[Comp: This is NOT an open access manuscript]

[Comp: for the online issue, please tag the article type as “research article”]

[DOI: 10.1525/cpcs.2025.2638313]

Text-as-Data Methods to Study Mass-Media Manipulations in Autocracies

Lanabi La Lova
European Institute
London School of Economics, London, United Kingdom
https://orcid.org/0000-0002-1918-5258

Abstract

This article explores how text-as-data methodologies can be used to reveal patterns of mass-media manipulation strategies employed in modern autocracies. First, it explains the importance of studying media strategies by reviewing literature on autocratic resilience, with a particular focus on scholarship addressing state-controlled mass-media tactics. Next, it outlines the computational methods currently used to study mass-media management, highlighting key academic contributions in the field of political communication, especially those focused on present-day autocracies. Finally, the article discusses the challenges of using text-as-data methods to study mass-media management strategies in autocracies.

KEYWORDS text as data, text analysis, media, autocracies, Russia

INTRODUCTION

Mass-media manipulations in autocracies—such as propaganda, censorship, and disinformation—are an important area of study in political science and communication due to their role in maintaining regime stability and support. This article explores the use of text-as-data methodologies to uncover patterns of mass-media manipulations.

First, the article reviews existing literature on autocratic resilience, highlighting the importance of studying state-controlled mass-media management strategies and discussing key arguments in contemporary scholarship on this matter. Second, it outlines text-as-data methods used to analyze media manipulations, emphasizing applications of these tools in political research and identifying key academic contributions in the field. Finally, it discusses the challenges and limitations of computational text analysis, acknowledging the complexities involved in working with large datasets from state-controlled media sources.

MASS-MEDIA MANIPULATIONS AND AUTOCRATIC RESILIENCE

Autocracies endure. Since the emergence of the earliest civilizations, most individuals have found themselves living under the governance of an autocratic[1] regime. Autocrats still rule more than one-third of the world's nations (Geddes, Wright, and Frantz 2018, 1; Ezrow and Frantz 2011, 27). As of 2023, most of the global population lives in "partly free" or "not free" countries (Freedom House 2024; V-Dem Institute 2025). The fastest-growing type of regime in the world is autocracy (Geddes, Wright, and Frantz 2018, 2). Despite the prevailing trend toward political and economic liberalization, numerous critical aspects within the realm of global politics continue to align with non-democratic paradigms (Weeks 2018, 3). At the same time, autocracies change as the world around them is transforming (Hallgarten 1954, 13) and continue to play an important role in global politics.[2] Above all, autocracies demonstrate notable resilience (Freedom House 2024).

Autocracies are still not well understood. Although scholars have made inroads into studying autocracies, there remain considerable gaps in our understanding of autocratic durability and

[1] For the purposes of this article, the terms *autocracy* and *dictatorship* are used interchangeably to refer to political systems characterized by centralized and absolute power held by a single ruler or a small group. This choice is made to simplify the discussion and to focus on the core features of these regimes.
[2] For instance, some possess substantial economic and military prowess or significant natural resources.

survival (Geddes, Wright, and Frantz 2018, 1–5; Weeks 2018, 2–4; Svolik 2012, 2–3; Tullock 1987, 1–17; Ezrow and Frantz 2011, xiii–xiv; Przeworski 2007). Ongoing research increasingly delves into the aspects of autocratic legislatures, bureaucracies, elections, party politics, as well as repression, leadership changes, and regime stability (Svolik 2012, 2; Acemoglu and Robinson 2019; Bueno de Mesquita et al. 2005; Bueno de Mesquita and Smith 2011; Gandhi 2008; Meng 2020; Mukherjee and Koren 2019; Boix and Svolik 2013; Nathan 2017). However, instead of exploring the variation among dictatorships themselves, political scientists often direct their attention toward contrasting democracies and dictatorships.[3] The factors that scholars frequently use to explain distinctions between democracies and dictatorships—such as the presence of genuinely free and fair elections—are not markedly diverse across authoritarian regimes. Consequently, because autocracies differ not only from democracies but also from each other (Geddes, Wright, and Frantz 2018, 1; Ezrow and Frantz 2011, 1), the existing body of research often struggles to deeply scrutinize the differences among them (Weeks 2018, 2).

There is a great heterogeneity among autocracies across many dimensions. Some autocracies are relatively stable and long-lasting, whereas others are short-lived (Wintrobe 1998, 3). Some rely on their capacity to repress the public, whereas others are genuinely supported by the populace (Guriev and Treisman 2022; Wintrobe 1998, 3). Some initiate international conflicts, whereas others live in peace with their neighbors (Geddes, Wright, and Frantz 2018, 1; Weeks 2018). Therefore, the field of autocratic (or authoritarian, depending on the choice of the definition) resilience (survival) is broad. In other words, the question "Why do

[3] Przeworski (2023) critiques dominant formal models of authoritarian regimes for their overemphasis on repression and information manipulation while neglecting the role of governance and service provision in sustaining popular support. He argues that autocracies do not survive solely by misleading or coercing citizens but also by delivering material and symbolic benefits that people value. This perspective challenges the notion that autocrats rule only through fear and deception, suggesting that governance effectiveness contributes to regime resilience.

some autocracies collapse while others display resilience and durability?"—an important question in political science—does not have a single and simple interpretation, way of answering, or even an answer (Bueno De Mesquita et al. 2005; Bueno de Mesquita and Smith 2011; Svolik 2012; Gallagher and Hanson 2013) because autocracies vary immensely in their freedom (V-Dem Institute 2025; Freedom House 2024) and across a range of socioeconomic, political, demographic, and temporal dimensions. Indeed, Kim Jong Un and Viktor Orbán use distinctly different tools to remain in office, and even Russia under Vladimir Putin and the Soviet Union under Joseph Stalin represent different societies not only economically and geographically but also in their regime's methods of repression, propaganda and censorship, tracing public opinion, and building relationships abroad.

Autocracies are uneasy subjects to study (Gel'man 2023; Smyth 2023; Lankina 2023; Dimitrov 2023). There is always a chance that disclosing information about the inside processes would bring to light not only the regime's strengths but also its weaknesses, thereby threatening its stability. Therefore, in autocracies, the state, acting as the thought police, may sometimes censor or modify information about itself, thus making the data released to the general public unreliable or unverifiable (Dimitrov 2023). Moreover, conducting original fieldwork in an autocracy not only requires demonstrating to its authorities that a researcher does not pose any risk to the regime but also involves the inconveniences of traveling to a part of the world that is often unfriendly to foreigners—and, in many cases, the need to learn the language and understand the culture, traditions, and religions of a region.

Autocracies depend on citizen support. Although a mechanism of autocratic resilience always entails a combination of various tools (Wintrobe 1998; Svolik 2012), autocracies can only survive by effectively rallying some critical share of the citizens to their side. On a

fundamental level, citizens support a dictatorship when provided with motivating factors (Wintrobe 1998). To maintain the support of the critical masses, leaders and elites in dictatorships work to ensure that these factors, or incentives, exist (Ezrow and Frantz 2011, 56). The incentives span a spectrum, encompassing economic benefits, such as the allocation of state resources, as well as security considerations rooted in the perception that dictatorship guarantees stability—given the potential for violence if the regime were to collapse (Ezrow and Frantz 2011, 55–56). Additionally, since a dearth of widespread public approval for the regime can serve as an advantage for opposition movements to gain momentum (Magaloni 2006), repression is frequently wielded as a deterrent against the withdrawal of support for the regime (Wintrobe 1998; Francisco 2005; Dimitrov 2023). Gehlbach et al. (2024) further highlight that autocrats face a fundamental dilemma in managing repression and information control, as excessive repression reduces the availability of reliable public opinion signals, forcing rulers to either mobilize their security apparatus preemptively or risk underestimating opposition threats.

Earlier studies of autocratic survival that focused on dictatorship paid significant attention to coercion, control, power sharing, economic redistribution, and policy toward other states,[4] considering these factors the cornerstones of autocratic systems (Wintrobe 1998; Svolik 2012). Many of such studies were largely descriptive in nature and predominantly focused on totalitarianism and fascism (Ezrow and Frantz 2011, 2, 28). The literature emphasized that the classical tyrants were feared rather than genuinely supported by the populace (Dikötter 2019, xiii; Wintrobe 1998). In other words, the old-school autocratic rulers, discussed in the earlier literature, aimed not to convince or persuade the populace, but rather to spread bewilderment, dismantle rationality, impose obedience, isolate individuals, and shatter their

[4] Open or autarkic, peaceful or threatening.

sense of dignity (Arendt 1951/2017). The populace in these regimes was obligated to construct the façade of consent. Those who did not conform were subject to fines, imprisonment, and sometimes lethal repercussions.

More recently, scholars have turned their attention to the institutional foundations of autocratic rule, recognizing that modern dictatorships do not rely solely on coercion but also on carefully designed governance structures. Among the literature on authoritarian institutions, four works require additional attention. Egorov and Sonin (2024a) argue that autocratic institutions serve not only as tools of repression but also as mechanisms for elite coordination and policy enforcement, highlighting the trade-offs between personalist rule and institutionalized governance. In a complementary analysis, Egorov and Sonin (2024b) emphasize the role of information flows in autocratic politics, showing that overly restrictive regimes risk bureaucratic inefficiency, while those allowing too much openness face elite defection. Gehlbach and Simpser (2015) further demonstrate that electoral manipulation is not just about securing victories but also about reinforcing bureaucratic loyalty and signaling regime strength to deter opposition. Similarly, Simpser (2013) argues that excessive and blatant electoral fraud functions as a tool of long-term political control, discouraging elite defection and solidifying authoritarian dominance even in cases where victory is already assured.

However, recent scholarship, focused on present-day dictators, or so-called *spin dictators* and *informational dictators*, suggests that many of these rulers place high importance on influencing public opinion and genuinely—and often successfully—aim to be popular in their countries (Guriev and Treisman 2015; 2019; 2022; Treisman 2018). This scholarship suggests that while authoritarian leaders lack direct public accountability akin to democratically elected officials, they care about popular approval much more than their

predecessors did (Guriev and Treisman 2022, 12). Similar arguments also appear in the general literature on autocratic politics. For instance, Weeks (2018) argues that many scholars underestimate the extent to which dictators can be held accountable by domestic actors and that most dictatorships are able to generate *audience costs* (Fearon 1994) on a level of democracies. Ezrow and Frantz (2011, 148–149) emphasize that although in autocracies the ways for citizens to penalize their leaders are limited, this by no means implies that dictators operate without any form of responsibility. Koesel and Bunce (2013) show that carefully tailored official narratives, combined with control over mass media, are used in autocracies to deter citizens from participating in popular mobilization against the rulers. The study argues that just as voters can punish politicians for failing to uphold their promises in democratic systems, dictators can exert a similar form of accountability. Consequently, according to Guriev and Treisman (2022, 75), instead of intimidating citizens, modern-day tyrants strive to gain popular support through projecting an image of competence, substituting the language of violence with that of accomplishment, and displaying skillful leadership and commitment.

The literature on mass-media manipulation strategies in present-day autocracies not only mirrors these arguments but also supports them with empirical evidence (e.g., Enikolopov, Petrova, and Zhuravskaya 2011; White, Oakes, and McAllister 2005). Carter and Carter (2023) in their fundamental study reveal and explain the variations in state-controlled news media manipulation strategies among non-free regimes. The authors argue that different tyrants use propaganda with different objectives in mind (Carter and Carter 2023, 1–36). Many autocratic rulers currently operate within nominally democratic institutions,

encompassing elections, national legislative bodies, and political opposition. However, the level of restraint imposed by these institutions significantly varies across non-free regimes.[5]

In cases where autocrats face few constraints—in other words, in cases where the rulers can almost entirely rely on repression—propaganda is used to signal the regime's strength and domination over the populace (Huang, Boranbaya-Akan, and Huang 2019; Little 2017). The power of these types of mass-media manipulations (often referred to as *propaganda for signaling* or *propaganda of fear*) stems from their inherent absurdity. By compelling citizens to consume content acknowledged by all of them as false, autocrats make their capacity for repression widely recognized (Svolik 2012, 81; Carter and Carter 2023, 9). In regimes where *propaganda for signaling* is at work, narratives regarding a nation's contemporary history are portrayed in absurd terms, citizens are told that their countries are envied worldwide, that their nation is thriving, and that the dictator is a champion of national sports (Dikötter 2019; Guriev and Treisman 2022) or a technology enthusiast (Jackson 2023). Simply put, *propaganda of fear* is far from surgical; it routinely issues explicit threats of repression to the domestic audience.

In the cases where institutions hold significant weight (such as Putin's regime before Russia's invasion of Ukraine in 2022), the rulers find themselves compelled to garner a certain degree of public support (Carter and Carter 2023, 1–36). Consequently, they use mass-media manipulations to sway citizens toward genuinely endorsing the merits of the rule and the ruler. In these cases, for propaganda to effectively persuade, news narratives need to exhibit a certain degree of neutrality, which, in turn, requires acknowledging some policy failures of

[5] Rozenas and Stukal (2019) show that in autocracies where direct censorship of economic news is impractical, state-controlled media instead employ selective attribution to manipulate public perceptions. Their analysis of Russian television news reveals that bad economic news is not suppressed but systematically blamed on external factors, while positive developments are credited to domestic leadership, particularly President Putin. This strategic framing is used most intensely during politically sensitive periods, such as elections and protests, to reinforce regime legitimacy.

the regime (Carter and Carter 2023, 9). Because viewers will inevitably turn away from the broadcasts that are insufficiently informative (Gehlbach and Sonin 2014), *spin dictators* are constrained in their ability to manipulate the news. Put simply, *persuasive propaganda* needs to be surgical; the effectiveness of mass-media manipulation strategies stems from their subtlety.[6]

It is reasonable to expect that state-controlled mass-media outlets in non-free regimes where nominally democratic institutions hold significant weight may wish to avoid coverage that could generate or exacerbate public discontent and/or emphasize perceived areas of the regime's strength, irrespective of the real or imagined nature of this strength. For this reason, a systematic analysis of the dynamics of state-controlled mass-media narratives may help reveal the changes in the regime's priorities and uncover its perceived areas of strength and weakness (King, Pan, and Roberts 2013; Lankina and Watanabe 2017). In other words, studying the variations in state-controlled information flows may help researchers find out *what an autocracy wants its citizens to think about,* whereas comparing state-controlled news to more independent media content can help uncover what the regime wants *to hide* from the public.

The case of Russia illustrates these dynamics, as Putin's propaganda strategy is shaped by both internal constraints and audience-specific considerations, influencing its persuasiveness inside and outside the country. Unlike purely repressive regimes, Russia has had to balance manipulation with credibility, especially in the digital sphere, where blatant censorship is less

[6] A related argument is presented by Gerschewski (2023), who suggests that autocracies structured according to *de-politicization* tend to produce propaganda that creates an illusion of competence and that, conversely, autocracies organized according to *over-politicization* more commonly employ ideological propaganda.

effective (Sanovich 2017). State-controlled media reinforce support among regime loyalists through belief-affirming propaganda, but their ability to persuade skeptics—domestically and abroad—is limited (Shirikov 2024a). Studies show that propaganda works best on those predisposed to its message, while opposition-minded individuals either reject it or become susceptible to counter-narratives (Peisakhin and Rozenas 2018; Shirikov 2024b).

Additionally, media trust in Russia and other post-communist autocracies is shaped more by contemporary political environments than by historical communist legacies, with trust in state media remaining high among supporters but low among critics (Shirikov 2024c). This suggests that Putin's regime relies less on broad persuasion and more on maintaining an ideological divide, ensuring that state media reinforce existing loyalties.

These constraints shape not only the content of propaganda but also the methods used to control the information environment. As Roberts (2018) highlights, authoritarian regimes do not rely solely on censorship but also employ *friction*—raising the cost of accessing alternative viewpoints—and *flooding*, overwhelming audiences with distracting content to drown out dissent. Similarly, Sanovich, Stukal, and Tucker (2018) show that the Russian government has moved from early engagement with online opposition toward more aggressive tactics, such as coordinated disinformation campaigns and digital suppression. These strategies help sustain Putin's propaganda model by ensuring that alternative narratives remain either inaccessible or discredited, reinforcing the ideological divide maintained by state media (Carter and Carter 2023).

Another studied example is Saudi Arabia, where the regime combines physical repression with computational propaganda to control public discourse. Unlike Russia, which must balance manipulation with credibility, Saudi Arabia relies more heavily on direct coercion and online disinformation. While imprisoning activists silences individual critics, it often

backfires by mobilizing their followers, leading to increased online dissent (Pan and Siegel 2020). At the same time, the Saudi government operates state-backed troll armies and bot networks to amplify pro-regime messages and flood the information space with distractions (Barrie and Siegel 2021). However, like Russia, Saudi Arabia faces limitations in fully controlling online discourse, as engagement with influence operations remains relatively low, and efforts to dominate the digital sphere often generate resistance rather than compliance. This highlights a broader challenge for autocratic regimes: while they adapt their strategies to different constraints, digital spaces remain difficult to fully control, requiring continuous adjustments in repression and propaganda tactics.

TEXT-AS-DATA TO STUDY MASS-MEDIA MANIPULATIONS

Understanding how autocracies manipulate mass media requires systematic analysis of large textual datasets. Traditional media studies relied on qualitative analysis of news coverage, but the scale and complexity of modern propaganda demand computational approaches. Text-as-data methods provide a means to systematically analyze patterns of censorship, propaganda, and narrative framing across large corpora of state-controlled media.

Text-as-data methods are not a novel approach to analyze mass-media strategies. Walter Lippmann's *Public Opinion* (1922) references a study conducted in the early 20th century, in which an individual named Wilcox collected 110 newspapers from 14 large American cities and classified the content of over 9,000 columns. For example, news columns were categorized into "War News," "General," and "Special," with "Special" further subdivided into "Business," "Sport," and "Society." In other words, until recently, text-as-data methods to study mass media relied on meticulous human reading, a method that is not scalable to the extensive corpora now accessible.

In today's world, technology enables machines not only to replace the laborious human classification once performed by Wilcox but also to undertake a wide range of text analysis functions. Enhanced web access allows social scientists to gather digital data from around the globe and connect with peers, thereby facilitating greater expertise, while advanced software and high computational power equip them with the tools necessary to analyze increasingly large datasets.

Text-as-data tools, also referred to as *quantitative text analysis* (in social sciences) or *natural language processing* (NLP, in computer science and data science), are a branch of artificial intelligence methods focused on the interaction between computers and human language. Among a wide variety of other applications, they offer innovative solutions for studying state-controlled mass-media manipulations. Over the last few decades, NLP has evolved from rule-based systems to data-driven approaches, fueled by advancements in machine learning and computational linguistics. Early NLP systems relied heavily on handcrafted rules, limiting their scalability and adaptability. However, with the advent of statistical methods and deep learning techniques, modern NLP models can now automatically learn complex patterns from large datasets, leading to significant improvements in tasks. By applying NLP techniques to analyze large volumes of textual data, researchers of mass-media manipulations in autocracies can uncover patterns, detect narratives, and identify instances of propaganda, misinformation, or censorship.

Because text-as-data methods are a recent development in the study of political communication, there is limited consensus on their optimal application. The field of text analysis exhibits significant methodological diversity, lacking both a unified framework and a common vocabulary to guide research design. This difficulty is exacerbated by the rapid advancement of NLP algorithms. Below, I offer a conceptual overview of the popular

methods of computational text analysis in mass-media communications. Additionally, I provide examples from key academic case studies focusing specifically on mass-media manipulations in autocracies.

Presently, one popular NLP method used in studying mass-media manipulations is ***sentiment analysis***. Sentiment analysis allows researchers to gauge the emotional tone of textual content (Liu 2022). In such analysis, the sentiment of text—such as a word, phrase, or sentence—is generally categorized as either positive, negative, or neutral, or quantified using a rating scale (e.g., valence from 1 to 5). In the context of state-controlled media, sentiment analysis can reveal attempts to manipulate public opinion by framing issues in a particular light or evoking specific emotional responses. There are two primary methodologies for assessing sentiment in text: lexical methodology and machine learning techniques.

Lexical methodology for sentiment analysis uses predefined lists of words, known as lexicons or dictionaries, with each word assigned a specific score reflecting the relevant emotion (Shapiro, Sudhof and Wilson 2022). A good example of this type of dictionary that I used personally—to analyze the sentiment of news stories from the Russian national media outlet Channel One—is a translation of the AFINN dictionary (Nielsen 2011), introduced by Rumshisky et al. (2017). The dictionary is available in open access and is implemented in Python. The algorithm relies on a lexicon comprised of 7,640 words with sentiment scores from −5 to 5. The overall sentiment score of a news report is calculated as a sum of the scores of individual words normalized by a document's word length. Based on my experience of applying the dictionary to a dataset of news transcripts, the accuracy and precision of the results were comparable to the performance of complex machine learning models, which I discuss below, while the computational time and memory requirements were lower.

An illustrative application of using dictionaries to analyze sentiment in news coverage from Russian media is Paskhalis, Rosenfeld, and Tertytchnaya (2022), who employ, compare, and validate the Lexicoder (Young and Soroka 2012) and RuSentiLex (Loukachevitch and Levchik 2016) dictionaries on a corpus of news reports to examine whether an independent outlet reoriented its coverage to favor regime interests. Similarly, Miller (2022) uses a lexicon-based methodology in his analysis of tweets to provide evidence of a backlash against Turkey's Twitter ban in March 2014.

The second, more recent approach to sentiment analysis involves ***machine learning techniques***, which predict the sentiment of a text probabilistically (Shapiro, Sudhof, and Wilson 2022). Two examples of such models that I employed in my research on Russian media are RuBERT for Sentiment Analysis[7] and Rubert-tiny[8]. Both models estimate the predicted probabilities of three classes (positive, negative, or neutral), select classes with maximal probability, and label each document in a corpus accordingly. Rubert-tiny automatically calculates one-dimensional estimates of a sentiment score for each document, which, in addition to sentiment-class labeling, can be used for comparative analysis (e.g., La Lova 2024). It should be noted, however, that tuning transformer models, such as BERT and LLaMA, for sentiment analysis is more challenging than using simple Python dictionaries due to their high computational requirements. These models rely on complex architectures that demand significant processing power and memory. For instance, in my experience, labeling a corpus of 400,000 news reports using a BERT-based model took 10 hours on a MacBook Air (M2, 2022, 8GB RAM).

[7] https://huggingface.co/blanchefort/rubert-base-cased-sentiment
[8] https://huggingface.co/cointegrated/rubert-tiny-sentiment-balanced

Another powerful NLP technique to study mass-media manipulations in autocracies is ***topic modeling***, which involves identifying recurring themes or topics within a corpus of text. By applying topic modeling to analyze state-controlled media content, researchers can identify the dominant narratives promoted by the regime, uncover hidden agendas, and track shifts in propaganda tactics over time. Traditional methods of topic modeling, such as latent semantic analysis (LSA) and latent Dirichlet allocation (LDA) have been foundational in identifying underlying topics by analyzing word co-occurrence patterns and statistical distributions. LSA, introduced by Deerwester et al. (1990), reduces dimensionality by capturing the latent structure in term-document matrices (Deerwester et al. 1990). LDA, proposed by Blei, Ng, and Jordan (2003), assigns documents to a mixture of topics, assuming each word is generated from a topic distribution (Blei, Ng, and Jordan 2003). More recent advancements, including BERT-based models, represent a significant evolution in topic modeling. BERT (Bidirectional Encoder Representations from Transformers), introduced by Devlin et al. (2018), and similar approaches leverage deep learning to generate contextual embeddings that capture nuanced semantic relationships between words, providing a more sophisticated and accurate understanding of topics in contemporary text corpora (Devlin et al. 2018).

An intrinsic difficulty with topic modeling algorithms, regardless of whether they rely on LSA, LDA, or transformers, is their inability to produce objective topic labels. Topics are characterized by a multitude of words, which are often distributed across several topics, leading to outputs that can be challenging to interpret. Even when the topics are interpretable, these unsupervised learning tools operate purely on data-driven principles and lack the capability to target specific concepts. Simply put, for instance, on what basis does the researcher determine that the corpus under study encompasses 10 topics rather than 110? In other words, topic modeling may be advantageous for researchers aiming to explore a corpus without preexisting knowledge; however, it becomes a limitation when the objective is to

align topics with distinct concepts, as the algorithm cannot facilitate this process (Ash and Hansen 2023).

Yet, likely due to their ease of implementation, topic models are among the most widely used methodologies by scholars examining mass media, including those investigating media manipulations within autocratic regimes. For instance, Alrababa'h and Blaydes (2021) used topic models to demonstrate how Syria strategically employed diversionary threats from 1987 to 2018. Additionally, King,Pan, and Roberts (2013) employed these models to show how censorship in China permits government criticism while silencing collective expression. Another example is Otlan et al. (2023), who relied on structural topic modeling to investigate the rhetoric employed by Russian federal TV in its coverage of foreign protests. Furthermore, topic modeling has been used to analyze the content of mass-media coverage in countries such as China (Zhang, Wang, and Hu 2022) and North Korea (Boussalis, Dukalskis, and Gerschewski 2023), among others.

Both sentiment analysis and topic modeling, when combined with network-based features and political event analysis, can uncover the presence of so-called computational propaganda and its strategic deployment in response to political dissent. For instance, Stukal et al. (2022) employ ensemble classifiers to identify pro-government Twitter bots in Russia, analyzing how they engage in negative campaigning against opposition figures and amplify state-approved narratives, particularly during protests. Similarly, Stukal et al. (2019) use a deep neural network to classify Russian bots by political orientation, demonstrating that pro-Kremlin bots follow distinct messaging strategies compared to opposition-aligned ones. Expanding beyond bot-driven manipulation, Munger et al. (2019) analyze how Venezuelan regime elites actively used Twitter during the 2014 anti-Maduro protests, employing topic

modeling and network analysis to show how government officials sought to fragment opposition narratives by flooding the platform with alternative discourses.

Named entity recognition (NER) is another NLP method that proves valuable in studying mass-media tactics. NER allows scholars to identify and classify named entities such as people, organizations, and locations mentioned in textual content (Jurafsky and Martin 2024). By analyzing the prevalence and context of named entities in state-controlled media, researchers can uncover connections between government officials, propaganda outlets, and other actors involved in shaping the narrative. A good example of application of NER in practice is VIINA, a near-real-time, multi-source event data system for the 2022 Russian invasion of Ukraine (Zhukov 2022). VIINA[9] draws on news reports from Ukrainian and Russian media, which are geocoded and classified into standard conflict event categories using machine learning. The data are GIS-ready and temporally precise to the minute.

Another, much simpler but still influential tool for NER is Newsmap (Watanabe 2018), a semi-supervised model that helps label the country-topic of news reports. The model is straightforward, implemented in R, can be improved by using a larger seed dictionary (La Lova 2024), and requires less computational power than more complex tools. Presently, based on Google Scholar, 58 academic publications have relied on Newsmap to classify news reports in terms of the country-topic of coverage.

Text similarity measures such as cosine similarity, Jaccard similarity, and Euclidian distance are widely used metrics for quantifying the similarity between two vectors in a high-dimensional space, commonly applied in text analysis (Manning, Raghavan, and Schütze 2008). Cosine similarity, for example, measures the cosine of the angle between two vectors,

[9] https://www.sungeo.org/node/15

which reflects how closely related they are, irrespective of their magnitude. This approach can be particularly useful in identifying thematic connections and trends within mass-media corpora. Cosine similarity can help in comparing news articles from different sources to assess how closely related their coverage of a particular event is. It can also be employed to detect shifts in media narratives over time by comparing the similarity of articles on the same topic across different periods. Additionally, cosine similarity can aid in clustering similar articles, thereby facilitating the categorization of news into distinct topics or identifying instances of potential media bias by comparing the content of articles from various outlets. By quantifying the textual similarity, researchers can better understand the discourse patterns and thematic emphasis within mass-media communications. A good example of using cosine similarity to address questions in mass-media management in autocracies is Paskhalis, Rosenfeld, and Tertytchnaya (2022), who estimated the similarity between independent and state-controlled TV channels in Russia.

Finally, the role of simple ***dictionaries***—structured collections of words or phrases pre-assigned with specific attributes such as categories, thematic labels, and sentiment labels already mentioned above—should not be underestimated, particularly during the data exploration phase, before designing and validating more accurate and precise tools. Dictionaries serve as a foundational tool for various text analysis tasks, including sentiment analysis, thematic extraction, and keyword identification (Grimmer and Stewart 2013). Dictionaries are constructed based on linguistic expertise and empirical data, with entries often including words along with their associated emotional or categorical values. For instance, sentiment dictionaries might score words on a scale reflecting their positive or negative sentiment (Liu 2022), while thematic dictionaries categorize terms according to specific topics or fields of study (Turney and Littman 2002).

In social science research, several dictionaries are widely used to analyze textual data. One of the most established is the LIWC (Linguistic Inquiry and Word Count), which categorizes words into various psychological and emotional dimensions, allowing researchers to explore affective, cognitive, and structural aspects of text (Pennebaker, Booth, and Francis 2001). Another influential resource is WordNet, a lexical database that groups words into sets of synonyms and defines their semantic relationships, which aids in understanding word meaning and context (Miller 1995). The SentiWordNet dictionary extends WordNet by providing sentiment scores for synsets, which helps quantify sentiment in text (Esuli and Sebastiani 2006). Additionally, the VADER (Valence Aware Dictionary and sEntiment Reasoner) is designed for analyzing sentiment in social media texts and other informal contexts, offering a nuanced approach to sentiment analysis (Hutto and Gilbert 2014).

The use of dictionaries allows researchers to systematically analyze and quantify text by mapping its content to predefined attributes. This method facilitates tasks such as identifying trends, measuring sentiment, and classifying thematic elements within large corpora. However, the effectiveness of dictionary-based approaches is dependent on the dictionary's comprehensiveness, relevance, and contextual fit with the text being analyzed.

A variety of examples of using dictionary-based analysis to examine mass-media coverage in autocracies is provided in the fundamental work by Carter and Carter (2023). The authors employ dictionaries across various subsets of propaganda corpora to analyze narratives about domestic and foreign events. Their study is based on a corpus comprising over eight million articles from 65 newspapers across 59 countries and in six major languages.

DRAWBACKS OF TEXT AS DATA METHODS

Despite its potential, the use of NLP to study mass-media manipulations in autocratic regimes also poses challenges and limitations. Language barriers, data availability, and algorithm biases can affect the accuracy and reliability of NLP analyses. Moreover, NLP methods may struggle to capture nuanced forms of manipulation, such as subtle propaganda techniques or context-dependent narratives. In other words, it is vital to acknowledge that while text-as-data may offer a practical approach for analyzing extensive collections of news transcripts, social media posts, and other texts from mass media, they generally yield less precise and accurate results compared to a detailed analysis conducted by human coders. As Grimmer and Stewart (2013) mention in their seminal paper, rigorous validation is vital when using automated content analysis methods. In the absence of validation, the results of NLP methods can be misleading or inaccurate.

CONCLUSION

In the study of mass-media manipulations within autocratic regimes, text-as-data methods offer a transformative approach for uncovering intricate patterns of propaganda, censorship, and disinformation. By leveraging advanced computational techniques, researchers can analyze vast datasets from state-controlled media sources with unprecedented scale and precision. These methods not only illuminate the strategies autocracies use to maintain control and shape public opinion but also enhance our understanding of how regimes adapt to changing political landscapes and technological constraints.

The application of sentiment analysis, topic modeling, named entity recognition, and other NLP techniques has already yielded valuable insights into the diverse and evolving methods of media manipulation employed by autocratic leaders. These tools allow scholars to detect shifts in propaganda narratives, uncover hidden agendas, and track how state-controlled media respond to political crises or public dissent. The growing body of research in this field

demonstrates the potential of computational methods to reveal patterns that might otherwise remain obscured in traditional qualitative media analysis.

However, significant challenges remain. The effectiveness of text-as-data methods depends on data accessibility, language resources, and algorithmic transparency. Many autocratic regimes actively restrict access to media archives, censor content, or flood digital spaces with noise, making it difficult to construct reliable datasets. Additionally, cultural and linguistic biases in NLP models may affect the accuracy of sentiment and topic analysis, particularly in non-English-language corpora. Addressing these challenges requires interdisciplinary collaboration between political scientists, computational linguists, and media researchers to refine methodologies and improve the robustness of automated text analysis in authoritarian contexts.

Future research should explore how emerging technologies—such as multimodal AI models, deep learning for image and video analysis, and real-time social media monitoring—can enhance the study of autocratic media strategies. Additionally, comparative studies across different autocratic regimes and media ecosystems can further illuminate the commonalities and divergences in propaganda tactics, shedding light on how regimes adjust their media strategies in response to international pressures and digital resistance movements.

Ultimately, the continued advancement of text-as-data methods provides a powerful analytical framework for studying authoritarian information control. These tools not only contribute to academic scholarship but also offer practical applications for journalists, policymakers, and civil society organizations seeking to understand, counteract, and expose state-controlled media manipulations. As computational methods evolve, so too does our ability to track, interpret, and challenge the increasingly sophisticated strategies autocracies use to shape information ecosystems in the digital age.

Acknowledgments

The author is grateful to Maxim Alyukov, Gulnaz Sharafutdinova, Tomila Lankina, Katerina Tertytchnaya, and Kenneth Benoit, as well as the audience at King's College London, UK, during the workshop "Social Science Research in the Context of Conflict, War, and Authoritarianism: Methodological Challenges and New Approaches to Data Collection and Analysis," where this work was presented, and to the anonymous reviewers for their insightful comments and valuable feedback.

Corresponding author email: l.lalova@lse.ac.uk

Published online

References

Acemoglu, Daron, and James A. Robinson. 2019. *The Narrow Corridor: How Nations Struggle for Liberty*. London: Penguin Books.

Alrababa'h, Ala', and Lisa Blaydes. 2021. "Authoritarian Media and Diversionary Threats: Lessons from 30 Years of Syrian State Discourse." *Political Science Research and Methods* 9, no. 4: 693–708.

Arendt, Hannah. 2017. *The Origins of Totalitarianism*. London: Penguin Books. First published 1951 by Schocken Books.

Ash, Elliott, and Stephen Hansen. 2023. "Text Algorithms in Economics." *Annual Review of Economics* 15, no. 1: 659–688.

Barrie, Christopher, and Alexandra A. Siegel. 2021. "Kingdom of Trolls? Influence Operations in the Saudi Twittersphere." *Journal of Quantitative Description: Digital Media* 1: 1–41.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.

Boix, Carles, and Milan W. Svolik. 2013. “The Foundations of Limited Authoritarian Government: Institutions, Commitment, and Power-Sharing in Dictatorships.” *The Journal of Politics* 75, no. 2: 300–316.

Boussalis, Constantine, Alexander Dukalskis, and Johannes Gerschewski. 2023. “Why It Matters What Autocrats Say: Assessing Competing Theories of Propaganda.” *Problems of Post-Communism* 70, no. 3: 241–252.

Bueno de Mesquita, Bruce, and Alastair Smith. 2011. *The Dictator’s Handbook: Why Bad Behavior Is Almost Always Good Politics*. London: Hachette.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson, and James D. Morrow. 2005. *The Logic of Political Survival*. Cambridge: MIT Press.

Carter, Erin Baggott, and Brett L. Carter. 2023. *Propaganda in Autocracies: Institutions, Information, and the Politics of Belief*. Cambridge, UK: Cambridge University Press.

Deerwester, Scott, Susan T. Dumais, Robert W. Furnas, Thomas K. Landauer, and Ginny Harshman. 1990. “Indexing by Latent Semantic Analysis.” *Journal of the American Society for Information Science* 41, no. 6: 391–407.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, MN: Association for Computational Linguistics.

Dikötter, Frank. 2019. *How to Be a Dictator: The Cult of Personality in the Twentieth Century*. New York: Bloomsbury Publishing.

Dimitrov, Martin K. 2023. *Dictatorship and Information: Authoritarian Regime Resilience in Communist Europe and China*. Oxford: Oxford University Press.

Egorov, Georgy, and Konstantin Sonin. 2024a. “Authoritarian Institutions.” In *Handbook of Institutional Economics*, 2nd ed., edited by Claude Menard and Mary Shirley, 209–236. Cham, Switzerland: Springer.

Egorov, Georgy, and Konstantin Sonin. 2024b. “The Political Economics of Non-Democracy.” *Journal of Economic Literature* 62, no. 2: 594–636.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. “Media and Political Persuasion: Evidence from Russia.” *The American Economic Review* 101, no. 7: 3253–3285.

Esuli, Andrea, and Fabrizio Sebastiani. 2006. “SentiWordNet: A Publicly Available Lexical Resource for Sentiment Analysis.” In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, 417–422. Genoa, Italy: European Language Resources Association (ELRA).

Ezrow, Natasha M., and Erica Frantz. 2011. *Dictators and Dictatorships: Understanding Authoritarian Regimes and Their Leaders*. New York: Bloomsbury Publishing.

Fearon, James D. 1994. “Domestic Political Audiences and the Escalation of International Disputes.” *American Political Science Review* 88, no. 3: 577–592.

Francisco, Ronald A. 2005. “The Dictator’s Dilemma.” *Repression and Mobilization* 64, no. 2: 58–81.

Freedom House. 2024. “Freedom in the World: The Mounting Damage of Flawed Elections and Armed Conflict.” Accessed June 12, 2025. https://freedomhouse.org/sites/default/files/2024-02/FIW_2024_DigitalBooklet.pdf

Gallagher, Mary, and Jonathan K. Hanson. 2013. “Authoritarian Survival, Resilience, and the Selectorate Theory.” In *Why Communism Did Not Collapse: Understanding Authoritarian Regime Resilience in Asia and Europe*, edited by Martin K. Dimitrov, 185–204. Cambridge: Cambridge University Press.

Gandhi, Jennifer. 2008. *Political Institutions Under Dictatorship*. Cambridge: Cambridge University Press.

Geddes, Barbara, Joseph George Wright, and Erica Frantz. 2018. *How Dictatorships Work: Power, Personalization, and Collapse*. Cambridge: Cambridge University Press.

Gehlbach, Scott, Zhaotian Luo, Anton Shirikov, and Dmitriy Vorobyev. 2024. "Is There Really a Dictator's Dilemma? Information and Repression in Autocracy." *SSRN*. April 2, 2025. Accessed June 12, 2025. https://ssrn.com/abstract=4782118

Gehlbach, Scott, and Alberto Simpser. 2015. "Electoral Manipulation as Bureaucratic Control." *American Journal of Political Science* 59, no. 1: 212–224.

Gehlbach, Scott, and Konstantin Sonin. 2014. "Government Control of the Media." *Journal of Public Economics* 118: 163–171.

Gel'man, Vladimir. 2023. "Exogenous Shock and Russian Studies." *Post-Soviet Affairs* 39, nos.1–2: 1–9.

Gerschewski, Johannes. 2023. *The Two Logics of Autocratic Rule*. Cambridge: Cambridge University Press.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21, no. 3: 267–297.

Guriev, Sergei, and Daniel Treisman. 2015. "How Modern Dictators Survive: An Informational Theory of the New Authoritarianism." *National Bureau of Economic Research* Working Paper No. w21136.

Guriev, Sergei, and Daniel Treisman. 2019. "Informational Autocrats." *Journal of Economic Perspectives* 33, no. 4: 100–127.

Guriev, Sergei, and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century*. Princeton, NJ: Princeton University Press.

Hallgarten, George Wolfgang Felix. 1954. *Why Dictators? The Causes and Forms of Tyrannical Rule, Since 600 BC*. New York: Macmillan Company.

Huang, Haifeng, Serra Boranbay-Akan, and Ling Huang. 2019. “Media, Protest Diffusion, and Authoritarian Resilience.” *Political Science Research and Methods* 7, no. 1: 23–42.

Hutto, Clayton J., and Eric Gilbert. 2014. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM 2014)*, 216–225. Ann Arbor, MI: AAAI Press (Association for the Advancement of Artificial Intelligence).

Jackson, Patrick. 2022. “Ukraine War: Vladimir Putin Gets a Tractor for His 70th Birthday.” *BBC News*, October 7, 2022. Accessed June 12, 2025. https://www.bbc.co.uk/news/world-europe-63176124

Jurafsky, Daniel, and James H. Martin. 2024. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd ed. Cambridge: Cambridge University Press.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. “How Censorship in China Allows Government Criticism but Silences Collective Expression.” *American Political Science Review* 107, no. 2: 326–343.

Koesel, Karrie J., and Valerie J. Bunce. 2013. “Diffusion-Proofing: Russian and Chinese Responses to Waves of Popular Mobilizations Against Authoritarian Rulers.” *Perspectives on Politics* 11, no. 3: 753–768.

Kuznetsova, Daria. 2024. “Broadcasting Messages via Telegram: Pro-Government Social Media Control During the 2020 Protests in Belarus and 2022 Anti-War Protests in Russia.” *Political Communication* 41, no. 4: 509–530.

La Lova, Lanabi. 2024. “Vladimir Putin on Channel One, 2000–2022.” *Political Communication* 42, no. 2: 1–19.

Lankina, Tomila. 2023. “Branching Out or Inwards? The Logic of Fractals in Russian Studies.” *Post-Soviet Affairs* 39, nos. 1–2: 70–85.

Lankina, Tomila, and Kohei Watanabe. 2017. "'Russian Spring' or 'Spring Betrayal'? The Media as a Mirror of Putin's Evolving Strategy in Ukraine." *Europe-Asia Studies* 69, no. 10: 1526–1556.

Lippmann, Walter. 1922. *Public Opinion*. London: George Allen & Unwin.

Little, Andrew T. 2017. "Propaganda and Credulity." *Games and Economic Behavior* 102: 224–232.

Liu, Bing. 2022. *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer Nature.

Loukachevitch, Natalia, and Alexander Levchik. 2016. "Creating a General Russian Sentiment Lexicon." Paper presented at the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, May. European Language Resources Association (ELRA).

Magaloni, Beatriz. 2006. *Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico*. Cambridge: Cambridge University Press.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Manning, Christopher, and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Meng, Anne. 2020. *Constraining Dictatorship*. Cambridge: Cambridge University Press.

Miller, Andrew Cesare. 2022. "#DictatorErdogan: How Social Media Bans Trigger Backlash." *Political Communication* 39, no. 6: 801–825.

Miller, George A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38, no. 11: 39–41.

Mukherjee, Bumba, and Ore Koren. 2019. *The Politics of Mass Killing in Autocratic Regimes*. Cham, Switzerland: Palgrave Macmillan.

Munger, Kevin, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2019. "Elites Tweet to Get Feet off the Streets: Measuring Regime Social Media Strategies During Protest." *Political Science Research and Methods* 7, no. 4: 815–834.

Nathan, Andrew J. 2017. "China's Changing of the Guard: Authoritarian Resilience." In *Critical Readings on the Communist Party of China*, Vol. 4, 86–99. Leiden: Brill.

Nielsen, Finn Årup. 2011. "AFINN." *Informatics and Mathematical Modelling*, Technical University of Denmark. Accessed June 12, 2025. http://www2.compute.dtu.dk/pubdb/pubs/6010-full.html

Otlan, Yana, Yulia Kuzmina, Aleksandra Rumiantseva, and Katerina Tertytchnaya. 2023. "Authoritarian Media and Foreign Protests: Evidence from a Decade of Russian News." *Post-Soviet Affairs* 39, no. 6: 391–405.

Pan, Jennifer, and Alexandra A. Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114, no. 1: 109–125.

Paskhalis, Tom, Bryn Rosenfeld, and Katerina Tertytchnaya. 2022. "Independent Media under Pressure: Evidence from Russia." *Post-Soviet Affairs* 38, no. 3: 155–174.

Peisakhin, Leonid, and Arturas Rozenas. 2018. "Electoral Effects of Biased Media: Russian Television in Ukraine." *American Journal of Political Science* 62, no. 3: 535–550.

Pennebaker, James W., Roger J. Booth, and Martha E. Francis. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum Associates.

Przeworski, Adam. 2007. "Capitalism, Democracy, and Science." In *Passion, Craft, and Method in Comparative Politics*, edited by Gerardo L. Munck and Richard Snyder, 456–503. Baltimore, MD: Johns Hopkins University Press.

Przeworski, Adam. 2023. "Formal Models of Authoritarian Regimes: A Critique." *Perspectives on Politics* 21, no. 3: 979–988.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton, NJ: Princeton University Press.

Rozenas, Arturas, and Denis Stukal. 2019. “How Autocrats Manipulate Economic News: Evidence from Russia’s State-Controlled Television.” *The Journal of Politics* 81, no. 3: 982–996.

Rumshisky, Anna, Mikhail Gronas, Paul Potash, Mikhail Dubov, Alexander Romanov, Sonal Kulshreshtha, and Alexey Gribov. 2017. “Combining Network and Language Indicators for Tracking Conflict Intensity.” In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part II*, Vol. 9, 391–404. Cham, Switzerland: Springer International Publishing.

Sanovich, Sergey. 2017. “Computational Propaganda in Russia: The Origins of Digital Misinformation.” *Oxford Internet Institute*. Accessed February 19, 2025. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Russia.pdf.

Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. 2018. “Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia.” *Comparative Politics* 50, no. 3: 435–454.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson. 2022. “Measuring News Sentiment.” *Journal of Econometrics* 228, no. 2: 221–243.

Shirikov, Anton. 2024a. “Fake News for All: How Citizens Discern Disinformation in Autocracies.” *Political Communication* 41, no. 1: 45–65.

Shirikov, Anton. 2024b. “Rethinking Propaganda: How State Media Build Trust Through Belief Affirmation.” *The Journal of Politics* 86, no. 4: 1319–1332.

Shirikov, Anton. 2024c. “Why Is Media Trust Low in Post-Communist Countries?” *Problems of Post-Communism* 71, no. 6: 544–556.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge: Cambridge University Press.

Smyth, Regina. 2023. “Plus ça Change: Getting Real About the Evolution of Russian Studies After 1991.” *Post-Soviet Affairs* 39, no. 1–2: 10–26.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2022. "Why Botter: How Pro-Government Bots Fight Opposition in Russia." *American Political Science Review* 116, no. 3: 843–857.

Stukal, Denis, Sergey Sanovich, Joshua A. Tucker, and Richard Bonneau. 2019. "For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia." *SAGE Open* 9, no. 2: 1–16.

Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. Cambridge: Cambridge University Press.

Treisman, Daniel. 2018. *The New Autocracy: Information, Politics, and Policy in Putin's Russia*. Washington, DC: Brookings Institution Press.

Tullock, Gordon. 1987. *Autocracy*. Dordrecht: Martinus Nijhoff Publishers.

Turney, Peter D., and Michael L. Littman. 2002. "Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus." Technical Report ERB-1094, NRC-44929. Ottawa: National Research Council of Canada.

V-Dem Institute. 2025. "Democracy Winning and Losing at the Ballot." Accessed June 12, 2025. https://www.v-dem.net/publications/democracy-reports/.

Watanabe, Kohei. 2018. "Newsmap: A Semi-Supervised Approach to Geographical News Classification." *Digital Journalism* 6, no. 3: 294–309.

Weeks, Jessica. 2018. *Dictators at War and Peace.* Ithaca, NY: Cornell University Press.

White, Stephen, Sarah Oates, and Ian McAllister. 2005. "Media Effects and Russian Elections, 1999–2000." *British Journal of Political Science* 35, no. 2: 191–208.

Wintrobe, Ronald. 1998. *The Political Economy of Dictatorship*. Cambridge: Cambridge University Press.

Young, Lori, and Stuart Soroka. 2012. "Affective News: The Automated Coding of Sentiment in Political Texts." *Political Communication* 29, no. 2: 205–231.

Zhang, Maggie Mengqing, Xiao Wang, and Yang Hu. 2022. “Strategic Framing Matters but Varies: A Structural Topic Modeling Approach to Analyzing China’s Foreign Propaganda About the 2019 Hong Kong Protests on Twitter.” *Social Science Computer Review* 41, no. 2: 397–414.

Zhukov, Yuri. 2022. “Near-Real Time Analysis of War and Economic Activity During Russia’s Invasion of Ukraine.” *Journal of Comparative Economics* 51, no. 4: 1232–1243.