A Longitudinal, Multinational, and Multilingual Corpus of News Coverage of the Russo-Ukrainian War

Dikshya Mohanty[1], Taisiia Sabadyn[1], Jelwin Rodrigues[1], Chenlu Wang[1], Abhishek Kalugade[1], Ritwik Banerjee[1,2]
[1]Department of Computer Science, [2]AI Innovation Institute
Stony Brook University, New York
{dimohanty, jelrodrigues, chenlwang, rbanerjee}@cs.stonybrook.edu
{taisiia.sabadyn, abhishek.kalugade}@stonybrook.edu

Abstract

We present DNIPRO, a corpus of 246K news articles from the Russo-Ukrainian war (Feb 2022 – Aug 2024) spanning eleven outlets across five nation-states (Russia, Ukraine, U.S., U.K., China) and three languages. The corpus features comprehensive metadata and human-evaluated annotations for stance, sentiment, and topical framing, enabling systematic analysis of *competing geopolitical narratives*. It is uniquely suited for empirical studies of narrative divergence, media framing, and information warfare. Our exploratory analyses reveal how media outlets construct incompatible realities through divergent attribution and topical selection without direct refutation of opposing narratives. DNIPRO empowers empirical research on narrative evolution, cross-lingual information flow, and computational detection of implicit contradictions in fragmented information ecosystems.

Keywords: multilingual corpus, media framing, geopolitical narratives, longitudinal analysis, information warfare

1. Introduction

The Russo-Ukrainian war, which escalated on February 24, 2022 with the declaration of a "special military operation" by the Russian government,[1] has significantly impacted the global geopolitical landscape (Derviş, 2023; Silvertsen, 2024). The war has redrawn traditional alliances, tested the resolve of international institutions, and created new geopolitical fault lines. Despite initial expectations of a swift Russian victory, Ukraine's resistance has resulted in a protracted war with substantial loss of life and human capital (Watling et al., 2023; Pancevski, 2024). This conflict has also been characterized by a quieter battle for narrative control, with news coverage from several nations playing a pivotal role in shaping global public opinion and influencing the war's trajectory (Bradshaw et al., 2024). These sources reflect their national priorities, strategic perspectives, and the perceptions they intend to create among information consumers (Johais and Meis, 2024). Yet, despite the centrality of media and narrative framing, there remains a lack of comprehensive datasets for longitudinal analysis of how nations have framed the war for their domestic and international audiences.

We introduce a collection of **D**IVERSE **N**ARRATIVES AND **I**NTERNATIONAL **P**ERSPECTIVES ON THE **R**USSO-**U**KRAINIAN **O**FFENSIVE (DNIPRO), a comprehensive longitudinal corpus of news coverage from February 1, 2022 to August 31, 2024 — the longest time span of any multinational or multilingual news corpus on the Russia-Ukraine war. The code-base and supporting materials are publicly available. The corpus encompasses eleven influential media outlets and news agencies aligned with five primary geopolitical actors with strategic interests in the conflict. Following methodological approaches established by Nygren et al. (2016) and Roman et al. (2017), the sources include media from direct participants—Russia and Ukraine—and influential global powers with economic, diplomatic, or military involvement: the USA, the UK, and China. The inclusion of state-affiliated media from Russia and Ukraine captures opposing conflict narratives that play a pivotal role in shaping assessments under uncertain information (Saletta et al., 2020). Media from China, the USA, and the UK represent perspectives of nations with substantial geopolitical and economic interests in the conflict's trajectory, as well as the military, diplomatic, and economic leverage to influence its outcome (Zhang et al., 2016; Thussu et al., 2017). These diverse perspectives and sources enable more comprehensive analysis of complex developments by helping analysts overcome the limitations of single-narrative framings (Werd, 2018).

DNIPRO offers rich annotations for topical framing, entity-level stance, and entity-level sentiment, enabling event tracking and other *longitudinal* analyses across geopolitical perspectives. The dataset does not attempt to verify the accuracy of any claims; rather, it provides a resource to study how wartime narratives are constructed and disseminated by major geopolitical actors.[2] The corpus

[1]Within 48 hours, media in Russia were banned from describing the developments as "war" (AFP, 2022).

[2]This aligns with established methodologies in discourse analysis (Entman, 1993; Fairclough, 2010) exam-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6452–6471
11–16 May 2026.

is uniquely suited to such studies, as it comprises 246,229 articles across 11 media sources and 3 languages, spanning 5 diverse nations, over a period of 31 months. Our **key contributions** include:

(1) a longitudinal, multinational, and multilingual corpus of news coverage spanning the first 31 months of the Russia-Ukraine war;
(2) inclusion of diverse media, enabling cross-cultural examination of conflict narratives; and
(3) annotations for comparative analyses of named entities, sentiment, stance, and topical framing.

Next, we place DNIPRO within the landscape of existing conflict-related corpora to highlight its distinct contributions to the field (§2), and describe how our design and construction choices (§3) make it a powerful resource for large-scale analyses of wartime narratives. We employ DNIPRO in several applications, using state-of-the-art natural language understanding systems and additional annotations (§4 and §5), demonstrating its utility.

2. Related Work

Research on wartime media coverage relies on datasets capturing narratives across national and institutional contexts, where competing claims and propaganda techniques pose unique challenges. Such resources are vital for training models to analyze geopolitical discourse and evaluating their ability to identify narrative strategies. As a reflection of diverse goals, existing resources vary in scope, modality, and annotations.

Datasets based on Twitter/X and Reddit are among the largest. Chen and Ferrara (2023) aggregate over 570 million tweets in English, Russian, and Ukrainian to study information propagation and influence operations, while Zhu et al. (2022) provide 8 million posts correlating post categories with pro-Ukrainian stance. Shevtsov et al. (2022) introduce 57.3 million English tweets annotated for sentiment and hate speech, and the Invasion@Ukraine corpus provides 8.7 million English tweets for content moderation and conspiracy narrative analysis (Susanne Pohl et al., 2023). The VoynaSlov collection (Park et al., 2022) bears similarities to our work through sentence-level media framing labels, though it focuses solely on Russian media. Alyukov et al. (2023) introduce one of the first cross-platform social media datasets, integrating 1.7 million posts to enable the analysis of propaganda and public opinion. Telegram-specific corpora include 4.1 million posts with pro/anti-Kremlin channel tags (Bawa et al., 2025) and 2.3 million posts with binary tags for human rights violations (Nemkova et al., 2023). Hakimov and Cheema (2024) introduce a corpus of 1.5 million tweets in 60 languages with automated annotations for sentiment, stance, and named entities. Their corpus is one of the widest multilingual social media datasets currently available.

News-based corpora are comparatively sparse. The RUWA corpus (Khairova et al., 2024) provides 16,500 news articles using semantic similarity for misinformation detection, but excludes the overarching war theme and curates only specific event-based articles. DNIPRO supersedes it by orders of magnitude in coverage, topics, multilinguality, temporality, and size. A longitudinal media analysis corpus is provided by Ibrahim et al. (2025), with two years of news data, but its coverage is restricted to nations with auxiliary involvement in the Russo-Ukrainian war.

Annotation practices across existing resources differ widely: manual schemata for framing (Park et al., 2022), channel stance (Bawa et al., 2025), and human rights violations (Nemkova et al., 2023); as well as automated large-scale labeling (Susanne Pohl et al., 2023; Hakimov and Cheema, 2024). Most resources rely on social media identifiers that degrade quickly, impeding reproduction.

DNIPRO addresses these gaps by providing longitudinally contiguous, multilingual news coverage from five primary geopolitical actors (Russia, Ukraine, U.S., U.K., and China), with comprehensive annotations for multiple downstream tasks. Unlike social media-focused datasets optimized for volume and engagement patterns, DNIPRO enables systematic comparative analyses of state-aligned narrative framing across languages, nations, and outlets, over a duration of 2 years and 7 months, making it a unique and powerful resource for empirical studies on the creation and dissemination of competing wartime realities (§5).

3. The DNIPRO Corpus

We present DNIPRO, a longitudinal corpus of 246,229 news articles spanning *31 months* of the Russo-Ukrainian war (February 1, 2022 – August 31, 2024). Our corpus captures competing narratives from 11 outlets across five nation states (Russia, Ukraine, U.S., U.K., and China) and three languages (English, Russian, and Mandarin Chinese). Its key descriptive statistics appear in Table 1.

Each article includes structured metadata following a unified schema (shown in Table 2) designed to facilitate large-scale longitudinal and comparative analyses across sources, languages, and geopolitical contexts. DNIPRO is constructed through a systematic multi-stage pipeline targeting news outlets and state-affiliated agencies selected on the basis of three criteria: (i) global influence and prominence within national media ecosystems, (ii) sus-

ining how language and framing construct social reality. Gamson and Modigliani (1989) demonstrated how this approach reveals the competition shaping public understanding of contentious events.

Corpus Scale	
Total articles	246,229
Avg. no. of articles/day	261.1
Title length (avg.)	11.7 words*
Article length (avg.)	332.5 words*
Coverage	
Temporal span	Feb 1, 2022 – Aug 31, 2024
No. of news sources	11 sources
Geopolitical entities	Russia, Ukraine, USA, UK, China
Linguistic Composition	
Languages covered	English, Russian, Mandarin Chinese
Non-English articles	61,637 (25.03%)

**After translation (of non-English sources to English).*

Table 1: DNIPRO at a glance: scale, temporal and geographic coverage, and multilingual composition.

Field	Description
Identifiers & Source	
`article_id`	Unique UUID v4 identifier
`url`	URL of the article
`source_name`	Normalized source name (e.g., *Xinhua*)
`source_country`	Country of origin (ISO 3166-1 alpha-3)
Content & Language	
`title`	Article headline
`article_length`	Word count (post-processing)
`language`	Article language (ISO 639-1)
Temporal Information	
`publication_date`	Publication date (ISO 8601)
`collection_date`	Date of URL access (ISO 8601)

Table 2: Core metadata schema for DNIPRO articles.

tained coverage of the Russo-Ukrainian war, and (iii) discernible editorial positioning ranging from independent journalism to state-aligned messaging. To the extent possible, access to full article archives was obtained through public RSS feeds, institutional subscriptions, or source-specific data collection pipelines. However, not all media outlets provided consistent or unrestricted access to their archives, and some selections reflect practical constraints in addition to topical relevance.

We first detail our collection methodology (§3.1), then describe the structure and access information for the data records (§3.2).

3.1. Data Acquisition

DNIPRO comprises *31 months* of continuous daily data collection (Feb. 1, 2022 – Aug. 31, 2024). To identify articles with substantive coverage of the Russo-Ukrainian conflict, we employ a tiered strategy. For outlets with dedicated sections or labels related to the war, we collect all articles published under those categories. Where available, we also extract articles tagged with conflict-specific metadata (e.g., "Ukraine conflict", "special military operation", "Russia-Ukraine war"). For sources lacking such taxonomies, we apply keyword-based filtering to isolate relevant content. However, some sources impose significant restrictions on retrospective API-based keyword searches. In these cases, we obtain the older articles directly through web browsers, operating within the scope of respective terms of service and copyright restrictions. To ensure quality control, we manually validate stratified random samples across sources, languages, and time periods throughout the acquisition process.

English-language "Western" media from the United States and the United Kingdom, namely *The Financial Times* (FT) and *Cable News Network* (CNN), were included as part of the Western bloc, reflecting the perspectives of countries with significant political, economic, and military involvement in the conflict. These outlets serve domestic and international audiences, and offer central platforms to help us understand how Western-aligned narratives are constructed and disseminated. No additional filtering of articles is done by search queries as these media houses offer dedicated sections for content related to the Russo-Ukrainian war. We collect a total of 5,760 and 4,218 articles from CNN and FT, respectively.

To ensure adequate representation of **Chinese state-aligned perspectives**, we collect articles from three leading outlets: *Xinhua*, *Global Times*, and *China Daily*. Xinhua publishes exclusively in Mandarin, whereas Global Times and China Daily provide parallel English-language coverage. Articles are collected from Xinhua using a bilingual keyword list, designed to minimize the loss of coverage while retaining topical precision. Mandarin Chinese queries comprised "乌克兰" (Ukraine), "俄罗斯" (Russia) and "俄乌" (Russia and Ukraine); English queries included "Russia–Ukraine", "Russia", "Ukraine", and "Ukrainian". These queries yield in 11,152 and 10,206 articles in Mandarin Chinese and English, respectively.

The **Ukrainian perspective** is represented by articles from *Censor.net*, *The Kyiv Independent*, and *European Pravda*, all of which provide English-language versions. While Censor.net influences Ukraine's domestic audience, the other two shape her global image. Their reporting styles vary—from Censor's rapid updates to the in-depth opinions and interviews found in European Pravda. As sources targeting the global Anglosphere, The Kyiv Independent and European Pravda offer sections dedicated

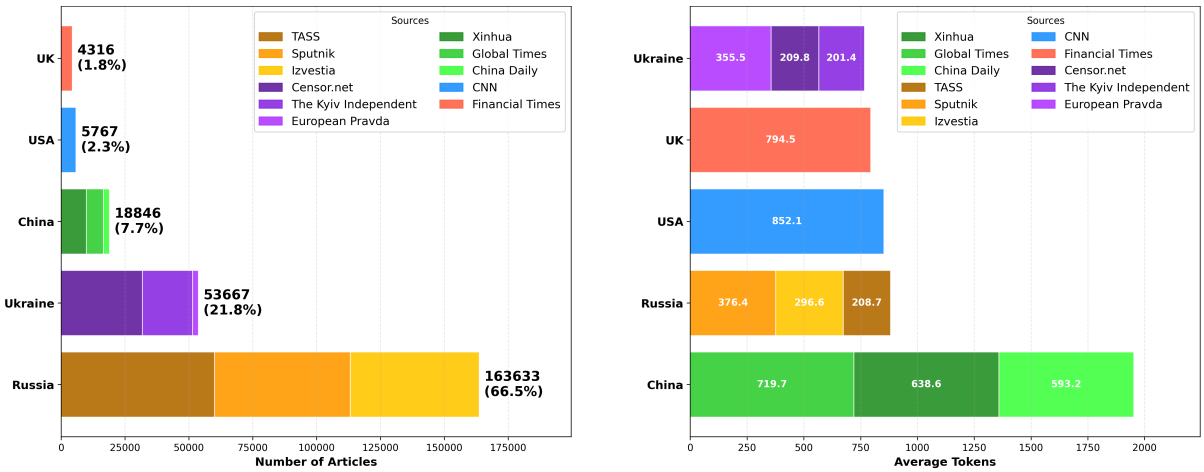


Figure 1: Composition of DNIPRO, across all sources, languages (English, Russian†, and Mandarin‡), and national affiliations: *Censor.net*, *The Kyiv Independent*, and *European Pravda* from Ukraine; *Sputnik*, *TASS*, and *Izvestia*† from Russia; *CNN* from the USA; *The Financial Times* from the UK; and *China Daily*, *Global Times*, and *Xinhua*‡ from China. Above, we see the distribution of articles (left) and the average word count of articles (right).

to the Russo-Ukrainian War, labeled respectively with "War" and "War with Russia". The news reporting approach of Censor.net largely covers the war, evidenced by the majority fraction of Ukrainian war-related tags in their "News" section. As such, no keyword search queries are applied to any of the Ukrainian sources. The collection from these sources amounts to 19,706 articles from The Kyiv Independent, 2,212 from European Pravda, and 31,748 from Censor.net.

The **Russian media** sources in our dataset comprise English articles from *TASS* and *Sputnik*, and Russian-language articles from *Izvestia*. No keyword search queries are required, as all articles are collected from the dedicated section "Special Operation of Russia in Ukraine". We collect 60,002 and 53,139 articles from TASS and Sputnik respectively, and 50,485 articles from Izvestia.

Figure 1 shows the distribution of articles and their average lengths across individual news sources and aggregated by geopolitical entities, providing an overview of the varying scales and styles of coverage across geopolitical perspectives.

3.2. Data Records

The DNIPRO corpus is available at no cost for non-commercial use under the `CC-BY-NC 4.0` license,[3] and the corresponding processing scripts and annotation prompts are available on GitHub.[4] It is structured for ease of use and distributed as separate downloadable files, comprising a primary metadata file, annotation files, and a documentation file. Due to copyright restrictions, we cannot directly include the texts. The provided URLs, however, serve as persistent identifiers enabling reproduction of DNIPRO, like most social media datasets. If an article becomes unavailable, the `collection_date` and `url` fields enable retrieval via the Internet Archive's WayBack Machine.[5]

1. **`metadata.parquet`** contains the complete set of 246,229 metadata records as described in Table 2. The Parquet format was chosen for its efficient compression and columnar storage, which enables fast querying and analysis using popular data science tools.
2. **`named-entities.parquet`** contains the named entity annotations for the entire corpus. Each record includes an `article_id` (see Table 2) and the list of named entities in that article.
3. **`topical-frames.parquet`** provides one row per `article_id` with its corresponding **topical frames** anchored in categories proposed by Habermas (1991) (Table 3).
4. **`sentiments.parquet`** provides sentiment scores for select entities and events across multiple news sources. For politically important and frequently occurring entities (*e.g.*, "Ukraine", "Russia", "Zelensky", "Kherson"), these scores are aggregated monthly. For entities tied to key events within this war (*e.g.*, "Bucha" or "Zaporizhzhia"), they are aggregated daily. The events used in this work are described in §5 (Table 4).
5. **`README.md`** is a comprehensive description of the dataset, the schemata, the license terms, and a code of conduct for ethical use.

We do not specify `train/test/dev` splits for DNIPRO. The corpus spans time, languages, and multiple sources; and supports diverse study designs (*e.g.*,

[3]The complete corpus is available on Zenodo: doi.org/10.5281/zenodo.18470677

[4]github.com/dikshyam/dnipro-codebase

[5]web.archive.org

temporal evolution or development of an entity/-topic/event; media source comparisons; transnational or cross-lingual reporting of events; or other types of aggregation). A preset split would constrain the diversity of possible applications. We encourage task-appropriate partitions instead.

4. Data Annotations

We now describe the enrichment of raw article metadata with annotations. Since copyright restrictions prevent the direct release of texts, we generate content-derived labels that form the dataset's core analytical value. The process involves:

- metadata and text normalization;
- translation of non-English text to English;
- labeling named entities;
- sentiment analysis;
- stance analysis;
- topical framing; and
- rigorous quality control with human evaluation.

4.1. Text Processing for Annotation

To prepare the corpus for annotation and analyses, we apply standardized processing steps to all data, across sources and languages. This ensures consistency across media sources with varying conventions in structure, formatting, and metadata.

Publication dates are parsed and normalized to ISO 8601 standard (YYYY-MM-DD), and multiple metadata fields—*e.g.*, "publication date", "published on"—are mapped to metadata fields in our unified schema (as shown in Table 2). The article text content is stripped of boilerplate elements like location or date prefixes commonly found in news articles (*e.g.*, "KYIV, March 3 –"), inline footnotes, bylines, and repeated source disclaimers. Duplicate articles are removed, to avoid redundancy in print syndication and multilingual reprints. To support downstream linguistic analyses, the text is then converted to UTF-8 encoding, segmented into paragraphs, and split into sentences.

4.2. Machine Translation

A quarter of DNIPRO consists of Mandarin Chinese and Russian texts, (`zh` and `ru`, respectively, in ISO 639-1). We translate these into English (`en`) to facilitate downstream analyses. We choose three models based on recent comparative studies on neural machine translation by Smirnov et al. (2022) and Alemayehu et al. (2024): OPUS-MT (Tiedemann et al., 2023), M2M100 (Fan et al., 2021), and NLLB-200-Distilled (Team, 2022). Since there were no readily available explicit benchmarks for `zh-to-en` translation of war or conflict narratives, we select a random set of ten articles from *Xinhua*, evaluated by two native speakers of Mandarin Chinese for semantic fidelity and fluency. The evaluators worked independently to score the output of these models on 1-5 Likert scale for fidelity and fluency, and `M2M100-418M` received the highest scores from both evaluators on both criteria. So, we employ this model for all `zh-to-en` translations. Russian articles from *Izvestia* are similarly evaluated by native Russian speakers, and translated using the `ru-to-en` path of `M2M100-418M`.

Frame	Definition
ECONOMY[1]	Macroeconomic issues and markets: sanctions/trade, inflation, growth, jobs, taxes, energy/commodity prices.
IDENTITY[3]	Group identity and belonging: nationality, ethnicity, religion, language, or gender; discrimination or identity politics.
MORALITY[2]	Explicit ethical/value framing: moral duty, fairness, dignity, humanitarian arguments.
LEGAL[1]	Courts and legal processes: rulings, lawsuits, prosecutions, designations, warrants, legal rights or entitlements.
POLICY[1]	Concrete government outlook or outreach, rules, bills, decrees, program design and details (what a policy does).
POLITICS[1]	Political proceedings and party competition: elections, horse-race strategy, reshuffles, coalition dynamics.
PUBLIC OPINION[1]	Public attitudes and moods: polls, surveys, and protests as expressions of public sentiment and interest.
SECURITY[1]	War/defense, violence/crime, policing, border enforcement, terrorism, cyberattacks.
WELFARE[1]	Benefits and human services: healthcare, pensions, housing aid, unemployment, assistance to refugees or displaced people.

Table 3: Topical frames in DNIPRO: an abridgment of the *Refugees and Migration Framing* schema (Yu and Fliethmann, 2022), with its *utilitarian*[1], *moral-universal*[2], and *identity-related*[3] categories derived from the discourse ethics of Habermas (1991).

4.3. Annotation Methodology

Named Entity Recognition: To enable fine-grained analysis of key actors and locations, we perform named entity recognition (NER) across the entire corpus (after translation to English). This annotation is carried out using spaCy (Honnibal et al., 2020) and its pretrained model `en_core_web_sm`, identifying entities and their types according to the OntoNotes 5.0 taxonomy (Weischedel et al., 2013). This identifies 85,155 distinct named entity mentions, and provides the foundational layer for downstream tasks such as tracking the prominence of specific actors over time, temporal studies of media focus across entities and events, and entity-specific sentiment and stance analyses.

Event (Date range)	# articles
Bucha Massacre (Apr 3–9, 2022)	273
Zaporizhzhia Nuclear Crisis (Mar 3–9, 2022)	219
Mariupol Hospital Attack (Mar 9–15, 2022)	138
Kursk Offensive (Aug 6–12, 2024)	180

Table 4: The four conflict events analyzed in our use case experiments, selected based on global coverage and the potential to generate narrative divergence.

Sentiment Analysis: To capture the sentiment surrounding the most prominent named entities, we provide *sentence-level* sentiment annotation using a pretrained model: `sentiment-roberta-large-english` (Hartmann et al., 2023). For each article, sentences containing such entities (*e.g.*, `Kursk`, `Bucha`, `Kherson`, `Zelensky`, `Ukraine`, `Russia`) are identified using spaCy, and passed to this classifier. The model outputs are then mapped to continuous sentiment scores, which are then aggregated on a daily and monthly basis at both the source and country level using source–country relations. These annotations thus provide time series data revealing how sentiment toward specific actors and events evolves across geopolitical perspectives.

Stance Detection: The dataset includes stance annotations obtained by few-shot instruction prompting with Phi-4 14B model (Microsoft, 2024), identifying the stance *towards a specified target*. The input comprises a sentence and the immediate context in its news article. To Phi-4, we provide the task instruction with a few examples, along with (a) a passage for contextual information, (b) a focal sentence, and (c) the target. The model is prompted to output a compact, formatted response:

```
Target:    <the identified target>
Stance:    favor|against|none
Rationale: <a short justification grounded in
           the given text>
```

Topical Framing: For topical framing, we use the same few-shot prompting with Phi-4 14B (Microsoft (2024)), where the input is an instruction describing the task, along with topical frame label definitions (shown in Table 3) and 1-2 examples per label. Then, for a given instance—comprising an article's title, subtitle (if present) and content—we prompt the model to generate a `json` object containing the list of applicable topical frames and their corresponding confidence scores. This supports analysis of how different sources emphasize humanitarian, military, diplomatic, or economic dimensions of the conflict.

4.4. Human Evaluation

To assess annotation quality, we conduct rigorous human evaluation on both stance detection and topical framing. Results demonstrate that automated labeling substantially agrees with expert judgment (PABAK = 0.71 for topical framing, Fleiss' κ = 0.643 for stance detection), validating their reliability for coarse-grain analyses while identifying specific categories that warrant careful interpretation.

Annotator-pair	Agreement (%)	Cohen's κ
Human annotators	86.00%	0.770
Annotator$_1$ and `phi-4`	72.68%	0.618
Annotator$_2$ and `phi-4`	70.00%	0.580

- Mean agreement: 76.23%
- **Fleiss'** κ across all three annotators: 0.643
- Mean pairwise Cohen's κ: 0.656 (range: 0.58-0.77)

Table 5: Annotation agreement on stance labels: agreement percentage, pairwise Cohen's κ, and Fleiss' κ.

4.4.1. Stance Annotations

The stance annotation task identifies whether a sentence (with context) expressed a stance (`favor` or `against`) towards any entity, or if no stance was present (`none`). Two human annotators with native proficiency in English annotated a sample of 150 instances. We evaluate their inter-annotator agreement, and then also with `phi-4` as the LLM annotator. Results in Table 5 present the pairwise and overall inter-annotator agreements, demonstrating strong human inter-annotator agreement (κ = 0.77) and moderate LLM-human agreements (κ = 0.62 and 0.58), showing consistent but less-than-human reliability. The overall Fleiss' κ = 0.643 indicates that the stance annotations are substantially reliable for large-scale analysis.

While the LLM shows a consistent gap with human judgment, this is expected for nuanced pragmatic language tasks like stance detection. The annotations remain valuable for trend analysis and comparative studies across sources—such as identifying systematic differences in how Russian vs. Ukrainian sources frame political leadership or specific events—while nuanced high-stakes stance analyses may benefit from human review.

4.4.2. Topical Frames

Two expert annotators independently label a stratified random sample of 200 articles, providing nine binary labels with corresponding [0,1] confidence scores for each article. Significant class imbalance is a common challenge in multi-label annotation, and the inter-annotator agreement must consider the high *base rate agreement* where annotators agree on a label being *absent* from an article. The average agreement on *label absence*, over all nine topical frames, accounts for 65.82% of our evaluation sample, with the positive class (for any label)

Frame	PABAK		Confidence	
	Human	All	Human	Phi-4
High reliability (PABAK > 0.8)				
Public Opinion	0.87 (93.4%)	0.865	0.80	0.89
Legal	0.85 (92.4%)	0.783	0.74	0.91
Welfare	0.81 (90.4%)	0.783	0.80	0.87
Substantial reliability (PABAK 0.69 - 0.8)				
Morality	0.75 (87.3%)	0.750	0.79	0.91
Identity	0.74 (86.8%)	0.783	0.79	0.92
Economy	0.70 (84.8%)	0.736	0.84	0.93
Security	0.69 (84.3%)	0.587	0.85	0.93
Moderate reliability (PABAK < 0.69)				
Policy	0.55 (77.7%)	0.526	0.78	0.89
Politics	0.48 (74.1%)	0.425	0.79	0.92
Average	0.71 (85.7%)	0.693	0.80	0.91

Table 6: Human inter-annotator agreement scores (PABAK and percentage), human-model agreement (PABAK), and annotator confidence scores ([0-1]) across all nine topical frames introduced in Table 3.

representing a minority of instances. Due to this prevalence bias, traditional measures like Cohen's or Fleiss' κ are overly pessimistic reliability measures for our data. For a more accurate assessment that accounts for this imbalance, we report the *prevalence-adjusted bias-adjusted kappa*, or PABAK (Byrt et al., 1993), which is especially crucial for evaluation of consensus on the rarer labels.

The **inter-annotator agreement** results, measured by PABAK, are presented in Table 6 and demonstrate substantial agreement on average (0.71, with 85.7% agreement). Three topical frame categories — Welfare, Public Opinion, and Legal — reveal near-perfect agreement ($>$ 0.8). Overall, seven of the nine labels achieve PABAK scores $\geq$ 0.7, indicating a consistent and shared understanding of the annotation framework. The two remaining classes, Policy and Politics, have moderate agreement, as they represent more abstract and often overlapping conceptual domains.

Further, we examine the correlation between the average confidence scores reported by the human experts and those reported by phi-4. The framing categories are complex and sometimes overlapping, resulting in a weak positive correlation ρ = 0.37. Notably, phi-4 exhibits systematically higher confidence (avg. 0.91) than human annotators (avg. 0.80) across all categories, regardless of the extent of agreement. This suggests that users should not interpret phi-4's self-reported confidence scores as reliable indicators of annotation quality.

Implications for Dataset Use: The strong agreement validates the overall reliability of our annotation schema. However, we recommend that users consider the following nuances in their studies:

1) Frames with *high/substantial reliability* (PABAK > 0.69) are suitable for fine-grained analysis such as tracking shifts in Welfare framing after major tragedies, or comparing Security or Legal discourse across state-aligned media.
2) The *moderately reliable* frames (PABAK < 0.69) — Policy and Politics — yield lower agreement due to their inherent conceptual overlap. These frame labels may be best suited for coarse-grained topic modeling or as supplementary features, with the understanding that they represent broader discourse categories that often intersect. As such, their unification may provide more reliable signals.
3) The expert-annotated sample of 200 documents, stratified across sources, languages, and time periods, provides a completely validated high-quality *gold-standard benchmark*. Our stratified sampling ensures it is suitable for evaluating the performance of topical frame classification systems on real-world data on contentious wartime narratives.

Together, these evaluations demonstrate that DNIPRO's annotations achieve a balance of scale and quality rare in wartime media corpora: automated methods enable comprehensive coverage, while stratified human validation ensures researchers can make informed decisions regarding labels, as per their task-specific goals. The annotation guidelines and task-specific prompts are provided in Appendix A and C, respectively.

5. Use Case Experiments

To demonstrate the analytical potential of DNIPRO, we present four use-case experiments exploiting its unique multi-perspective structure. These are not exhaustive analyses, but proof-of-concept demonstrations showing how the corpus enables diverse research in computational journalism, political communication, discourse analysis, and information warfare studies. Each application highlights specific affordances of the corpus while suggesting broader research directions.

1. A Thematic Study of Conflict Events: Our analysis focuses on the four major conflict events shown in Table 4, which received substantial global coverage, generated clear geopolitical divisions, and invited transnational comparisons. For each event, we filter articles using specific location identifiers (*e.g.*, Bucha) to ensure relevance.

We examine how conflict events are framed by analyzing patterns of thematic salience across news outlets. By identifying which concepts, actors, and subtopics are underscored in coverage, we reveal how outlets differentially emphasize aspects such as *historical context*, *humanitarian concerns*, or *institutional responses*.

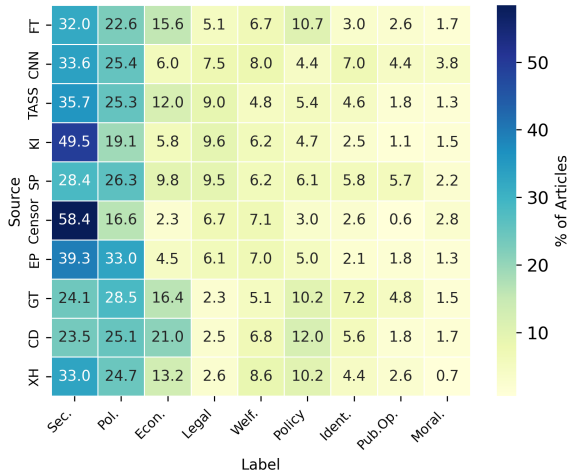


Figure 2: Topical framing across news outlets (for each frame, % of articles with confidence > 0.5): `Security` dominates all outlets, while `Economy` is most prominent in Chinese media and *Financial Times* (FT). Other abbr.: *Kyiv Independent* (KI), *Sputnik* (SP), *European Pravda* (EP), *Global Times* (GT), *China Daily* (CD), *Xinhua* (XH).

Word clouds visualizing these thematic patterns for each conflict event and outlet country are provided in Appendix B. Our prelimnary analysis shows that Russian and Chinese outlets consistently emphasize institutional terminology (*special military operation*), security concerns (*nuclear weapons*), and counter-narratives (*media manipulation*, *ceasefire*), reflecting strategic adversarial framing. Ukrainian outlets predominantly highlight humanitarian dimensions (*civilian casualties*) with frequent references to specific conflict locations. Media from the USA and the UK balance humanitarian topics with institutional responses (*sanctions*, *international criminal court*) and policy discussions (*no fly zone*).

Thus, DNIPRO captures differences in thematic salience and agenda-setting across geopolitical actors narrating the same underlying events.

2. Topical Framing: We conduct a corpus-wide analysis using the schema introduced in Table 3. Article-level topical frames provide a comparable map of what each media outlet emphasizes, revealing systematic priorities that contextualize downstream analyses such as stance, sentiment, or discourse. Articles are annotated with `phi-4` using a few-shot instruction prompt, producing multi-label assignments along with confidence scores (as described in §4.4.2).

Figure 2 shows that `Security` dominates across all outlets, with the highest shares in Ukrainian media, followed by Russian state media and Western outlets like CNN. The second-most prevalent category is `Politics`, elevated in Russian and Chinese media. It is substantially present in other outlets too. `Economy` is most visible in Chinese media and, naturally, The Financial Times, while `Policy` appears steadily across outlets, with higher emphasis in Chinese media and the UK. Relative to others, Russian media leans more on `Legal` topics. Notably, discussions of `Morality` are extremely rare despite the focal topic of this dataset being an extended war, indicating that operational, strategic, and material considerations far outweigh explicit ethical rhetoric in war coverage.

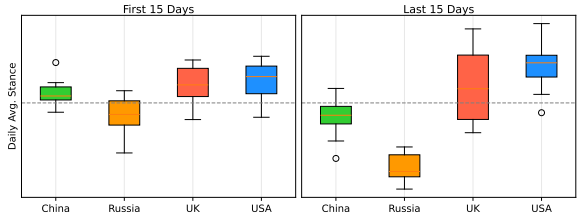


Figure 3: Daily average stance scores toward `Ukraine`, grouped by nation, comparing the first 15 days of the war (left) with the last 15 days of data collection (right). Note the dramatic intensification of Russian negativity and American support, increased variability in the UK, and the gentle negative shift in Chinese media.

3. Stance Analysis: We compare stance toward `Ukraine` between the war's first 15 days and the final 15 days of data collection (Figure 3), revealing striking temporal shifts. Chinese media shifted from mildly pro-Ukrainian to mildly oppositional, likely reflecting China's deepening strategic alignment with Russia. Russian coverage moved from a mildly negative stance to overwhelmingly hostile – possibly correlating with battlefield changes and political rhetoric. The US coverage became even more supportive, while the UK exhibited greater variability, possibly due to the transition from early solidarity to war fatigue, domestic economic pressures and/or leadership changes.

The temporal dimension of DNIPRO proves crucial for these insights: China's pivot, Russian and American intensification in opposite directions, and growing ambivalence in the UK.

4. Divergent Narratives: To investigate how media outlets construct and contest narratives (Gamson et al., 1992; Takeshita, 1997), we analyze claims to identify conflicting representations of key events. Building on prior work in conflict journalism (Entman, 2010; Thussu et al., 2017), this analysis unearths how opposing perspectives are embedded in individual claims.

Using weekly data for each event, we extract sentence-level claims and categorize them into six types, motivated by epistemology (Chafe and Nichols, 1986), subjectivity and stance analysis (Wiebe et al., 2005; Ferreira and Vlachos, 2016), and speech acts (Austin, 1962; Searle, 1969):

factual: direct statements of observable or confirmed events;
attributed: claims attributed to specific actors;
disputed: explicit denials or rebuttals;
reported: indirect accounts from third-parties;
command/policy: official statements or directives;
analytical/assessment: interpretive or evaluative judgments by experts or institutions.

With GPT-3.5 plus human-in-the-loop validation, we label claims via structured prompts, and group semantically related claims through agglomerative clustering on Sentence Transformer representations (Reimers and Gurevych, 2019).

Our key findings demonstrate DNIPRO's utility in the study of *information warfare*, *perception*, and *distortion* (Galtung and Ruge, 1965) — not as explicit propaganda, but as systematic divergence of constructed realities across geopolitical divides:

- **Contradictions are usually implicit:** Contradictions predominantly emerge across geopolitical divides as competing interpretations. Outlets rarely offer explicit rebuttals, instead they construct incompatible narratives through different topical frames, omissions, or counter-attributions and selective emphasis (Table 7).
- **Alignment with geopolitical blocs:** Entailment occurs primarily within aligned media ecosystems: Ukrainian and Western outlets reinforce each other through overlapping `factual`, `reported`, or `command/policy` claims, while Russian media maintains internal consistency through `attributed` and `analytical` claims citing official statements.
- **Framing substitutes for rebuttal:** When covering the same event (e.g., an airstrike), Russian media might characterize it as precision targeting of military assets (`factual/attributed`), while Western media offer contrary eyewitness accounts (`disputed/reported`).

	Bucha Massacre
RU	Russia had testimony from residents of Bucha and Mariupol, who reported brutal crimes, such as torture, committed by Ukrainian nationalists. *(TASS)*
UA	All eyewitnesses noted the systematic nature of the killings in Bucha by the Russian army. *(Censor)*
	Sumy Oblast Attacks
RU	Russian forces jointly with border guards repelled these attacks and hit enemy positions near the border and in the Sumy Region. *(TASS)*
UA	Russian forces attacked six border areas and settlements of Sumy Oblast on June 7, firing 14 times and causing at least 54 explosions . . . *(Kyiv Independent)*

Table 7: Competing realities narrated by Russian and Ukrainian media: Russian sources attribute atrocities to Ukrainian forces (yellow), while Ukrainian sources blame Russian forces (red). Narratives are opposed in attribution and framing, absent direct refutation or negation.

6. Conclusion

We present DNIPRO, a longitudinal corpus of 246K news articles capturing competing geopolitical narratives of the Russo-Ukrainian war across 31 months (Feb 2022 – Aug 2024). Unlike prior corpora focused on social media volume or single-language coverage, DNIPRO enables comparative analysis across nation-states through structured, multilingual news data with human-evaluated annotations for stance, sentiment, and topical framing.

Our experiments demonstrate DNIPRO's capacity to support diverse analytical approaches and reveal systematic patterns, from thematic emphasis patterns and temporal stance tracking to fine-grained claim-level contradiction analysis. This makes DNIPRO uniquely suited for studying information ecosystems where competing narratives coexist without direct engagement, and *contradictions are constructed implicitly*: media outlets build parallel realities through divergent attribution and selective emphasis rather than direct refutation.

This has methodological implications: traditional approaches assume explicit falsehoods or rebuttals (Scantamburlo and Zollo, 2026; Guo et al., 2022), but competing narratives operate through *incompatible framing*. Thus, data-driven research needs new methods that map how attribution choices and temporal patterns construct divergent realities across information ecosystems.

DNIPRO enables scalability in several research directions: (a) longitudinal narrative evolution (e.g., what are the predictors of geopolitical realignments); (b) cross-lingual information flow (e.g., which media outlets serve as bridges, and which ones serve as echo chambers); (c) event-driven polarization: (e.g., can we predict which events trigger lasting narrative shifts); and (d) computational detection of incompatible narrative frames, in the absence of explicit disagreement markers.

As international conflicts unfold across the fragmented media landscape, understanding these dynamics becomes essential for researchers, policymakers, journalists, and citizens. DNIPRO is publicly available under CC-BY-NC 4.0:

Dataset: doi.org/10.5281/zenodo.18470677
Code: github.com/dikshyam/dnipro-codebase

We invite researchers to utilize DNIPRO, and develop new methods for detecting, tracking, and understanding narrative divergence: work that bridges computational analysis with urgent questions about truth, persuasion, and power in the digital age.

7. Limitations

Temporal scope: DNIPRO covers February 2022 through August 2024, but the Russo-Ukrainian War has not concluded. Thus, significant developments after this period — including shifts in foreign policy, territorial control, or international alliances — are not captured. Researchers studying more recent events should consider this temporal boundary when interpreting findings, or extend our corpus to include later developments.

Geographic coverage: Our selection stresses on geopolitical actors directly involved (Russia and Ukraine) and influential despite no direct engagement in the war (the United States, the United Kingdom, and China). NATO representation, however, is limited to two of its thirty-two member states. Perspectives from several influential NATO states (e.g., Germany and its state-own media, Deutsche Welle) are absent. Regional actors and nations with complex positions may also be included in the analyses we have underscored in our work. Such actors include Hungary and Romania, both of which border Ukraine; Belarus, also bordering Ukraine, but currently allied with the Russian state; Türkiye, a powerful regional actor, especially in the Black Sea; Georgia, a NATO contender that continues to maintain reasonably strong relations with Russia; and India, which maintains nuanced multi-alignments. The inclusion of these actors may reveal more sophisticated patterns of alignment and hedging than the polarized perspectives currently represented.

Annotation reliability: While our human evaluations demonstrate substantial inter-annotator agreement (§4.4), automated ground-truth annotations are never perfect. This is not a limitation for qualitative analyses, mixed-method approaches, or empirical analyses for coarse-grained signals; but research requiring finer labels for specific nuanced tasks may benefit from additional human validation of their task-relevant subsets of DNIPRO.

Scope of use-case experiments: Our exploratory analyses demonstrate analytical possibilities rather than provide definitive empirical claims. These demonstrations are not limitations of the corpus. Rather, they serve as proof-of-concept for the types of research DNIPRO enables: longitudinal stance tracking, framing comparison across media outlets and nation states, and contradiction detection. But we emphasize that these are not exhaustive treatments of these topics. Each use case is illustrative, and intended for deeper investigations.

Translation and linguistic nuances: Non-English articles are machine-translated to English for cross-lingual analysis. Even as this enables systematic comparisons, translations may lose culturally specific rhetoric or connotations that carry meaning in the original language. Researchers with relevant language expertise may benefit from analyzing the untranslated Russian and Chinese Mandarin articles for such discursive phenomena.

8. Ethical Considerations

DNIPRO documents an ongoing conflict with significant human costs and geopolitical implications. We acknowledge several potential ethical dimensions:

Dual-use considerations. This corpus enables research into how narratives shape perception, with applications in computational journalism, political communication, and information literacy. However, the same analytical methods could be misused to refine propaganda or manipulate opinion. We urge researchers to consider the societal implications of their work and to strive for transparency, cross-cultural understanding, and informed discourse.

Source inclusion and representation. DNIPRO inherently reflects the biases present in global media coverage. The inclusion of opposing narratives—including state-controlled media—is a deliberate choice, to empower comparative analyses. **Inclusion does not imply endorsement of content**. Some sources may contain disinformation or inflammatory rhetoric. We include them because the research possibilitied outlined earlier require access to the full spectrum of discourse, not sanitized versions. Researchers should contextualize sources appropriately using the provided metadata (media outlet ownership, editorial stance, etc.). We strongly recommend any derived datasets or findings to retain all the provided metadata, to ensure that articles can be interpreted within their proper editorial and temporal contexts.

Sensitive content. This corpus pertains to an armed conflict. Many articles may describe violence and other distressing aspects of armed conflicts. Researchers working with this data should be aware of this, particularly when conducting qualitative analysis or annotation tasks. We recommend appropriate support structures for research teams.

Data provenance and access, All content was gathered from publicly available news sources without breaching access restrictions or terms of service. We provide URLs rather than full text to respect intellectual property rights and enable verification, and do not include any copyright-protected content or any private communications. The CC-BY-NC 4.0 license permits research and educational use while prohibiting commercial exploitation.

We designed DNIPRO to support research that illuminates how information ecosystems function during contested events — work that ultimately serves democratic accountability, informed citizenship, and conflict resolution.

We earnestly welcome community input on additional safeguards or ethical considerations.

9. Bibliographical References

AFP. 2022. Russia Bans Media Outlets From Using Words 'War', 'Invasion'. The Moscow Times. Date: Feb 6, 2022; Retrieved: Nov 29, 2024.

Abu Ubaida Akash, Ahmed Fahmy, and Amine Trabelsi. 2025. Can Large Language Models Address Open-Target Stance Detection?

Hizkiel Mitiku Alemayehu, Hamada M Zahera, and Axel-Cyrille Ngonga Ngomo. 2024. Error analysis of multilingual language models in machine translation: A case study of English-Amharic translation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 19758–19768, Miami, Florida, USA. Association for Computational Linguistics.

Maxim Alyukov, Maria Kunilovskaya, and Andrei Semenov. 2023. Wartime Media Monitor (WarMM-2022): A Study of Information Manipulation on Russian Social Media during the Russia-Ukraine War. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 152–161, Dubrovnik, Croatia. Association for Computational Linguistics.

John Langshaw Austin. 1962. *How to Do Things with Words*. Oxford University Press.

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.

Samantha Bradshaw, Mona Elswah, Monzima Haque, and Dorian Quelle. 2024. Strategic Storytelling: Russian State-Backed Media Coverage of the Ukraine War. *International Journal of Public Opinion Research*, 36(3):edae028.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

Ted Byrt, Janet Bishop, and John B. Carlin. 1993. Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5):423–429.

Wallace Chafe and Johanna Nichols, editors. 1986. *Evidentiality: The Linguistic Coding of Epistemology*, volume 20 of *Advances in Discourse Processes*. Ablex, Norwood, NJ.

Emily Chen and Emilio Ferrara. 2023. Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War between Ukraine and Russia. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):1006–1013.

Leyla Derviş. 2023. Transformation of Geopolitical Perceptions in the Russian-Ukrainian War: Impact on Regional Relations in the Future. *Futurity of Social Sciences*, 1(1):21–34.

Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Robert M. Entman. 2010. Media framing biases and political power: Explaining slant in news of Campaign 2008. *Journalism*, 11(4):389–408.

Norman Fairclough. 2010. *Critical Discourse Analysis: The Critical Study of Language*, 2 edition. Routledge.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Michael Auli, and Armand Joulin. 2021. Beyond English-Centric Multilingual Machine Translation. *Journal of Machine Learning Research*, 22(107):1–48.

William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1163–1168, San Diego, California. Association for Computational Linguistics.

Johan Galtung and Mari Holmboe Ruge. 1965. The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers. *Journal of Peace Research*, 2(1):64–91.

William A. Gamson, David Croteau, William Hoynes, and Theodore Sasson. 1992. Media Images and the Social Construction of Reality. *Annual Review of Sociology*, 18(1):373–393.

William A. Gamson and Andre Modigliani. 1989. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *American Journal of Sociology*, 95(1):1–37.

Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. 2022. A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

Jürgen Habermas. 1991. *Erläuterungen zur Diskursethik*. Number 975 in Suhrkamp Taschenbuch Wissenschaft. Suhrkamp, Frankfurt am Main.

Sherzod Hakimov and Gullal S. Cheema. 2024. Unveiling Global Narratives: A Multilingual Twitter Dataset of News Media on the Russo-Ukrainian Conflict. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, ICMR '24, page 1160–1164, New York, NY, USA. Association for Computing Machinery.

Jochen Hartmann, Mark Heitmann, Christian Siebert, and Christina Schamp. 2023. More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, 40(1):75–87.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Majd Ibrahim, Bang Wang, Minghua Xu, and Hang Xu. 2025. A multidimensional analysis of media framing in the Russia-Ukraine war. *Journal of Computational Social Science*, 8(34).

Eva Johais and Mareike Meis. 2024. 'Unleash the hounds!': NAFO's memetic war narrative on the Russo-Ukrainian conflict. *Critical Studies on Security*, 13(1):1–13.

Nina Khairova, Andrea Galassi, Fabrizio Lo Scudo, Bogdan Ivasiuk, and Ivan Redozub. 2024. Unsupervised approach for misinformation detection in Russia-Ukraine war news. In *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume IV: Computational Linguistics Workshop*, pages 21–36. CEUR Workshop Proceedings.

Microsoft. 2024. Phi-4 Technical Report. Accessed: Oct 17, 2025.

Poli Nemkova, Solomon Ubani, Suleyman Olcay Polat, Nayeon Kim, and Rodney D. Nielsen. 2023. Detecting Human Rights Violations on Social Media during Russia-Ukraine War.

Gunnar Nygren, Michal Glowacki, Jöran Hök, Ilya Kiria, Dariya Orlova, and Daria Taradai. 2016. Journalism in the Crossfire: Media Coverage of the War in Ukraine in 2014. *Journalism Studies*, 19(7):1059–1078.

OpenAI. 2023. Gpt-3.5 language model. `https://developers.openai.com/api/docs/models/gpt-3.5-turbo`.

Bojan Pancevski. 2024. One Million Are Now Dead or Injured in the Russia-Ukraine War. The Wall Street Journal. Date: Sep 17, 2024; Retrieved: Feb 3, 2025.

Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. 2022. Challenges and Opportunities in Information Manipulation Detection: An Examination of Wartime Russian Media. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5209–5235. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 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 3982–3992. Association for Computational Linguistics.

Nataliya Roman, Wayne Wanta, and Iuliia Buniyak. 2017. Information wars: Eastern Ukraine military conflict coverage in the Russian, Ukrainian and U.S. newscasts. *International Communication Gazette*, 79(4):357–378.

Morgan Saletta, Ariel Kruger, Tamar Primoratz, Ashley Barnett, Tim van Gelder, and Robert E. Horn. 2020. The role of narrative in collaborative reasoning and intelligence analysis: A case study. *PLoS One*, 15(1):e0226981.

Teresa Scantamburlo and Fabiana Zollo. 2026. A Scoping Review of Misinformation Research: Definitions, Conceptualizations, and Measurement Approaches. *ACM Trans. Web*.

John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Alexander Shevtsov, Christos Tzagkarakis, Despoina Antonakaki, Polyvios Pratikakis, and Sotiris Ioannidis. 2022. Twitter Dataset on the Russo-Ukrainian War. *arXiv*, 2204.08530.

Julia Silvertsen. 2024. Projecting Geopolitical Power and Acting as One: The EU's Foreign Policy in the Russian-Ukrainian Conflict [2014

and 2022-23]. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Alexander V. Smirnov, Nikolay Teslya, Nikolay Shilov, Diethard Frank, Elena Minina, and Martin Kovacs. 2022. Quantitative Comparison of Translation by Transformers-Based Neural Network Models. In *International Conference on Enterprise Information Systems*.

Janina Susanne Pohl, Simon Markmann, Dennis Assenmacher, and Christian Grimme. 2023. Invasion@Ukraine: Providing and Describing a Twitter Streaming Dataset That Captures the Outbreak of War between Russia and Ukraine in 2022. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):1093–1101.

Toshio Takeshita. 1997. Exploring the Media's Roles in Defining Reality: From Issue-Agenda Setting to Attribute-Agenda Setting. In Maxwell McCombs, Donald L. Shaw, and David Weaver, editors, *Communication and Democracy: Exploring the Intellectual Frontiers in Agenda-Setting Theory*, pages 15–27. Lawrence Erlbaum Associates, New York.

NLLB Team. 2022. No Language Left Behind: Scaling Human-Centered Machine Translation.

Daya Kishan Thussu, Hugo De Burgh, and Anbin Shi, editors. 2017. *China's Media Go Global*. Routledge, London.

Jörg Tiedemann, Mikko Aulamo, Daria Bakshandaeva, Michele Boggia, Stig-Arne Grönroos, Tommi Nieminen, Alessandro Raganato, Yves Scherrer, Raul Vazquez, and Sami Virpioja. 2023. Democratizing Neural Machine Translation with OPUS–MT. *Language Resources and Evaluation*, 58:713–755.

Jack Watling, Oleksandr V. Danylyuk, and Nick Reynolds. 2023. Preliminary Lessons from Russia's Unconventional Operations During the Russo-Ukrainian War, February 2022–February 2023. Royal United Services Institute (RUSI) for Defence and Security Studies. Date: Mar 29, 2023; Retrieved: Nov 28, 2024.

P. G. de Werd. 2018. *Critical Intelligence: Analysis by Contrasting Narratives*. Doctoral thesis, Universiteit Utrecht.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. In *Language Resources and Evaluation*, volume 39, pages 165–210. Springer.

Qi Yu and Anselm Fliethmann. 2022. Frame Detection in German Political Discourses: How Far Can We Go Without Large-Scale Manual Corpus Annotation? *Journal for Language Technology and Computational Linguistics*, 35(2):15–31.

Xiaoling Zhang, Herman Wasserman, and Winston Mano. 2016. China's Expanding Influence in Africa: Projection, Perception and Prospects in Southern African Countries. *Communicatio*, 42(1):1–22.

Yiming Zhu, Ehsan ul Haq, Lik-Hang Lee, Gareth Tyson, and Pan Hui. 2022. A Reddit Dataset for the Russo-Ukrainian Conflict in 2022. *arXiv*, 2206.05107.

10. Language Resource References

Ralph Weischedel and Martha Palmer and Mitchell Marcus and Eduard Hovy and Sameer Pradhan and Lance Ramshaw and Nianwen Xue and Ann Taylor and Jeff Kaufman and Michelle Franchini and Mohammed El-Bachouti and Robert Belvin and Ann Houston. 2013. *OntoNotes Release 5.0 LDC2013T19*. Linguistic Data Consortium. Linguistic Data Consortium, ISLRN 151-738-649-048-2.

Appendix A. Annotation Methodology

This appendix provides complete prompt specifications and implementation details for the large-scale automated annotations described in §4.3. All prompts follow structured output formats to enable reliable parsing and post-processing.

Models and APIs: We used the following language models for automated annotations:

- `microsoft/phi-4` (14B parameters) (Microsoft, 2024) for stance detection and topical framing;
- a fine-tuned RoBERTa-large model by Hartmann et al. (2023)[6] for sentiment analysis; and
- GPT-3.5 Turbo, via the OpenAI API (OpenAI (2023); Brown et al. (2020)), for thematic tagging and claim typology.

Stance detection proceeds in two stages: (1) *target generation* identifies stance-worthy entities or events in a sentence, and then (2) *stance classification* determines the expressed attitude toward each target. This decomposition improves accuracy by separating entity recognition from sentiment interpretation (Akash et al., 2025). We apply stance detection to articles from: (a) the first and last 15 days of data collection, and (b) four major conflict events listed in Table 4: the Bucha massacre, the Kursk offensive, Mariupol hospital attack, and Zaporizhzhia nuclear crisis.

Target Generation. The model identifies all stance-worthy targets – actors, organizations, or events that can be supported or opposed – mentioned in a sentence. The prompt template for target generation is shown next:

```
You are performing Target Detection for Stance
Classification.

Task:
1. Read the sentence carefully.
2. Identify ALL the stance-worthy targets
   (entities or events) mentioned in the
   sentence.
3. A stance-worthy target can be:
   - An actor (government, organization,
     leader) whose responsibilities or
     actions can be supported or opposed.
   - An event/claim that people can express
     favor, opposition, or neutrality toward.
4. Return multiple targets if present,
   separated by commas.
5. If no stance-worthy targets are found,
   return exactly: "Targets: NONE".
6. Output only the targets, do not repeat
   the sentence.
Examples:
1) Input:
   Sentence: "Ukrainian forces downed 22 of
   the 27 Shahed-type drones launched by
   Russia overnight on June 5, the Air Force
   said in its morning update."
   Output:
   Targets: Ukrainian forces, Russia

2) Input:
   Sentence: "Russian Defense Ministry said
   it also shot down Ukraine's Western-
   provided missiles."
   Output:
   Targets: Russian Defense Ministry, Ukraine,
            Russia

Now analyze the following sentence:
   Sentence: {sentence}
   Output format:
   Targets: <comma-separated targets or NONE>
```

Stance Classification. For each detected target, the model classifies the stance expressed in the sentence using the surrounding passage as context, thereby resolving pronouns and implicit references.

```
You are performing Stance Classification,
with context.

Task:
1. Read the full passage carefully.
2. Focus only on the highlighted sentence.
3. Given the detected target, classify the
   stance expressed in the specified sentence
   towards that target.
4. Additionally, provide a short reasoning
   (1-2 sentences) explaining why the stance
   is FAVOR, AGAINST, or NONE.

Labels:
  - FAVOR   = supportive or positive stance
  - AGAINST = negative stance
  - NONE    = neutral, descriptive, or
              unclear stance

Important:
  - Use the passage to resolve ambiguous
    references.
  - If the target is NONE, stance will also
    be NONE.
  - Keep the reasoning factual and grounded
    in the text.

---

### Examples:
1) Input:
     Passage: Ukrainian forces downed ...
     Sentence: "Ukrainian forces downed 22 of
     the 27 Shahed-type drones launched by
     Russia overnight on June 5, the Air
```

[6] `siebert/sentiment-roberta-large-english` (Hugging Face).

```
    Force said in its morning update."
    Target: Ukrainian forces
  Output:
    Stance: FAVOR
    Reason: The sentence highlights a
    successful action by Ukrainian forces,
    emphasizing their effectiveness in
    defense. It frames them positively by
    reporting their achievement in downing
    most of the drones.

---

Passage: {passage}
Specified Sentence: {sentence}
Target: {target}

Output format:
  Stance: <FAVOR / AGAINST / NONE>
  Reason: <brief reason for the stance>
```

Sentiment analysis is performed at sentence-level, for key entities like `Russia`, `Ukraine`, `Zelensky`, `Bucha`, `Kursk`, etc. For each entity, we (a) extract all sentences mentioning the entity using the English-language model of spaCy (Honnibal et al., 2020); (b) pass each sentence to the fine-tuned RoBERTa-large model (Hartmann et al., 2023); (c) map model outputs to continuous scores in `[-1, 1]`; and (d) aggregate the scores daily *and* monthly, by source *and* country.

```
Identify the sentiment (positive, negative,
or neutral) towards {entity} in this context:
"{sentence}".
```

This simple format provides sufficient context for the pretrained model, without requiring extensive prompt engineering.

Topical framing is done using the nine categories described in the schema presented in Table 3. We perform this multiclass classification using `phi-4` (Microsoft, 2024). The model receives an article's title, subtitle (if present), and the full text content. As output, it then provides applicable frames along with confidence scores.

```
You are a careful policy text classifier.
Your task is to identify the topical frames
expressed in the input text using a fixed set
 of predefined labels. You may assign one or
more labels if multiple topics are present.

Use only the following labels: ECONOMY,
IDENTITY, LEGAL, MORALITY, POLICY, POLITICS,
PUBLIC OPINION, SECURITY, WELFARE.

Definitions:
- ECONOMY: Macroeconomy and markets,
  including sanctions, trade, inflation,
  growth, employment, taxation, and energy
  or commodity prices.
- IDENTITY: Group identity and belonging,
  including nationality, ethnicity, religion,
  language, gender, discrimination, or
  identity politics.
- LEGAL: Courts and legal processes,
  including rulings, lawsuits, prosecutions,
  designations, warrants, and legal rights or
  entitlements.
- MORALITY: Explicit ethical or value-based
  framing, including moral duty, fairness,
  dignity, and humanitarian arguments.
- POLICY: Concrete government actions,
  including rules, bills, decrees, programs,
  and policy design details describing what a
  policy does.
- POLITICS: Elections, party or elite
  competition, political strategy, reshuffles,
  and coalition dynamics.
- PUBLIC OPINION: Expressions of public
  sentiment, such as polls, surveys,
  attitudes, or protests.
- SECURITY: War and defense, violence or
  crime, policing, border enforcement,
  terrorism, or cyberattacks.
- WELFARE: Social benefits and human services,
  including healthcare, pensions, housing aid,
  unemployment support, and refugee or
  internally displaced person assistance.

Return STRICT JSON only, with no additional
text:
    {
      "labels": [...],
      "confidences": {
        "<LABEL>": <0...1>
      }
    }

If no labels apply, return an empty list for
"labels".

---

Article:
  Title: {title}
  Content: {content}
```

Thematic tagging (see §5, use case 1: "A Thematic Study of Conflict Events") is done by extracting article-level tags using GPT-3.5 Turbo with one-shot prompting. The model is instructed to extract a set of specific tags summarizing each article's content. The tags were intended to capture salient persons or people, locations, military units, political or narrative themes, and ideological frames surrounding major conflict events.

To ensure a consistent structure that can support post-processing, the prompt was instructed to

produce outputs in a strict JSON format.

```
**Instruction:**
You are a geopolitical analyst trained to
extract concise topical tags from news
articles related to international conflict,
especially the Russo-Ukrainian war.

Your task is to read the content of a single
article and extract **5 to 12 relevant tags**
that capture the article's key locations,
actors, events, topics, or narrative themes.

Each tag should follow these rules:
- Tags must be **1-4 words long**.
- Use **lowercase**, unless the tag contains a
  **proper noun**.
- Tags should be **specific and informative**,
  not generic.
- Avoid vague tags such as "news", "fake", or
  "war" unless they are central to the
  article.
- Tags may include **people**, **institutions**,
  **military units**, **locations**, **investigations**,
  **political events**, **legal processes**, or
  **disinformation narratives**.

Use **only the article content** to determine the
tags.

Return only **valid JSON** with no explanations
or additional text.

**Tagging Guidelines:**

Tags may represent:
- Key **locations** (e.g., Bucha, Donbas, Odessa)
- Key **actors** (e.g., Sergei Lavrov, Budanov)
- **Events or incidents** (e.g., Bucha massacre,
  Kramatorsk attack)
- **Institutions or investigations** (e.g.,
  International Criminal Court, Bellingcat,
  NATO, Wagner)
- **Narratives or themes** (e.g., false flag
  allegations, media manipulation)
- **Evidence or technologies** (e.g., satellite
  imagery, intercepted communications)

**Output Format:**
Return the result in valid JSON using the
following structure:

    {
      "article_ID": [
        "tag 1",
        "tag 2",
        "tag 3",
        "tag 4",
        "tag 5"
      ]
    }

- Replace "article_ID" with the actual
  article identifier.
- Return **5-12 tags** per article.
- Do not include explanations or additional
  text.

**Example 1:**

**Title:** One tragic Ukraine story told with
drones, satellites and social media

**Content:**
Tragic individual stories of 21st century war
in Ukraine are coming into focus in 21st
century ways. [...] When Russia invaded,
Filkina stayed behind helping people in Bucha
and cooked for the Ukrainian military. [...]
Der Spiegel reported that Germany's foreign
intelligence agency intercepted Russian radio
chatter about the killing of civilians in
Bucha. [...] Russia is also apparently trying
to tell its own story by hacking into the
social media accounts of Ukrainian soldiers,
according to Meta. [...] The United Nations
voted Thursday to suspend Russia from the UN
Human Rights Council. [...] Organizations
like Bellingcat are using satellite and
social media to document war crimes and
identify perpetrators. [...] International
criminal investigations rely on
authenticated images, witness testimonies,
and other evidence concerning killings,
torture, and crimes against humanity. [...]

**Example Output:**

    {
      "article_example_1": [
        "Bucha killings",
        "Russian invasion",
        "war crimes evidence",
        "social media manipulation",
        "Ukrainian journalists",
        "UN human rights council suspension",
        "war crimes trials",
        "satellite imagery",
        "Bellingcat investigations",
        "International Criminal Court",
        "documentation of atrocities"
      ]
    }

---

Now extract tags from the following article.

Article ID: {{article_id}}
Title: {{title}}
Subtitle: {{subtitle}}
Content: {{content}}

Return only valid JSON.
```

Claim typology was established with the six types

of claims (see §5, use case 4: "Divergent Narratives"): factual, attributed, disputed, reported, command/policy, and analytical/assessment. We use a structured two-shot prompting strategy with GPT-3.5 Turbo to extract and classify claims.

Definition: A **claim** is a statement asserting a fact, report, allegation, policy decision, interpretation, or prediction related to the conflict.

Tag Definitions:
1. **Factual:** A statement presented as an objective fact without attribution to a specific actor or source.
2. **Attributed:** A claim explicitly attributed to a person, organization, government, or authority.
3. **Disputed:** A claim that is challenged, denied, or contradicted by another actor or source.
4. **Reported:** A claim framed as being reported by media, unnamed officials, or sources without direct confirmation.
5. **Command/Policy:** A directive, official order, government decision, or announced policy action.
6. **Analytical/Assessment:** Interpretation, analysis, speculation, prediction, or expert evaluation about the conflict.

Instruction:
You are a geopolitical analyst trained to extract claims from news articles about the Russia-Ukraine war.

Your task is to read the content of a single article and extract every distinct claim explicitly stated in the text.

Follow these rules carefully:
- Extract only claims explicitly present in the article text.
- Do not invent or infer claims not supported by the text.
- Claims may span one or multiple consecutive sentences if they form a single coherent statement.
- Do not split a single claim unnecessarily.
- Do not merge unrelated statements into one claim.
- Preserve the exact wording of the original sentence(s) in the "sentence" field.
- The "summary" must be concise, neutral, and factual.
- Use ONLY the article content for analysis.
- Extract all distinct claims present in the article.
- Return only valid JSON and no additional explanation.

Output Format:
Return the result in the following JSON structure:

```
    {
      "article_ID": {
        "claim_1": {
          "sentence": "Original sentence(s)
from the article expressing the claim.",
          "summary": "Short neutral summary
of the claim.",
          "claim_type": "One of the
predefined claim types"
        }
      }
    }
```

- Replace "article_ID" with the actual article identifier.
- Number claims sequentially (claim_1, claim_2, ...).
- The "sentence" field must contain the original sentence(s) from the article.
- The summary must be concise and neutral.
- The claim_type must match exactly one of the predefined categories.
- Return only JSON with no additional text or explanations.

Example 1:

Title: Russia Hits Macron, Shifting Political Dynamics Ahead of Elections in France

Content:
48.7 million French voters are officially registered to cast their ballot [...] Emmanuel Macron announced that a new round of sanctions against Russia was needed and said there are very clear clues pointing to war crimes in Bucha. [...] Green party's Yannick Jadot called for a full embargo on Russian gas and oil, criticising Macron's relationship with Russian President Vladimir Putin. [...] Support for Macron has now fallen to 27%, according to Politico's Poll of Polls, while Marine Le Pen's voting intentions have surged from 17% to 22% since the start of the war. [...]

Example Output:
```
{
  "article_example_1": {
    "claim_1": {
      "sentence": "Green party's Yannick
      Jadot called on Thursday for a full
      embargo on Russian gas and oil,
      criticising Macron's seemingly-warm
      relationship with Russian President
      Vladimir Putin.",
      "summary": "Jadot calls for embargo on
      Russian gas and oil and criticizes
      Macron's stance on Putin.",
      "claim_type": "Attributed"
```

```
    },
    "claim_2": {
      "sentence": "Support has now fallen to
      27%, according to Politico's Poll of
      Polls. Marine Le Pen has seen voting
      intentions in her favour surge from 17%
      to 22% since the start of the war.",
      "summary": "Macron's support decreases
      to 27% while Le Pen's rises to 22% in
      the polls.",
      "claim_type": "Reported"
    },
    "claim_3": {
      "sentence": "Emmanuel Macron announced
      that a new round of sanctions against
      Russia was needed and said there are
      very clear clues pointing to war crimes
      in Bucha.",
      "summary": "Macron calls for new
      sanctions against Russia and accuses
      them of war crimes in Bucha.",
      "claim_type": "Command/Policy"
    }
  }
}
```

Example 2:

Title: The horror of the Kramatorsk station attack

Content:
The Russian missile attack on Kramatorsk railway station that killed at least 50 people [...] It will fuel suspicions that Russia is targeting the civilian population to sow fear and break resistance after the setbacks its army has experienced in six weeks of war. [...] The reaction of Moscow officials and media followed a pattern of obfuscation, denial and outright fiction. [...] The different versions echoed the conflicting stories Russia used to try to distance itself from the downing in 2014 of Malaysia Airlines flight MH17. [...]

Example Output:

```
{
  "article_example_2": {
    "claim_1": {
      "sentence": "It will fuel suspicions
      that Russia is targeting the civilian
      population to sow fear and break resis-
      tance, after the setbacks its army has
      experienced in six weeks of war.",
      "summary": "The attack may reflect a
      strategy of targeting civilians to
      intimidate Ukraine.",
      "claim_type": "Analytical/Assessment"
    },
    "claim_2": {
      "sentence": "The different versions
      echoed the conflicting stories Russia
      used to try to distance itself from the
      downing in 2014 of Malaysia Airlines
      flight MH17.",
      "summary": "Russia has previously used
      conflicting narratives to distance
itself
      from incidents like the MH17 downing.",
      "claim_type": "Reported"
    }
  }
}

---

Now extract claims from the following article.


Article ID: {{article_id}}
Title: {{title}}
Subtitle: {{subtitle}}
Content: {{content}}

Return only valid JSON.
```

Appendix B. Use Case Experiments

Thematic Study of Conflict Events: We present word clouds in Figure 4 to visually support our thematic analysis (described in § 5. Using GPT-3.5 (OpenAI, 2023) with structured prompts, we tag articles from weekly time windows surrounding each major event to identify key concepts, actors, and subtopics. For the events described in Table 4, we obtain 810 articles, with 10.5 such tags (avg.) per article. We aggregate these tags by event and country to find the distinct thematic emphasis patterns across media outlets.

In summary, each word cloud aggregates GPT-3.5-tagged concepts, actors, and subtopics by outlet country for each of the four conflict events, with term size reflecting relative frequency. These visualizations highlight the varied thematic focus across geopolitical actors.

Temporal Sentiment Analysis: Figure 5 shows 5-day rolling average sentiment scores toward Russia and Ukraine in the immediate aftermath of the Bucha massacre (Mar - Apr 2022), aggregated by source country. Ukrainian media exhibits consistently negative sentiment toward Russia, as expected given their direct involvement in the conflict. Occasional positive spikes likely reflect optimistic coverage of military resistance or international support pledges. Russian media also shows a deeply negative sentiment – counterintuitive, given state alignment. This pattern suggests that war reporting employs vocabulary (violence, casualties, destruction) that sentiment models flag as inherently negative, regardless of political framing.

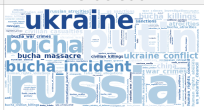
Event	CHINA	RUSSIA	UKRAINE	WEST	Wordcloud
Bucha Massacre	• bucha incident • china • peace talks • sanctions • un security council • civilian casualties • media manipulation • russia-ukraine conflict	• media manipulation • un security council • bucha incident • russian defense ministry • bucha • false flag allegations • sergey lavrov • ukraine	• bucha • war crimes • russian occupiers • ukraine • genocide • kyiv region • civilian casualties • bucha massacre	• war crimes • vladimir putin • bucha atrocities • bucha massacre • volodymyr zelensky • media manipulation • russian forces • bucha	
Mariupol Hospital Attack	• humanitarian corridors • mariupol • ceasefire • russian demands • vladimir putin • emmanuel macron • olaf scholz • ukraine crisis	• vladimir putin • humanitarian corridors • russia • russian military operation • donetsk people's republic • mariupol • denazification • russian defense ministry	• mariupol • civilian casualties • russian forces • donetsk oblast • azov battalion • sergei lavrov • mariupol bombing • war crimes	• humanitarian crisis • russian invasion • ukraine conflict • vladimir putin • russia-ukraine war • war crimes • civilian casualties • ukraine	
Zaporizhzhia Nuclear Crisis	• ukraine • china • nuclear facilities • russia-ukraine conflict • vladimir putin • russia • zaporizhzhia nuclear power plant • nato expansion	• russia • ukraine • nuclear weapons • vladimir putin • special military operation • russian military operation • western sanctions • fire incident	• international atomic energy agency • zaporizhzhia plant • nuclear safety • war in ukraine • russian aggression • intelligence reports • general staff • combat flights	• russia-ukraine war • vladimir putin • russian invasion • russia • no-fly zone • sanctions • russian forces • ukrainian resistance	
Kursk Offensive	• kursk • ukrainian army • russian armed forces • evacuation • kursk state • us military • russian defense ministry • civilian casualties	• kursk region • civilian casualties • russian defense ministry • ukrainian armed forces • ukraine • russia • fsb • valery gerasimov	• sumy oblast • russian forces • rocket attack • ukrainian authorities • mortar attacks • evacuations ordered • civilian casualties • border communities	• kursk region • ukraine • ukrainian incursion • russian territory • russia • evacuations • vladimir putin • russian response	

Figure 4: Thematic emphasis patterns across news outlets for four major conflict events, derived from GPT-3.5-tagged concepts, actors, and subtopics aggregated by outlet country.

Media in the USA and the UK exhibit pronounced volatility, with positive sentiment spikes corresponding to perceived diplomatic wins such as the US President's Warsaw speech or high-level visits to Kyiv. Chinese media maintains characteristically moderate sentiment, consistent with their "strategic partnership" framing, though scores decline during peak media coverage of Bucha in early April.

These patterns illustrate how sentiment analysis captures both explicit editorial stance and the underlying semantic content of war reporting, which may skew negative even if events are framed favorably.

Appendix C. Annotation Guidelines

We share the guidelines provided to human annotators for the validation experiments (see §4.4).

Topical Framing Annotation: The **task** is to identify the topical frames expressed in an article using the full text as context. Articles may receive multiple labels reflecting co-occurring themes. The **labels** are defined as per Table 3. Annotators are instructed as follows:

- assign labels based on explicit or implicit framing cues in the text;
- multiple labels are permitted and encouraged when themes co-occur;
- provide confidence scores `[0, 1]` for each assigned label; and
- avoid ranking labels by importance.

Stance Annotation: The **task** is to determine the stance expressed in a sentence toward a specified target, using the surrounding text as context. Each sentence receives exactly one of these **labels**:

(1) FAVOR: positive stance (support, praise, endorsement);
(2) AGAINST: negative stance (criticism, blame, opposition); or
(3) NONE: No clear stance, or purely descriptive text, or ambiguous stance.

The annotators are provided these instructions:

- *Use the full passage* to resolve ambiguous pronouns or implicit references.
- If the target is NONE, the stance is also NONE.
- Focus on the stance expressed in the sentence, not your personal opinion.
- Implicit stances (e.g., through selective vocabulary) should be labeled.

Table 8 and Table 9 provide annotated examples from the gold-standard human-annotated validation set, illustrating the topical framing and stance annotation tasks, respectively. For stance annotation, each sentence is paired with a stance label assigned toward a target entity. For topical framing, each passage may be associated with one or more topical frames, reflecting the co-occurrence of themes within the same article. The highlighted spans indicate the textual evidence that motivated the assigned labels. These highlights are intended to make the annotation rationale explicit and to ease reader interpretation; they were not provided to human annotators, and as such, the annotations are *not* derived from any span-level supervision.

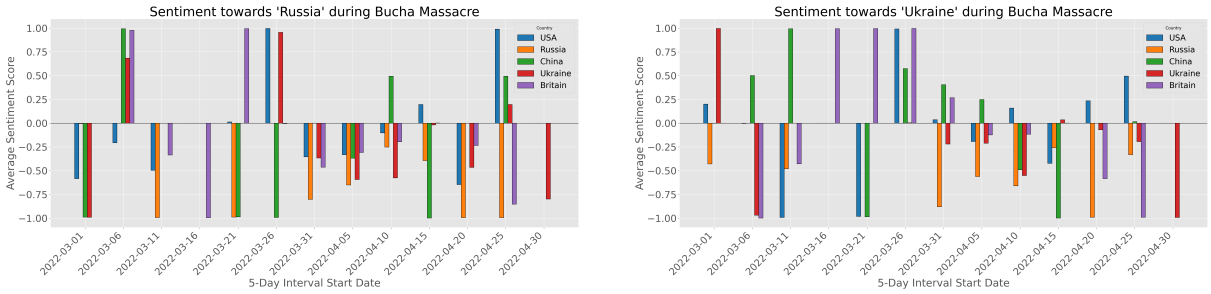


Figure 5: The average sentiment scores at 5-day intervals, for Russia (left) and Ukraine (right), during the aftermath of the Bucha massacre (March - April 2022). News outlets are grouped by their national affiliations.

Passage	**Frames**
Title: *Putin gives honorary title to Russian brigade accused of war crimes in Bucha* **Excerpt:** *A Russian brigade accused of committing war crimes in Bucha was awarded an honorary title by President Vladimir Putin. Putin congratulated the unit for its "great heroism and courage" and awarded it the title of "Guards" for "protecting Russia's sovereignty." International leaders condemned the alleged atrocities, while investigators cited "reasonable grounds" to believe war crimes were committed.*	• MORALITY: 0.9 • POLITICS: 0.6 • LEGAL: 0.6 • SECURITY: 0.6

Table 8: Topical framing annotation examples showing multi-label assignments with confidence scores.

Sentence (with target)	**Stance**
Russia has failed to capture Kyiv or Kharkiv, Ukraine's two biggest cities, and suffered heavy losses against a far stronger **Ukraine** than even its allies expected.	FAVOR
Aurel Braun, a professor at the University of Toronto who has written presciently on how western governments and companies misjudged the irrationalities of the **Putin regime**, points out that international relations is full of theories about how the interplay of powerful corporate interests and government produces policy.	AGAINST
When we see **Russia**'s unprovoked and unjustified invasion of Ukraine and massive disinformation campaigns and information manipulation, it is essential to separate lies from facts.	AGAINST
US officials have calculated that **Putin**, facing crippling sanctions across other sectors of his economy, can't afford to weaponize his energy supplies by cutting off sales to Europe.	NONE
Vladimir Putin awarded Heroes of the Russian Federation in the Yekaterininsky Hall.	NONE

Table 9: Stance annotations from the gold-standard human validation set, shown with highlighted target entities.