examples from narrative-active to non-narrative channels.

Message (English translation)	Share count	Narrative	Sub-narrative
In Pokrovsk, almost 15 million were allocated to support the media, 73 million to the water utility, 34 million to beautification, and almost 30 million to the functioning of heat energy, — Telegraf, citing the community budget for 2025. Fifty million is provided for officials' salaries, 98 million for the elimination of emergency situations, and 147 million for education. According to DeepState, the Russian Armed Forces are already 2–3 km from the city. As of the end of the year, 70% of residential buildings, 80% of social facilities, and 95% of industrial facilities had been damaged or destroyed. There is no electricity, gas, heating, or water supply in the city.	3	Discrediting or ridiculing representatives of Ukrainian authorities	Corruption in Ukraine's political and military leadership
Slovakia threatens Ukraine with stopping electricity supplies because of the halt in gas supplies. "If necessary, we will stop electricity supplies, which Ukraine critically needs during power outages," Fico said.	2	Ukraine is left alone and will lose	Partners abandoned Ukraine
The Biden administration informed Trump's team that Ukraine will have to "resolve the issue of lowering the mobilization age," — Sullivan. In his opinion, the shortage of people remains an acute problem while the United States is providing "a huge amount of ammunition and military equipment."	2	The West controls Ukraine and uses it for its own goals	Ukraine is an instrument of the United States

Table 8: Representative cross-group semantic-similarity examples from non-narrative to narrative-active channels.

	Ukrainian version
System instructions	Ти уважний український класифікатор. Поверни ЛИШЕ валідний JSON. Без зайвого тексту. НЕ вигадуй. НЕ використовуй markdown. НЕ перекладай і не перефразовуй текст наративів. У відповіді повертай ТІЛЬКИ narrative_id та sub_narrative_id (без тексту наративів).
Task instructions	Завдання: Визнач, чи повідомлення ПРЯМО або НЕПРЯМО ('натякаючи') просуває будь-який наратив зі списку. Правила: • Спочатку вирішi, чи є хоча б один відповідний наратив. • Відповідай «так» лише якщо твердження у повідомленні прямо підтримує наратив. • Якщо це загальні новини або нейтральний опис подій — НЕ став жодних наративів. • Може бути таке, що повідомлення може містити кілька наративів. Обери той, що найбільше підходить, або найбільш чіткий. • Читай уважно. Дивись на повідомлення з української перспективи. • Використовуй ТІЛЬКИ narrative_id та sub_narrative_id зі списку. • Якщо жоден наратив не підходить — поверни null для narrative_id і sub_narrative_id та confidence 0. • Додай confidence від 0 до 1. • Вивід має бути валідним JSON. • Наприклад: – "кличко заявив що україна стикається з браком солдатів і пропонує знизити вік мобілізації деталі за посиланням" – це повідомлення нав'язує думку, що Кличко хоче знизити мобілізаційний вік. Це страшить громадян і у них виникає погана думка про політика Кличка, тому тут є наратив "Дискредитація чи висміювання представників української влади" – "мужики красавчики понимая что ударные дроны вылетают з контейнера на прицепе фуры начали забрасывать его камнями рискуя своей жизнью конечно они понимали об опасности было бы хорошо если бы их нашли и наградили просто красавчики neoficialniybezsonov" – повідомлення стимулює 'захоплюватись' сміливістю росіян і хоче переконати, що перемога росіян однозначна, тому тут присутній наратив "Перемога України неможлива"

Table 9: Original version of final prompt.

	English version
System instructions	You are a diligent Ukrainian classifier. Return ONLY valid JSON. No extra text. Do NOT make anything up. Do NOT use Markdown. Do NOT translate or paraphrase the narrative text. In your response, return ONLY the narrative_id and sub_narrative_id (without the narrative text).
Task instructions	Task: Determine whether the message DIRECTLY or INDIRECTLY ('by implication') promotes any of the narratives listed below. Rules: • First, decide whether there is at least one relevant narrative. • Answer 'yes' only if the statement in the post directly supports the narrative. • If it is general news or a neutral description of events — DO NOT assign any narratives. • It may be the case that a post contains several narratives. Choose the one that is most appropriate or the clearest. • Read carefully. View the post from a Ukrainian perspective. • Use ONLY narrative_id and sub_narrative_id from the list. • If no narrative is suitable — return null for narrative_id and sub_narrative_id and confidence 0. • Set confidence to a value between 0 and 1. • The output must be valid JSON. • For example: – "Klitschko has stated that Ukraine is facing a shortage of soldiers and is proposing to lower the conscription age – details via the link" – this message seeks to suggest that Klitschko wants to lower the conscription age. This frightens citizens and gives them a negative impression of the politician Klitschko; therefore, there is a narrative "Discrediting or ridiculing representatives of the Ukrainian authorities" – "A bunch of cool guys, realising that attack drones were taking off from a container on the back of a lorry, started pelting it with stones, risking their lives. Of course, they knew the danger involved. It would have been great if they'd been found and rewarded – just a bunch of cool guys. neoficialniybezsonov" – The post encourages people to 'marvel' at the Russians' bravery and aims to convince them that a Russian victory is inevitable, so the narrative here is "A Ukrainian victory is impossible"

Table 10: English version of the prompt used for weak labeling (translated from Ukrainian via DeepL). To ensure reproducibility, please refer to the original Ukrainian prompt in the Table 9.