EUvsDisinfo: A Dataset for Multilingual Detection of Pro-Kremlin Disinformation in News Articles

João A. Leite
The University of Sheffield
Sheffield, United Kingdom
jaleite1@sheffield.ac.uk

Olesya Razuvayevskaya
The University of Sheffield
Sheffield, United Kingdom
o.razuvayevskaya@sheffield.ac.uk

Kalina Bontcheva
The University of Sheffield
Sheffield, United Kingdom
k.bontcheva@sheffield.ac.uk

Carolina Scarton
The University of Sheffield
Sheffield, United Kingdom
c.scarton@sheffield.ac.uk

Abstract

This work introduces EUvsDisinfo, a multilingual dataset of disinformation articles originating from pro-Kremlin outlets, along with trustworthy articles from credible / less biased sources. It is sourced directly from the debunk articles written by experts leading the EUvsDisinfo project. Our dataset is the largest to-date resource in terms of the overall number of articles and distinct languages. It also provides the largest topical and temporal coverage. Using this dataset, we investigate the dissemination of pro-Kremlin disinformation across different languages, uncovering language-specific patterns targeting certain disinformation topics. We further analyse the evolution of topic distribution over an eight-year period, noting a significant surge in disinformation content before the full-scale invasion of Ukraine in 2022. Lastly, we demonstrate the dataset's applicability in training models to effectively distinguish between disinformation and trustworthy content in multilingual settings.

CCS Concepts

• **Computing methodologies → Natural language processing**;
• **Applied computing → Law, social and behavioral sciences**.

Keywords

Disinformation, Dataset, pro-Kremlin, Classification, News Articles

ACM Reference Format:
João A. Leite, Olesya Razuvayevskaya, Kalina Bontcheva, and Carolina Scarton. 2024. EUvsDisinfo: A Dataset for Multilingual Detection of Pro-Kremlin Disinformation in News Articles. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3627673.3679167

1 Introduction

Information warfare is a special case of disinformation that aims at manipulating public opinion to achieve military or political objectives [11, 22, 24]. Specifically, pro-Kremlin information warfare has been a topic of study across different disciplines [1, 14, 15, 18]. Since the annexation of Crimea, the spread of pro-Kremlin disinformation in Europe has intensified, encompassing many languages and even influencing search engines [7, 8, 12, 28].

Combating pro-Kremlin disinformation articles presents significant challenges due to its dissemination across various languages, during an extensive period of time, and covering a wide range of different topics. Current disinformation datasets are not suitable for training machine learning models aimed to address this task since they lack the combination of such characteristics. Most datasets focus on analysing short claims [23, 25] or social media posts [17, 29], whilst the majority of the existing article-level disinformation corpora consist of monolingual data [4–6, 9, 16, 19, 20, 20, 26]. A significantly smaller amount of multilingual datasets is available [2, 3, 10, 13], and they contain the following limitations: (i) representing parallel translated data [2, 3], (ii) covering short periods of time ($\leq$ 1 year) [2, 10, 13], (iii) containing only few articles (around $1-2K$) [2], and (iv) focusing on narrow topics [2, 10, 13]. To the best of our knowledge, the work by Solopova et al. [21] is the only dataset addressing the theme of pro-Kremlin *information warfare*. Nevertheless, their dataset is targeted towards the detection of a pro-Western vs pro-Kremlin stance rather than disinformation.

In this work, we leverage the journalistic investigation carried out by specialists to generate a large and diverse multilingual dataset containing disinformation articles from pro-Kremlin outlets and trustworthy counterparts from reliable / less biased sources. To produce such a dataset, we use the debunk articles written by EUvsDisinfo[1], who since 2015 has been providing an EU-wide response to information warfare by debunking disinformation narratives across EU-member states. Our dataset is the largest (18,249 articles) to-date, most topically diverse (508 topics manually assigned by EUvsDisinfo), spans over the longest period of time (8.5 years), and is the most diverse in terms of languages (42 languages) compared to other article-level multilingual disinformation detection datasets (see Table 1 for an overview of related datasets in comparison to ours). Our key contributions are: (i) The novel multilingual disinformation dataset[2], becoming the largest and most diverse article-level set in terms of language, topics, and time periods. (ii) The analysis of the prominence of pro-Kremlin disinformation topics across different languages and time periods. (iii) The evaluation of several

[1]https://euvsdisinfo.eu
[2]Dataset: https://doi.org/10.5281/zenodo.10514307

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679167

Table 1: Overview of multilingual article-level content verification datasets.

Dataset	Size	Labelling	# Languages	Multilinguality	Topic(s)	Period
TALLIP-FakeNews [2]	2, 304	Manual	3	Translation-based	COVID-19	4 months
MM-Covid [13]	11, 565	Fact-checkers + Manual	6	Inherent	COVID-19	6 months
FakeCovid [10]	5, 182	Fact-checkers	40	Inherent	COVID-19	4 months
Multiverse [3]	3, 009	Source-based	5	Semi-translation-based	Celebrities, Science, Politics, Culture, World	Unknown
EUvsDisinfo (ours)	18, 249	Fact-checkers	42	Inherent	508 topics originating from Pro-Kremlin disinformation	8.5 years

baseline models using our novel dataset for the task of multilingual binary disinformation detection. Our dataset and code are made publicly available, along with documentation, as supplementary material to facilitate reproducibility[3].

2 Methodology

Our dataset leverages the structured information of the debunking articles published by EUvsDisinfo (see Figure 1 for an example). It consists of two classes of articles: disinformation and trustworthy. Disinformation articles are acquired directly from the links mentioned on the left hand-side of the EUvsDisinfo debunk page. Trustworthy articles are derived from the links to reliable sources within the response section of the article, in which EUvsDisinfo directly debunks the false narrative. Such URLs can lead to news articles, official documents, statements, books, encyclopedia articles, social media posts, and fact-checking articles, among others.

Figure 1: Example of debunk article by EUvsDisinfo.

To verify that EUvsDisinfo does not reference disinformation articles in the response section, we manually inspected a randomly sampled set of 30 debunk articles, and annotated all the URLs mentioned within their response section. In total, 350 URLs were labelled as either *trustworthy* or *potential disinformation*. To do so, we considered the context in which the URL is referenced in the response section. For example, if the URL is referenced after a sentence such as "see similar cases", it is marked as potential disinformation (for reference, see the last paragraph in the response section of Figure 1). We identified that all of the URLs marked as *potential disinformation* actually correspond to other debunk articles by EUvsDisinfo or other fact-checking agencies such as Bellingcat and StopFake, instead of directly links to disinformation.

2.1 Data Collection

Given the wide variety of websites mentioned in EUvsDisinfo debunks, with varying HTML structure, extracting textual content is not trivial. For this task, we employ the Diffbot API[4], which is a proprietary tool that uses machine learning to extract the content of a given web page. Although Diffbot is a closed-source service, it is free of charge for academic purposes. We retrieve articles that are no longer available on the web, by using the digital library Wayback Machine[5]. Also, the language of each trustworthy article is inferred using Polyglot[6], as EUvsDisinfo only specifies the languages of the disinformation articles.

Upon collecting the contents of all the URLs mentioned on EUvsDisinfo debunks, resulting in 35, 839 articles, we apply three filtering strategies to ensure consistency in the dataset. First, we remove 12, 875 URLs whose domain is not of a news outlet type. This is done to ensure consistency in the theme of disinformation detection of news articles published by news outlets. To do so, we first manually inspect the most frequent domains of the URLs, and label them as one of the following: “News Outlet”, “Organisation”, “Social Media”, “Fact-checker”, and “Other”. Next, we remove 4, 048 instances referring to error messages, log-in prompts, and paywalls through lexicon-bases rules, and through removing articles with less than 700 characters, which is significantly shorter than the average article (6, 346 characters). Lastly, we remove 667 URLs cited within sentences containing n-grams referring to other fact-checking articles (e.g. "See earlier disinformation cases"). The full list of n-grams can be found in the supplementary materials[3].

2.2 Dataset Overview

The full dataset has 18, 249 news articles with 10, 682 (59%) and 7, 567 (41%) marked as disinformation and trustworthy, respectively. It contains articles by 2, 946 different publishers on 508 unique topics, spanning across 8.5 years, between 06/01/2015 and 01/08/2023. On average, the articles are 6, 346 characters in length and cover 3.8 different topics. The dataset contains 42 different languages, of which 17 are not present on existing multilingual article-level

[3]Supplementary material: https://github.com/JAugusto97/euvsdisinfo

[4]https://www.diffbot.com
[5]https://web.archive.org
[6]https://github.com/aboSamoor/polyglot

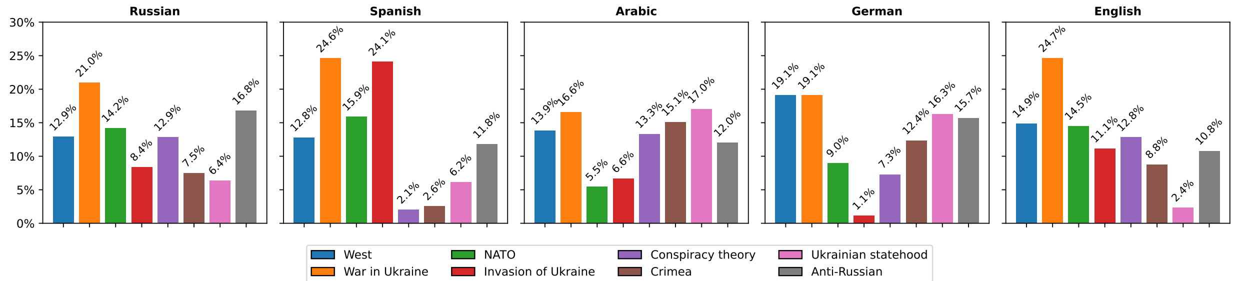

Figure 2: Top 8 most frequent topics across disinformation articles for the top 5 languages.

disinformation datasets presented in Table 1. Please refer to the supplementary material for a detailed breakdown of all languages. Out of all 2, 946 publishers in the dataset, 1, 187 and 1, 759 are associated with disinformation and trustworthy articles, respectively. The top five publishers of disinformation articles are known pro-Kremlin outlets: Sputnik (15.7%), RT (11.3%), RIA Novosti (4.7%), Tsargrad (2.1%), and Ukraina.ru (1.6%). For trustworthy articles, the top five publishers include Reuters (6.7%), BBC (6.5%), The Guardian (5.3%), Deutsche Welle (3.9%), and Radio Free Europe (3.2%).

2.3 Analysis of Topics

Figure 2 shows the distribution of the eight most frequent topics across the whole dataset for the five most common languages. Notably, the topic of "War in Ukraine" is the most prevalent across all five languages. However, certain topics appear more frequently in specific languages. The topic "Invasion of Ukraine" is over twice more common in Spanish than in other languages. "Conspiracy theory" and "Crimea" are frequently discussed in four out of the top five languages, except Spanish, where they account for only 2.1% and 2.6% of articles respectively, compared to an average of 11.6% and 11% in other languages. A similar discrepancy is noted for "Invasion of Ukraine" in German, appearing in just 1.1% of articles, while averaging 12.5% in Russian, Spanish, Arabic, and English. These disparities suggest a language-specific targeted dissemination of disinformation contingent on the topics.

To examine the emergence of related themes over time, we group similar high-frequency topics into four overarching themes: "COVID-19", "West", "Russia", and "Ukraine". These themes are manually defined by merging similar topics from the top 50 most frequent in the dataset. Figure 3 shows the temporal distribution of disinformation topics under each theme. From the first quarter of 2021 onwards, there is a noticeable increase in disinformation articles across all themes, coinciding with Russia's escalating military presence along the Ukrainian border. Similar surge in the number of "COVID-19"-related topics can be observed in the months following the start of the pandemic. After the full-scale invasion of Ukraine in the first quarter of 2022, disinformation articles on all topics decline significantly, except for "War in Ukraine" and "Invasion of Ukraine", which become the dominant disinformation topics.

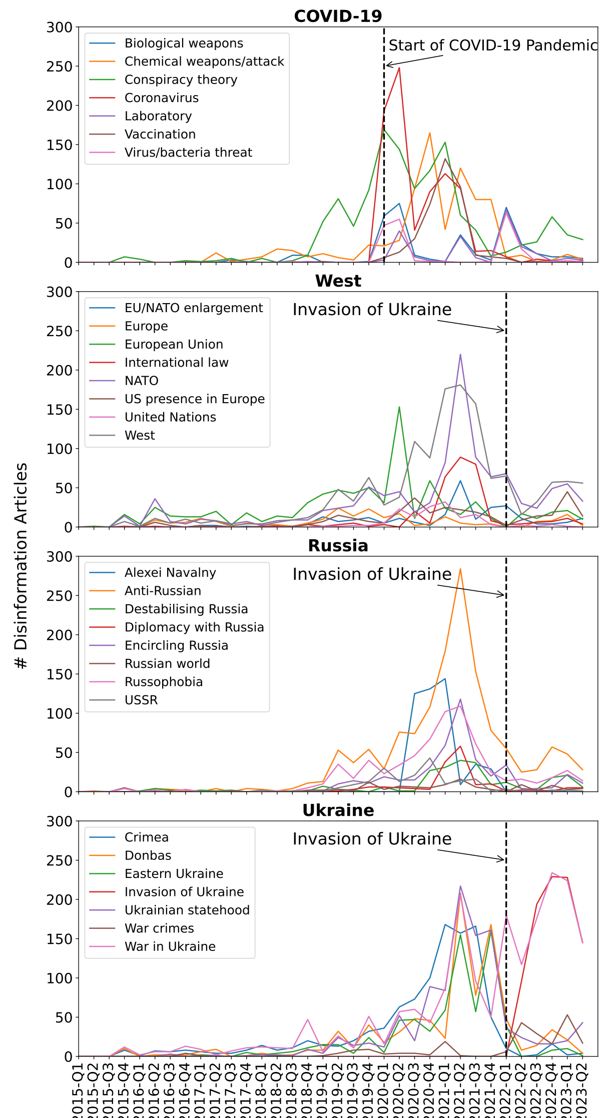

Figure 3: Temporal distribution of disinformation topics.

3 Experimental Setup

We employ the EUvsDisinfo dataset to perform binary disinformation detection (i.e. classify news articles as either disinformation or trustworthy). While the full dataset contains articles in 42 unique languages, several languages are severely underrepresented and/or highly skewed towards one of the classes. Therefore in order to draw reliable conclusions, we discard underrepresented languages (fewer than 25 articles) and those in which more than 95% of the articles belong to only one of the two classes. The resulting dataset used in the experiments comprises 14, 063 articles distributed across 14 languages (% of disinformation articles is also shown)[3]: English (6, 546, 6%), Russian (5, 825, 92%), German (313, 69%), French (292, 57%), Spanish (287, 85%), Georgian (156, 94%), Czech (152, 73%), Polish (147, 30%), Italian (103, 83%), Lithuanian (78, 36%), Romanian (68, 25%), Slovak (35, 91%), Serbian (31, 87%), Finnish (30, 27%).

We experiment with four models that have been used in prior work on article-level disinformation classification. We train commonly adopted baselines from previous work [4, 13, 20, 21]: Multinomial Naive Bayes (MNB), and Support Vector Machine (SVM), both using bag-of-words to encode the textual features. We also train transformer-based models, mBERT and XLM-RoBERTa, as three of the relevant multilingual datasets discussed in Section 1 use mBERT as a baseline [2, 10, 21], and XLM-R is used by Li et al. [13]. We conduct a hyperparameter search for each model by: (i) splitting the dataset into train and development sets with 90% and 10% of the data, respectively. The highest scoring hyperparameter configuration with respect to the development set is used throughout the experiments. (ii) discarding the development set, and splitting the remaining 90% of the dataset into train and test sets using a 10-fold cross-validation strategy. The folds are stratified by both language and class. Model performance for each language is measured using the $F1\text{-}Macro$ score to account for the skewed distribution of classes. To obtain a unified metric that encapsulates the overall system performance across all languages, we further average the $F1\text{-}Macro$ scores per language to obtain the $F1\text{-}Macro_{AVG}$ score. Further details such as hyperparameters and hardware configurations can be found in the supplementary material[3].

4 Results

Table 2: Classification results (Mean ± STD F1-Macro). Best scores are in bold.

Language	MNB	SVM	mBERT	XLM-R
EN	0.49±0.01	0.82±0.03	**0.89±0.03**	0.86±0.03
RU	0.49±0.02	0.76±0.05	**0.83±0.04**	0.82±0.03
DE	0.46±0.11	0.83±0.08	**0.92±0.05**	0.86±0.05
FR	0.65±0.13	**0.85±0.07**	0.83±0.08	0.83±0.10
ES	0.46±0.03	0.80±0.15	**0.92±0.11**	0.85±0.12
KA	0.73±0.27	**0.85±0.26**	0.73±0.22	0.66±0.22
CZ	0.42±0.04	0.82±0.12	0.83±0.12	**0.88±0.10**
PO	0.71±0.19	0.78±0.14	**0.88±0.08**	0.82±0.12
IT	0.56±0.24	0.80±0.23	**0.81±0.15**	0.78±0.15
LT	0.65±0.22	0.72±0.14	**0.89±0.15**	0.78±0.18
RO	0.57±0.30	0.52±0.32	0.73±0.25	**0.88±0.16**
SK	0.83±0.29	0.83±0.29	0.86±0.25	**0.92±0.19**
SR	0.76±0.32	0.76±0.32	**0.82±0.29**	0.73±0.29
FI	0.71±0.33	0.63±0.34	0.66±0.30	**0.77±0.31**
AVG	0.61	0.77	**0.83**	0.82

Table 2 shows the classification results for the proposed baselines. mBERT achieves the highest average score (0.83), and the highest per-language score in 8 out of the 14 languages: English, Russian, German, Spanish, Polish, Italian, Lithuanian, and Serbian. Next, XLM-R achieves an $F1\text{-}Macro_{AVG}$ score of 0.82, which is 1.2% lower than that of mBERT, despite XLM-R having roughly 20 times more parameters. The standard deviations for mBERT and XLM-R show that their scores largely overlap for most languages. Nonetheless, XLM-R achieves the highest scores for Czech, Romanian, Slovak, and Finnish. The SVM baseline achieves an $F1\text{-}Macro_{AVG}$ score of 0.77, a decrease of 7.2% compared to mBERT. Surprisingly, the SVM baseline achieves the highest scores for two languages: French and Georgian. Lastly, the MNB model scores the lowest with an $F1\text{-}Macro_{AVG}$ score of 0.61, a significant 26.5% decrease compared to mBERT.

5 Conclusions and Future Work

This paper introduced EUvsDisinfo, a large, linguistically, temporally, and topically diverse dataset of disinformation and trustworthy articles originating from pro-Kremlin and reliable / less biased outlets, respectively. Using the dataset, we found evidences of language-specific targeting of specific topics, and revealed a surge in disinformation content related to the war in Ukraine right before its full-scale invasion in 2022. Lastly, we proposed classification baselines using our dataset for the task of binary disinformation detection in a multilingual setting. In future work, we plan to leverage the structure of our dataset to explore evidence-aware fact-checking approaches by linking disinformation and trustworthy articles referring to the same narrative.

Ethical Statement

The dataset is seeded from the publicly available debunks published by EUvsDisinfo. The authors of the manuscript are not part of the EUvsDisinfo organisation. The dataset content originates from news articles, thus it does not contain personal data. The dataset specifically targets pro-Kremlin disinformation; while pro-Western disinformation is not within the scope of this work. We recognise that the dataset is susceptible to misuse for malicious purposes (e.g., used by originators of disinformation to improve their techniques), and we strongly urge researchers to use it in accordance with best practice ethics protocols. Our dataset is compliant with FAIR principles [27]. It is made fully available with a unique digital object identifier, in a CSV format that can be processed by most widely used tools, and is released under an Apache 2.0 license. Since the textual content of news articles may be copyrighted, we do not include them in the dataset, but instead provide a software (https://doi.org/10.5281/zenodo.10492913) to allow researchers to collect the content themselves.

Acknowledgments

This work has been co-funded by the UK's innovation agency (Innovate UK) grant 10039055 (approved under the Horizon Europe Programme as vera.ai, EU grant agreement 101070093) under action number 2020-EU-IA-0282.

References

[1] Steve Abrams. 2016. Beyond propaganda: Soviet active measures in Putin's Russia. *Connections* 15, 1 (2016), 5–31.

[2] Arkadipta De, Dibyanayan Bandyopadhyay, Baban Gain, and Asif Ekbal. 2021. A Transformer-Based Approach to Multilingual Fake News Detection in Low-Resource Languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 1, Article 9 (nov 2021), 20 pages. https://doi.org/10.1145/3472619

[3] Daryna Dementieva, Mikhail Kuimov, and Alexander Panchenko. 2023. Multiverse: Multilingual Evidence for Fake News Detection. *Journal of Imaging* 9, 4 (2023), 77.

[4] Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jennine B Everett, et al. 2018. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science*. 17–21.

[5] Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In Search of Credible News. In *Artificial Intelligence: Methodology, Systems, and Applications*, Christo Dichev and Gennady Agre (Eds.). Springer International Publishing, Cham, 172–180.

[6] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, and Sudipta Kar. 2020. BanFakeNews: A Dataset for Detecting Fake News in Bangla. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 2862–2871. https://aclanthology.org/2020.lrec-1.349

[7] Štefan Ižak. 2019. Using the topic of migration by pro-kremlin propaganda: case study of Slovakia. *Journal of Comparative Politics* 12, 1 (2019), 53–70.

[8] Andrzej Jarynowski, Łukasz Krzowski, and Stanisław Maksymowicz. 2023. Biological mis (dis)-information in the Internet as a possible Kremlin warfare.

[9] Georgi Karadzhov, Pepa Gencheva, Preslav Nakov, and Ivan Koychev. 2017. We Built a Fake News / Click Bait Filter: What Happened Next Will Blow Your Mind!. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, Ruslan Mitkov and Galia Angelova (Eds.). INCOMA Ltd., Varna, Bulgaria, 334–343. https://doi.org/10.26615/978-954-452-049-6_045

[10] Gautam Kishore Shahi and Durgesh Nandini. 2020. FakeCovid – A Multilingual Cross-domain Fact Check News Dataset for COVID-19. *arXiv e-prints*, Article arXiv:2006.11343 (June 2020), arXiv:2006.11343 pages. https://doi.org/10.48550/arXiv.2006.11343 arXiv:2006.11343 [cs.CY]

[11] Alexander Lanoszka. 2016. Russian hybrid warfare and extended deterrence in eastern Europe. *International affairs* 92, 1 (2016), 175–195.

[12] Tobias Lemke and Michael W Habegger. 2022. Foreign Interference and Social Media Networks: A Relational Approach to Studying Contemporary Russian Disinformation. *Journal of Global Security Studies* 7, 2 (04 2022), ogac004. https://doi.org/10.1093/jogss/ogac004 arXiv:https://academic.oup.com/jogss/article-pdf/7/2/ogac004/43510245/ogac004.pdf

[13] Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. 2020. Toward A Multilingual and Multimodal Data Repository for COVID-19 Disinformation. In *2020 IEEE International Conference on Big Data (Big Data)*. 4325–4330. https://doi.org/10.1109/BigData50022.2020.9378472

[14] Darren L Linvill, Brandon C Boatwright, Will J Grant, and Patrick L Warren. 2019. "THE RUSSIANS ARE HACKING MY BRAIN!" investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior* 99 (2019), 292–300.

[15] James Pamment and Corneliu Bjola. 2016. Digital containment: Revisiting containment strategy in the digital age. *Global Affairs* (2016).

[16] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In *Proceedings of the 27th International Conference on Computational Linguistics*, Emily M. Bender, Leon Derczynski, and Pierre Isabelle (Eds.). Association for Computational Linguistics, Santa Fe, New Mexico, USA, 3391–3401. https://aclanthology.org/C18-1287

[17] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A Stylometric Inquiry into Hyperpartisan and Fake News. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, Melbourne, Australia, 231–240. https://doi.org/10.18653/v1/P18-1022

[18] Bettina Renz. 2016. Russia and 'hybrid warfare'. *Contemporary Politics* 22 (06 2016), 1–18. https://doi.org/10.1080/13569775.2016.1201316

[19] Gautam Kishore Shahi, Julia Maria Struß, and Thomas Mandl. 2021. Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection. *Working Notes of CLEF* (2021).

[20] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data* 8, 3 (2020), 171–188.

[21] Veronika Solopova, Christoph Benzmüller, and Tim Landgraf. 2023. The Evolution of Pro-Kremlin Propaganda From a Machine Learning and Linguistics Perspective. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, Mariana Romanyshyn (Ed.). Association for Computational Linguistics, Dubrovnik, Croatia, 40–48. https://doi.org/10.18653/v1/2023.unlp-1.5

[22] Philip M Taylor. 2013. Munitions of the mind: A history of propaganda from the ancient world to the present era. In *Munitions of the Mind*. Manchester University Press.

[23] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a Large-scale Dataset for Fact Extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 809–819. https://doi.org/10.18653/v1/N18-1074

[24] Rod Thornton. 2015. The changing nature of modern warfare: responding to Russian information warfare. *The RUSI Journal* 160, 4 (2015), 40–48.

[25] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Regina Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, Vancouver, Canada, 422–426. https://doi.org/10.18653/v1/P17-2067

[26] Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. 2020. Weak Supervision for Fake News Detection via Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 516–523. https://doi.org/10.1609/aaai.v34i01.5389

[27] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.

[28] Evan M Williams and Kathleen M Carley. 2023. Search engine manipulation to spread pro-Kremlin propaganda. *Harvard Kennedy School Misinformation Review* (2023).

[29] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one* 11, 3 (2016), e0150989.