# Wartime Propaganda in Social Media

# Introduction

#### Purpose of this study

This thesis aims to examine how propaganda is used in social media during a state of war. In particular, I evaluate multimodal – text and image - Twitter data from Russian and Ukrainian governmental organizations and state-affiliated politicians concerning the 2022 Russian invasion into Ukraine and the ongoing Russo-Ukrainian War. This examination is done on multiple levels. First, the data is manually annotated according to the propaganda types found within the text or image. A manual analysis follows, examining what types of propaganda are most common, whether there are any differences on the Ukrainian – Russian or organization – individual level, and whether the content of the propaganda has changed over the course of time. Finally, several state-of-the-art textual, visual, and multimodal models are used to further evaluate which pieces of propaganda are easy to discern, and whether they are comparable to another multimodal type of media – memes.

This paper is structured as follows. First, an introductory section, detailing what propaganda is, as well as its relation to war, social media, and cognition, is given. It is followed by a section on the particular types of propaganda examined in this thesis, as well as background on computational propaganda, and a further examination of what types of propaganda might be expected to be found in Russian and Ukrainian media. Next comes the methods section, detailing the data, how it was was collected, cleaned and annotated, as well as the manual and statistical analysis. This is followed by experimental analysis with machine learning models. Finally, the results of the manual and computational analyses are examined and discussed. The paper is concluded with an overview of the work done here and considerations for further work.

[particular research questions here or later? Hypotheses?]

#### What is propaganda and why is it important?

*Propaganda* is a term with many interpretations. In general, it means to disseminate or promote particular ideas, and the word itself is often used in a negative sense. When we use *propaganda* to denote purpose, it is regarded as a way of conveying an ideology to an audience, seeking to reinforce or modify its attitudes, behavior, or both. It is used by all kinds of individuals and organizations – governmental and private, reactionary and revolutionary, left- and right-leaning, militaristic and pacifistic. It is a way of delivering a message that is easily understood and has a quick, strong impact on the audience (Jowett & O’Donnell, 2011).

Propaganda has existed throughout all ages. Confucius, in his *Analects,* emphasizes “skillfulness in speech”, and warns against what might be perceived as propaganda in modern days: “You show approval of him because he seems sincere in what he says. But is he truly a gentleman, or does he simply put on a dignified appearance?”. In ancient civilizations, institutional propaganda was often produced in the form of monuments for the rulers, such as the Pyramids in Ancient Egypt. Because war was a common activity, establishing the basis for war was partly done by exemplifying the differences between the warring civilizations, and the unique culture, such as statues of the gods popular in the specific area, were used both to signify the strength of the state as well as a propaganda piece to further the “us versus them” narrative (Jowett & O’Donnell, 2011). In the 20th and 21st centuries, propaganda evolved, with governments around the world forming organizations tasked with producing their own, as well as discrediting enemy propaganda. This evolution came with a change in form as well: mass-produced posters, radio, films, television, and finally, social media, became the primary forms of communication, and thus propaganda.

Propaganda is a powerful tool that is often used to deceive individuals or, in some form, shroud the truth. This can range from relatively harmless actions, such as advertisements for products, to productions with huge consequences, such as media campaigns intent on justifying war. It is therefore imperative that we understand how propaganda works, which types of it are used, which ones are useful, and what action can be taken to counteract it. It has been shown that when hate speech is exposed as such, it tends to lose its power (Denton, 2000, as cited in Bondarenko, 2020) – such is the case with propaganda as well.

#### Propaganda and War

A state of war requires a shift in social dynamics. This takes on various forms: an increase in spending on the military, the introduction of conscription, food rationing, an increase in media censorship. All of these choices require a justification to the public – otherwise, support for the war may go down, which would require the state to withdraw. Propaganda is an effective way of establishing why these changes are required, and it has thus been used widely during modern wars.

Purpose of propaganda during war

The actions taken by governments during wars are often detrimental to an individual, reducing both their rights and their quality of life. Because of this, an important purpose of propaganda during wartime is to convince people that the sacrifice of individualism for the purpose of a “greater good” is a worthy endeavor (Laskin, 2019). Laskin states that J.F. Kennedy’s famous statement “ask not what your country can do for you, ask what you can do for your country” is “the quintessence of propaganda”. McCrann (2009) agrees, noting the common use of the word “we” in propaganda, which has two purposes: firstly, it establishes the individual as part of the group, and notes that people identifying with a group may do activities that help the group, even if they are detrimental to the individual, as well as validate the group’s behavior, while the group validates theirs, making the social identity of a person act as a “social glue” (Van Vugt & Hart, 2004); secondly, it implies an antagonistic group of “them”, using the social dynamics of in- and outgroups to, among other things, dampen empathetic feelings towards the enemy (Cikara et al., 2011).

Forms of propaganda during war

Another important factor to consider is the forms of propaganda used during war. In the two world wars, propaganda posters issued by governments were used for communicating public policy (McCrann, 2009). Posters are an useful form of communication, because they can deliver a multimodal message: both the image, and the (optional, but often used) accompanying text can have meanings that reinforce each other. Images allow for a communication based on shared history and culture, while text may be used for a more direct message or an implied course of action. McCrann notes how using an image of a strong man, wearing working clothes, with dirt on his hands, surrounded by a range of middle-class men sitting and listening to him, “reinforces the concept of a historical commitment to American values, character, and citizen governance”. The accompanying text, “Save Freedom of Speech – Buy War Bonds”, while not being directly related to the image, suggests to the reader a course of action for retaining this image of a free nation and becoming part of it. As another example, McCrann (2009) notes how the Japanese government used the image of a cherry blossom, a well known historic symbol of the samurai, in combination with images of kamikazes, painting these fatal missions as noble, respectable efforts.

In the last 20 years, the Internet, in particular social media, have become a popular source of information, and they have been exploited to deliver propaganda during wartime. The 2011 revolution in Egypt initially started as a protest event on the social networking website Facebook, and the Internet was widely used by protesters to exchange information, organize events, as well as deliver real-time updates on the ongoing revolution (El-Khalili, 2013). Governmental organizations were not initially active online, but after seeing the power this new form of communication can have, moved quickly to take over. Various state-affiliated groups, including the parliament, the army, as well as individual ministries created social media accounts, and started using them more often. In particular, the Supreme Council of the Armed Forces (SCAF) began using social media for communicating directly with the citizens. With the unique role SCAF played in the revolution, starting out in support of the revolutionaries, but later taking over power and becoming rulers, a change in their communication could be observed (Naguib, 2011), initially coming out with messages stating that the army would protect the revolutionaries and act merely as “guardians of the revolution”, shifting over time to acting as “bearers of stability” and promising to lead the country back to a normal life, and finally branding the revolutionaries as thugs and painting the revolutionary youth as enemies of the nation.

#### Propaganda and Social Media

As mentioned before, propaganda took on new forms with the advent of the Internet and social media. The Internet offers a unique environment for the spread of disinformation, unverified facts, and a mix between news and rumours, sometimes called “paranews” (Dowe, 1997). It did not take long for political propaganda to make way to it. In fact, an entire ecosystem of propaganda exists solely within the Internet – media companies created to spread particular ideas, real and fake accounts of individuals spreading these same ideas on social media, thus appearing more credible, as well as websites made for the sole purpose of divulging the political biases – both in content and control – the aforementioned companies have (Jowett & O’Donnell, 2011). While propaganda is strongest when it comes from a centralized, authoritative source, social media allows any individual the power to spread their message, as there is little to no upfront cost and no journalistic ethics to consider. New platforms for engagement emerge often, and many people use social media as a source of news (Walker & Matsa, 2021).

Even though social media allows smaller organizations and individuals to spread the message, state-affiliated groups have been noted to disseminate propaganda via the Internet as well. An extensive research into how Russia uses social media found that there are multiple points of attack: social media operations focusing on the near abroad, where a significant amount of the population speaks Russian and there are historical and cultural ties; operations focusing on the far abroad, such as the 2016 U.S. election; as well as pro-Russian social media accounts that may not be directly affiliated with the state, but receive support from it (Helmus et al., 2018). The research also separates their means into three different levels: “white” outlets, such as official governmental accounts and state-affiliated news websites, “gray” outlets, such as consipracy websites, radicalised political accounts and news aggregators, as well as “black” outlets, such as bot farms and fake accounts. These efforts have countable effects: people that engaged with Russia-based media tended to be more accepting of Russian narratives, such as the U.S. government’s fault in the Russo-Ukrainian war (Gerber & Zavisca, 2016).

[chinese propaganda?]

[targets of propaganda]

#### Propaganda and Cognition

[propaganda and war return here]

[lakoff here perhaps]

Propaganda should be considered from the cognitive perspective as well. As there are a lot of different types of propaganda (discussed more thoroughly in section []), it would be useful to know the psychological reasons why some types work, and whether some work better than others.

In her research on the cognitive underpinnings of propaganda, Bondarenko (2020) discerns four different types of explicit propaganda and considers their uses. First, the purpose of converting an individual construal of the world to a collective one is reiterated, as mentioned in section [purpose of propaganda section]. However, differently than Laskin and McCrann, Bondarenko considers this involvment in a group an illusion, as the opinions and ideas of the group are not initially inherent in or natural for the individual targeted by this propaganda. The types of propaganda used for this task are *reference to authority* and *support group*. Second, Bondarenko underlines the use of emotion in propaganda. Emotions allow an instinctual or socially conditioned shift in mood for the targets of propaganda (Forgas & Koch, 2013). In political propaganda, negative emotions, such as fear, anger, or sense of guilt are most common, as they can easily be connected to an enemy; however, emotions such as compassion are sometimes appealed to as well. The emotional is used over the rational, as it allows for a conceptual entrenchment in memory, and thus, an effect that lasts longer (Bondarenko, 2020).

Third, reality is distorted by means of transforming or substituting some aspects presented in the propaganda piece. Here Bondarenko refers to joint attention and how methods such as *false analogies, re-scaling* and *local absence of alternative* are used to, in cognitive terms, substitute the anchor of joint attention (Tomasello, 1995) for an alternative, distorted one. This can be done in a number of ways: *false analogies* use some form of real relations as an anchor (such as a large number of Ukrainians knowing how to speak Russian) and replace that with a propagandistic anchor (a large number of people living in Ukraine *are* Russian). *Re-scaling* intensifies or diminishes the importance of a real anchor (such as the debt a country has accrued); *local absence of alternative* supposes that a proposed idea is inherently correct since there are (locally) no realistic alternatives.

Finally, Bondarenko (2008) considers words-triggers: words, phrases or statements that trigger a picture or concept in the target’s minds. These can be culturally related concepts, like the cherry blossoms mentioned above, but they can also be normal words that have become conceptually entwined with other words and concepts by the use of propaganda itself. As an example, Snake Island, a small Ukrainian island in the Black Sea, became a concept of resilience and hope for Ukraine after a piece of audio from the military personnel stationed there, recording a conversation in which the forces refused to surrender to an attacking Russian warship, was distributed in the media. An interesting case occurs when both sides in a military conflict use the same word-trigger for differing concepts: for example, the phrase “special military operation”, used by Russian president Vladimir Putin to refer to the invasion of Ukraine in 2022, may trigger a sense of solidarity among Russian people, as it stresses that the conflict is not a war or invasion, while the word “operation” implies a sense of professionalism. On the other hand, for Ukrainian people, it may be used as an example of the double-speak and propaganda that Russia employs towards the war.

[peers book here maybe]

# Methods

# Data

Data was gathered from Twitter, using the twarc (Summers et al., 2023) Python package. 10710 tweets were gathered from 28 different accounts in the timeframe of 2022-02-24 - 2023-02-08. The tweets were then further filtered to contain only those that were written in English and had both an image and text. Replies and retweets were excluded to prevent repetitions, as a lot of the accounts are naturally related and retweet each other often. These filters led to the final dataset of 5133 tweets. Note that at the time of writing, Twitter API only allows access to about 800 of the most recent tweets from a single account’s timeline, which means that the recency of the tweets in the final dataset depends on how active the account is, as tweets that do not contain exactly an image with text were filtered out. Figure 1 shows the distribution of the tweets with regards to time. As can be seen and mentioned above, the data is skewed in time, with more tweets collected as their creation date approaches the collection date, 2023-02-08. A small number of datapoints from the start of the invasion can be noted, with a large gap between them and the rest of the data. Upon manual inspection of the data, these outliers did not seem out of the ordinary in their content or authorship, but for the purposes of this thesis, the 24 tweets created between the 24th and 28th of February 2022 were excluded (are they?).

The 28 accounts were chosen manually. The accounts of the foreign ministries of both of the countries (*mfa\_russia* and *mfa\_ukraine*) were chosen as starting points. From these, an inspection of the accounts they followed was done. While the accounts were ultimately chosen subjectively, some criteria was set: they should have a Twitter-assigned label of one of *Russian government organization*, *Russian government official*, *Ukrainian government organization*, or *Ukrainian government official;* they should communicate either completely, or primarily, in English; and they should have some relevance to the Russo-Ukrainian war. Additionally, one account labeled as *state-affiliated media* was chosen for both Ukraine and Russia. Note that these labels existed at the time of data collection; after that, Twitter reworked its account functionality, and accounts now have gray check marks to denote their status as a government organization or official, though it does not note which government the account is affiliated to. Additionally, the media accounts lost their *state-affiliated media* label. A list of the accounts used, a short description, and the number of their tweets in the dataset are shown in Tables X and X. [filter for irrelevant posts and update]

[not using media agencies anymore – rework here]

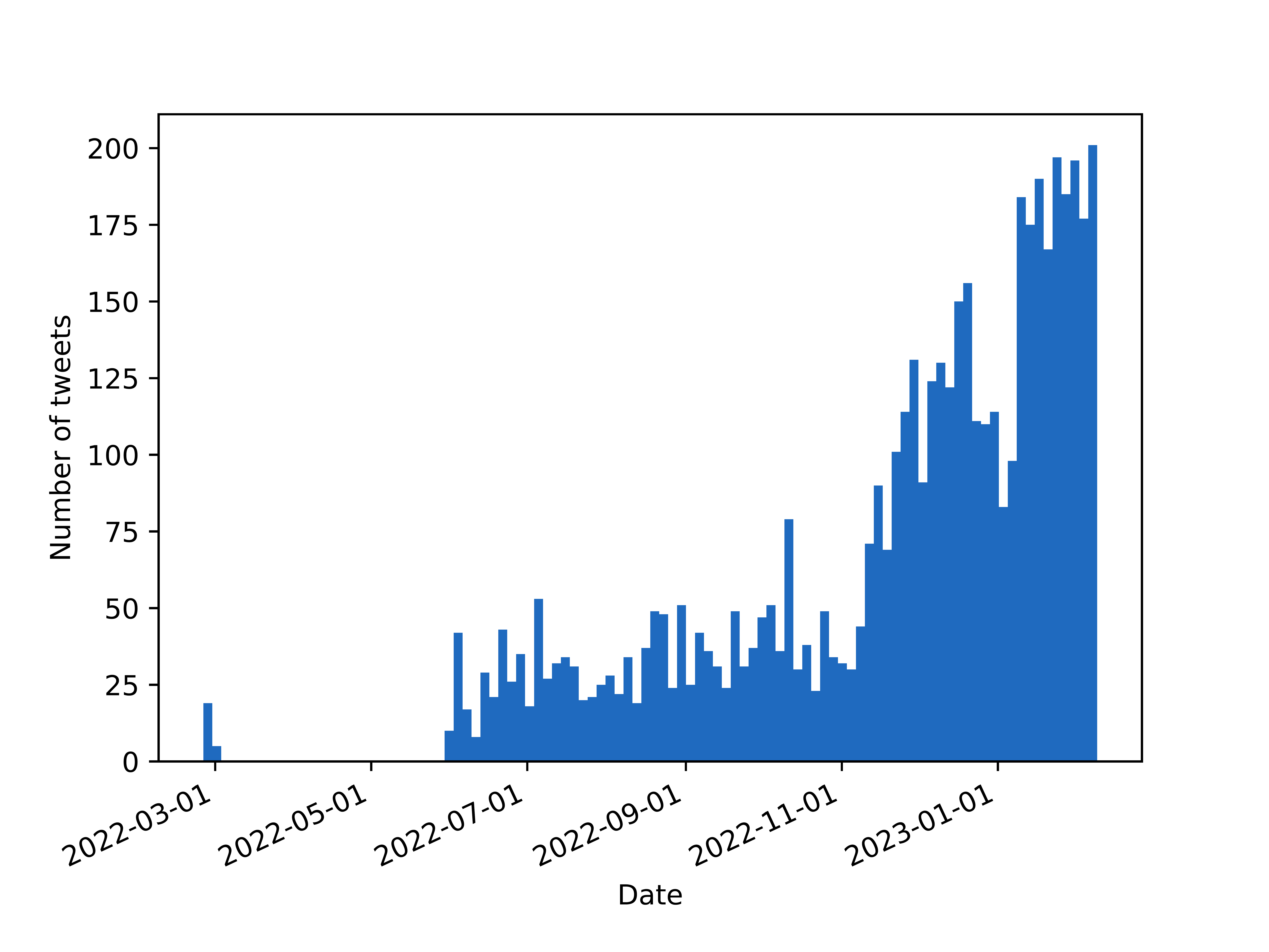


Figure 1: Distribution of the dataset over time.

|  |  |  |
| --- | --- | --- |
| Account handle | Description of account | Number of tweets in dataset |
| Mfa\_russia | Ministry of Foreign Affairs of Russia | 416 |
| Mission\_rf | Permanent Mission of Russia to the International Organizations in Vienna | 191 |
| Armscontrol\_rus | Russian Arms Control Delegation in Vienna | 210 |
| RusMission\_EU | Permanent Mission of Russia to the EU | 55 |
| russiaun | Russian Mission to the UN in NY | 243 |
| rusembusa | Russian Embassy in the USA | 341 |
| Mission\_russian | Russian Mission to the UN and other International Organizations in Geneva | 179 |
| russianembassy | Russian Embassy in London | 455 |
| Natomission\_ru | Russian Mission to NATO | 10 |
| Coe\_russia | Russian Embassy to France | 54 |
| Amb\_ulyanov | Mikhail Ulyanov, representative of Russia to the International Organizations in Vienna | 25 |
| Dpol\_un | Dmitry Polyanskiy, representative of Russia to the UN | 9 |
| MedvedevRussiaE | Dmitry Medvedev, deputy chair of the Security Council of Russia | 5 |
| GGatilov | Gennady Gatilov, representative of Russia to the UN in Geneva | 18 |
| tassagency\_en | Russia-affiliated media agency | 838 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Mfa\_ukraine | Ministry of Foreign Affairs of Ukraine | 199 |
| UKRinUN | Permanent Mission of Ukraine to the UN | 94 |
| Ukraine | Official account of Ukraine | 24 |
| DefenceU | Ministry of Defence of Ukraine | 340 |
| ukrintheusa | Ukrainian Embassy in the USA | 101 |
| dmytrokuleba | Dmytro Kuleba, Minister of Foreign Affairs of Ukraine | 155 |
| oleksiireznikov | Oleksii Reznikov, Minister of Defence of Ukraine | 241 |
| sergiykyslytsya | Sergiy Kyslystya, representative of Ukraine to the UN | 239 |
| denys\_shmyhal | Denys Shmyhal, Prime Minister of Ukraine | 219 |
| olegnikolenko\_ | Oleg Nikolenko, spokesperson for the Ukrainian MFA | 13 |
| eminedzheppar | Emine Dzheppar, First Deputy Minister of Foreign Affairs of Ukraine | 252 |
| zelenskyyua | Volodymyr Zelenskyy, president of Ukraine | 16 |
| United24media | Ukraine-affiliated media agency | 191 |

#### Propaganda techniques [might move above other methods]

As propaganda does not have a consensus on its meaning, there are differing opinions on the number of techniques propagandists employ. The Institute for Propaganda Analysis (IPA), in a book published around the start of World War 2, proposed seven techniques (Lee & Lee, 1939); some other publications discern twenty argumentation fallacies (Weston, 2018) or use eighteen (Martino et al., 2019), seven (Abd Kadir & Abu Hasan, 2014) or twenty-two (Dimitrov et al., 2021) propaganda techniques. Moreover, Dimitrov et al. split these techniques: the first 20 were used both for image and text labeling, while the last 2 were only used for annotating images. However, a lot of these works build on each other, and thus the techniques in them have a lot of overlap.

In this paper, I adapt the 20 techniques used for multimodal labeling in Dimitrov et al. (2021), as well as one more technique from the Institute for Propaganda Analysis. Furthermore, I add three more labels: *text*, *portrait*, and *irrelevant*, for a total of 24 labels. Below I provide the definitions of these labels, which are used as a guideline for annotation.

[review the techniques for rewording, ask Peer]

1. **Irrelevant**: a data point which is not related to the Russo-Ukrainian War. This could be, for example, a foreign minister writing about their trip to Mali.
2. **Text**: a data point where the image consists primarily of text. This is a popular way of communication on social media, especially Twitter. Note that if the text in the image and the text of the tweet are similar, it may be treated as a **Repetition**.
3. **Portrait**: a data point where the image consists of a picture of a person. On Twitter, this is usually done when the person in the picture is cited in the text of the tweet. This can be closely related to **Appeal to authority**.
4. **Presenting irrelevant data (Red Herring):** Introducing irrelevant material to the issue being discussed, so that everyone’s attention is diverted away from the points made.
5. **Misrepresentation of someone’s position (Straw Man):** When an opponent’s proposition is substituted with a similar one, which is then refuted in place of the original proposition.
6. **Whataboutism:** A technique that attempts to discredit an opponent’s position by charging them with hypocrisy without directly disproving their argument.
7. **Causal oversimplification:** Assuming a single cause or reason when there are actually multiple causes for an issue. It includes transferring blame to one person or group of people without investigating the complexities of the issue.
8. **Obfuscation, Intentional vagueness, Confusion:** Using words which are deliberately not clear so that the audience may have their own interpretations. For example, when an unclear phrase with multiple definitions is used within the argument and, therefore, it does not support the conclusion.
9. **Appeal to authority:** Stating that a claim is true simply because a valid authority or expert on the issue said it was true, without any other supporting evidence offered. I consider the special case in which the reference is not an authority or an expert in this technique.
10. **Black-and-white Fallacy:** Presenting two alternative options as the only possibilities, when in fact more possibilities exist. We include dictatorship, which happens when we leave only one possible option, i.e., when we tell the audience exactly what actions to take, eliminating any other possible choices.
11. **Name Calling or Labeling:** Labeling the object of the propaganda campaign as either something the target audience fears, hates, finds undesirable or loves, praises.
12. **Loaded Language:** Using specific words and phrases with strong emotional implications (either positive or negative) to influence an audience.
13. **Exaggeration or Minimisation:** Either representing something in an excessive manner: making things larger, better, worse (e.g., the best of the best, quality guaranteed) or making something seem less important or smaller than it really is (e.g., saying that an insult was just a joke).
14. **Flag-waving:** Playing on strong national feeling (or to any group, e.g., race, gender, political preference) to justify or promote an action or idea.
15. **Doubt:** Questioning the credibility of someone or something.
16. **Appeal to fear/prejudice:** Seeking to build support for an idea by instilling anxiety and/or panic in the population towards an alternative. In some cases the support is built based on preconceived judgments.
17. **Slogans:** A brief and striking phrase that may include labeling and stereotyping. Slogans tend to act as emotional appeals.
18. **Thought-terminating cliche:** Words or phrases that discourage critical thought and meaningful discussion about a given topic. They are typically short, generic sentences that offer seemingly simple answers to complex questions or that distract attention away from other lines of thought.
19. **Bandwagon:** Attempting to persuade the target audience to join in and take the course of action because “everyone else is taking the same action”.
20. **Reductio ad hitlerum:** Persuading an audience to disapprove an action or idea by suggesting that the idea is popular with groups hated in contempt by the target audience. It can refer to any person or concept with a negative connotation.
21. **Repetition:** Repeating the same message over and over again so that the audience will eventually accept it.
22. **Smears:** A smear is an effort to damage or call into question someone’s reputation, by propounding negative propaganda. It can be applied to individuals or groups.
23. **Glittering generalities:** These are words or symbols in the value system of the target audience that produce a positive image when attached to a person or issue. Peace, hope, happiness, security, wise leadership, freedom, “The Truth”, etc. are virtue words. Virtue can be also expressed in images, where a person or an object is depicted positively.
24. **Plain folks:** Plain folks can be seen as the method by which a propagandist encourage their audience that his or her ideas are good because they are the people of the plain folks. Other than that, the propagandist will often attempt to use the specific phrase or expression or even jokes with a specific intonation of a specific audience.

# References

Abd Kadir, S., & Abu Hasan, A. S. (2014). A content analysis of propaganda in Harakah newspaper. Journal of Media and Information Warfare (JMIW), 5, 73–116.  
, Bondarenko, I. (2020). Tools of Explicit Propaganda: Cognitive Underpinnings. Open Journal of Modern Linguistics, 10(01), 23–48. https://doi.org/10.4236/ojml.2020.101003  
, Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and Them: Intergroup Failures of Empathy. Current Directions in Psychological Science, 20(3), 149–153. https://doi.org/10.1177/0963721411408713  
, Dimitrov, D., Ali, B. B., Shaar, S., Alam, F., Silvestri, F., Firooz, H., Nakov, P., & Martino, G. D. S. (2021). Detecting Propaganda Techniques in Memes (arXiv:2109.08013). arXiv. http://arxiv.org/abs/2109.08013  
, Dowe, T. (1997). News You Can Abuse. Wired. https://www.wired.com/1997/01/netizen-6/  
, El-Khalili, S. (2013). Social media as a government propaganda tool in post-revolutionary Egypt. First Monday.  
, Forgas, J. P., & Koch, A. S. (2013). Mood effects on cognition. In Handbook of cognition and emotion (pp. 231–251). The Guilford Press.  
, Gerber, T. P., & Zavisca, J. (2016). Does Russian Propaganda Work? The Washington Quarterly, 39(2), 79–98. https://doi.org/10.1080/0163660X.2016.1204398  
, Helmus, T. C., Bodine-Baron, E., Radin, A., Magnuson, M., Mendelsohn, J., Marcellino, W., Bega, A., & Winkelman, Z. (2018). Russian social media influence: Understanding Russian propaganda in Eastern Europe. Rand Corporation.  
, Jowett, G. S., & O’Donnell, V. (2011). Propaganda & Persuasion. SAGE Publications.  
, Laskin, A. V. (2019). Defining propaganda: A psychoanalytic perspective. Communication and the Public, 4(4), 305–314.  
, Lee, A., & Lee, E. B. (1939). The fine art of propaganda.  
, Martino, G. D. S., Yu, S., Barrón-Cedeño, A., Petrov, R., & Nakov, P. (2019). Fine-Grained Analysis of Propaganda in News Articles (arXiv:1910.02517). arXiv. http://arxiv.org/abs/1910.02517  
, McCrann, G.-E. (2009). Government wartime propaganda posters: Communicators of public policy. Behavioral & Social Sciences Librarian, 28(1–2), 53–73.  
, Naguib, R. (2011, December 30). A year in review: The SCAF rules in 93 letters. Egypt Independent. https://www.egyptindependent.com/year-review-scaf-rules-93-letters/  
, Summers, E., Brigadir, I., Hames, S., Kemenade, H. van, Binkley, P., tinafigueroa, Ruest, N., Walmir, Chudnov, D., Thiel, D., Betsy, Chartier, R., celeste, Lin, H., Alice, Chosak, A., Lenz, M., McCain, R. M., Milligan, I., … Elwert, F. (2023). DocNow/twarc: v2.14.0. Zenodo. https://doi.org/10.5281/zenodo.7799050  
, Tomasello, M. (1995). Joint attention as social cognition. Joint Attention: Its Origins and Role in Development, 103130, 103–130.  
, Van Vugt, M., & Hart, C. M. (2004). Social Identity as Social Glue: The Origins of Group Loyalty. Journal of Personality and Social Psychology, 86(4), 585–598. https://doi.org/10.1037/0022-3514.86.4.585  
, Walker, M., & Matsa, K. E. (2021, September 20). News Consumption Across Social Media in 2021. Pew Research Center’s Journalism Project. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/  
, Weston, A. (2018). A rulebook for arguments. Hackett Publishing.