ThreatGram101: Extreme Telegram Replies Data with Threat Levels

Kamalakkannan Ravi(✉) and Jiann-Shiun Yuan

University of Central Florida, Orlando, FL 32816, USA
{kamalakkannan.ravi,jiann-shiun.yuan}@ucf.edu

Abstract. With the growth of social media, threats in comments targeting public officials, entities, or organizations have become increasingly common. Previous research on threat detection has typically focused on broad categories such as normal speech, hate speech, and offensive speech, lacking a more focused approach to identify calls for harm explicitly. To address this gap, we present a comprehensive dataset of user replies from Telegram channels associated with political extremism or the cyberbullying of public officials in the United States. Using keywords, we identified Telegram channels with extreme ideological leanings, high rates of grievances, and threatening language.

We employed expert annotation to label a subset of replies from this dataset, creating a labeled set of 15,076 replies categorized as no threat, judicial threat, and non-judicial threat. This paper releases two datasets: 2 million unlabeled replies and 15,076 labeled replies from 17 Telegram channels. This dataset aims to enhance proactive monitoring and mitigation strategies for negative content, threats, and abusive language in social media comments. It provides a valuable resource for threat detection, political extremism analysis, countering violent extremism, and the study of cyberbullying dynamics on social media platforms, addressing current limitations in data diversity and enabling more effective responses to online threats.

Keywords: countering violent extremism · cyberbullying detection · natural language processing · political violence · radicalization behavioral indicators · social media · telegram · user-generated content

1 Introduction

The rise of social media has dramatically transformed how individuals communicate and express their views, but it has also facilitated a troubling increase in ideologically motivated violence, particularly against public officials and institutions. High-profile incidents, such as the January 6, 2021, U.S. Capitol attack [34], alongside cases of anarchist violent extremism (AVE) [16], starkly illustrate the shift from online extremism to real-world violence. These events highlight the pressing need for robust mechanisms to detect and mitigate emerging threats in digital spaces.

J. A. Lossio-Ventura et al. (Eds.): SIMBig 2024, CCIS 2496, pp. 275–291, 2025.
https://doi.org/10.1007/978-3-031-91428-7_20

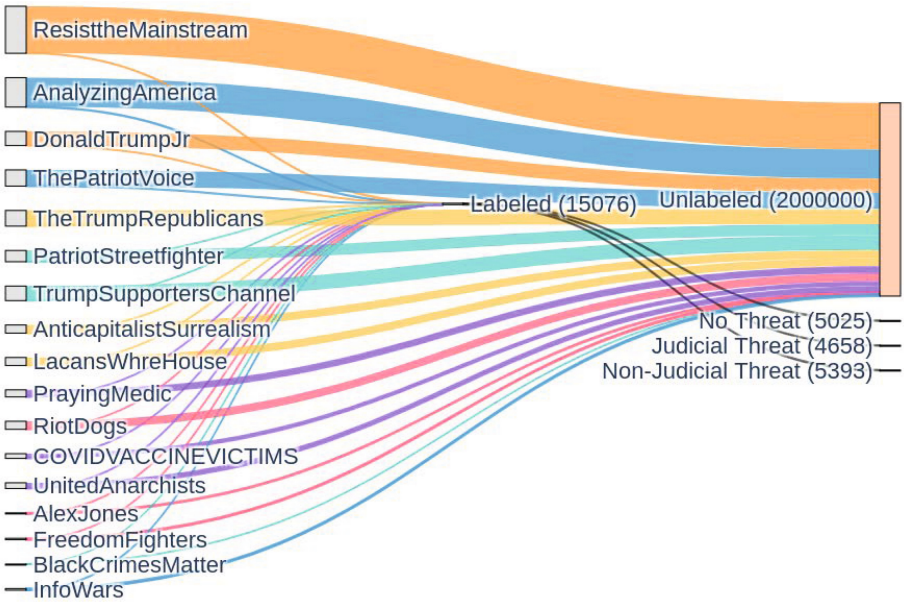

Fig. 1. Data collection from public Telegram channels and labeling

Identifying and categorizing the various forms of domestic violent extremism (DVE) presents challenges, including distinctions between anti-government violent extremism (AVE), far-right extremism, sovereign citizen violent extremism (SCVE), militia violent extremism (MVE), and racially/ethnically motivated violent extremism (RMVE) [16]. While this dataset is designed to enhance the understanding of threats targeting public officials, its primary contribution lies in providing a labeled resource that aids in the classification and detection of specific threat levels rather than distinguishing between these broader categories of extremism. Each of these groups, driven by grievances and inflammatory rhetoric, propagates violent ideologies that pose significant risks to societal stability [33,37]. While this paper does not aim to differentiate among these forms of extremism, it highlights the necessity of a nuanced approach to threat detection to address the overarching threats posed by DVE effectively.

Traditional models, such as the U.S. Secret Service's threat assessment framework, emphasize behavioral indicators and contextual factors rather than simplistic profiles or specific threats [9]. This proactive model focuses on identifying patterns of planning and behavior, which can be adapted to the digital realm through the integration of Natural Language Processing (NLP) and other advanced technologies. These technologies are particularly relevant for social media platforms like Telegram, where extremist content and preparatory behaviors-actions that indicate planning or intent beyond mere verbal expression-are prevalent [42]. This understanding underscores the importance of

analyzing not only the language used but also the context and potential implications of user interactions on such platforms.

Similarly, the work of Vossekuil et al., which focuses on preventing targeted violence against judges and courts, highlights the importance of early detection of radicalization signs, such as unusual interest in targets or communication about violent intentions [41]. This reinforces the need for sophisticated methods of threat identification, which can be improved through well-designed datasets and a refined focus on threat classification.

The evolution of cyber harassment, as explored by Danielle Keats Citron [11], further underscores the intersection between online abuse and political violence. Harassment on platforms like Telegram can escalate into more serious abuse, necessitating more advanced detection systems. These systems must be capable of identifying harassment patterns, threats, and abusive language in real time, making the role of cybersecurity and NLP techniques increasingly crucial [1,18].

Distinguishing between different forms of violence, such as affective and predatory violence, is another key factor in addressing online threats [24]. Social media platforms play a significant role in intelligence gathering related to radicalization and violent extremism, which calls for more precise detection methods. As political polarization grows, especially on social media, platforms like Telegram further complicate the threat landscape by reinforcing radical views [22]. This emphasizes the need for interdisciplinary approaches that combine behavioral analysis with advanced analytics to counter the effects of polarization on extremism.

Recent research, such as that by McBride et al. [23] and the Composite Violent Extremism (CoVE) framework by Gartenstein-Ross et al. [15], highlights the challenges posed by ideologically ambiguous extremists. These frameworks point to the need for more precise data and advanced NLP tools to analyze fragmented ideologies and predict threats [13]. Furthermore, Don Rassler's work [27] on the integration of data analysis in counterterrorism, and the challenges faced by election workers [14], illustrate the growing importance of sophisticated tools for monitoring and preventing violence on platforms like Telegram.

Most studies on online threats have relied on broad categories such as normal speech, hate speech, and offensive speech [12]. While useful, these frameworks often lack the granularity necessary to capture subtle distinctions in threat severity, particularly when identifying indicators of potential violence [30–32]. As threats evolve, more fine-grained classification models are needed to effectively monitor and address emerging risks. This challenge becomes even more pronounced when analyzing user behavior on platforms like Telegram [42], where threats may manifest in a variety of forms-ranging from violent rhetoric to preparatory behaviors that existing classification systems struggle to capture.

In parallel, while NLP techniques have shown promise in detecting abusive language and categorizing threats, there is still a significant need for well-labeled datasets, comparable to those used in domains such as video analysis [3], signal analysis [10,19,39], power flow analysis [35,36], and biomedical research [5,17,21,25,28,40]-to effectively capture the complexities of cyber harassment and its intersections with various forms of online abuse, including political violence

and extremism. The development of such datasets is essential to advance threat detection models.

In response to these challenges, our study presents a comprehensive dataset of user replies from Telegram channels associated with political extremism and the cyberbullying of public officials in the United States. By adopting a more focused, three-class threat system-differentiating between no threat, judicial threats, and non-judicial threats-we address the limitations of existing data diversity and classification methods. This dataset serves as a valuable resource for researchers and practitioners, contributing to more effective monitoring and proactive threat mitigation strategies. Through this work, we aim to enhance the capacity to respond to the growing risks posed by online threats, particularly those targeting public figures and institutions.

2 Problem Statement

The primary goal of this study is to improve the detection and classification of online threats by addressing key shortcomings in current research, as outlined in the introduction. Specifically, we aim to improve existing threat classification frameworks and introduce a comprehensive dataset that enables more effective threat detection across diverse online environments.

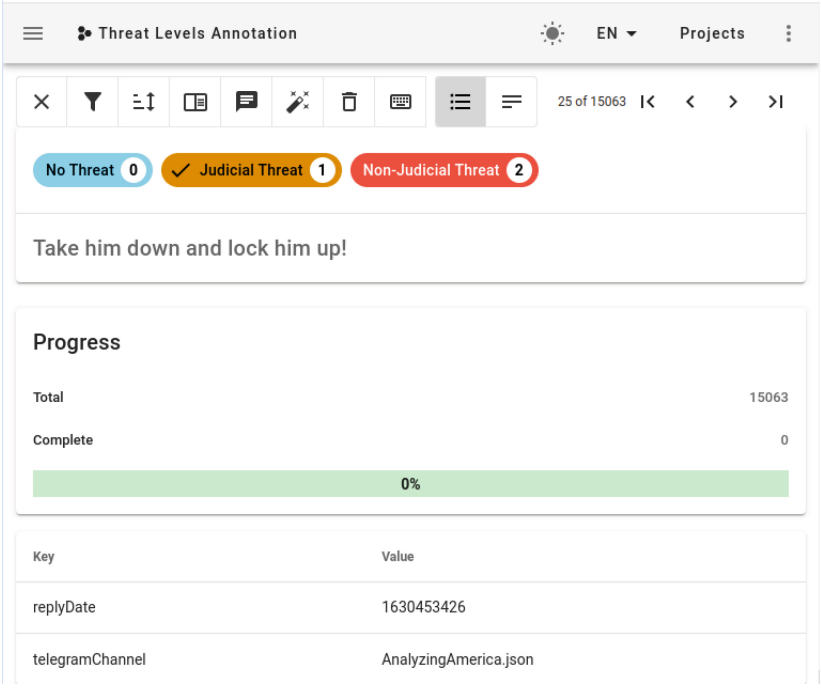

Fig. 2. Threat level annotation using Doccano.

2.1 Contribution 1: Threat Levels

Threat detection models in the field have predominantly classified content into broad categories such as normal speech, hate speech, and offensive speech. However, this approach lacks the ability to capture varying threat levels, particularly those that involve explicit calls for violence [12]. Existing systems have struggled to identify the nuanced indicators of potential threats, creating a gap in the ability to effectively monitor online extremism. For example, models like those developed by Ravi et al. underscore the limitations of current frameworks, which fail to account for subtle distinctions between legal rhetoric and true incitements to violence [32].

To address this issue, we propose a more focused classification system that introduces three distinct threat categories: no threat, judicial threat, and non-judicial threat. This system enhances the ability to detect subtler forms of extremism, especially in instances where violent rhetoric is concealed within legal discourse. By refining the definition of threat levels, we aim to clarify the intent behind user interactions on platforms like Telegram, where extremist rhetoric often manifests in ambiguous forms. The definitions of these threat levels and their implications for threat detection are elaborated in Sects. 3 and 4.

2.2 Contribution 2: Labeled Dataset

One of the significant barriers to advancing threat detection is the lack of comprehensive, well-labeled datasets that reflect the complexity of online threats [12,30–32]. Existing datasets have fallen short of capturing the intricate nature of online abuse, particularly in the context of extremist content. While previous works, such as those by Wulczyn et al. [44], have utilized a combination of crowdsourcing and machine learning to analyze personal attacks, these studies did not extend to the more complex identification of violent threats. Similarly, Ashraf et al. [4] examined the detection of violent threats on YouTube, but their binary classification approach lacked the granularity necessary to distinguish among different types of threats.

These limitations underscore the urgent need for more specialized datasets that can accurately reflect the diverse range of threats across different platforms. Telegram, with its unique structure and widespread use by extremist groups, provides an ideal data source for this research. Unlike traditional social media platforms like Reddit or Twitter, Telegram supports threaded discussions and message comments, which allow for more interactive and extremist content to proliferate. The prevalence of violent rhetoric from both far-right and far-left groups on Telegram [42] further emphasizes the importance of studying emerging social media platforms in the context of threat detection [8].

To address this gap, we selected Telegram as our primary data source due to its significant role in hosting extremist content and facilitating political and ideological discourse. Unlike traditional social media platforms like Reddit and Twitter, Telegram's support for threaded discussions and message comments in public groups [42] provides a unique opportunity to capture interactive and

extremist content. Our dataset comprises 2 million unlabeled replies and 15,076 expertly labeled replies from Telegram channels (illustrated in Fig. 1) associated with political extremism and the cyberbullying of public officials in the United States between 2019 and 2024, as depicted in Fig. 3 and a breakdown by class/label for each channel is given in Table 2. The rationale behind choosing Telegram and the specifics of our data collection methodology are discussed in Sect. 4.

Section 5 offers a detailed overview of the collected data, including channel specifics, labeled and unlabeled data counts, and the unique contributions and value of the dataset. Additionally, Sect. 6 summarizes our findings and addresses the ethical considerations associated with the development of this dataset and the definition of threat levels.

In summary, this paper seeks to advance the field of threat detection by introducing a refined classification system and providing a comprehensive dataset.

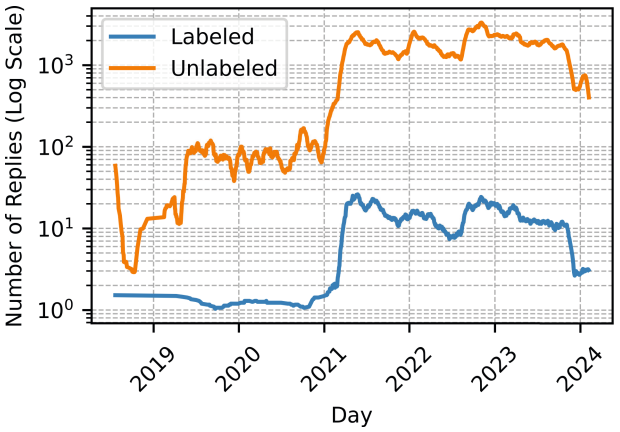

Fig. 3. Aggregate user activity in selected Telegram channels

3 Definition of Threat Levels

The study presents a refined threat classification system designed to enhance clarity and specificity in threat detection, addressing the limitations of existing research that typically categorizes harmful content into broad categories such as normal speech, hate speech, and offensive speech [12].While these frameworks provide a foundational understanding of harmful online behavior [6,20], they often lack the specificity needed to differentiate between varying threat levels [7]. Building on this foundation, Ravi et al. proposed a more granular six-level threat classification system, which included concepts such as fighting words, incitement, and true threats [32]. However, this approach faced challenges in effectively capturing nuanced threat indicators due to overlapping categories that hindered practical applications.

To rectify these issues, our study introduces a judicial/non-judicial threat classification system that distinguishes between legally framed grievances and explicit threats of violence, particularly in the context of extremist rhetoric. Judicial threats, while often couched in legal terminology, can incite actions that contribute to radicalization and social disruption. Extremist groups frequently leverage legal rhetoric to evade detection while mobilizing their followers or disseminating ideologies. Consequently, it is essential to identify these subtler risks. Non-judicial threats, in contrast, involve direct incitements to violence or harm and are generally easier to detect.

This classification framework improves threat detection by capturing both overt and legally masked threats, thereby enhancing the effectiveness of counter-violent extremism (CVE) strategies [9,42]. This new approach focuses on explicit calls for harm and provides a clear framework for identifying and responding to online threats. The three-category scheme has been developed with expert guidance from criminologists [32] to ensure a clear and actionable distinction between different types of threats. While it may appear coarse-grained, these categories reflect the legal and practical definitions of threats. The classification system simplifies threats into three distinct classes, which are elaborated upon as follows:

Table 1. Telegram channels with count of unlabeled and labeled replies

Telegram Channel Name	Unlabeled Replies	Labeled Replies
AlexJones	11,063	72
AnalyzingAmerica	300,406	4,255
AnticapitalistSurrealism	84,400	192
BlackCrimesMatter	1,211	7
COVIDVACCINEVICTIMS	49,342	153
DonaldTrumpJr	146,081	634
FreedomFighters	15,080	95
InfoWars	17,602	143
LacansWhreHouse	82,923	131
PatriotStreetfighter	116,114	387
PrayingMedic	73,388	176
ResisttheMainstream	482,596	4,505
RiotDogs	83,182	170
ThePatriotVoice	166,986	825
TheTrumpRepublicans	160,603	1,774
TrumpSupportersChannel	149,506	1,414
UnitedAnarchists	59,517	143
Total	**2,000,000**	**15,076**

3.1 No Threat

The "No Threat" category includes statements that do not suggest or indicate any form of physical harm, imprisonment, or threats against any individual, group, or organization. This category encompasses both non-threatening comments and statements with ambiguous language that does not clearly advocate for harm or illegal action. Examples of statements in this category include: "Hillary was just seething. That look on her face was priceless," "You'll regret this," "Live free or die," and "They are traitors to the party." By defining this category, we aim to separate benign or ambiguous language from more direct threats.

3.2 Judicial Threat

The "Judicial Threat" category includes statements that explicitly call for or threaten legal action, such as civil lawsuits, arrests, or criminal prosecutions within the bounds of standard legal norms. Although these threats involve legal consequences, they still represent a significant level of threat due to their potential to incite action through legal channels. Examples of statements in this category include: "Lock her up" and "Wake the hell up and put that POS in jail!!" This categorization helps differentiate threats that operate within a legal context from those that advocate for extralegal or violent actions.

3.3 Non-Judicial Threat

The "Non-Judicial Threat" category represents the most severe form of threat, encompassing statements that explicitly advocate for non-legal actions or harm, such as physical violence or vigilante justice. This category includes statements that call for or suggest unlawful actions, posing the highest level of threat. If a statement includes elements of both judicial and non-judicial threats, it is classified as a non-judicial threat due to its extremity. Examples include: "Time to start a civil war!" "Hang Mike Pence!" "Fuck the judicial system and shoot these assholes in the face instantly." "Try Pelosi for treasonous actions against America and hang her." By clearly defining this category, we provide a framework to identify the most dangerous forms of online speech, thereby enhancing the capability to monitor and mitigate such threats effectively.

This focused classification system offers a more precise methodology for identifying and categorizing threats. By concentrating on explicit calls for harm and differentiating between legal and extralegal threats, our approach serves as a robust tool for analyzing and responding to online extremism and radicalization. Through this refined framework, researchers and practitioners can more effectively engage with the complexities of threat detection and enhance their capacity to address the challenges posed by online violence.

Table 2. Labeled replies with breakdown by label for each channel.

Telegram Channel Name	Labeled	No Threat	Judicial Threat	Non-Judicial Threat
AlexJones	72	23	27	22
AnalyzingAmerica	4,255	936	1,721	1,598
AnticapitalistSurrealism	192	130	7	55
BlackCrimesMatter	7	2	1	4
COVIDVACCINEVICTIMS	153	88	38	27
DonaldTrumpJr	634	290	168	176
FreedomFighters	95	25	16	54
InfoWars	143	51	29	63
LacansWhreHouse	131	107	5	19
PatriotStreetfighter	387	199	65	123
PrayingMedic	176	124	32	20
ResisttheMainstream	4,505	1,519	1,199	1787
RiotDogs	170	116	0	54
ThePatriotVoice	825	341	205	279
TheTrumpRepublicans	1,774	536	668	570
TrumpSupportersChannel	1,414	440	472	502
UnitedAnarchists	143	98	5	40
Total	**15,076**	**5,025**	**4,658**	**5,393**

4 Experimental Design, Materials and Methods

To collect relevant data from Telegram, we began by identifying 25 channels using specific keywords, including "jan 6," "anarchy," "proud," "riot," "patriot," "freedom," "crime," "maga," "late-stage capitalism," and "conspiracies." These keywords were chosen to reflect both far-right and far-left perspectives, as identified in previous studies on extremist and threatening online content [2,42,43]. The selected channels were known for frequently using grievance-driven and threatening language [42], making them suitable for our research on online threats.

The data collection process was methodically planned to ensure compliance with ethical guidelines and privacy concerns. We created a Telegram account using a virtual phone number obtained through Google Voice. This approach allowed us to join public Telegram groups without revealing personal information. Once inside the groups, we utilized the "view discussions" option to access the chat histories, which were subsequently exported as .json files. Our collection efforts were strictly limited to publicly available content, and we avoided any interactions with Telegram users to maintain ethical standards. Additionally, no personally identifiable information was collected from users or annotators during this process.

To maintain the dataset's quality and relevance, channels with fewer than 1,000 replies were excluded due to insufficient user activity, aligning with established research practices [29]. After applying this criterion, 17 Telegram channels were selected for detailed analysis. Figure 3 illustrates the aggregate user activity across these channels. The data collection covered the entire history of these channels, from their inception to February 6, 2024, using Telegram's chat export tool. This comprehensive collection included both messages and their replies. To maintain the focus on text-based content, non-text elements such as URLs and empty responses were removed during data preprocessing. This refinement resulted in a robust dataset comprising a total of 2,301,110 replies.

The annotation categories, developed with guidance from criminology experts, were based on Ravi et al. [32], focusing on three specific threat levels: no threat, judicial threat, and non-judicial threat, as described in Sect. 3. For the annotation process, a subset of 301,110 replies was selected and annotated by an expert graduate student (trained by U.S.-based criminology experts, as referenced in Ravi et al. [32]) using the open-source annotation tool Doccano [26], as shown in Fig. 2. To streamline the annotation process and ensure consistency, we employed a two-step approach: (1) an initial categorization using a fine-tuned RoBERTa large language model, and (2) a manual review and refinement of these annotations. This detailed labeling process resulted in a balanced and well-defined dataset of 15,076 labeled replies. Additionally, we have provided a more detailed breakdown of the class/label distributions across each channel in the dataset, which is now included in Table 2. The balanced nature of the labeled dataset ensures a more nuanced understanding of threat levels and improves the model's ability to detect various forms of online threats.

5 Discussion

The final dataset comprises 2 million unlabeled replies and 15,076 labeled replies, providing a comprehensive resource for studying online threats and extremist content. The dataset is publicly accessible on Mendeley and can be downloaded using this link: https://data.mendeley.com/datasets/tm9s68vgxd/1. The Telegram channels used for data collection are detailed in Table 1, which includes the number of replies posted (both labeled and unlabeled) and the reply counts per day, shown in aggregate in Fig. 3.

For ease of use, Table 3 provides an overview of the file names and structure within the .json files. Table 4 further details the dataset's specifications, emphasizing the focus on labeled replies to create a comprehensive dataset of

Table 3. Open-Sourced Data Details

Folder	Filename	Structure
Data 1 - Raw and unlabeled	Unlabeled.json	replyDate reply telegramChannel
Data 2 - Raw and labeled	Labeled.json	replyDate reply telegramChannel Label

Table 4. Specifications Table

Specification	Details
Subject	Artificial Intelligence
Specific subject area	Computational Social Science, Social Computing, Computational Linguistics
Data format	Data 1: Raw and Unlabeled Data 2: Raw and Labeled
Type of data	Tables (.json)
Data collection	The data were collected using the Telegram Chat Export Tool built within the application.
Data source location	Telegram
Data accessibility	Repository name: Threatgram 101: Extreme Telegram Replies Data with Threat Levels Direct URL to data: https://data.mendeley.com/datasets/tm9s68vgxd/1
Instructions for accessing these data	Data is hosted publicly in Mendeley and can be downloaded by visiting the above given link.
Related research article	Ravi, K. and Yuan, J.S., 2024. Ideological Orientation and Extremism Detection in Online Social Networking Sites: A Systematic Review.

user-generated threats on Telegram. This dataset offers several key contributions to the field:

- **Enhanced Threat Labels:** The dataset includes more focused threat levels (no threat, judicial threat, non-judicial threat), enabling researchers to develop and evaluate advanced threat detection models. This granularity supports more precise analyses of radicalization, extremism, and online propaganda, moving beyond general categorizations.
- **Advancing NLP Research:** Spanning a 4-year period from 2019 to 2024, this dataset-with its 2 million unlabeled and 15,076 labeled replies-provides a rich resource for advancing Natural Language Processing (NLP). It enables researchers to develop and refine algorithms for detecting extremist and threatening language, thereby enhancing social media monitoring systems.
- **Diverse Data Source:** By focusing on data from Telegram, this dataset offers unique insights into extremist and threatening content on a platform distinct from traditional social media like Twitter and Reddit. This diversity provides a novel context for analysis, broadening the scope of research into online extremism.
- **Benchmark for Comparative Studies:** The labeled dataset serves as a benchmark for comparing and assessing the effectiveness of various threat detection and classification systems. It aids in the validation and improvement of these technologies, fostering advancements in threat detection methodologies.

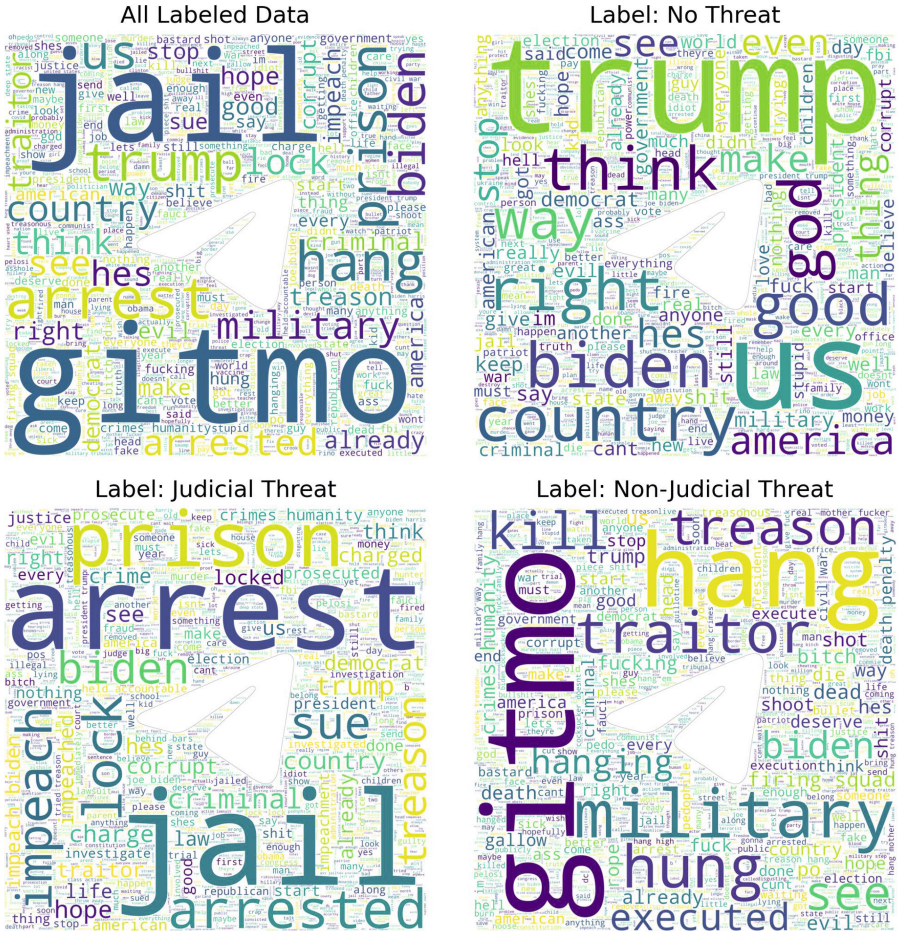

Fig. 4. Word cloud of all labeled replies combined, and for each label individually

– **Supporting Behavioral and Political Research:** The dataset's focus on political extremism and cyberbullying provides valuable insights into the behaviors and patterns of online threats and abusive language. It supports studies on countering violent extremism (CVE) and radicalization, contributing to broader efforts to understand and mitigate online threats (as illustrated in Fig. 4).

6 Conclusion

ThreatGram 101 introduces a refined approach to online threat detection by providing a detailed dataset of extreme Telegram replies. By categorizing threats into three distinct levels-no threat, judicial threat, and non-judicial threat-this dataset offers a more precise framework for threat classification (as detailed in (Sect. 3). This refinement enhances the accuracy of threat detection compared to traditional broad classifications.

The dataset, comprising 2 million unlabeled replies and 15,076 labeled examples from 17 Telegram channels, provides a comprehensive view of user-generated content related to political extremism and cyberbullying. This resource (discussed in Sects. 4 and 5) not only supports the development of more effective monitoring and response strategies but also advances the field of Natural Language Processing in threat detection.

Despite these advancements, limitations remain, including the lack of inter-annotator agreement measures, as only a single expert annotator was used. While expert training and strict adherence to criminological definitions reduced bias, political subjectivity is challenging to eliminate entirely in such tasks. Future work will involve expanding the classification scheme for greater granularity and incorporating multiple annotators to ensure consistency and reduce subjectivity.

Overall, ThreatGram 101 represents a valuable asset for researchers and practitioners, offering new insights into extremist behavior and online abuse while enhancing threat detection models. As online threats continue to evolve, this dataset equips us with the necessary tools for more accurate and specific responses to emerging challenges.

Ethical Considerations. Quantifying the threat level of public Telegram replies through expert annotation does not fully capture the diversity of Telegram communities in the United States, where the platform is less widely used compared to other regions [38]. In 2023, Telegram had approximately 10 million monthly users in the U.S., representing less than 2% of its global user base [8]. Our data collection focused exclusively on public Telegram groups and channels that are accessible without requiring invitations. By excluding private groups, we aimed to protect user privacy and minimize the risk of misuse of our methodology. Our data collection and classification processes received formal ethical clearance, ensuring that we followed responsible research practices.

We implemented a standardized annotation framework developed in consultation with social science and criminology experts [32], ensuring a consistent and rigorous process to mitigate the potential unintended bias that may arise from relying on a single expert annotator, particularly in classifications involving social and political content. Nonetheless, we acknowledge that the potential for bias remains, and we recognize this as a limitation of the study.

Regarding messages labeled as "Judicial Threat," we wish to clarify that our classification schema does not imply any punitive actions or judgment against individuals discussing legal matters. This category intends to identify messages that could potentially incite action within legal frameworks, as defined in collab-

oration with social science and criminology experts. The goal is to highlight the potential for certain legal discussions to escalate into actions that may pose risks, while respecting the right to express grievances and pursue lawful solutions.

Our research was approved by the Institutional Review Board (IRB) at the University of Central Florida under IRB ID STUDY00006200, titled Data Collection and Labeling for Social Media Language Research. The IRB determined that our study qualifies as human subjects research that is exempt from regulation. This assessment confirmed that the activities described, including the collection and labeling of publicly available content, do not involve any private or identifiable information from individual users. The IRB reviewed our data collection methods, which explicitly excluded private content, and determined that our study adheres to ethical guidelines for secondary data analysis. By adhering to these ethical protocols, we ensure that our research remains both responsible and respectful of privacy while contributing to the development of advanced threat detection methodologies.

Conflicts of Interest. The authors affirm that they have no known financial interests or personal relationships that could have influenced the research presented in this paper. There are no conflicts of interest that could potentially bias the results or interpretations of the findings.

References

1. Agatston, P., Kowalski, R., Limber, S.: Youth views on cyberbullying. In: Cyberbullying Prevention and Response, pp. 57–71. Routledge (2012)
2. Alava, S., Chaouni, N., Charles, Y.: How to characterise the discourse of the far-right in digital media? Interdisciplinary approach to preventing terrorism. Procedia Comput. Sci. **176**, 2515–2525 (2020)
3. Alijanpour, M., Raie, A.: Video event recognition using two-stream convolutional neural networks. In: 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), pp. 1–5. IEEE (2021)
4. Ashraf, N., Mustafa, R., Sidorov, G., Gelbukh, A.: Individual vs. group violent threats classification in online discussions. In: Companion Proceedings of the Web Conference 2020, pp. 629–633 (2020)
5. Azizi, S., Soleimani, R., Ahmadi, M., Malekan, A., Abualigah, L., Dashtiahangar, F.: Performance enhancement of an uncertain nonlinear medical robot with optimal nonlinear robust controller. Comput. Biol. Med. **146**, 105567 (2022)
6. Bahador, B.: Monitoring hate speech and the limits of current definition. 86272 **12**, 291–298 (2023)
7. Basile, P., Caputo, A., Semeraro, G.: An enhanced Lesk word sense disambiguation algorithm through a distributional semantic model. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1591–1600 (2014)
8. Bianchi, T.: Telegram app downloads in selected american countries 2023 (2024). https://www.statista.com/statistics/1234567/telegram-app-downloads-by-country/. Accessed 04 Mar 2024

9. Borum, R., Fein, R., Vossekuil, B., Berglund, J.: Threat assessment: defining an approach for evaluating risk of targeted violence. Behav. Sci. Law **17**(3), 323–337 (1999)
10. Chowdhury, R.I., Sun, A.K., Tamir, A., Shahnaz, C., Fattah, S.A.: Brain-drive: a smart driver for controlling digital appliances using cognitive command. In: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 851–855. IEEE (2017)
11. Citron, D.K.: Hate Crimes in Cyberspace. Harvard University Press (2014). https://digitalcommons.law.umaryland.edu/books/91
12. Davidson, T., Warmsley, D., Macy, M., Weber, I.: Automated hate speech detection and the problem of offensive language. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, pp. 512–515 (2017)
13. Elfardy, H., Diab, M., Callison-Burch, C.: Ideological perspective detection using semantic features. In: Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics, pp. 137–146 (2015)
14. Fischler, J.: State and local election workers quitting amid abuse, officials tell U.S. senate panel. Penn Capital-Star (2023)
15. Gartenstein-Ross, D., Zammit, A., Chace-Donahue, E., Urban, M.: Composite violent extremism: conceptualizing attackers who increasingly challenge traditional categories of terrorism. Stud. Conflict Terrorism 1–27 (2023)
16. GW Program on Extremism: Anarchist/left-wing violent extremism in America: trends in radicalization, recruitment, and mobilization (2021). https://extremism.gwu.edu/anarchist-violent-extremism. Accessed 12 Aug 2024
17. Hoogenboom, S.A., et al.: Missed diagnosis of pancreatic ductal adenocarcinoma detection using deep convolutional neural network. Gastroenterology **160**(6, Suppl.), S–18 (2021). https://doi.org/10.1016/S0016-5085(21)00794-0. https://www.sciencedirect.com/science/article/pii/S0016508521007940
18. Islam, M.S., Rafiq, R.I.: Comparative analysis of GPT models for detecting cyberbullying in social media platforms threads. In: Annual International Conference on Information Management and Big Data, pp. 331–346. Springer, Cham (2023)
19. Kamalakkannan, R., Rajkumar, R., Raj, M.M., Devi, S.S.: Imagined speech classification using EEG. Adv. Biomed. Sci. Eng. (ABSE) **1**(2), 20–32 (2014). https://doi.org/10.5281/zenodo.13888262. https://www.researchgate.net/publication/309967859_Imagined_Speech_Classification_using_EEG
20. Khan, F.H., Qamar, U., Bashir, S.: eSAP: a decision support framework for enhanced sentiment analysis and polarity classification. Inf. Sci. **367**, 862–873 (2016)
21. Kumar, S., Ravi, K., Mulay, S., Ram, K., Sivaprakasam, M.: Deep residual network based automatic image grading for diabetic macular edema. In: Research Poster Papers of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2018). https://doi.org/10.13140/RG.2.2.24611.02082/1. https://www.researchgate.net/publication/374471910_Deep_Residual_Network_based_Automatic_Image_Grading_for_Diabetic_Macular_Edema
22. Levin, S.A., Milner, H.V., Perrings, C.: The dynamics of political polarization (2021)
23. McBride, M.K., Carroll, M., Mellea, J.L., Savoia, E.: Targeted violence. Perspect. Terrorism **16**(2), 24–38 (2022)
24. Meloy, J.R., Hoffmann, J.: International handbook of threat assessment. Oxford University Press (2021)

25. Molaei, S., Ghorbani, N., Dashtiahangar, F., Peivandi, M., Pourasad, Y., Esmaeili, M.: Fdcnet: presentation of the fuzzy CNN and fractal feature extraction for detection and classification of tumors. Comput. Intell. Neurosci. **2022**(1), 7543429 (2022)
26. Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., Liang, X.: doccano: text annotation tool for human, vol. 34 (2018). https://github.com/doccano/doccano
27. Rassler, D.: A view from the CT foxhole: ravi satkalmi, director of intelligence, united states capitol police. CTC Sentinel **16**(7) (2023). https://ctc.westpoint.edu/wp-content/uploads/2023/07/CTC-SENTINEL-072023.pdf
28. Ravi, K., Selvaraj, S., Mulay, S., Ram, K., Sivaprakasam, M.: Breast cancer histology classification using deep residual networks. In: Research Poster Papers of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2018). https://doi.org/10.13140/RG.2.2.22094.43840. https://www.researchgate.net/publication/374471906_Breast_cancer_histology_classification_using_Deep_Residual_Networks
29. Ravi, K., Vela, A.E.: Comprehensive dataset of user-submitted articles with ideological and extreme bias from reddit. Data Brief **56**, 110849 (2024). https://doi.org/10.1016/j.dib.2024.110849. https://www.sciencedirect.com/science/article/pii/S2352340924008138
30. Ravi, K., Vela, A.E.: Rico: reddit ideological communities. Online Soc. Netw. Media **42**, 100279 (2024). https://doi.org/10.1016/j.osnem.2024.100279
31. Ravi, K., Vela, A.E., Ewetz, R.: Classifying the ideological orientation of user-submitted texts in social media. In: Proceedings of the 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 413–418. IEEE (2022). https://doi.org/10.1109/ICMLA55696.2022.00066. https://ieeexplore.ieee.org/document/10069289
32. Ravi, K., Vela, A.E., Jenaway, E., Windisch, S.: Exploring multi-level threats in telegram data with AI-human annotation: a preliminary study. In: 2023 International Conference on Machine Learning and Applications (ICMLA), pp. 1520–1527. IEEE (2023). https://doi.org/10.1109/ICMLA58977.2023.00229. https://ieeexplore.ieee.org/document/10459792
33. Ravi, K., Yuan, J.S.: Ideological orientation and extremism detection in online social networking sites: a systematic review. Intell. Syst. Appl. **24**, 200456 (2024). https://doi.org/10.1016/j.iswa.2024.200456. https://www.sciencedirect.com/science/article/pii/S2667305324001303
34. Rozenshtein, A.Z., Shugerman, J.H.: January 6, ambiguously inciting speech, and the overt-acts rule. Const. Comment. **37**, 275 (2022)
35. Sarkar, R.: Power flow computation under k-line removal. IEEE Trans. Power Syst. **37**(2), 1653–1656 (2021)
36. Sarkar, R.: An analytical approach for reducing k-line failure analysis and load shed computation. IET Gener. Transm. Distrib. **16**(13), 2623–2641 (2022)
37. Simi, P., Ligon, G., Hughes, S., Standridge, N.: Rising threats to public officials: a review of 10 years of federal data (2024)
38. Stocking, G., et al.: The role of alternative social media in the news and information environment. Pew Research Center (2022)
39. Tamir, A., Salem, M., Lin, J., Alasad, Q., Yuan, J.S.: Multi-tier 3D IC physical design with analytical quadratic partitioning algorithm using 2D P&R tool. Electronics **10**(16), 1930 (2021)
40. Tamir, A., Salem, M., Yuan, J.S.: Protec: a transformer based deep learning system for accurate annotation of enzyme commission numbers. IEEE/ACM Trans. Comput. Biol. Bioinform. (2023)

41. Vossekuil, B., Borum, R., Fein, R., Reddy, M.: Preventing targeted violence against judicial officials and courts. Ann. Am. Acad. Pol. Soc. Sci. **576**(1), 78–90 (2001)
42. Walther, S., McCoy, A.: Us extremism on telegram. Perspect. Terrorism **15**(2), 100–124 (2021)
43. Withers, K.L., Parrish, J.L., Terrell, S., Ellis, T.J.: The relationship between the "dark triad" personality traits and deviant behavior on social networking sites (2017)
44. Wulczyn, E., Thain, N., Dixon, L.: Ex machina: personal attacks seen at scale. In: Proceedings of the 26th International Conference on World Wide Web, pp. 1391–1399 (2017)