The Russia–Ukraine War in Chinese Social Media

LLM Analysis Yields a Bias Toward Neutrality

Rogers, R.; Zhang, X.

DOI
10.1177/20563051241254379

Publication date
2024

Document Version
Final published version

Published in
Social Media + Society

Article

SM+S
social media + society

Social Media + Society
April-June 2024: 1–12

DOI: 10.1177/20563051241254379
journals.sagepub.com/home/sms
Sage

The Russia–Ukraine War in Chinese Social Media: LLM Analysis Yields a Bias Toward Neutrality

Richard Rogers[1] and Xiaoke Zhang[2]

Abstract

This study is a cross-platform analysis of the discourses surrounding the Russia–Ukraine war in Chinese social media. Making use of both manual as well as automated classification of discussion about the war, we found most significantly the mass amplification of Russian state positions on Weibo and the reframing of the war as being in the Chinese national self-interest on Douyin. We situate what we call cross-national amplification as well as the national self among other notions that seek to capture the broad discursive power of the Chinese state including digital nationalism, soft propaganda, and playful patriotism. A second set of findings include some agreement between the manual and automated classification, albeit with the artificial intelligence (AI)-assisted platforms showing what we call a bias toward neutrality. We also emphasize the importance of a cross-platform analysis (rather than a single-platform analysis) when seeking to capture public sentiment on social media and the types of orchestrated, state discursive power on display.

Keywords

digital nationalism, soft propaganda, playful patriotism, cross-national amplification, and national self

Introduction: Studying the Russia–Ukraine War Discourse in Chinese Social Media

Ever since Russia's full-scale invasion of Ukraine in February 2022, a growing body of literature, especially from policy and think tank circles, has reported on how much of the related content on Chinese social media takes what could be called a pro-Russian stance, occasionally mixed with framings concerning Chinese national interests (Repnikova & Zhou, 2022). In the following, we explore the shaping of this discourse about the Russia–Ukraine war through an empirical study of Weibo and Douyin, two leading Chinese social media platforms.

Our initial point of departure for the study of the shaping of public discourse on social media is various forms of state "discourse power" (Thibaut, 2022). One is the result of top-down propaganda, intended to "control narratives around key public issues and assert the Chinese Communist Party's discourse power" through official media channels and accounts (A. Zhang et al., 2021, p. 2). This "digital nationalism" (Schneider, 2018) is complemented by softer, more bottom-up strategies, where seemingly everyday users adopt the same viewpoints and amplify the state discourse (X. Chen et al., 2020; Creemers, 2017; Hyun et al., 2014; Zou, 2021). Once referred to as 50c party posts (King et al., 2013, 2017), these contributions appear to be from genuine users. A more recent characterization of how state-orchestrated discourse forms on social media has been dubbed playful patriotism, where the use of hashtags such as #positiveenergy is encouraged, particularly on Douyin (X. Chen et al., 2020). Videos tagged with #positiveenergy have been found to promote the "Chinese state's political agenda" (Chen et al., 2020, p. 97) through "consensus-building" for the sake of "national unity and cohesion" (Fung & Hu, 2022, p. 140).

The aim here is to contribute to the study of state-orchestrated discourse online by examining the origins and substance of the narratives surrounding the Russia–Ukraine war on Chinese social media. We are employing the term narrative in

[1]University of Amsterdam, The Netherlands
[2]Renmin University of China, China

Corresponding Author:
Richard Rogers, University of Amsterdam, Turfdraagsterpad 9, 1012 XT Amsterdam, The Netherlands.
Email: rogers@uva.nl

the sense of "a representation of a particular situation . . . in such a way as to reflect or conform to an overarching set of aims" (New Oxford American Dictionary, 2010). We are interested in how accounts of the Russia–Ukraine war are represented in Chinese social media and how these representations may be characterized.

The research focuses on two of the three most prominent social media platforms in China, Weibo and Douyin (WeChat being the other). We situate our research within the broader literature on how the state shapes discourse on social media in China, as mentioned above. After a brief overview of the characteristics of Weibo and Douyin, we discuss our cross-platform approach as distinctive from a single-platform analysis, with the benefit of finding discursive and source commonalities across platforms while still considering platform specificity, given each platform's affordances and vernacular culture (Rogers, 2023).

In the analysis, we also take advantage of recent developments in computational methods, particularly the proposition of large language models (LLMs) for classification tasks (Törnberg, 2023). Utilizing these tools, and checking them against manual classification, we map out the most significant war-related talking points and larger narratives on Weibo and Douyin, discussing their origins through engagement analysis. Empirically, the combination of the computational analysis with the qualitative examination allows us to make observations about the power and limitations of language models in clustering and classification tasks. The LLMs produce rather broad, seemingly apolitical narratives, we found, while the manual investigation picked out more specifically the Russian talking points and some larger narratives that could have been overlooked if relying solely on automated classification.

Our findings provide compelling evidence of the amplification of Russian state narratives by Chinese state-sponsored accounts as well as political influencers on Weibo, with a relative absence of nationalist Chinese content. The notable exceptions are mentions of multipolarity and U.S. hegemony, which could be considered a narrative that overlaps with the Russian state interests (Gabuev & Kovachich, 2023). On Douyin we note the nationalization of the war discourse, where national self-interests are predominant. We dub the discourse as the national self, where the war becomes a prism through which (and opportunity) to discuss its benefits for China, such as the potential for the return of Vladivostok (Baptista, 2020). Unlike on Weibo, on Douyin the Russian state narratives are not prominent; Chinese state-sponsored accounts are posting more Chinese nationalist content.

These findings not only point up the importance of cross-platform analysis (rather than a single-platform approach) to the study of discourse on Chinese social media, for the dominant storylines on each platform may be distinctive. They also demonstrate how additional forms of discourse power alongside digital nationalism, soft propaganda as well as playful patriotism may be present. Our analysis found what we term cross-national amplification of Russian talking points and narratives as well as the reframing of the war discourse in the national self-interest.

Conceptualizing Discourse Power in Chinese Social Media

Despite heavy censorship, surveillance, and state media control, social media in China has been seen as a platform where individuals could voice certain concerns, including participating in anti-corruption campaigns and "exert[ing] pressure on the regime to become responsive" to other issues, including environmental and financial ones (Stockmann & Luo, 2017, p. 189; Guan, 2019) as well as frustrations with COVID-19 restrictions imposed by the authorities (Yang & Zhang, 2021). For example, during the COVID-19 pandemic, when restrictions were framed by the state as a time for "strengthening national belonging," one line of argumentation found that this "correct collective sentiment" was not all-enveloping owing to "alternative articulations of grief and rage" that challenged the party line (C. Zhang, 2022, p. 219).

Another set of scholarship about critical voice-giving in Chinese social media has focused on satirical and coded expressions of concern (Luqiu, 2017). This "adaptive agency" can assume the form of "subversive satire" (Xi, 2023, p. 2), including running commentary on such issues as gender inequality as well as the urban–rural divide, developed through stand-up comedy and available through online streaming (D. Chen & Gao, 2023).

Taken together, these and similar studies emphasize how "public discourse under authoritarian rule is not monolithic" in China (D. Chen & Gao, 2023, p. 1). The point is made even for online environments where state viewpoints and user engagement with state and state-sympathetic accounts seem pervasive and hegemonic. Audiences are active, it is often argued, and "it is entirely possible for them to reject the main discourse, as they often do" (Schneider, 2018, p. 227).

Thus, we situate our study within this larger anti-monolithic purview of public discourse on Chinese social media, looking out for alternative expression and counter-voice. At the same time, we recognize that in an environment with heavy censorship and surveillance (Cairns & Carlson, 2016), the study of public sentiment expressed online, especially those that rely on most engaged-with content in social media, as we do, are apt to foreground the state's discourse power. In such studies, it is the specific orchestration of that power that becomes of special interest. That is, the power manifests itself in multiple forms. As such this study has both an anti-monolithic starting point as well as an interest in the type of discourse power orchestrated by the state.

Discourse power has been defined as the capacity for "narrative agenda-setting . . . focused on reshaping global governance, values, and norms to legitimize and facilitate the expression of state power" (Thibaut, 2022, p. 2). Within

the study of such discourse power, at least three main strategies (and terms) have been outlined, each capturing how a different set of actors drive what could be understood as the hegemonic (yet not monolithic) rhetoric pervading the Chinese social media landscape: digital nationalism (Schneider, 2018), soft propaganda (Zou, 2021), as well as playful patriotism (X. Chen at al., 2020).

Schneider's point of departure for digital nationalism is the study of the state's mass communication online, especially how it is achieved in a medium once known for its allegedly liberating potential (2018). He conceives of the state as a political technology that strategically constructs online networks that wield the power to shape discourses as nationalist. While emphasizing that much online content is mundane and apolitical, that which veers into the political is met by networks that watch, suppress, and shape it. As we also have found and report below, these may be networks of state social media accounts that amplify particular talking points and narratives, or journalists and political influencers who do the same. They also may be "armies" of seemingly everyday users, as put forward by King et al. (2013, 2017), and further explored by Zou (2021) in terms of soft propaganda. Zou (2021) writes of a "heterogeneous 'thought work' network," referring to how ideological inculcation (*sixiang gongzuo*) is orchestrated by a multiplicity of actors (p. 202).

One of the larger points concerns the repackaging of political content in entertainment media. The discussion recalls how the internet and social media have become significant forums for "restyling" state propaganda, like cinema, television, comics, and other media and forms of cultural expression preceding them (Cai, 2016; Zou, 2021). Indeed, it is also in line with the notion of playful patriotism, as put forward by X. Chen et al. (2020), where users of the Douyin (short-form video) platform promote the party line through online fun. They tag content with #positiveenergy in keeping with the upbeat national unity called for by the state. Zhao and Ye (2023) report on how state accounts (rather than everyday users or influencers) also make users laugh on Douyin, thereby restyling state points of view.

We found that on Douyin influencers and other state-affiliated users recast world events (in this case, the Russia–Ukraine war) as being in the national self-interest or affecting it. The war becomes a means to make new territorial claims or worry about relations with certain countries who are not aligned with or neutral toward Chinese or Russian positions.

These conceptualizations all relate to discourse power, as noted above, though each emphasizes a different set of actors and orchestration or operative vector of power. In the following, we explore further these actor sets and orchestrations in organizing public discourse in Chinese social media concerning the Russia–Ukraine war. Specifically, we seek to uncover which talking points and related narratives dominate public discourse in the social media platforms and situate them conceptually. The research question guiding this study is how to characterize the public discourse about the war taking shape on Chinese social media. Is it primarily through the mechanisms of digital nationalism, soft propaganda, playful patriotism, or are there other orchestrations at work? As stated, we found two variations on discourse power that we report. One we term cross-national amplification. It is the massive amplification or reposting of Russian talking points and narratives by Chinese state accounts on Weibo, belying the official claim to neutrality in the war (Carter Center China, 2022). On Douyin, we found something different from playful patriotism by everyday users. The material that animates the most (in the war space on Douyin) is by state-affiliated accounts refashioning war themes in the national self-interest, creating opportunities for China or complicating diplomatic relations.

The Styles of Engagement on Douyin and Weibo

Weibo, sometimes known as the Chinese Twitter, is a microblogging website with nearly 600 million monthly active users as of the first quarter of 2023. Douyin, often referred to as the Chinese TikTok, is the popular online short-video blogging platform, with over 700 million monthly active users during the same period. Since the outbreak of the Russia–Ukraine war, discussions, debates and other content surrounding the clash have been prevalent on these two social media platforms. For example, the hashtag, #RussiaUkraineConflict, has generated significant traction on Weibo, accumulating over 400 million views and sparking tens of thousands of discussions as of July 2023. Notably, it ranks as the second most popular hashtag on the topic page, indicating substantial interest and engagement. On Douyin, the same hashtag has accumulated over 100 billion views as of July 2023. This places it among the top 10 most impactful topics on Douyin for 2022 (and beyond), alongside subjects like Covid-19 (200 billion views), World Cup (70 billion views), the Winter Olympics in Beijing (29 billion views), and Liu Genghong, the fitness influencer (20 billion views). These figures demonstrate the extensive reach and influence of the Russia–Ukraine war discourses on both platforms.

Each platform's distinctive cultures and features are worthy of mention, given how they shape the methods by which each platform is studied. Weibo is typically thought of as an "older, more mainstream" platform, while Douyin is a more "youth-oriented space" (Meng & Literat, 2023, p. 2). In terms of content format (and the structure of the data, as we come to), Weibo primarily features text-based posts, occasionally accompanied by pictures and videos. Douyin contains short-form videos, with trending sounds, filters, and hashtags, and they aim to entertain and attract attention. In terms of their recommendation algorithms, Weibo is more centralized, and their users tend to actively search for information. In contrast, Douyin filters content based on individuals' previous consumption, tailoring user experience to

specific interests, leading to a more passive mode of entertainment consumption and a content space which could be called a wartok (or stream of war-related content). In other words, having viewed war-related content, users will receive more of it.

Of particular relevance to this study are the actors driving the distinctive discursive content observed on these two platforms. To introduce them, we refer initially to recent studies on the COVID-19 pandemic as seen through the two social media platforms. On Weibo, state media accounts assumed a significant role in disseminating pandemic-related content, often serving as the primary sources of reposted content (Yang & Vicari, 2021). The accounts circulated narratives that highlight the Chinese government's efforts in epidemic prevention and control. In contrast, on Douyin, users often shared personal experiences, from a first-person perspective, that aligned with state policies (Yang et al., 2022). The platform showcased nationalist narratives, promoting a sense of national pride and unity.

These differences in platform communications align with what we have found when studying the Russia–Ukraine war: a preponderance of reposted content on Weibo by state media accounts and state-aligned influencers and a first-person perspective on Douyin that concerns seeing the war through the lens of national self-interest.

Methodology: A Cross-Platform Approach With Manual and AI-Assisted Labeling

To answer the question concerning how public discourse about the war is being shaped in Chinese social media, we undertake a cross-platform approach, studying Weibo and Douyin, two of the top three (by monthly user counts) Chinese social media platforms. We analyze the content of the posts as well as the accounts driving the most engaged-with ones.

The cross-platform approach has benefits over a single-platform approach, for it allows for the study of commonalities in public discourse across platforms (Rogers, 2018). It does so by pursuing data commensurability (studying classified content and post engagement on both platforms), enabling a form of triangulation. At the same time, it acknowledges platform particularities and distinctive users, demonstrating the differences platforms make. It thereby provides a sense of how single-platform studies would skew generalizing accounts of public discourse on social media (Rogers, 2023).

This study examines posts on Weibo and Douyin about the Russia–Ukraine War from February 2022 to July 2023. It employs a process of querying, clustering, and labeling. First, for collecting data, we relied on the research affordances of the platforms (Rogers, 2019). For Weibo, we made keyword queries, resulting in posts that contained the keywords or keywords in hashtags, as users frequently post without hashtags. For Douyin, we undertook hashtag-based querying, given how users post videos and tag them with hashtags.

The date range of our data collection is February 24, 2022 (the start of the full-scale Russian invasion) to July 1, 2023 (the time of the study), resulting in a total of 2,340 unique posts from Douyin and 14,237 unique posts from Weibo. To ensure the inclusion of relevant posts, we followed a two-step query design, where we added significant co-keywords and co-hashtags that we found after the first iteration (Rogers, 2019). The final set was grouped with the stances pro-Russian, pro-Ukrainian, or more general keywords descriptive of the war. Examples of the keywords (and their contrasts) that we used include 乌纳 (Ukraine Nazi), 乌克兰右翼 (Far-right politics in Ukraine), 对俄制裁 (Sanctions against Russia), 俄罗斯入侵乌克兰 (Russian invasion of Ukraine), 俄乌战争 (Russia–Ukraine War), 俄乌冲突 (Russia–Ukraine Conflict), 乌克兰危机 (Ukraine Crisis), 俄乌谈判 (Russia–Ukraine Negotiation), 乌克兰难民 (Ukraine Refugee), 瓦格纳事件 (Wagner's Incident), and 瓦格纳叛变 (Wagner's Mutiny).

With the data collected, we performed thematic clustering analyses both by automated and manual means. For Weibo, we undertook an automated classification of posts, grouped into stances, as discussed above. We then examined the stances, classifying them anew by manual means, where we sought to check whether the automated technique had mischaracterized posts as pro-Russian, pro-Ukrainian, or neither. For Douyin, we performed a co-hashtag analysis with visual network analysis and labeled clusters (Venturini et al., 2021). The manual labels were made through a close reading of the top hashtags per cluster. We then thematically classified each data set (per cluster) with an automated classification system. The two sets of labels were subsequently placed side by side to analyze the complementarities and incongruities between the war discourses. We also zoomed in on the posts per cluster with the highest engagement scores and examined the accounts most significantly driving that engagement.

For the computational analysis of the posts, we chose LLMs, given the opportunities they offer for analyzing textual data, especially text classification (Goyal et al., 2022). We deployed two distinct LLMs, GPT by OpenAI and Claude by Anthropic, to analyze and label data sets acquired from Weibo and Douyin, respectively, given the characteristics of each data source. The Weibo data set consists of a medium-sized corpus with hundreds of lines of the individual posts. For the Douyin data set, we concatenated the hashtags from all posts into one line, resulting in a substantial text corpus. We decided to utilize GPT-3.5 (GPT-3.5 turbo) for Weibo's posts, for its proven performance in annotating tweets (Tekumalla & Banda, 2023). For Douyin we utilize Claude-2, which is capable of processing 100k tokens, in order to handle Douyin's larger corpus of hashtags.

Table 1. Main Discursive Themes Concerning the Russia–Ukraine War on Chinese Social Media, Weibo and Douyin, February 22, 2022 to July 1, 2023.

Themes	Weibo	Douyin
Platform-specific	Ukrainian corruption, Western Russophobia, Russian domestic support of the war, and the risks and dangers of Ukrainian refugees	Historical border dispute (Vladivostok's return to China) and China's newly complicated relationships with other key countries
Overlapping	U.S. hegemony/multipolarity	U.S. hegemony/multipolarity
Alternative	Criticism of humanitarian crisis caused by Russia's invasion	Grieving for the victims of war on both sides of the conflict

Of course, the use of LLMs for analysis is in its infancy and is the source of uncertainty about how well they perform and can be trusted. Some studies have found that LLMs outperform humans in annotation and classification tasks (Gilardi et al., 2023; Törnberg, 2023), while others note their limitations compared with fine-tuned classifiers (Ziems et al., 2023). Furthermore, the capacity of LLMs to label nuance—a central objective of interpretative analysis—is only beginning to be explored (Nelson et al., 2021).

So, in our work, we compare our manual labeling and grouping with that done by the LLMs, leading to additional research questions about the kinds of thematic descriptions put forward by the two methods. In comparing them, we consider specifically the differences the LLMs make, both generally for chatty LLMs as well as for each. In all, we found that the LLMs produced more generic, even apolitical labels compared with the qualitative interpretations. We describe this LLM labeling output as a bias toward neutrality, a characteristic shared by both.

We designed prompts to have the LLMs generate the classifications (Egami et al., 2023). Specifically, we designed the prompts through what is referred to as "prompt perturbation" (Mishra et al., 2023), a strategy that is used to improve the performance of an LLM through multiple iterations with small adjustments. In addition, our approach asked the LLMs to adopt a research persona (Child et al., 2019; Ziems et al., 2023). The prompts followed the suggestions outlined by Törnberg (2023), defining (1) the research persona for the language model to adopt, enhancing focus and determinism; (2) the stances using common keywords for each stance; (3) the relevant contextual cues, such as internet slangs or code languages; and (4) the desired format of the output.

The final prompt used is as follows: *You are a narratologist tasked with mapping out five narratives of the provided text and categorizing the text about the Russia–Ukraine war. Please output a CSV table with two columns: (1) the narrative in English (summarized within 20 words) and (2) the stance of the narrative (pro-Russian, pro-Ukraine, or neutral). Below are some examples of the potential coded languages of pro-Russian, pro-Ukraine, and neutral narratives for your reference: (1) Pro-Russian refers to Ukraine soldiers as* 乌纳 *(UkNazi) or* 乌贼 *(squid); (2) Pro-Ukraine criticizes Russian army with sarcasm* 菜鹅 *(Veggie Goose/Weak Russian Army),* 晋军 *(Jin Dynasty/Putin's Army),* 晋凉 *(Cold Putin/Putin is over),* 鹅粉 *(Fans of the Goose/Fans of Putin); and (3) Neutral provides factual updates and describes battlefield developments without assigning blame.*

Analysis: The War Discourses on Weibo and Douyin

In this section, we first describe the findings we made overall concerning the posts about the Ukraine-Russia war on each platform, before turning to the comparison between the manual and automated techniques, which yielded additional findings about the differences LLMs make. In the end, we return to the overall platform-specific findings concerning Weibo's cross-national amplification of Russian points and Douyin's rendering of the war in the Chinese self-interest, discussing the importance of a cross-platform analysis over a single-platform approach, given the distinctive differences and as well as a key commonality.

Cross-National Amplification on Weibo

In the war discourse on Weibo, generally, there is a proliferation of Russian stories or content that originates from Russian state-affiliated media and is reposted by Chinese state media as well as state-aligned influencers. These are talking points and narratives predominantly revolving around U.S. hegemony, Ukrainian corruption, Western Russophobia, Russian domestic support of the war, and the risks and dangers of Ukrainian refugees (see Table 1). One notable example, which gained significant traction on Weibo, concerns the story that the United States has installed and operates biolabs throughout Ukraine. Propagated by Russian sources and reposted by Chinese state outlets, this story was initially used to attribute the origin of COVID during the pandemic and has resurfaced during the war. One illustration of this reposting dynamic (in Figure 1) provides a glimpse of how this story circulated on Weibo. On August 15, 2022, Sputnik quoted Robert Kennedy Jr.'s statements as evidence of biolab operations in Ukraine, and 2 days later, CCTV.com (China Central Television) shared a remarkably similar post. In another comparable instance, Russia Today referenced a report from a Russian official on February 1, 2023 and the

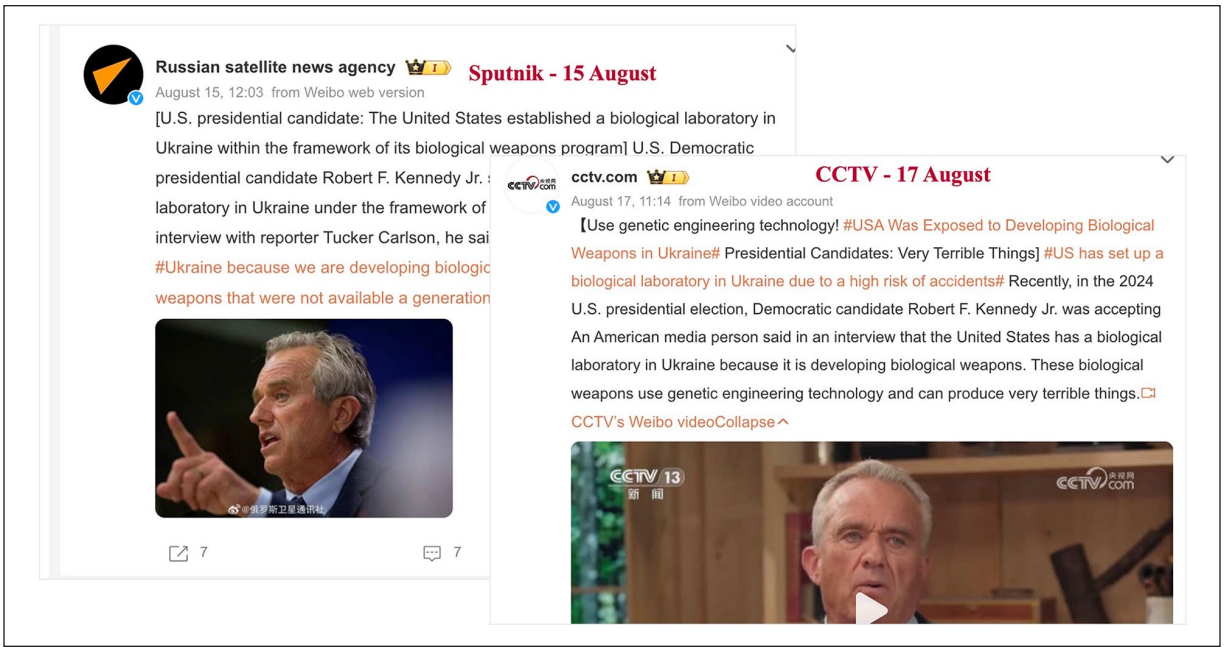

Figure 1. Cross-national amplification of the story of U.S. biolabs in Ukraine on Weibo. Dates: August 15 and 17, 2022.

CCP tabloid Global Times followed suit thereafter, disseminating the same quotes and posing the question, "What is the U.S. hiding in the biolabs discovered in Ukraine?"

We observe this pattern with the posts criticizing unilateral sanctions imposed by the United States, discussing Ukrainian President Zelensky's luxury Italian property, and showing polls that indicate the vast majority of Russians trust President Putin and his military decisions. These stories fit larger narratives about U.S. hegemony, Ukrainian corruption, and Russian domestic support of the war, respectively.

While much of the most engaged-with content about the war is shaped by reposted Russian content, there is also a strand of material posted by Chinese (and Russian) sources individually, which could be characterized as a common theme, or "overlapping narrative" (Gabuev & Kovachich, 2023). It not only concerns U.S. hegemony but also calls for the development of a multipolar world, either in abstraction or occasionally with respect to the rise of new geopolitical configurations and alliances, which likely references BRICS, the informal confederation seeking economic cooperation made up of Brazil, Russia, India, China, and South Africa. These findings point to both the cross-national amplification (of Russian points by Chinese state media as well as influencers) as well as a convergence of narratives between the two countries.

Amplification is a particular form of discourse power insofar as it directs communicative attention toward specific points, effectively aggregating or snowballing them (Musgrave, 2017; Y. Zhang et al., 2017). In the context of the Russia–Ukraine war discourse on Chinese social media, the retweeting, recreation, and referencing of Russian talking points amplify and grow the larger Russian narratives.

The main Russian media outlets in China—Russia Today, Sputnik, and the Russian Embassy in China—have official Weibo accounts with a combined total of tens of millions of followers. As illustrated above, their posts about U.S. hegemony, Ukrainian corruption, Russophobia, Russian domestic support of the war, and the dangers and risks of Ukrainian refugees (to name the ones we found most significantly) have a significant reach on Chinese social media platforms in and of themselves. In our data set, the presence of the Russian media outlets Sputnik and RT in the Weibo war discourse is significant.

The propagation of the Russian discourse, however, is aided significantly by the propagation of Russian media posts by Chinese state accounts and political influencers. Among the most present media organizations in the discursive space are the Chinese state-affiliated outlets, Reference News and Global Times as well as Guancha, which scholars have described as the "locus of nationalist sentiment in Chinese cyberspace" (Sullivan & Wang, 2022, p. 81). In addition to state-affiliated and new nationalist sources, independent bloggers and influencers, known as the Weibo big Vs, embrace the same content. Together, the state-affiliated and the state-aligned influencers retweet, recreate, and reference the Russian points, amplifying their presence by directly sharing the content with their broader audience and contributing (algorithmically) to the engagement metrics of the posts, ultimately bringing more attention and traffic. Hence, both Chinese state-affiliated accounts and Weibo political influencers propagate Russian talking points through manifold amplification.

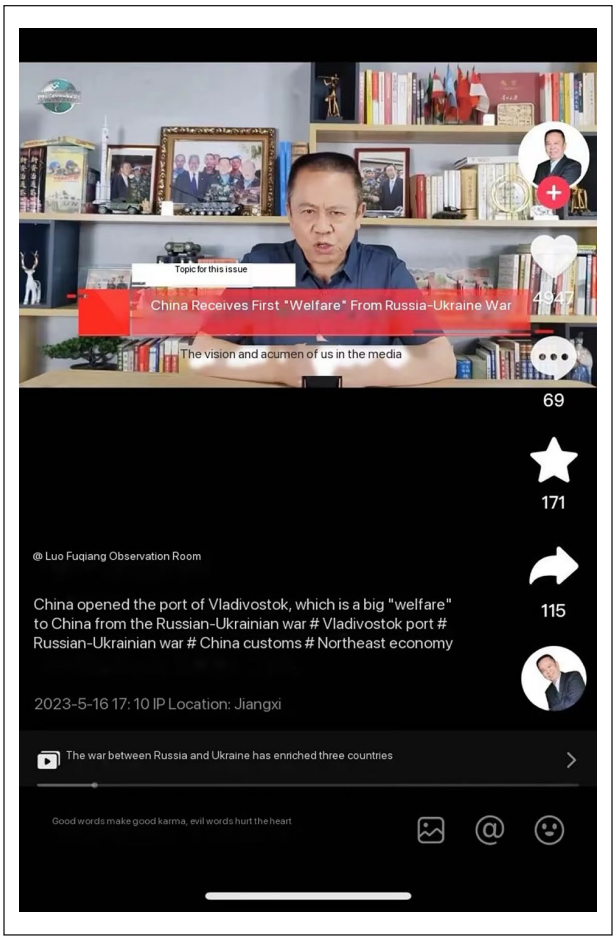

Figure 2. A national self-interest talking point on Douyin. Returning Vladivostok to China as one of the benefits of the Russia–Ukraine war for China.
Source: LuoFuqiangObserversation@Douyin, May 16, 2023.

Nationalization of the War Discourse on Douyin

In contrast to the reposted Russian narratives prevalent on Weibo, on Douyin a central prism through which the war is discussed is the national, or more specifically, the national self-interest (see Table 1). There are at least three framings that are directly related to it. One is U.S. hegemony and the Chinese desire for multipolarity or the end of American domination of geopolitics. Another is the focus on the prospect of Vladivostok's return, turning what was a settled state of affairs into a historical border dispute between China and Russia (see Figure 2). There is also a set of posts about how the war is affecting China's relations with several countries key to Chinese interests. The countries are unlikely to support Chinese positions when they overlap with or reinforce Russia's.

It should be noted that, as with Weibo, there is a small set of content that grieves for those suffering and could be viewed as an alternative narrative, not so unlike the one against the COVID-19 restrictions as discussed above. It laments the distant suffering, but it also portrays both sides as victims, which could be said to be more of a correct phrasing, should one wish to discuss the matter of the effects of the war on everyday people.

The nationalization of the war discourse on Douyin is not an organic phenomenon, as evidenced by the top 20 most popular accounts in the wartok, where 75% of them are essentially government-affiliated. These media accounts contribute to over half of the total likes in the sample, highlighting their significant role in shaping the platform sentiment. The adaptation to the social media environment suggests the state's capability to use the new media to its own advantage (Hyun et al., 2014; Hyun & Kim, 2015), not to mention to restyle and soften Chinese discourse power (Zou, 2021). As a result, Douyin's co-optation is congruent with the idea that the latest Chinese social media platforms are considered "a vast laboratory in which to brew digital nationalism" (Liang, 2019, p. 2349).

The majority of found narratives views the war through the lens of the national self. Douyin, remarkably, is thus devoid of the dominance of reposted Russian narratives that we found on Weibo, with one exception. There is overlap between the Russian point of view and the Chinese with respect to a geopolitical reset, or a departure away from U.S. hegemony in favor of a multipolar world, found on both platforms. Thus, the cross-platform analysis resulted in not just the predominance of platform narrative peculiarities but also a commonality.

Automated Classification and a Bias Toward Neutrality

For the automated analysis of Weibo, we utilize GPT-3.5, which, once prompted as described above, outputted leading narratives. For Douyin, we turned to Claude-2, which did the same. We utilized different LLMs because of the constitution of the data sets as well as the capabilities of the models, as we note above. In the following, we describe the agreement between the automated and the manual analysis as well as an LLM effect which may be attributed to their value alignment. That effect, which we call a bias toward neutrality, is in evidence in the tonality of the labeling. It also can be seen in the stance classifications of a few narratives the automated approach outputted that were not in agreement with the manual ones, as we discuss.

For both platforms under study, when grouping posts as narratives and classifying stances, the LLMs exhibited what appears to be an embedded value alignment as well as a broad interpretation of a narrative which may have been affected by the alignment. Aligning artificial intelligence (AI) agents with certain human values is considered an important component in the development of LLMs, particularly chatty ones or those designed for public use (Russell, 2021). For example, OpenAI's guidelines suggest the AI agent break down complex politically loaded questions into

simpler, informational questions when possible (OpenAI, 2022). It also emphasizes the importance of producing more "factually grounded claims" (Goldstein et al., 2023, p. 44). Anthropic's alignment efforts primarily concentrate on values such as being helpful, honest, and harmless, often referred to as the HHH principle (Askell et al., 2021; Bai et al., 2022). As a result of pre-training and fine-tuning, such alignment interventions seemingly result in a tendency toward neutral responses. In our cases, the LLMs produced more neutral terminology as well as narratives compared with the manual analysis.

The LLM effect became apparent when examining Weibo posts that were thematically grouped and given the stance neutral (see Figures 3 and 4). Certain posts classified as such by the LLM were seeming statements of facts that disguised a pro-Russian point of view, which also have been described as part of a disinformation or influence campaign against North Atlantic Treaty Organization (NATO) expansion (Paul & Matthews, 2016). For example, the statement, "Finnish parliament's support for NATO is uncertain," is incorrect or speculation at best, as the majority of parliamentarians actually supported the country's membership, as reported by Finland's largest newspaper (Teivainen, 2022). The uncertainty, however, would be favorable to a position favoring Russia, which is against NATO expansion. Another example of a post classified as neutral expresses regret for the war, suggesting it could have been avoided if the Soviet Union had not dissolved.

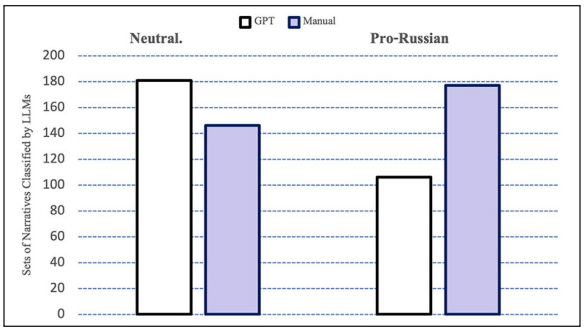

Figure 3. Weibo stances classified manually versus automatically by LLM.
Source: Weibo war-related posts, February 22, 2022, to July 1, 2023.

In these cases, the LLM alignment effect manifests itself in two senses. First, statements presented in a factual manner may be more likely to be labeled as neutral by the AI platform. Second, when value alignment strives to output apolitical or neutral point-of-view phrasing, it would be more likely to classify it as neutral rather than as stance-taking. In any case, the LLM grouped and labeled more content as neutral than the manual classification.

Turning to Douyin, briefly, the AI-assisted analysis using Claude-2 yielded results rather similar to the co-hashtag analysis and manual labeling, though there are differences in tonality as well as the interpretation of one particular cluster (see Table 2 and Figure 5). As is shown, there is some agreement between the automated and manual classifications with respect to historical analogy, though the manual classification found a certain Chinese opportunism absent in the

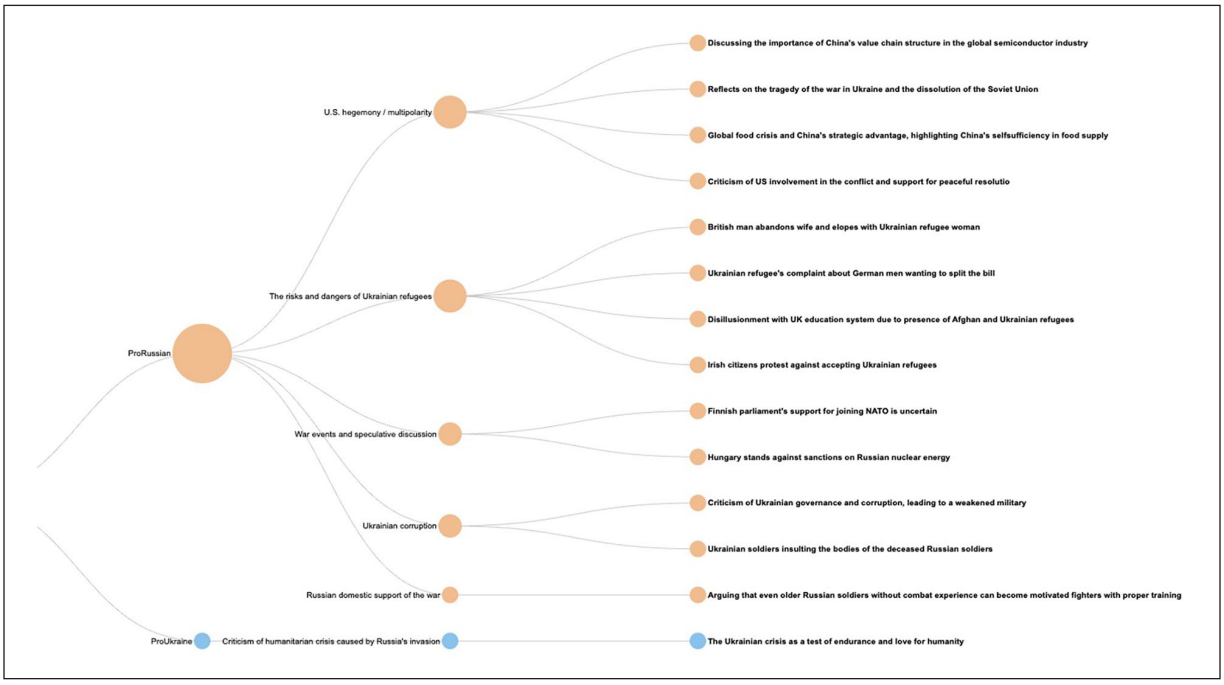

Figure 4. Word Tree of the subset of Weibo narratives that were classified as neutral by LLM and partially reclassified manually.
Source: Weibo war-related posts, February 22, 2022, to July 1, 2023.

Table 2. Douyin Co-Hashtag Analysis With Manual Labeling Versus AI (Claude-2) Grouping and Labeling Hashtags.

Douyin—co-hashtag analysis with manual labeling	Douyin—automated hashtag clustering and labeling
Real combatants: War is between US/NATO and Russia	Combatants and key figures
Peace instead of (American) hegemony	Anti-war sentiments
The war complicates relationships between China and other key countries	Geopolitics and international reactions
War events and speculative discussion	Military operations and battlefield updates
Historical analogy and opportunity (return of Vladivostok)	Historical context and analogies

Source: Douyin war-related posts, February 22, 2022, to July 1, 2023.

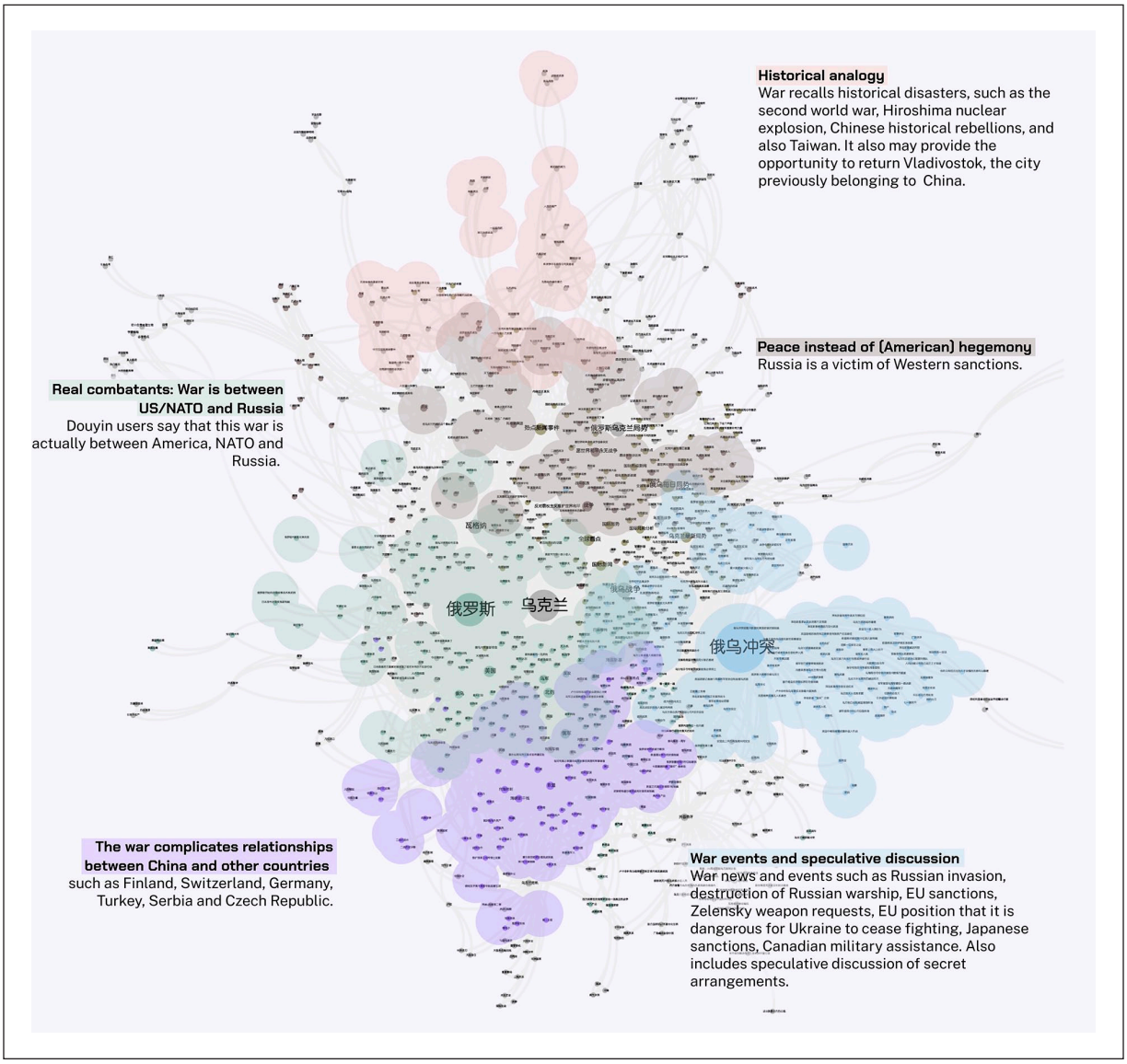

Figure 5. Douyin co-hashtag analysis with manual classifications.
Source: Douyin war-related posts, February 22, 2022, to July 1, 2023.

automated labeling. When turning to the other agreements, it is of interest to discuss the tonality of each. Whereas the manual classification identified such narratives as the "real combatants" of the war (US/NATO vs. Russia) and "peace instead of American hegemony," the automated method outputs are more muted. Those narratives are classified as being about "geopolitics" and "anti-war sentiment," rather than NATO aggression and U.S. hegemony. The one narrative

missing from the automated labeling is how the war is thought to complicate China's relationships with other countries such as Finland, Switzerland, Germany, Turkey, Serbia, and the Czech Republic. Each country has taken positions that would be incompatible with Russian justifications for the war and implicit support of them, thus affecting Chinese relations with each.

Conclusions: The Study of Chinese Discourse Power With and Without Automated Theme Classification

Existing studies on discourse power exerted on Chinese social media describe forms of digital nationalism. These range from the state-orchestrated media sources narrating the party line to 50c-style, seemingly organic users and playful youth doing the same, now on both older, mainstream as well as newer and younger social media platforms. Much of the work also emphasizes that discourse on social media is not monolithic. There are alternative points of view and ways of framing that are not correct but also not seen as vulgar by the state, such as certain expressions of grief.

With respect to how state discourse power is orchestrated, we have made findings that are somewhat distinctive from existing scholarship, when studying the posts related to the Russia–Ukraine war on Weibo and Douyin, two of the three largest social media platforms in China. Each platform represents the Russia–Ukraine war rather differently, with Weibo suffused with Russian content and Douyin with the national self-interest.

On Weibo, the discourse is dominated by reposted Russian content. This cross-national amplification, as we term it, puts in wide circulation Russian narratives about the war such as U.S. hegemony, Ukrainian corruption, Western Russophobia, Russian domestic support of the war, and the risks and dangers of Ukrainian refugees. One narrative—counter-hegemony and multipolarity—overlaps with an overtly state-driven Chinese framing of contemporary geopolitics. Whether overlapping with Chinese interests or in keeping with Russian state framings, much of the top content in the war space on Weibo is more from the "Russian world" than something else (Meister & Jilge, 2023).

On Douyin, we found that the war is reframed as potentially in the national self-interest, given the opportunities it presents for China. It reopens the question of what was phrased as the historical border dispute over Vladivostok and sees relations with key countries as being complicated. As on Weibo, there is also a discussion of how to counter U.S. hegemony.

So, we found substantive platform peculiarities and one commonality. On Weibo and less so on Douyin, there is also a smattering of content that one could describe as alternative, or not befitting the main, state-driven narratives. The Russia–Ukraine war has resulted in human suffering and users grieve the victims as they did during the COVID-19 pandemic.

We made these findings using both manual classification of social media content together with AI-assisted ones, which led to another set of findings about LLM-related analysis. The most significant one is what we call a bias toward neutrality. We observed such a bias both in the tonality of the labeling provided by LLMs but also in that they found more neutral narratives than the manual classification. We discuss these findings as LLM effects, or how value alignment may be affecting both the content summaries as well as their classification.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author(s) received the Horizon Europe project, vera.ai, under grant agreement no. 101070093.

ORCID iD

Richard Rogers https://orcid.org/0000-0002-9897-6559

References

Askell, A., Bai, Y., Chen, A., Drain, D., Ganguli, D., Henighan, T., . . .Kaplan, J. (2021). *A general language assistant as a laboratory for alignment* (arXiv:2112.00861). https://doi.org/10.48550/arXiv.2112.00861

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., Elhage, N., Hernandez, D., Hume, T., Johnston, S., Kravec, S., . . .Kaplan, J. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback (arXiv:2204.05862). https://doi.org/10.48550/arXiv.2204.05862

Baptista, E. (2020, July 4). *Why Russia's Vladivostok celebration prompted a backlash in China*. South China Morning Post.

Cai, S. (2016). *State propaganda in China's entertainment industry*. Routledge.

Cairns, C., & Carlson, A. (2016). Real-world islands in a social media sea: Nationalism and censorship on Weibo during the 2012 Diaoyu/Senkaku crisis. *The China Quarterly*, *225*, 23–49. https://doi.org/10.1017/S0305741015001708

Carter Center China. (2022, April 19). *Chinese public opinion on the war in Ukraine*. US-China Perception Monitor. https://uscnpm.org/2022/04/19/chinese-public-opinion-war-in-ukraine/

Chen, D., & Gao, G. (2023). The transgressive rhetoric of standup comedy in China. *Critical Discourse Studies*, *20*(1), 1–17. https://doi.org/10.1080/17405904.2021.1968450

Chen, X., Kaye, D. B. V., & Zeng, J. (2020). #PositiveEnergyDouyin: Constructing "playful patriotism" in a Chinese short-video application. *Chinese Journal of Communication*, *14*(1), 97–117.

Child, R., Gray, S., Radford, A., & Sutskever, I. (2019). *Generating long sequences with sparse transformers* (arXiv:1904.10509). https://doi.org/10.48550/arXiv.1904.10509

Creemers, R. (2017). Cyber China: Upgrading propaganda, public opinion work and social management for the twenty-first century. *Journal of Contemporary China*, *26*(103), 85–100.

Egami, N., Jacobs-Harukawa, M., Stewart, B., & Wei, H. (2023). *Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models* (arXiv:2306.04746). https://doi.org/10.48550/arXiv.2306.04746

Fung, A., & Hu, Y. (2022). Douyin, storytelling, and national discourse. *International Communication of Chinese Culture*, *9*, 139–147. https://doi.org/10.1007/s40636-022-00259-z

Gabuev, A., & Kovachich, L. (2023, June 3) *Comrades in tweets? The contours and limits of China-Russia cooperation*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2021/06/03/comrades-in-tweets-contours-and-limits-of-china-russia-cooperation-on-digital-propaganda-pub-84673

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). *ChatGPT outperforms crowd-workers for text-annotation tasks* (arXiv:2303.15056). https://doi.org/10.48550/arXiv.2303.15056

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative language models and automated influence operations: Emerging threats and potential mitigations* (arXiv:2301.04246). https://doi.org/10.48550/arXiv.2301.04246

Goyal, T., Li, J. J., & Durrett, G. (2022, December). *Falte: A toolkit for fine-grained annotation for long text evaluation*. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Abu Dhabi, UAE. https://doi.org/10.18653/v1/2022.emnlp-demos.35

Guan, T. (2019). The "authoritarian determinism" and reductionisms in China-focused political communication studies. *Media, Culture & Society*, *41*(5), 738–750.

Hyun, K. D., & Kim, J. (2015). Differential and interactive influences on political participation by different types of news activities and political conversation through social media. *Computers in Human Behavior*, *45*, 328–334. https://doi.org/10.1016/j.chb.2014.12.031

Hyun, K. D., Kim, J., & Sun, S. (2014). News use, nationalism, and Internet use motivations as predictors of anti-Japanese political actions in China. *Asian Journal of Communication*, *24*(6), 589–604.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), 484–501.

Liang, S. (2019). Florian Schneider, China's Digital Nationalism. *International Journal of Communication*, *13*, 2348–2350.

Luqiu, L. R. (2017). The cost of humour: Political satire on social media and censorship in China. *Global Media and Communication*, *13*(2), 123–138. http://doi.org/10.1177/1742766517704471

Meister, S., & Jilge, W. (2023). *After Ostpolitik: A new Russia and Eastern Europe Policy based on lessons from the past* (DGAP analysis, 2). Forschungsinstitut der Deutschen Gesellschaft für Auswärtige Politik e.V. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-86641-6

Meng, X., & Literat, I. (2023). #AverageYetConfidentMen: Chinese stand-up comedy and feminist discourse on Douyin. *Feminist Media Studies*, 1–17. https://doi.org/10.1080/14680777.2023.2267781

Mishra, A., Bai, Y., Soni, U., Arunkumar, A., Jinbin, H., Kwon, B., & Bryan, C. (2023). *PromptAid: Prompt exploration, perturbation, testing and iteration using visual analytics for large language models* (arXiv:2304.01964). https://doi.org/10.48550/arXiv.2304.01964

Musgrave, S. (2017, August 9). I get called a Russian bot 50 times a day. *Politico*. https://www.politico.com/magazine/story/2017/08/09/twitter-trump-train-maga-echo-chamber-215470/

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*, *50*(1), 202–237. https://doi.org/10.1177/0049124118769114

New Oxford American Dictionary. (2010). *Narrative*. Oxford University Press.

OpenAI. (2022). *Snapshot of ChatGPT model behavior guidelines*. https://cdn.openai.com/snapshot-of-chatgpt-model-behavior-guidelines.pdf

Paul, C., & Matthews, M. (2016). *The Russian "firehose of falsehood" propaganda model: Why it might work and options to counter it*. RAND Corporation.

Repnikova, M., & Zhou, W. (2022, March 14). What China's social media is saying about Ukraine. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2022/03/china-xi-ukraine-war-america/627028/

Rogers, R. (2018). Digital methods for cross-platform analysis. In J. Burgess, A. Marwick, & T. Poell (Eds.), *SAGE handbook of social media* (pp. 91–110). Sage.

Rogers, R. (2019). *Doing digital methods*. Sage.

Rogers, R. (ed.). (2023). *The propagation of misinformation in social media: A cross-platform analysis*. Amsterdam University Press.

Russell, S. (2021). Human-compatible artificial intelligence. In S. Muggleton, N. Chater, & N. Chater (Eds.), *Human-like machine intelligence* (pp. 3–23). Oxford University Press.

Schneider, F. (2018). *China's digital nationalism*. Oxford University Press.

Stockmann, D., & Luo, T. (2017). Which social media facilitate online public opinion in China? *Problems of Post-Communism*, *64*(3–4), 189–202. https://doi.org/10.1080/10758216.2017.1289818

Sullivan, J., & Wang, W. (2022). China's "wolf warrior diplomacy"": The interaction of formal diplomacy and cyber-nationalism. *Journal of Current Chinese Affairs*, *52*(1), 68–88.

Teivainen, A. (2022, March 24). HS: Over half of Finns are in favour of joining NATO. *Helsinki Times*. https://www.helsinkitimes.fi/finland/finland-news/domestic/21233-hs-over-half-of-finns-are-in-favour-of-joining-Nato.html

Tekumalla, R., & Banda, J. M. (2023, July 23–28). Leveraging large language models and weak supervision for social media data annotation: An evaluation using COVID-19 self-reported vaccination tweets. In H. Mori, Y. Asahi, A. Coman, S. Vasilache,

& M. M. Rauterberg (Eds.), *HCI International 2023—Late breaking papers. HCII 2023. Lecture notes in computer science*. Springer. https://doi.org/10.1007/978-3-031-48044-7_26

Thibaut, K. (2022, December 13). Chinese discourse power: Ambitions and reality in the digital domain. *Atlantic Council*. https://www.atlanticcouncil.org/in-depth-research-reports/report/chinese-discourse-power-ambitions-and-reality-in-the-digital-domain/

Törnberg, P. (2023). *ChatGPT-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot learning* (arXiv:2304.06588). https://doi.org/10.48550/arXiv.2304.06588

Venturini, T., Jacomy, M., & Jensen, P. (2021). What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts. *Big Data & Society*, *8*(1), 1–16. https://doi.org/10.1177/20539517211018488

Xi, Y. (2023). Adaptive agency: The satire genre and the motives behind its use in the era of social media in China. *Humanities and Social Sciences Communications*, *10*(1), Article 277. https://doi.org/10.1057/s41599-023-01768-x

Yang, Z., Luo, X., Jia, H., Xie, Y., & Zhang, R. (2022). Personal narrative under nationalism: Chinese COVID-19 vaccination expressions on Douyin. *International Journal of Environmental Research and Public Health*, *19*(19), Article 12553.https://doi.org/10.3390%2Fijerph191912553

Yang, Z., & Vicari, S. (2021). The pandemic across platform societies: Weibo and Twitter at the outbreak of the covid-19 epidemic in China and the West. *Howard Journal of Communications*, *325*, 493–506.

Zhang, A., Wallis, J., & Bogle, A. (2021). *Trigger warning: The CCP's coordinated information effort to discredit the BBC*. Australia Strategic Policy Institute. https://www.aspi.org.au/report/trigger-warning

Zhang, C. (2022). Contested disaster nationalism in the digital age: Emotional registers and geopolitical imaginaries in COVID-19 narratives on Chinese social media. *Review of International Studies*, *48*(2), 219–242. https://doi.org/10.1017/S0260210522000018

Yang, S. X., & Zhang, B. (2021) Gendering Jiang Shanjiao: Chinese feminist resistance on Weibo during the COVID-19 lockdown. *International Feminist Journal of Politics*, *23*(4), 650–655. https://doi.org/10.1080/14616742.2021.1927134

Zhang, Y., Wells, C., Wang, S., & Rohe, K. (2017). Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election. *New Media & Society*, *20*(9), 3161–3182. https://doi.org/10.1177/1461444817744390

Zhao, L., & Ye, W. (2023). Making laughter: How Chinese official media produce news on the Douyin (TikTok). *Journalism Practice*, 1–25. https://doi.org/10.1080/17512786.2023.2199720

Ziems, C., Held, W. A., Shaikh, O. A., Chen, J., Zhang, Z., & Yang, D. (2023). *Can large language models transform computational social science?* (arXiv:2305.03514). https://doi.org/10.48550/arXiv.2305.03514

Zou, S. (2021). Restyling propaganda: Popularized party press and the making of soft propaganda in China. *Information, Communication & Society*, *26*(1), 201–217. https://doi.org/10.1080/1369118X.2021.1942954

Author biographies

Richard Rogers (Ph.D., University of Amsterdam) is a Professor of New Media & Digital Culture, Department of Media Studies, University of Amsterdam. His research interests include digital methods, new media, and digital culture.

Xiaoke Zhang (M.S., Carnegie Mellon University) is a Ph.D. student in Sociology at the Center for Studies of Sociological Theory and Method and the Department of Sociology at Renmin University of China. Her research interests include economic sociology, platform studies, and computational social science.