BelElect: A New Dataset for Bias Research from a ”Dark” Platform

Sviatlana Höhn,[1*] **Sjouke Mauw,** [1] **Nicholas Asher** [2]
[1] University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2] IRIT Computer Science Research Institute of Toulouse, France
sviatlana.hoehn@uni.lu, sjouke.mauw@uni.lu, Nicholas.Asher@irit.fr

Abstract

New social networks and platforms such as Telegram, Gab and Parler offer a stage for extremist, racist and aggressive content, but also provide a safe space for freedom fighters in authoritarian regimes. Data from such platforms offer excellent opportunities for research on issues such as linguistic bias and toxic language detection. However, only a few corpora from such platforms exist, and only in English. This article presents a new Telegram corpus in Russian and Belarussian languages tailored for research on linguistic bias in political news. In addition, we created a repository to make all currently available corpora from so-called ”dark” platforms accessible in one place.

Introduction

Detection of linguistic bias in documents usually starts with a dataset. Creating and sharing such datasets saves researchers' resources and enables study reproducibility. Nevertheless, linguistic bias is usually not annotated. One of the reasons for that is how linguistic bias is defined. Most NLP-based works on bias detection (see (Blodgett et al. 2020) for a critical review) operate with the definition of linguistic bias taken from the Oxford Research Dictionary saying that linguistic bias is a *“systematic asymmetry in word choice that reflects the social-category cognitions that are applied to the described group or individual(s)”* (Beukeboom and Burgers 2017). Thus, bias is seen as a likelihood that some linguistic categories are associated with other terms or attributes (e.g. stereotypes *male – smart*, *female – pretty*). This definition makes reasonable using word embeddings that offer convenient calculation of term proximity based on standard similarity metrics (Ferrer et al. 2021). However, as (Blodgett et al. 2020) and (Höhn, Asher, and Mauw 2021) argue, bias research currently lacks understanding of how bias harms. This is why we suggest to add to the frequent association analysis also the analysis of *word choices that justify actions towards described social categories*.

While bias detection literature largely focuses on public data, such as Wikipedia (Hube 2017), online newspapers (Kiesel et al. 2019) and “traditional” social media such as Twitter (Chun et al. 2019; Guimarães, Figueira, and Torgo 2021), Facebook (Choi et al. 2017; Moore 2021) and YouTube (Jiang, Robertson, and Wilson 2019), instant messengers such as Telegram, and new social networks, such as TikTok only recently became the subject of scholarly investigation, see for instance (Salikov 2019; Kermani 2020; Urman and Katz 2020; Walther and McCoy 2021) for Telegram and (Medina Serrano, Papakyriakopoulos, and Hegelich 2020) for TikTok. Only rarely do researchers publish the datasets which they use in their studies. This is usually explained by the public availability of the data (anyone can go and download them). However, the media themselves are not static; channels and users can choose to delete posts from public channels and groups, and even channels can disappear forever, which is problematic for the academic requirements of research reproducibility also called *replication crisis* (Cockburn et al. 2020). In addition, messengers such as Telegram are often used by actors who prefer to be “in the shadow” - such as extreme right political movements and COVID-19 deniers (Jarynowski et al. 2021; Guhl and Davey 2020) but also oppressed political organisations in non-pluralistic systems and democracy defenders (Ameli and Molaei 2020; Kermani 2020; Salikov 2019). Zeng and Schäfer (2021) describe such platforms as “dark” - they are less regulated and promote very much extreme, and even illegal content, but can also provide a safe area to fight for human rights in authoritarian regimes.

By compiling and sharing corpora from “dark” platforms researchers enable new scientific discussion and new insights about the specific discourse on these platforms and facilitate research reproducibility. This in turn may facilitate discussions about content regulation on “dark” platforms. This is why we share a new, public Telegram corpus of Telegram posts from seven public channels in Russian and Belarussian languages. Further, we summarise currently available corpora from ”dark” platforms in a GitHub repository.

Related Work

In this section we inspect existing Telegram studies and corpora. Such corpora are rare. For instance the repository of hate speech corpora maintained by Vidgen and Derczynski (2020) contains 63 corpora in 16 languages, and none of them is built from Telegram. Nevertheless, an impressing number of scholarly publications is dedicated to Telegram data analysis. Some of the publications mention how

*Funded by FNR Luxembourg INTER-SLANT 13320890.

researchers collected their Telegram data, however, only few authors share the datasets. Although the data from public Telegram channels are available for everyone, we observed during our study of political bias that some posts were deleted and some channels were sanctioned by the government and removed from the messenger. As a consequence, hyperlinks to examples in our corpus stopped working. This is why, a static copy of the data ensures research reproducibility and has a value on its own.

The next section shows the variety of research topics related to misinformation, bias and similar based on Telegram data. Further, we give an overview of the Telegram corpora currently available for the research community (mostly, no annotation is done before publication of those corpora). We describe available, also mostly unannotated, corpora from other "dark" platforms. Our Github repository[1] contains hyperlinks to all these resources hosted in different repositories, to make them findable all on one place.

Analysis of Telegram Data

Because Telegram content moderation policies only prohibit distribution of violence and illegal pornographic content, the platform is heavily used by all sorts of extremist groups, such as jihadist, right-extremist and white-supremacist (Guhl and Davey 2020). However, only public channels and chats are regulated by this policy, private chats are by design not accessible for surveillance, and Telegram sees its key value in this feature (see Durov's Telegram channel https://t.me/durov/176). The platform became a "safe space to hate" (Guhl and Davey 2020) as the analysis of 208 channels shows: 87,9% of the channels contained anti-minority language, and 60,1 % supported terrorist organisations and explicitly called for violence. A recent analysis of 230 Telegram channels posting about COVID-19 topics disclosed that 33 of them contain more than 250 appeals to kill politicians, public persons and medical doctors (Wiebe 2022).

Researchers observe that the content on Telegram channels is very much influenced by political events, such as elections, political conflicts and policy discussions (Kermani 2020). Khaund, Shaik, and Agarwal (2020) argue that Telegram data provide insights into political discussions (in terms of bias and public opinion) that other platforms do not offer. The Telegram corpus presented in this work has been used in (Höhn, Asher, and Mauw 2021) to validate a game theoretic bias model proposed in (Asher, Hunter, and Paul 2021).

In their analysis of the 2017 Iranian presidential election discourse on Telegram, Kermani (2020) show that alternative opinions are almost never discussed, and most popular viewpoints are reinforced and reproduced. Osadchuk (2019) describes how fake and biased messages played a role in the process of a Russian-Ukrainian prisoners exchange in August-September 2019.

In their analysis of US extremism on Telegram, (Walther and McCoy 2021) show that more Telegram channels publish far-right content than far-left; and multiple channels are connected to well-known extremists and hate groups. Urman and Katz (2020) demonstrate that far-right networks are highly decentralised and very heterogeneous. The authors argue that the growth of the far-right networks on Telegram is connected to the ban of these movements on Facebook and Instagram in 2019. Urman and Katz (2020) also show the relationships to other far-right networks on other platforms, such as 4chan.

Thus, the intention of the Telegram creator to enable privacy and free speech on the Internet has been successfully used to promote democracy, but also misused by extremist organisations and criminals who cannot publish their content on other, better regulated media channels. In addition to that, Salikov (2019) observes that being the communication channel of the opposition at the beginning, and despite being declared as illegal in Russia, the Telegram platform is heavily used by the Russian authorities in order to monitor public mood and influence (or manipulate) the population.

Telegram Corpora

We found only two publicly released Telegram corpora. Both are mainly in English and cover topics relevant for the Anglo-Western society, in contrast to studies discussed in the preceding section that cover multiple political topics from the Post-Soviet states and Middle-Eastern political events.

The Pushshift Telegram dataset (Baumgartner et al. 2020) is compiled from (mostly English) 27.800 channels and 2, 2 M unique users. The initial seed for the dataset creation was compiled manually from 124 channels known as right-wing extremist and 137 channels known as cryptocurrency-related. Other channels were added following a "snow-ball" principle (see. (Baumgartner et al. 2020) for details). This corpus does not contain any bias annotation.

Solopova, Scheffler, and Popa-Wyatt (2021) compiled a corpus from all messages from a Telegram channel supporting Donald Trump from end of 2016 till January 2021. The corpus was annotated automatically and manually. Automatic annotations cover offensive language, and manual annotations mark up harmful language and cover only a part of the corpus related to the Capitol riot. Manual labeling contained five classes: incitement; pejoratives; insulting, offensive and abusive uses; divisive speech; and codes.

Corpora from Other "Dark" Platforms

Although Telegram is one of the most popular "dark" platforms, the less prominent members of this class also need researchers' attention. Only few datasets in relation to these alternative platforms have been published, and the number of studies about bias-related topics on these data is very small. However, their outcomes show that researchers and policy-makers need to have a closer look at the activities in the "dark" zone (Bagavathi et al. 2019; Mathew et al. 2020).

Fair and Wesslen (2019) created and published a dataset from **Gab** that contains over 37 M posts, more than 24, 5 M comments, and over 800.000 user profiles. The dataset was collected between August 2016 and December 2018.

4chan is another influential network that did not receive much attention from researchers. Created in 2003, the

[1] https://github.com/sviatlanahoehn/darkplatforms/

anonymous imageboard attracted more than 27 M users per month. [2], but was not able to monetize the traffic. Nevertheless, it became the place for hate and harassment (Hine et al. 2017). Specifically, the board for political discussion /pol/ dedicated to politically incorrect content (antisemitic, racist, white-supremacist and similar) is known to distribute hateful and toxic content (Hine et al. 2017). To support the academic research on 4chan data, (Papasavva et al. 2020) produced and published a 4chan dataset of $134,5$ M posts from over $3,3$ M /pol/ conversation threads from a period between June 2016 and November 2019.

Zignani et al. (2019) compiled and released a **Mastodon** dataset. This distributed micro-blogging service offers an option for its users to mark their posts as inappropriate or sensitive. Posts on Mastodon are called *toots*. The dataset contains information about 363 instances and nearly 8.900 toots.

Parler is an emerging social network recently favoured by extreme right users. Along with the statistical analysis of users, flags, badges and followers, (Aliapoulios et al. 2021) also released a dataset from Parler. It includes 183 M posts authored by 4 M users from August 2018 to January 2021.

All these datasets contain biased, racist, extremist, toxic and hateful posts, and are useful resources for research on these topics. Nevertheless, datasets from these platforms in languages other than English would provide further insights in the nature of biases, because values of the recipient play a role in how biased messages are framed and interpreted (Höhn, Asher, and Mauw 2021). This is why our dataset brings an added value in the variety of already published data. Specifically, the corpus is created for research on linguistic bias, but can be re-purposed for other directions.

Other Russian and Belarussian Corpora

Current academic publications mention large Russian and Belarussian corpora for more traditional language research (texts from literature and newspapers, tree banks, parallel corpora) (Zakharov 2013; Sitchinava et al. 2012). Social media corpora are analysed for Russian language, however, Belarussian is underrepresented (Miller 2019; Bodrunova, Blekanov, and Kukarkin 2019). Our work contributes to the field of Russian and Belarussian corpus research by adding a conceptually new language resource from an under-researched but heavily used communication channel.

BelElect Dataset

The creation of the dataset was motivated by the need for examples from social media posts that describe the same events from different angles, framing those events differently (qualitatively similar to examples in (Asher, Hunter, and Paul 2021) but more in quantity). As we started this research, presidential elections in Belarus took place, and the reports on social media about the events after the elections offered an interesting case. The protests after the elections were motivated by the accusation of falsification of the election results. These events split the country in state and opposition supporters. We chose Telegram as a platform for analysis because this messenger became the main communication channel in countries such as Belarus where free speech can be sanctioned.

[2] *https://omr.com/de/4chan-pleite/*

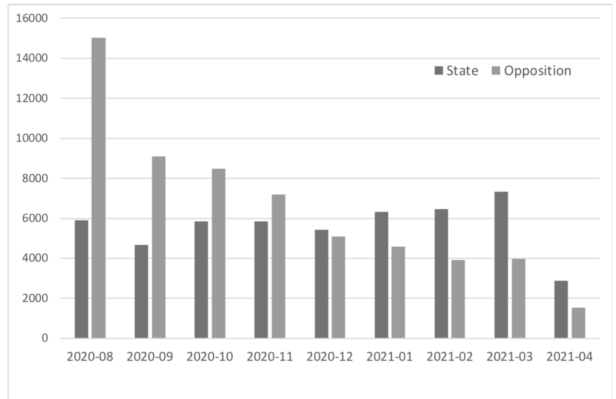


Figure 1: Downloaded posts by month

Composition

The dataset contains Telegram posts from 1.08.2020 to 14.04.2021 covering the time of the presidential elections in Belarus and the protests in the country after the elections.

Our aim was to have a balanced dataset this is why we included four state-supporting channels ONT NEWS, BelTA, Zheltye Slivy and Pool 1 (the president's channel), and four popular opposition channels BelSAT, Belarus Seychas (En. *Belarus Today*), Belarus Golovnogo Mozga (En. *Belarus of the Brain*) and TUT.BY with posts in Russian and Belarussian languages. These channels also contain reposts from other channels such as NEXTA. The entire dataset was downloaded on April 15, 2021 using the built-in function of the desktop version of Telegram. It contains 140.388 Telegram posts (76.918 opposition and 63.470 state); 109.721 posts contain non-empty text (58.976 opposition and 50.745 state). We limited the size of images and media files to 8 GB. The media files exceeding this limit were not included into download. They can be added using the message ID and channel ID. Figure 1 shows the number of downloaded posts per month grouping all pro-state (black) and all pro-opposition (gray) channels.

Although the dataset is quite balanced between opposition and state channels, we found that biases occur at the selectivity (a decision of a channel to report or not to report an issue) and coverage (how much space is allocated to an event or fact) level. The opposition reports mainly about protests (selectivity bias), other topics are scarcely covered (coverage bias). In contrast, state media start reporting about protests ca. 12 hours after they began, and not immediately (selectivity bias), and they mainly report about many other facts and events in the country and the world making the protests in the country to a less important issue (coverage bias).

Format

The `.tgz` archive contains 8 sub-archives, each of them contains two parts: an automatically generated `result.json` file containing the annotations and media directories with all media files, each media type in a separate directory. The `JSON`-files start with channel metadata:

```
"name": "channel_name",
"type": "public_channel",
"id": channel_id_number,
"messages": [list of messages]
```

Every message in the message list is structured as follows:

```
"id": message_id_number,
"type": "message",
"date": message_date,
"edited": date_edited,
"from": channel_name,
"from_id": channel_id,
"text": message_text
```

Some messages can also contain photos, videos, stickers, voice messages, video messages, GIFs and/or files. Each of them would be annotated only if it is part of the message. If a media file was part of a message but was not downloaded, it will be marked in the following way:

```
...
"file": "(File not included.
        Change data exporting
        settings to download.)",
"thumbnail": "(File not included.
        Change data exporting
        settings to download.)",
        "media_type": "video_file",
        "mime_type": "video/mp4",
"duration_seconds": seconds_number,
"width": 640,
"height": 480,
...
```

FAIR Principles

The publication of the BelElect dataset meets the FAIR principles[3] in the following way:

1. The dataset is *findable*: it can be downloaded from currently three open-access, indexed and searchable repositories: Zenodo (DOI: *10.5281/zenodo.5844350*), GitHub (*https://github.com/sviatlanahoehn/darkplatforms/*) and Orbilu (*http://hdl.handle.net/10993/49661*)
2. It is *accessible* by a simple download from currently three open-access repositories. The download does not require any special procedure.
3. The data *interoperability* is guaranteed via JSON format that is widely used in many programming languages and libraries.
4. The main purpose of the publication of the BelElect dataset is to make the data *reusable* and to support research reproducibility. This corpus is a snapshot of the data on the chosen eight Telegram channels at a particular date. After that, some of the opposition channels have been declared extremist by the Belarussian government and have been deleted completely. Thus, without this snapshot, any further research on these data could not be possible.

[3]https://www.go-fair.org/fair-principles/

Further, automated translation can be employed by researchers who do not speak Russian and Belarussian to analyse the content of the messages.

Use of the BelElect Dataset for Research

This dataset has been successfully used to validate a game theoretic approach to bias described in (Höhn, Asher, and Mauw 2021). However, there is more to discover due to the differently biased formulations of descriptions of the same events. For example, we found new types of bias in the dataset. We illustrate it on a new bias type that we call *timing bias*. In addition we found that messages framed with one type of bias may have a differently biased function. The examples in this paper were translated into English, original messages are written in Russian or Belarussian languages, or a mixture of both.

Timing Bias One of the events leading to a peak in the number of posts on all channels was the inauguration of the president on September 23. Example 0.1 is the first mention of the inauguration of Lukashenko on the state news agency BelTA. The picture in the message is quite generic.

Example 0.1 *BelTA 23.09.2020 10:31*

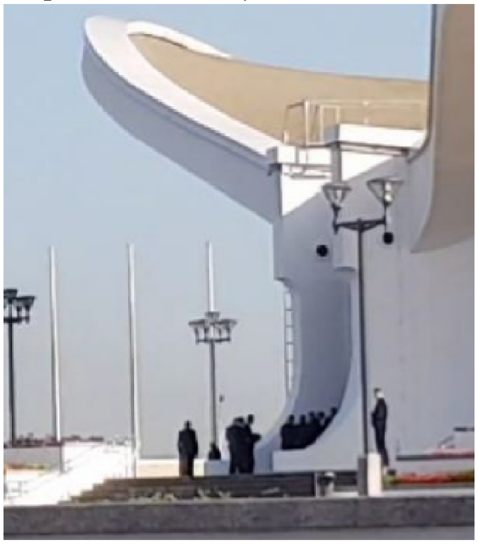
Lukashenko took office as President of Belarus

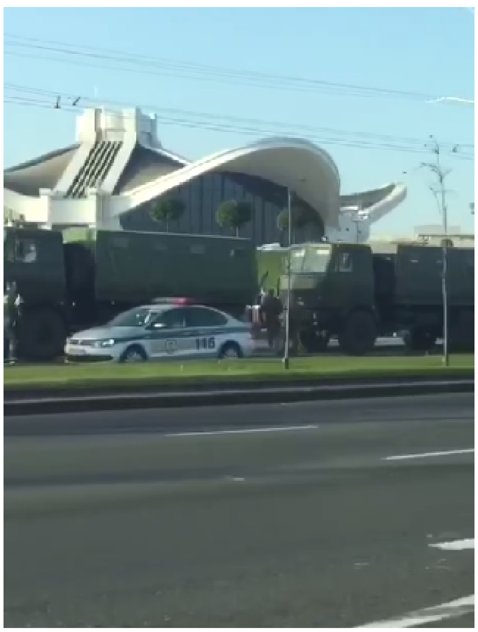
The inauguration ceremony is taking place these minutes at the Palace of Independence.

Putting his right hand on the Constitution, Alexander Lukashenko recited the oath in the Belarusian language. Afterwards he signed the oath, and afterwards Chairman of the Central Commission for Elections and National Referenda Lidia Yermoshina handed in the presidential credentials to Alexander Lukashenko.

Opposition channels also report about this event, for instance Example 0.2, however blaming the state of making a secret of this event. Apparently, nothing was known about the inauguration one hour before the ceremony started which is very unusual for this type of ceremonies. The next example shows excerpts of the original photo (first picture) and screenshot of the original video (second picture).

Example 0.2 *Belarus Seychas 23.09.2020 09:18*





Lukashenko's press secretary did not announce the date of the inauguration. She said that the date will be known "closer to the time" of the ceremony.

Today in the morning in Minsk. People in suits were brought there in minibuses, security measures were strengthened, and the roads were blocked.

The reason for the morning gathering is still unclear. It is possible that it has something to do with the upcoming inauguration.

These examples show that media purposefully delay reports about important planned events in order to manipulate the population's behaviour. This type of bias has never been annotated, neither it is detectable based on models that rely on lexicon, whether word lists or word embeddings. The timing bias, however, becomes analysable by a comparative analysis of posts from channels representing different parties, and reporting about the same events from different angles.

Is the *timing bias* different from the *selectivity bias*? In our Examples 0.1 and 0.2, the subject of the reports is the planned inauguration of the president. Is is to expect that this event will be celebrated by the state and the citizens, and it is quite expectable that in a democracy, an elected president is supported by the majority of the population, and the inauguration would be welcomed by the majority. It is expectable that such events are announced some time before they are going to take place. However, Belarussian government chose to officially publish the information about it *at the moment it has started*. The delayed reports about the protests by the state channels differ from this delayed report about the planned event in at least three points: (1) the protests were observable, state channels simply ignored them; (2) the protests were not planned and not under the state control; and (3) it was not clear how long they continue. Therefore, we can see that the delayed reports about the protests and delayed information about the planned inauguration of the president are of a similar but not exactly the same nature. Consequently, if timing bias is a sub-type of the selectivity bias, then there must be other sub-types of the selectivity bias that better describe cases similar to delayed reports about protests. We include this question in our future research on bias typology.

Why is timing bias relevant for linguistic bias? Our working definition of bias includes the notion of justification of actions towards social categories. By comparing Examples 0.2 and 0.1 we can see that Ex. 0.2 argues to mistrust the government while Ex. 0.1 projects the feeling of security and control over the situation. Such dissonances are only detectable by comparing the linguistic formulations of the messages. However, more research is needed to define the (linguistic) timing bias formally.

Bias Form vs. Function Other types of bias (such as gender or racial bias) can be used in order to discredit a political actor. Example 0.3 shows how a gender-biased statement is used to create a negatively loaded image of a presidential candidate. The term *meatball fairy* refers to the profession of Sviatlana Tikhanovskaya who was a cook before she started her political career.

Example 0.3 *Zheltye Slivy, 18.01.2021, (excerpt)*

The meatball fairy is a peculiar lady. The worse her home country is, the happier she gets. She has a strange hormonal imbalance: as soon as she gets into trouble, she has an overabundance of endorphins.

Thus, researchers working on bias need to differentiate between *bias form* and *bias function*. The formulation in Example 0.3 has a form of a gender-biased statement while its function is politically biased.

Limitations of the Work

This research presents our first result in creation of a Telegram corpus that contains a balanced distribution of descriptions of political events from opposite perspectives. While this corpus provides a solid ground for linguistic bias research, we need to keep in mind that every corpus is only a static snapshot of language. Linguistic studies need to be placed in relationships with the socio-cultural context in which that language has been produced, in order to make valid conclusions about bias and harm.

Similar corpora from other languages and other political events of a larger (global or local) impact are needed to make valid generalisations about the nature and the mechanisms of linguistic bias and its effect on social categories. More specifically, based on this corpus, it is only possible to draw connections between linguistic formulations in media and actual actions towards particular parts of a population for these specific languages (Russian and Belarussian) and this specific region (Belarus) at this specific time (2020-2021). Generalisation from this corpus about other regions would be invalid, however, insights about biases form this corpus can be transferred and investigated in other languages.

Conclusions and Future Work

We described and shared a new dataset that we compiled to study linguistic bias in political news. The dataset contains contrasting formulations of descriptions of the same events after Belarussian presidential elections 2020.

The positive intention to provide a safe space for free speech on Telegram and similar platforms has an undesired side effect: these platforms promote hate and harassment. While more established social media such as Facebook and Twitter react to political pressure and make at least attempts to monitor their content and sanction publishers of toxic content, platforms such as Telegram insist on free speech rights and support them by technical solutions such as end-to-end encryption. Nevertheless, appeals to kill people have nothing to do with free speech, they are illegal and even criminal.

By publication of corpora from so-called “dark” platforms we hope to contribute to the discussion on the regulation of messengers as media, and especially to support researchers working on linguistic bias in political news. The BelElect dataset can be also used for research in social sciences, socio-linguistics, and natural language processing.

Ethical Statement

The data have been downloaded only from public channels acting as or even maintained by news agencies. We did not collect any personal messages or personal data from individual users. However, some messages contain media materials that contain images and video recordings of children which need special protection. In addition, the media files contain recordings of physical violence against people, scenes of injuries and suffering people, verbal harassment against people, and in some cases also appeals to kill police officers and politicians. Like other datasets from “dark” platforms, these data need to be handled with caution.

Acknowledgments

We thank the ANR PRCI grant SLANT, the Luxembourgish National Research Fund, INTER-SLANT 13320890 and the 3IA Institute ANITI funded by the ANR-19-PI3A-0004 grant for research support.

References

Aliapoulios, M.; Bevensee, E.; Blackburn, J.; Bradlyn, B.; De Cristofaro, E.; Stringhini, G.; and Zannettou, S. 2021. An early look at the Parler online social network. *arXiv preprint arXiv:2101.03820* .

Ameli, S. R.; and Molaei, H. 2020. Election Journalism: Investigating Media Bias on Telegram during the 2017 Presidential Election in Iran. *Digital Journalism* 8(8): 975–991.

Asher, N.; Hunter, J.; and Paul, S. 2021. Bias in Semantic and Discourse Interpretation. *Linguistics and Philosophy* .

Bagavathi, A.; Bashiri, P.; Reid, S.; Phillips, M.; and Krishnan, S. 2019. Examining untempered social media: analyzing cascades of polarized conversations. In *Proc. 2019 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM'19)*, 625–632.

Baumgartner, J.; Zannettou, S.; Squire, M.; and Blackburn, J. 2020. The Pushshift Telegram Dataset. In *Proc. 14th Int. AAAI Conf. on Web and Social Media (ICWSM'20)*, 840–847.

Beukeboom, C. J.; and Burgers, C. 2017. Linguistic Bias. In *Oxford Research Encyclopedia, Communication*. Oxford University Press.

Blodgett, S. L.; Barocas, S.; Daumé III, H.; and Wallach, H. 2020. Language (Technology) is Power: A Critical Survey of ”Bias” in NLP. In *Proc. 58th ACL Meeting*, 5454–5476. ACL.

Bodrunova, S. S.; Blekanov, I. S.; and Kukarkin, M. 2019. Topics in the Russian Twitter and relations between their interpretability and sentiment. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 549–554. IEEE.

Choi, I.; Milne, D. N.; Glozier, N.; Peters, D.; Harvey, S. B.; and Calvo, R. A. 2017. Using different Facebook advertisements to recruit men for an online mental health study: Engagement and selection bias. *Internet Interventions* 8: 27–34.

Chun, S. A.; Holowczak, R. D.; Dharan, K.; Wang, R.; Basu, S.; and Geller, J. 2019. Detecting Political Bias Trolls in Twitter Data. In *Proc. 15th Int. Conf. on Web Information Systems and Technologies (WEBIST'19)*, 334–342.

Cockburn, A.; Dragicevic, P.; Besançon, L.; and Gutwin, C. 2020. Threats of a replication crisis in empirical computer science. *Communications of the ACM* 63(8): 70–79.

Fair, G.; and Wesslen, R. 2019. Shouting into the void: A database of the alternative social media platform Gab. In *Proc. 13th Int. AAAI Conf. on Web and Social Media (ICWSM'19)*, 608–610.

Ferrer, X.; van Nuenen, T.; Such, J. M.; and Criado, N. 2021. Discovering and Categorising Language Biases in Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 140–151.

Guhl, J.; and Davey, J. 2020. A Safe Space to Hate: White Supremacist Mobilisation on Telegram. Institute for Strategic Dialogue.

Guimarães, N.; Figueira, Á.; and Torgo, L. 2021. Profiling Accounts Political Bias on Twitter. In *Proc. 16th Iberian Conf. on Information Systems and Technologies (CISTI'21)*, 1–4. IEEE.

Hine, G.; Onaolapo, J.; De Cristofaro, E.; Kourtellis, N.; Leontiadis, I.; Samaras, R.; Stringhini, G.; and Blackburn, J. 2017. Kek, cucks, and god emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. In *Proc. 11th Int. AAAI Conf. on Web and Social Media (ICWSM'17)*. AAAI.

Hube, C. 2017. Bias in wikipedia. In *Proc. 26th Int. Conf. on World Wide Web Companion (WWW'17)*, 717–721.

Höhn, S.; Asher, N.; and Mauw, S. 2021. Examining Linguistic Biases in Telegram with a Game Theoretic Analysis. In *Proceedings of the 3rd Multidisciplinary International Symposium on Disinformation in Open Online Media (MISDOOM)*, 16–32.

Jarynowski, A.; Semenov, A.; Kaminski, M.; and Belik, V. 2021. Mild Adverse Events of Sputnik V Vaccine Extracted from Russian Language Telegram Posts via BERT Deep Learning Model. *medRxiv* .

Jiang, S.; Robertson, R. E.; and Wilson, C. 2019. Bias misperceived: The role of partisanship and misinformation in Youtube comment moderation. In *Proc. 13th Int. AAAI Conf. on Web and Social Media (ICWSM'19)*, 278–289.

Kermani, H. 2020. Telegram Discourse on The 2017 Iran's Presidential Election. *Asian Journal of Media and Communication (AJMC)* 4(1).

Khaund, T.; Shaik, M.; and Agarwal, N. 2020. Data Collection and Sensemaking from Telegram: A Case Study of Ukrainian Political Leaders Channels and Chat Groups. In *Proc. Int. Conf. on Social Computing, Behavioral-Cultural Modeling & Prediction, and Behavior Representation in Modeling and Simulation (SBP-BRiMS'20)*.

Kiesel, J.; Mestre, M.; Shukla, R.; Vincent, E.; Adineh, P.; Corney, D.; Stein, B.; and Potthast, M. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proc. 13th Int. Workshop on Semantic Evaluation (SemEval'19)*, 829–839.

Mathew, B.; Illendula, A.; Saha, P.; Sarkar, S.; Goyal, P.; and Mukherjee, A. 2020. Hate begets hate: A temporal study of hate speech. *Proc. of the ACM on Human-Computer Interaction* 4(CSCW2): 1–24.

Medina Serrano, J. C.; Papakyriakopoulos, O.; and Hegelich, S. 2020. Dancing to the partisan beat: A first analysis of political communication on TikTok. In *Proc. 12th ACM conference on web science (WebSci'20)*, 257–266.

Miller, D. T. 2019. Topics and emotions in Russian Twitter propaganda. *First Monday* .

Moore, G. A. 2021. *Word usage bias in Facebook communities: A psycholinguistic analysis of political commentary*. Ph.D. thesis, Fielding Graduate University.

Osadchuk, R. 2019. Ukrainian media jump the gun on Russia-Ukraine prisoner swap. https://medium.com/dfrlab/ukrainian-media-jump-the-gun-on-russia-ukraine-prisoner-swap-3216f860bc04. Accessed: 2022-01-11.

Papasavva, A.; Zannettou, S.; De Cristofaro, E.; Stringhini, G.; and Blackburn, J. 2020. Raiders of the lost Kek: 3.5 years of augmented 4chan posts from the politically incorrect board. In *Proc. 14th Int. AAAI Conf. on Web and Social Media (ICWSM'20)*, 885–894.

Salikov, A. 2019. Telegram as a Means of Political Communication and its use by Russia's Ruling Elite. *Politologija* 95(3): 83–110.

Sitchinava, D. V.; et al. 2012. Parallel corpora within the Russian National Corpus. *Prace Filologiczne* (63): 271–278.

Solopova, V.; Scheffler, T.; and Popa-Wyatt, M. 2021. A Telegram corpus for hate speech, offensive language, and online harm. *Journal of Open Humanities Data* 7(9).

Urman, A.; and Katz, S. 2020. What they do in the shadows: examining the far-right networks on Telegram. *Information, communication & society* 1–20.

Vidgen, B.; and Derczynski, L. 2020. Directions in abusive language training data, a systematic review: Garbage in, garbage out. *PloS one* 15(12): e0243300.

Walther, S.; and McCoy, A. 2021. US Extremism on Telegram. *Perspectives on Terrorism* 15(2): 100–124.

Wiebe, J.-H. 2022. Täglich Tötungsaufrufe auf Telegram. https://www.tagesschau.de/investigativ/funk/todesdrohungen-telegram-101.html. Accessed: 2022-01-11.

Zakharov, V. 2013. Corpora of the Russian language. In *International conference on text, speech and dialogue*, 1–13. Springer.

Zeng, J.; and Schäfer, M. S. 2021. Conceptualizing "Dark Platforms". Covid-19-Related Conspiracy Theories on 8kun and Gab. *Digital Journalism* 1–23.

Zignani, M.; Quadri, C.; Galdeman, A.; Gaito, S.; and Rossi, G. P. 2019. Mastodon content warnings: Inappropriate contents in a microblogging platform. In *Proceedings of the International AAAI Conference on Web and Social Media*, 639–645.