“A Special Operation”: A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems’ Coverage of the Russo-Ukrainian War

Hans W.A. Hanley, Deepak Kumar, Zakir Durumeric
Stanford University
hhanley@stanford.edu, kumarde@stanford.edu, zakird@stanford.edu

Abstract
The coverage of the Russian invasion of Ukraine has varied widely between Western, Russian, and Chinese media ecosystems with propaganda, disinformation, and narrative spins present in all three. By utilizing the normalized point-wise mutual information metric, differential sentiment analysis, word2vec models, and partially labeled Dirichlet allocation, we present a quantitative analysis of the differences in coverage amongst these three news ecosystems. We find that while the Western press outlets have focused on the military and humanitarian aspects of the war, Russian media have focused on the purported justifications for the “special military operation” such as the presence in Ukraine of “bio-weapons” and “neo-nazis”, and Chinese news media have concentrated on the conflict’s diplomatic and economic consequences. Detecting the presence of several Russian disinformation narratives in the articles of several Chinese media outlets, we finally measure the degree to which Russian media has influenced Chinese coverage across Chinese outlets’ news articles, Weibo accounts, and Twitter accounts. Our analysis indicates that since the Russian invasion of Ukraine, Chinese state media outlets have increasingly cited Russian outlets as news sources and spread Russian disinformation narratives.

1 Introduction
On February 24, 2022, Russian Federation President Vladimir Putin announced a “special military operation to demilitarize and denazify” Ukraine (Thompson and Myers 2022). Following the initial invasion, media outlets in various parts of the world covered the war in drastically different lights. For instance, the Western press (e.g., CNN, Fox News, New York Times) labeled the “special operation” a “war crime” laden “unprovoked invasion” perpetrated by the Russian government (Phillip 2022). Russian outlets (e.g., Sputnik News, Russia Today), in turn, have largely denied any war crimes, placing fault for the necessity of the “special operation” on Western countries (Thompson and Myers 2022). Chinese state media outlets (e.g., China Today, People’s Daily) meanwhile have advocated for diplomacy while simultaneously blaming Western powers for sparking the conflict. However, despite the evident differences and the large impact these differences have had on different populations’ perceptions of the war, there has not been a systematic analysis of the narratives present in these three media ecosystems (McCarthy and Xiong 2022). In this work, we present one of the first quantitative analyses of the key differences and similarities between the narratives touted by Chinese, Russian, and Western outlets about the Russo-Ukrainian War. Specifically, we propose an approach to determine and compare which topics media ecosystems report on, how each ecosystem reported on those topics, and finally whether there was influence between different media ecosystems (e.g., Russian to Chinese).

To perform our analysis, we first curate a dataset of 11,359 different articles about Ukraine from Western (4,536 articles from eight outlets), Russian (3,572 articles from ten outlets), and Chinese (3,251 articles from seven outlets) news ecosystems. Utilizing differential sentiment analysis and word2vec models, we then detail how each ecosystem has largely covered aspects of the Russo-Ukrainian War. Using a normalized scaled pointwise mutual information metric (NPMI) and partially labeled Dirichlet allocation (PLDA), we then extract the most characteristic words and topics from each ecosystem’s articles. We quantitatively show that while Western outlets have consistently labeled the “invasion” as a “war” and described the “crimes” committed throughout Ukraine, both Russian and Chinese news outlets have largely characterized the invasion as a “crisis” or a “conflict.” Similarly, while the Western press has focused on the humanitarian and day-to-day military aspects of the war, Russian outlets have focused on justifications for the “special military operations” like the presence of “bio-weapons” and “neo-nazis” in Ukraine and Chinese news outlets have concentrated on the diplomatic and economic fallout of the invasion.

After performing our comparative topic analysis, we identified the repeated presence of several Russian disinformation narratives (Price 2022) within the Chinese news ecosystem, particularly about US-funded Ukrainian biological weapons facilities. We thus measure the degree to which Russian news outlets have influenced Chinese news outlets’ coverage. Specifically, we document the frequency that seven different Chinese outlets use Russian news outlets as sources and their reuse of Russian-sourced images within their coverage of the war on their websites, Twitter accounts, and Weibo accounts (a Chinese version of Twitter). Observ-
ing a marked increase in Chinese state media citations of Russian sources beginning in early February 2022, we finally measure how an extended group of 39 Chinese media outlets interacted with and promoted Russian disinformation narratives on Weibo. Looking at the popularity of Russian outlets’ posts about Russian disinformation topics on Weibo, we find that these posts enjoyed higher levels of popularity compared to posts that do not reference these disinformation stories. Finally, fitting a negative binomial regression to model the number of Weibo posts from Chinese news outlets about different Russian disinformation narratives, we find that as Chinese news outlets cite more Russian sources, they are more likely to post disinformation.

Our work underscores the importance of performing analyses across multiple platforms and media ecosystems in understanding the nuances of how global events are framed, how different populations interpret and digest world events, and how disinformation originates and spreads. The Russo-Ukrainian War is a global event with global implications every country must consider ranging from skyrocketing international food prices, the resettlement of refugees, and threats of nuclear fallout (Treisman 2022; Thompson and Myers 2022); focusing only on news and campaigns targeted at Western to understand how populations are processing these implications can only go so far. We hope our quantitative approach can serve as the basis for future studies.

2 Methodology

To perform our comparative analysis of the attitudes, narratives, and topics discussed by the Western press, Chinese state media, and Russian propaganda websites, we collect a total of 11,359 unique news articles published between January 1, 2022, and April 15, 2022 (Table 1). To later understand the degree of Russian influence on Chinese media reports and discussions surrounding the Russo-Ukrainian War, we also collect posts from Weibo.

Table 1: We gather a set of English-language articles about Ukraine from Western, Russian, and Chinese media ecosystems.

<table>
<thead>
<tr>
<th>Western</th>
<th>Articles</th>
<th>Russian</th>
<th>Articles</th>
<th>Chinese</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>cnn.com</td>
<td>477</td>
<td>tass.com</td>
<td>861</td>
<td>chinadaily.com.cn</td>
<td>1044</td>
</tr>
<tr>
<td>nytimes.com</td>
<td>571</td>
<td>sputniknews.com</td>
<td>743</td>
<td>cgtv.com</td>
<td>966</td>
</tr>
<tr>
<td>washingtonpost.com</td>
<td>414</td>
<td>news-front.info</td>
<td>508</td>
<td>globaltimes.cn</td>
<td>367</td>
</tr>
<tr>
<td>theguardian.com</td>
<td>636</td>
<td>geopolitica.ru</td>
<td>80</td>
<td>ecns.cn</td>
<td>702</td>
</tr>
<tr>
<td>yahoo.com</td>
<td>774</td>
<td>southfront.org</td>
<td>241</td>
<td>xinhuanet.com</td>
<td>430</td>
</tr>
<tr>
<td>reuters.com</td>
<td>716</td>
<td>katehon.com</td>
<td>65</td>
<td>pdnews.cn</td>
<td>286</td>
</tr>
<tr>
<td>foxnews.com</td>
<td>566</td>
<td>journal-neo.org</td>
<td>90</td>
<td>english.cctv.com</td>
<td>107</td>
</tr>
<tr>
<td>nbcnews.com</td>
<td>382</td>
<td>rt.com</td>
<td>689</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>strategic-culture.org</td>
<td>102</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>waronfakes.com</td>
<td>193</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>4536</td>
<td>3572</td>
<td>3251</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For our list of Western outlets, we manually selected eight highly popular mainstream news websites from across the political spectrum (Zannettou et al. 2017). In addition to a set of nine Russian websites identified by the US State Department (Rus 2020), for our Russian dataset, we include the recently launched waronfakes.com. Since its initial appearance in March 2022, the New York Times and others have investigated the site as a hub of Russian disinformation (Thompson and Myers 2022; Hanley, Kumar, and Durumeric 2023). For our list of Chinese media news websites, we utilize seven English-language news websites identified by the US State Department as “foreign missions” (Ortagus 2020). We recognize that these lists do not incorporate all articles circulated in each media ecosystem and thus are naturally biased. However, our selection of websites do represent a cross-section of some of the most widely circulated news sources in each ecosystem and thus provide indications of reporting for Western (You 2022), Russian (Rus 2020), and Chinese (Ortagus 2020) news media.

We utilize a breadth-first scraping algorithm and the Python Selenium package to collect the set of English-language articles that each of our websites published about Ukraine. Specifically, for each website, we scrape 5 hops from the root page (i.e., we collect all URLs linked from the homepage [1st hop], then all URLs linked from those pages [2nd hop], and so forth). To get Ukraine-related articles, for each website page, we use the Python newspaker3k library to collect article contents and to determine if the article mentions “Ukraine”. We further supplement this corpus by using Google’s API to find and add articles indexed in 2022 that mention Ukraine. We note that due to the lack of precision in acquiring the publication date of each article with newspaker3k, we utilize the Python library htmlparser3k to extract each article’s publish date. Altogether, between January 1, 2022, and April 15, 2022, we collect 11,359 articles about Ukraine; 4,536 from Western Press outlets, 3,572 from Russian propaganda websites, and 3,251 from Chinese state media (Table 1).

Weibo Dataset. To understand the degree of Russian influence on Chinese media reports and discussions surrounding the Russo-Ukrainian War, we also collect posts from Weibo,
a Chinese Mandarin-language version of Twitter (McCarthy and Xiong 2022). We collect the posts of the accounts of the seven different Chinese state media organizations from our news article dataset (for the CGTN news organization, we collect the Weibo posts of its CGTN and CGTN journalist group/CGTN记者团 accounts). To help quantify the connection of each of these media organizations to the Russian government and Russian state media, we further scrape the accounts of the Russian Embassy/@RussianEmbassy, Russia Today/@RT, and Sputnik News/@SputnikInt. Lastly, in addition to our Chinese news organizations’ Weibo accounts and Russian state-sponsored Weibo accounts, we collect the Weibo posts of the 200 users who most prominently discussed the Russo-Ukrainian conflict at the end of February as labeled by Fung et al. (Fung and Ji 2022). This list was manually created from users who “actively posted about and ranked among the top posts of trending hashtags related to the Russo-Ukrainian war.” After combining our lists of Weibo users, and removing inactive and duplicate accounts, we had a total of 191 distinct accounts. For each account in our dataset, we scraped the account on four occasions (March 14, March 28, April 06, and April 16) to ensure our dataset was comprehensive. To scrape each Weibo account, we utilize the Python weibo-scraper tool. Ultimately, our dataset consists of 191 different accounts and 343,435 distinct Weibo posts from between January 1 and April 15, 2022.

**Twitter Dataset.** In addition to our Weibo dataset, we further collect the tweets of the seven different news Chinese news outlets within our news article dataset (China Daily, CGTN, Global Times, Chinese News Service, Xin Hua, People’s Daily, and CCTV). Unlike for our Weibo dataset, we do not collect the set of Chinese users who most prominently discussed the Russo-Ukrainian conflict on Twitter (Twitter has been banned in China since 2009 (Barry 2022)), limiting our Twitter analysis to these seven major state-sponsored Chinese outlets who also regularly tweet. To investigate these accounts’ connection to the Russian government and Russian news media, we again collect the tweets of the RussianEmbassy/@RussianEmbassy, Russia Today/@RT.com, and Sputnik News/@SputnikInt. We collect the tweets of each account using the Tweepy API (Roesslein 2009) on four different instances (March 06, March 13, April 02, and April 16). Ultimately, our Twitter dataset consists of 62,717 unique tweets from 10 different accounts from January 1 and April 15, 2022.

**Pointwise Mutual Information.** To determine different news ecosystems’ associations with distinct words, we utilize the normalized pointwise mutual information metric. Pointwise mutual information (PMI) is an information-theoretic measure for discovering associations amongst words (Bouma 2009). However, as in Kessler et al., rather than finding the pointwise mutual information between different words, we utilize this measure to understand words’ association with different categories (Kessler 2017). In this way, we seek to identify the characteristic words of each ecosystem’s coverage of the Russo-Ukrainian War (i.e., Western, Chinese, and Russian media). We utilize the normalized and scaled version of PMI to prevent our metric from being biased towards rarely occurring words and to increase interpretability. Scaled normalized PMI (NPMI) for a word $i$ and each category $C_j$ is calculated as follows:

$$PMI(word_i, C_j) = \log_2 \frac{P(word_i, C_j)}{P(word_i)P(C_j)}$$

$$NPMI(word_i, C_j) = \frac{PMI(word_i, C_j)}{-\log_2(P(word_i, C_j))}$$

where $P$ is the probability of occurrence and a scaling parameter $\alpha$ is added to the counts of each word. NPMI ranges between (-1,1). We choose $\alpha = 50$ given the size of our dataset (Turney 2001). An NPMI value of $-1$ represents that the word and the category never occur together (given that we utilize the scaled version this never occurs), 0 represents independence, and $+1$ represents perfect co-occurrence (Bouma 2009). Finally, before computing NPMI on our dataset, we first lemmatize and remove stop words as in prior work (Zannettou et al. 2020).

**Partially Labelled Latent Dirichlet Allocation.** In addition to identifying words characteristic of each news ecosystem, we also extract the set of topics that are distinctive to each ecosystem. To do this, we utilize Partially Labelled Dirichlet Allocation (PLDA). PLDA is an extension of the widely-used topic analysis algorithm Latent Dirichlet Allocation (LDA) (Ramage, Manning, and Dumais 2011). PLDA, like LDA, assumes that each document is composed of a distribution of different topics (which themselves are composed as a distribution of different words). However, unlike LDA, each document can form topics from a pool associated with one or more of its specific labels. For example, a newspaper article from nytimes.com, which is labeled as “Western”, can draw from a set of labeled topics associated with “Western” (as opposed to an article from chinadaily.com.cn which can draw from a set of labeled topics associated with “Chinese”). In addition to drawing from the distribution of topics associated with its labels, documents also further draw from a pool of latent topics that are associated with every document in the dataset. PLDA can thus model the topics that are common to every document while also identifying discriminating topics for each label (i.e., topics specific to “Western”, “Russian”, “Chinese”).

Again when fitting our PLDA model, we first lemmatize and remove stop words. When computing topics, we further weight words using term-frequency inverse document frequency (TF-IDF). Previous work has shown that this weighting leads to more accurate topics (Zannettou et al. 2020). To find the appropriate amount of topics, we optimize the word2vec topic coherence score $c_v$ that measures the semantic similarity among extracted topic words (Zannettou et al. 2020). We utilize a baseline number of 300 latent topics, varying the number of topics per label from 1 to 20. We achieve the best coherence score of 0.46 with 15 topics associated with each label (345 total topics).
### 3 Comparing The Coverage of the War

In this section, we perform a quantitative study of the shared and distinctive narratives and topics discussed within the Russian, Chinese, and Western ecosystems.

#### Media Ecosystems’ Shared Topics

We begin our analysis by highlighting the topics (Table 2), as elicited by our PLDA model, that are shared between our three distinct media ecosystems from articles published between January 1 and April 15, 2022.

The most common shared topic—by article count—refers to military activities in the build-up to the war (137 Western articles, 90 Russian articles, 70 Chinese articles). Starting in the Spring of 2021 and continuing into February of 2022, Russia began to amass troops on their shared border with Ukraine (Schmollinger and Singh 2022). Each ecosystem wrote a correspondingly large amount about the growth of the Russian military presence. Each ecosystem further discusses the fallout following the invasion. 214 different articles (Topic 2) focused on US and EU sanctions placed on Russia (81 Western articles, 43 Russian articles, 91 Chinese articles). Each ecosystem also discusses the North Atlantic Treaty Organization’s (NATO’s) role in the conflict (47 Western articles, 81 Russian articles, 58 Chinese articles), with Russian articles, in particular, emphasizing its role. The North Atlantic Treaty Organization/NATO was founded in 1949 as a defensive military alliance between the United States, Canada, and several Western European countries against the threat of the Soviet Union. The organization

<table>
<thead>
<tr>
<th>Topic</th>
<th>Keywords</th>
<th>Western Articles</th>
<th>Russian Articles</th>
<th>Chinese Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>border, troops, exercises, belarus, eastern, invade, tensions, imminent, attack, borders</td>
<td>137 (3.0%)</td>
<td>90 (2.5%)</td>
<td>70 (2.2%)</td>
</tr>
<tr>
<td>2</td>
<td>sanctions, biden, tuesday, thursday, statement, address, announced, meeting, impose, united</td>
<td>81 (1.7%)</td>
<td>43 (1.2%)</td>
<td>91 (2.8%)</td>
</tr>
<tr>
<td>3</td>
<td>city, shelling, killed, forces, destroyed, reported, residents, building, southern, near</td>
<td>74 (1.6%)</td>
<td>63 (1.8%)</td>
<td>60 (1.8%)</td>
</tr>
<tr>
<td>4</td>
<td>refugees, refugee, poland, million, children, fleeing, ukrainians, fled, border, polish</td>
<td>98 (2.1%)</td>
<td>22 (0.6%)</td>
<td>69 (2.1%)</td>
</tr>
<tr>
<td>5</td>
<td>nato, alliance, europe, expansion, security, soviet, west, treaty, european, eastward</td>
<td>47 (1.0%)</td>
<td>81 (2.3%)</td>
<td>58 (1.8%)</td>
</tr>
</tbody>
</table>

Table 2: The top five shared latent topics (in terms of the number of articles) in the Western, Russian, and Chinese ecosystems.
adopted the agreement, called Article 5, that an attack on one member of the alliance would be treated as an attack on all members (NAT 2022). Since the fall of the Soviet Union, several new countries have joined the defensive pact including Hungary, Poland, and the Czech Republic, with North Macedonia being the newest member (2020), leading to increased tensions with Russia (NAT 2022).

Finally, each ecosystem has written extensively about the Ukrainian refugee crisis, with 98 western articles, 22 Russian articles, and 58 Chinese articles documenting the displacement of more than 6 million Ukrainians (Wlodarczak-Semczuk, Chopra, and Porter 2022).

Comparing Shared Coverage with Differential Sentiment Analysis. We note that despite having many articles that share topics, each ecosystem often described these topics differently. To capture how each ecosystem generally wrote about the war, we thus run sentiment analysis using the labMT sentiment analysis toolkit (Dodds et al. 2011). labMT grades individual words on a scale from 1 (most negative) to 9 (most positive).

We find that each ecosystem wrote about the Russo-Ukrainian War somewhat neutrally. Western coverage was the most negative with an average article sentiment of 5.42; Russian coverage came next at 5.43 and Chinese coverage was the most positive at 5.62. We confirm this ranking by further extracting the average article sentiment using the Vader sentiment Python library (scores on a scale of -1 [most negative] to +1 [most positive]) getting scores of -0.089, -0.072, and 0.017 for the Western, Russian, and Chinese ecosystems respectively (Hutto and Gilbert 2014).

To help explain the differences in sentiment scores between individual ecosystems, we next identify which negatively and positively polarized words contributed the most to each dataset’s sentiment score; we visualize this in pairwise graphs in Figures 1, 2, 3. These graphs compare the relative shift in the labMT sentiment score caused by the frequency of negatively and positively polarized words. Looking at the contribution of words to each ecosystem’s sentiment score, we see that the Western media’s consistent reference to the Russo-Ukrainian conflict as a “war” and an “invasion” contributes the most to its relative negativity (Figures 1 and 2). In contrast, the Russian ecosystem, largely calls the Russo-Ukrainian War a “special operation”, a “conflict”, or a “crisis” (Figure 1).

The Western usage of more violent and negative language is further mirrored in differential sentiment between the Western press and Chinese media (Figure 2). The Western press commonly refers to “attacks”, “killing”, “crimes”, “assaults”, and “refugees.” The Chinese media ecosystem, in contrast, is more likely to write about prospects of “peace”, talk about “security”, or suggest “cooperation” (Figure 2). Furthermore, the Chinese media ecosystem has largely avoided using harsh language to describe the Russo-Ukrainian War, using words like “war”, “attacks”, and “invasion”, even less frequently than in the Russian media ecosystem (Figure 3). This largely explains why the Chinese media ecosystem remained more positive than both the Western and Russian media, despite all outlets discussing similar war-related topics.

Dissecting Shared Coverage with Word2Vec. To further understand the differences in coverage of similar topics/events between the media ecosystems, we now focus on how each discusses three distinct locales: “Bucha”, “Mariupol”, and “Ukraine”. We choose “Bucha” and “Mariupol” to cover two of the major cities/events in the Russo-Ukrainian War. In early April 2022, mass graves were discovered in Bucha, leading to outcries about Russian atrocities. The Russian government has denied these claims even as Ukrainian officials have claimed that over 8,000 war crimes were committed in Bucha (Cheng 2022). We choose Mariupol given its pivotal role in the Russo-Ukrainian War. Mariupol became the epicenter of a humanitarian disaster (Kirby 2022). We finally choose “Ukraine” as “Ukraine” covers attitudes of each ecosystem towards the country and the war generally.

To uncover how each ecosystem has covered our three lo-
Table 3: Top twenty most similar terms (by cosine similarity) in each news ecosystem to “Bucha”, “Mariupol”, and “Ukraine.”

<table>
<thead>
<tr>
<th>Rank</th>
<th>Western</th>
<th>Russian</th>
<th>Chinese</th>
<th>Western</th>
<th>Russian</th>
<th>Chinese</th>
<th>Western</th>
<th>Russian</th>
<th>Chinese</th>
<th>Western</th>
<th>Russian</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>atrocity</td>
<td>massacre</td>
<td>killing</td>
<td>besiege</td>
<td>volnovakha</td>
<td>volnovakha</td>
<td>country</td>
<td>country</td>
<td>underway</td>
<td>prepare</td>
<td>aim</td>
<td>postpone</td>
</tr>
<tr>
<td>2</td>
<td>slay</td>
<td>killing</td>
<td>atrocity</td>
<td>city</td>
<td>city</td>
<td>besiege</td>
<td>russia</td>
<td>demiilitarize</td>
<td>disarm</td>
<td>continue</td>
<td>creation</td>
<td>russia</td>
</tr>
<tr>
<td>3</td>
<td>suburb</td>
<td>atrocity</td>
<td>grave</td>
<td>devastated</td>
<td>kharkiv</td>
<td>shelling</td>
<td>continue</td>
<td>creation</td>
<td>russia</td>
<td>prepare</td>
<td>aim</td>
<td>postpone</td>
</tr>
<tr>
<td>4</td>
<td>gruesome</td>
<td>mayor</td>
<td>exhumed</td>
<td>port</td>
<td>sunny</td>
<td>bombardment</td>
<td>rearm</td>
<td>denazify</td>
<td>dangerous</td>
<td>hasten</td>
<td>kiev</td>
<td>minimal</td>
</tr>
<tr>
<td>5</td>
<td>killing</td>
<td>town</td>
<td>disturbing</td>
<td>besieged</td>
<td>liberate</td>
<td>mykolayiv</td>
<td>impeding</td>
<td>wake</td>
<td>calm</td>
<td>uninterrupted</td>
<td>pledge</td>
<td>regard</td>
</tr>
<tr>
<td>6</td>
<td>massacre</td>
<td>footage</td>
<td>incident</td>
<td>boychenko</td>
<td>azov</td>
<td>kherson</td>
<td>impending</td>
<td>wake</td>
<td>calm</td>
<td>uninterrupted</td>
<td>pledge</td>
<td>regard</td>
</tr>
<tr>
<td>7</td>
<td>body</td>
<td>body</td>
<td>dead</td>
<td>boichenko</td>
<td>blockade</td>
<td>melitopol</td>
<td>impending</td>
<td>wake</td>
<td>calm</td>
<td>uninterrupted</td>
<td>pledge</td>
<td>regard</td>
</tr>
<tr>
<td>8</td>
<td>grisly</td>
<td>image</td>
<td>scene</td>
<td>siege</td>
<td>afu</td>
<td>berdyansk</td>
<td>monthslong</td>
<td>prompt</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
<td>regard</td>
</tr>
<tr>
<td>9</td>
<td>discover</td>
<td>dead</td>
<td>massacre</td>
<td>trap</td>
<td>militia</td>
<td>rouble</td>
<td>unabated</td>
<td>pledge</td>
<td>regard</td>
<td>uninterrupted</td>
<td>pledge</td>
<td>regard</td>
</tr>
<tr>
<td>10</td>
<td>atrocity</td>
<td>alleged</td>
<td>atrocity</td>
<td>volnovakha</td>
<td>surround</td>
<td>northeastern</td>
<td>Russia</td>
<td>flare</td>
<td>none</td>
<td>prepare</td>
<td>aim</td>
<td>postpone</td>
</tr>
<tr>
<td>11</td>
<td>fedorov</td>
<td>evidence</td>
<td>strew</td>
<td>berdiansk</td>
<td>corridor</td>
<td>theater</td>
<td>continue</td>
<td>creation</td>
<td>russia</td>
<td>hasten</td>
<td>kiev</td>
<td>minimal</td>
</tr>
<tr>
<td>12</td>
<td>corpse</td>
<td>murder</td>
<td>mayor</td>
<td>kherson</td>
<td>retreat</td>
<td>boychenko</td>
<td>continue</td>
<td>creation</td>
<td>russia</td>
<td>hasten</td>
<td>kiev</td>
<td>minimal</td>
</tr>
<tr>
<td>13</td>
<td>sickness</td>
<td>torture</td>
<td>torture</td>
<td>mangush</td>
<td>hospital</td>
<td>sumy</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>14</td>
<td>unspeakable</td>
<td>accusation</td>
<td>alleged</td>
<td>vershynin</td>
<td>militant</td>
<td>odesa</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>15</td>
<td>unearth</td>
<td>staged</td>
<td>image</td>
<td>berdyansk</td>
<td>kharkov</td>
<td>energodar</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>16</td>
<td>horrifying</td>
<td>picture</td>
<td>theater</td>
<td>encircle</td>
<td>regiment</td>
<td>toksak</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>17</td>
<td>hostomel</td>
<td>fake</td>
<td>staged</td>
<td>remained</td>
<td>village</td>
<td>southeastern</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>18</td>
<td>borodyanka</td>
<td>film</td>
<td>prosecutor</td>
<td>battered</td>
<td>kherson</td>
<td>mayor</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>19</td>
<td>rapist</td>
<td>appear</td>
<td>incident</td>
<td>izium</td>
<td>uaf</td>
<td>mykolayiv</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
<tr>
<td>20</td>
<td>izium</td>
<td>lie</td>
<td>protester</td>
<td>mykolayiv</td>
<td>battalion</td>
<td>deport</td>
<td>unloading</td>
<td>ongoing</td>
<td>mount</td>
<td>continue</td>
<td>uninterrupted</td>
<td>pledge</td>
</tr>
</tbody>
</table>

Table 3: Top twenty most similar terms (by cosine similarity) in each news ecosystem to “Bucha”, “Mariupol”, and “Ukraine.”
on crimes committed in Ukraine and the plight of Ukrainian refugees. The most discriminating topic specifically concerns the city of Bucha where mass graves were discovered in early April (Cheng 2022).

In contrast to the larger themes particular to the Western media, using our NPMI metric, we see the most characteristic words of the Western ecosystem to have do the with seemingly innocuous spelling of Ukrainian cities (e.g., Kyiv) and the president of Ukraine, Volodymyr Zelensky’s, name (Table 5). We note, however, that the spelling of these entities has taken on political meaning (Dickinson 2019). Different ecosystems, in their reporting on events in Ukraine, have decided either to utilize the Ukrainian or Russian spelling of words. Here we see that the Western ecosystem, in contrast to both the Russian and Chinese ones, utilizes Ukrainian spellings. We further see this mirrored in the most uncharacteristic words within this dataset (Table 6), where we see “Kiev, Lugansk, and Kharkiv” rather than their Ukrainian spellings of “Kyiv, Luhansk, and Kharkiv.”

**Russian Media Coverage.** In contrast to Western media, Russian media articles focus on the Donbas region of Ukraine and the portrayal of the Ukrainian government and its military (AFU) as being filled with “neo-nazis.” Notably, as seen in Table 4, Russian media (39 articles) have a keen focus on the Donbas region of Ukraine, which is partly controlled by two separatist regimes: the Donetsk People’s Republic (DPR) and the Luhansk People’s Republic (LPR). While still parts of Ukraine, significant efforts from the Russian media ecosystem have gone into demonstrating a distinction between these area’s regimes and internationally recognized Ukrainian government. The day before the Russian invasion of Ukraine, the leaders of DPR and LPR called for Russia to help “repel Kiev’s aggression”, further ostensibly prompting the Russian government to invade and “liberate” the regions. With Russian media recognizing these regimes, the DPR and LPR are some of the most characteristic words in the Russian news ecosystem; and simultaneously the least descriptive of the Western news ecosystem (Tables 5 and 6).

We further observe a focus on Ukraine being filled with “neo-nazis”. The word “neo-nazi” is one of the most characteristic words utilized by Russian media (Table 5) and a part of the most discriminating Russian topic (Table 4). As previously noted, a major aspect of the Russian claim that Ukraine required “denazification” was that the Azov battalion was a part of Ukraine’s military; (again) this despite Azov’s depoliticization and Ukraine’s low level of antisemitism (Shekhovtsov 2020; Masci 2018). Ukraine’s current president Volodymyr Zelensky is even Jewish (Thompson and Myers 2022). Lastly, we note the disinformation narrative about biological weapons that appeared as the third discriminating topic

---

**Table 4: The top five (in terms of the number of articles) discriminating topics for each media ecosystem.**

**Table 5: The most characteristic words for each media ecosystem.**

**Table 6: The most uncharacteristic words for each media ecosystem.**

---

3https://web.archive.org/web/20220501202030/https://tass.com/politics/1409091


---

345
for Russian articles. On March 6, 2022, the Russian news reporting website Tass reported that the US had been funding biological weapons development within Ukraine and had subsequently destroyed the facilities upon the Russian invasion. This narrative was debunked by the US State Department (Price 2022) and the New York Times (Qiu 2022).

**Chinese Media Coverage.** As seen in Table 4 and Table 5, the most discriminating aspects of the Chinese media are its concerns with the economic and diplomatic fallout of the war in Ukraine. The largest discriminating topic that was uncovered by PLDA concerned “negotiations” and attempts to broker peace. We see this echoed further in the second, fourth, and fifth most discriminating topics. This largely matches the view found through the sentiment analysis approach in Section 3.

In addition to emphasizing “negotiations”, we further observe a heavy emphasis on the economic ramifications of the Ukraine crisis on the economic health of China and the rest of the globe. Several of the most distinctive words, and the sixth most discriminating topic discovered (not shown), discuss primarily the effect that the Russo-Ukrainian conflict will have on food prices. Because Russia and Ukraine together comprise about 30% of global wheat exports and 20% of maize exports, the war has caused major disruptions to commodity prices. The United National Food and Agricultural Organization found that food commodity prices went up 12.6% between February and March of 2022 (Treisman 2022). This particular worry about food prices was echoed widely in the Chinese news ecosystem with a word like “percent”, “growth”, and “sustainable” being characteristic of its news reporting on Ukraine.

We additionally note that the Russian disinformation narrative of US-funded Ukrainian biological laboratories found significant traction within Chinese media. The Chinese spokesman for the Foreign Ministry even tweeted a video directly from Russia Today about these supposed biological weapons laboratories and we see this largely throughout the Chinese news ecosystem (Table 4) with 103 articles.

**Summary**

The Western, Russian, and Chinese news ecosystems have largely utilized divergent narratives in covering the Russo-Ukrainian war. In describing the invasion, Western media has utilized Ukrainian spellings of Ukrainian towns and officials and have highlighted the military and humanitarian crises taking place in Ukraine. Chinese outlets have avoided describing some of the more grisly aspects of the invasion, referring to Russo-Ukrainian War as a “conflict” and “crisis”, causing the overall sentiment score of their articles to be somewhat higher than the scores of Russian the Western media. Finally, the Russian media ecosystem has largely focused on the purported reasons for the “special military operation”, “the denazification” of Ukraine, and the supposed “liberation” of Ukraine’s people.

---


**Figure 4:** Citations of Russian news sources (Tass, RIA, Sputnik News, and RT) by Chinese media outlets.

**4 Russian Influence on Chinese State Media**

Despite differences in the topics discussed by Chinese and Russian ecosystems, we observed that Chinese media outlets echoed narratives from Russian news sources. In particular, Chinese news outlets amplified the disinformation narrative that the US had funded Ukraine-based biological weapons laboratories. We saw a similar Russian influence in the coverage of the massacre at Bucha (Table 3). In this section, we document the degree to which Russian state media have influenced Chinese media coverage and discussions of the Russo-Ukrainian War in their news articles, on Weibo, and on Twitter. We take a multi-modal approach, quantifying influence along two axes: (1) citations of Russian outlets as news sources and (2) reuse of Russian-sourced images.

**Chinese Media’s Use of Russian News Sources**

We first examine citations of Russian news sources within Chinese state media reporting. Specifically, we look at Chinese media’s citations of the Russian state-controlled news agencies Tass, Russia Today, Sputnik News, and RIA Novosti (the prominent news agency that started Sputnik News (Feinberg 2017)) as sources for reporting about Ukraine across online news articles, on Weibo, and on Twitter. As an example of what we consider a citation, on May 13th, the Twitter account of CGTN tweeted: *Leader of South Ossetia sets referendum on joining Russia on July 17, TASS reported.* For each mention of the Russian outlets across each platform, we manually verify that the Chinese state outlet utilized them as a news source rather than just mentioning the outlet (for our Weibo, one of the authors can read Mandarin Chinese).

The number of citations of Russian news sources in Chinese news reporting on Ukraine saw a sharp uptick beginning in February (Figure 4). We note that this peak occurred across all platforms. We further note that across every platform, each Chinese news source utilized a Russian news organization at least once in their reporting about Ukraine (Table 7). Global Times on Weibo, CGTN on Twitter, and China Daily within their news articles most consistently relied on Russian news sources. As seen in Table 8, this is particularly true for news sources like Tass and RIA Novosti, which...
Weibo | Russian Citations | Russian “#” | Russian News Outlet | Weibo | Twitter | News Articles
--- | --- | --- | --- | --- | --- | ---
中国人民网(China News Service) | 142 | 63 | Echinanews | 36 | 804 | ecns.cn 26
中央日报(China Daily) | 65 | 76 | ChinaDaily | 12 | 2178 | chinadaily.com.cn 68
CGTN/CCTV记者团 | 69 | 87 | CGTNOfficial | 468 | 3487 | cgtv.com 45
人民日報(People’s Daily) | 6 | 20 | PDCChina | 65 | 826 | pdnews.cn 14
央視新聞(CCTV) | 11 | 68 | CCTV | 3 | 34 | cctv.com 6
环球时报(GLOBAL TIMES) | 395 | 83 | globaltimesnews | 33 | 2729 | globaltimes.cn 27
新华网(Xinhua News) | 35 | 27 | XHNews | 21 | 804 | xinhuanet.com 35

Table 7: The number of posts for each Chinese outlet per platform referencing a Russian news organization as a news source or posting a hashtag utilized by Russian state-sponsored accounts between January 1 and April 15, 2022.

are consistently utilized across different platforms. Looking at Twitter and Weibo, every outlet similarly also extensively tweeted and posted to Weibo utilizing hashtags utilized by Russian state-sponsored accounts (Sputnik News, Russia Today, and the Russian Embassy).

Looking at our extended set of users on Weibo, beyond our core set of Chinese media outlets and Russian entities (RT, Russian Embassy, and Sputnik News), we identify another 4,478 references to Russian news outlets among the remaining users. Among these users are several other Chinese news outlets: Reference News/参考消息 585 times, Sina News/新浪军事 369 times, and Observer News/观察者网 310 times. These same outlets utilize hashtags on Weibo shared by RT, Russian Embassy, and Sputnik News 96, 113, and 75 times within the same period respectively (the 2nd, 1st, and 5th most often in our dataset respectively). Overall, across every non-Russian Weibo account (188 accounts), we see that 132 (70.2%) reference a Russian news outlet, and all use hashtags also shared by Russian state-sponsored Weibo accounts. Collectively, the posts that cited Russian sources received 10,792,782 reactions (Weibo version of likes), with a median of 114 reactions per post. As an aside, the most popular post (579,146 reactions), submitted by Sina News/新浪军事 and citing the RIA news agency, described the destruction of Ukrainian ships in Odesa.

Chinese Media’s Reuse of Russian-sourced Images

To understand the degree to which Russian-sourced imagery has influenced the Chinese media and users, we determine what percentage of their Ukraine-related images came from Russian sources. We utilize the Google Cloud Vision API’s reverse search functionality to determine the set of an image’s possible sources. Returned with this search are the entities associated with the photo and a list of other website pages that posted an identical image. We consider an image to be Ukraine-related if one of the entities returned with the image is “Ukraine.” We consider a Ukrainian-related image to be of Russian origin if it was first posted on a .ru website or one of our Russian news websites prior to when it was posted on Weibo. We visit each page returned with our reverse-image search with Selenium and utilize the Python library `htmlDate` to acquire the publish date to ensure that the image was posted on a Russian page first.

---

*https://cloud.google.com/vision/

Table 8: Citations by Chinese news outlets to Russian news sources between January 1 and April 15, 2022.

We first perform a reverse image search on the images posted within the tweets and Weibo posts of our set of seven Chinese state media outlets. Altogether this consists of 9,393 images across 29,052 Weibo posts and 7,809 images across 44,804 unique tweets. After running our set of Weibo images through Google’s Cloud API, we identify only 18 (0.2%) different images that were of Russian origin, 6 of them coming from rt.com and 4 from sputniknews.com. 10 of these images were posted by Global Times/环球时报, 4 by CGTN, 2 from People’s Daily, 1 from China Daily, and 1 from CCTV. Among the 7,809 images posted by Chinese state media Twitter accounts, only 6 (0.07%) photos were of Russian origin. For Chinese state media, we thus see that they largely did not utilize imagery from Russian sources.

While Chinese media outlets did not heavily utilize Russian-sourced/associated images, we do see elevated usage of Russian-sourced imagery amongst the most prominent Ukraine-related Weibo accounts. Of the 96,614 images posed by these accounts between January 1, 2022, and April 15, 2022, 13,854 (14.3%) of images were Ukraine-related, and 1,196 (8.6%) of images came directly from a Russian source. Similarly, among the 16,063 images about Ukraine, 2,313 (14.4%) of these images came from Russian sources. For images not about Ukraine nor Russia, only 1.2% of images were from Russian sources. 109 (58%) of the 188 accounts not directly associated with Russia, posted at least one Russian-sourced image about Ukraine, with 14 (7.4%) users posting a higher percentage of Russian-sourced images than the Russian embassy account (21%).

Russian Disinformation Narratives on Weibo

Having shown the increased influence of Russian news sources within online Chinese communities, we finally look at if this increased influence led to the increased spread of Russian disinformation narratives by Chinese news outlets. As shown in the previous section, outlets like the Global
Tables 9: References to Russian disinformation campaigns.

<table>
<thead>
<tr>
<th>Account</th>
<th>Bio-Weapons Posts</th>
<th>Ukrainian Nazis Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>中国新闻网</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>中国日报</td>
<td>138</td>
<td>10</td>
</tr>
<tr>
<td>CGTN/记者团</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>人民日报</td>
<td>109</td>
<td>3</td>
</tr>
<tr>
<td>央视新闻</td>
<td>121</td>
<td>0</td>
</tr>
<tr>
<td>环球时报</td>
<td>165</td>
<td>96</td>
</tr>
<tr>
<td>新华网</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

Times have a higher level of connection to Russian outlets than others like CCTV. While conducting our previous analyses, we observed that these same outlets often posted about the Russian disinformation campaign about US-funded Ukrainian biological weapons laboratories and Ukrainian neo-nazis (Table 9). For example, 环球时报/Global Times had 261 different Weibo posts concerning these stories (165 about Ukrainian biological weapons and 96 about Ukrainian neo-nazis) while also referencing Russian sources 415 times and posting 10 Russian-sourced images. Similarly, the news outlet with the most posts about these disinformation narratives (182 about biological weapons, and 87 about Ukrainian neo-nazis) Sina News/新浪军事 also had elevated amounts of references to Russian outlets as sources. In this section, we thus model the connection of different Chinese news outlets to the Russian government and Russian state media and their tendency to publish Russian disinformation stories.

**Experimental Setup.** For this experiment, we utilize our extended Weibo dataset, including the additional 32 Chinese state media entities that were in Fung et al.’s dataset; altogether 39 outlets. To model each news outlet’s promotion of Russian disinformation about Ukraine, we use the combined count of their mentions of the US-funded Ukrainian biological weapons laboratories and their mentions of the Ukrainian government being dominated by neo-nazis. We utilize these two disinformation narratives as they were heavily promoted by Russian websites as seen in Section 3. Furthermore, among our 39 Chinese outlets, posts that mentioned these disinformation narratives were fairly popular (the posts for each collectively received 2,756,638 and 409,368 Weibo reactions respectively). Running Mann-Witney U-tests comparing the popularity of posts (based on the number of reactions) mentioning these stories to posts not mentioning either, we find in both cases that these stories enjoyed higher popularity than posts not mentioning them. Posts mentioning these two narratives received an average of 3,132 reactions and 1,347 reactions respectively.

To properly model the interplay between our variables indicating an outlet’s connection to Russian entities and the counts of its publication of Russian disinformation, we utilize a negative binomial regression. We utilize a negative binomial rather than Poisson regression for data given that negative binomial models do not make the strict assumption that the mean of the data is equivalent to the variance; this assumption has been shown to be unrealistic for many real-world scenarios (Morina and Bernstein 2022). As input variables, we utilize the number of citations of Russian news agencies, the number of shared images from Russian sources, and finally the number of shared hashtags with the Weibo version of RT, Sputnik News, and the Russian Embassy. Looking initially at these variables we see that have 0.716, 0.600, and 0.671 Pearson correlations respectively with the number of Russian disinformation stories that each individual Chinese outlet published. This already illustrates that outlets with close ties to Russian entities tend to post more Russian disinformation narratives. Finally, we note that we regularize disinformation story counts by the number of posts each account made during our collection period.

**Results.** As seen in Table 10, Russian news source utilization is a predictor of posting about Russian disinformation with a coefficient of 0.0168 (1 more Russian citation leads to an expected 1.6% increase in Russian disinformation stories [e.g., for every 42 Russian news citation, the number of expected disinformation stories doubles]). With a p-value cutoff of 0.05, we see that it is the only one of our variables that is a predictor of Chinese state media accounts’ usage of Russian disinformation stories. The other two variables that we considered were largely accounted for after taking into account how often a given account cited Russian news outlets. We thus see that the increased influence of Russian news on Chinese outlets is correlated with increased amounts of Russian disinformation stories within Chines media ecosystems.

### 5 Related Work

We are not the first to perform a differential analysis of different text corpora to understand their treatment of given topics. Olteanu et al. (Olteanu et al. 2015) compare differential discourse surrounding climate change. Similarly, Galvez et al. (Galvez, Tiffenberg, and Altszuler 2019) compare the differential word associations of different genders.

Our work further builds several others’ in quantifying the use of images in the spread of narratives. One of the first to take an image-based approach in examining the contents of social media misinformation, Seo et al. characterized the use of images as propaganda during the 2012 Israeli–Hamas conflict (Seo 2014). More recently, Zannettou et al. found that Russian state-sponsored Twitter accounts had a marked influence on the memes or images that appeared on Reddit and the alternative social media platform Gab (Zannettou et al. 2018). Like our approach that considers both images, metadata, and text of Russian origin, Zhang et al. (Zhang et al. 2019) utilize a multi-modal approach that takes into account both images and text to track misinformation and rumors on Twitter. By looking at images and hyperlinks, in their papers, Wilson and Starbird et al. look at the influence of Syrian White Helmets across different platforms and Hanley et al. look at the spread of QAnon (Wilson and Starbird...
invasion” colors crisis while the Western press refers to it as a “war” and Chinese outlets often refer to the war as a “conflict” or peace negotiations. Similarly, our finding that Russian systems is different is pivotal. For example, we found that the overall sentiment and coverage of different ecosystems surrounding events in Ukraine belonged to personal Russian non-state-sponsored accounts (Golovchenko, Hartmann, and Adler-Nissen 2018). As a whole, 1,811 different Twitter profiles belonging to individual users rather than to large state media or journalist accounts had the biggest effect in generating pro-Russian propaganda in their study.

6 Discussion and Conclusion
In this work, we employ a quantitative approach to understand the distinctive narratives about the Russo-Ukrainian War being spread by Western, Russian, and Chinese media. Utilizing PLDA, the NMPI information metric, word2vec models, and differential sentiment analysis, we find that each ecosystem, in turn, has entirely different and nuanced perspectives. Looking at the topics written and performing a cross-platform study, we show a spike in the influence of Russian news outlets on the Chinese news ecosystem since the beginning of the Russo-Ukrainian War.

When conducting such studies, we argue that uncovering why the overall sentiment and coverage of different ecosystems is different is pivotal. For example, we found that the reason why Chinese ecosystems had a more positive sentiment about the war is due to their emphasis on cooperation and peace negotiations. Similarly, our finding that Russian and Chinese outlets often refer to the war as a “conflict” or “crisis” while the Western press refers to it as a “war” and “invasion” colors why different ecosystems have distinctive perspectives on the ongoing military operations in Ukraine. Only by taking an approach that incorporates several methods (e.g., PLDA, word2vec) can these insights be uncovered.

Lastly, we note that it is imperative to acknowledge that events like the Russo-Ukrainian War are global with no country in the world not affected (Treisman 2022). We recognize that Russo-Ukrainian War is an ongoing conflict and that information about the war is changing day to day. We hope to remain objective and sensitive about the issues discussed here.

Acknowledgements
This work was supported in part by the National Science Foundation under grant #2030859 to the Computing Research Association for the CIFellows Project, a gift from Google, Inc., and NSF Graduate Fellowship DGE-1656518.

References
Barry, E. 2022. These Countries Have Bans on Twitter, Facebook and TikTok — Time.
Bouma, G. 2009. Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL.


Kirby, P. 2022. Donbas: Why Russia is trying to encircle Ukraine’s east - BBC News.


McCarthy, S.; and Xiong, Y. 2022. As the world reacts in horror to Bucha, China’s state media strikes a different tone — CNN.


Price, N. 2022. The Kremlin’s Allegations of Chemical and Biological Weapons Laboratories in Ukraine - United States Department of State.


Ramage, D.; Manning, C. D.; and Dumais, S. 2011. Partially labeled topic models for interpretable text mining. In *17th ACM Conf. on Knowledge Discovery and Data Mining*.


Treisman, R. 2022. War in Ukraine sends global food prices surging to record high : NPR.


