Insights into the Monitoring and Research Behind the Report

Eli Lewien

A. Towards a Telegram Monitoring Methodology and Tool Development

The Center's team in Myanmar began to focus resources on monitoring Telegram in January of this year when it became apparent the platform was increasingly being weaponized by pro-military elements, including to share harmful content. This involved the team designing a monitoring methodology and tools, including:

- *Defining the monitoring universe*. One of the most challenging aspects of monitoring Telegram is "defining the universe" or identifying channels and groups on the platform, and then which to monitor. This is a manual process that requires a lot of effort (unfortunately there's no Crowd Tangle for Telegram). In Myanmar, the monitoring team effectively used crowdsourcing and other investigative techniques to identify over 100 channels and groups sharing pro-military and pro-democracy content.
- *Determining a monitoring focus.* The monitoring team quickly discovered that monitoring 100 plus channels and groups, some of them that posts hundreds of times a day, was to feasible nor useful. This led the team to narrow its monitored universe to 10 of the 72 most followed and most active channels and groups that were sharing pro-military content. In most monitoring contexts resources are limited and tradeoffs must be made. In this situation the team chose the accounts most likely to be sharing the most harmful content as this was of most interest to stakeholders at the time.
- *Tool development*. To increase monitor ability to search and review content (instead of being limited through searches on phone interfaces), a set of tools were developed to scrape and analyze relevant posts and archive them for future use, this includes conducting keyword searches for different forms of content like hate speech. This was a time-intensive exercise and worth the investment because the tools have been repackaged and made available to be deployed on other Carter Center missions (See III).
- *Creating a workflow and report structure*. Another post (or a book) could be written on the importance ensuring that a proper workflow is in place to guide all aspects of the monitoring process, including tool development, data collection, analysis, and writing, and that the workflow is designed with informing the intended report or product structure. In Myanmar, the workflow and resulting monitoring infrastructure has led all data and ongoing project work to be accessible in the cloud. This has made it possible for the team to work collaboratively, including on translations (which are time intensive), and to share feedback throughout the entire monitoring and writing processes.

The above steps, and tools developed on the way, can be utilized in other contexts the Carter Center works to begin monitoring and reporting on activity in Telegram (and other platforms), and were used to inform the research behind the most recent Telegram report.

B. Research Behind the Telegram Report

During the initial two months of Telegram monitoring the team determined that there was value in continued monitoring, but that there were also problematic actors and content on the platform that merited more focused research that could both be of interest to relevant stakeholders and possibly provide insight into the scope and scale of problems on Telegram in Myanmar, and lead to recommendations that could be used to advocate for change. The team met to determine some major questions it wanted to answer, including:

- How much and what kind of harmful content is being shared?
- Who are the most offending actors?
- What is the intention of pro-military actors?
- Is there coordination happening between actors on the platform?
- What content violates Telegram policies and what should be done?

Given the high volumes of constantly new content encountered during monitoring (upwards of 100,000 posts per month), the team decided to focus research efforts to a specific timeframe. The one-month period of March was selected for a variety of reasons, including that it would allow the team to monitor and better understand trends through the month while collecting post data and to develop research approaches for eventual data analysis.

C. Development of Content Analysis Frameworks in March

While the team waited through March to pull data to be used for the report on April 1, it continued to monitor activity on Telegram and developed two data analysis frameworks to be used during the analysis phase in April and May, this included the:

Harmful content analysis framework. This framework includes definitions of types of harmful content, guidance to monitors for how to manually review and code posts by one or more types of harmful content, and datasheets and coding systems designed for monitors to review and code high volumes of posts. During creation of this framework the team identified harmful content as content or approaches that target individuals or groups through hate speech, harassment, disinformation, or speech which could lead to violence against these individuals or groups, including doxxing.

Content trend analysis framework. This framework includes the five major "themes" or kinds of content monitors identified as being shared by pro-military accounts in Myanmar, developed from manual monitor review and analysis of high volumes of posts across and from individual groups and channels through February and March. This included posts that:

- Sought to build support for the military
- Create an online pro-military community
- Praise international supporters of the regime and condemn detractors
- Doxx pro-democracy supporters
- Blame the resistance for the ongoing conflict and problems across Myanmar

The harmful content framework was essential for guiding monitoring coding and gaining a better estimate of the volume and kinds of harmful content being shared. The content trend analysis framework was essential for eventually answering questions about the intention of pro-military actors on the platform and coordination between them.

D. Data Collection, Review, Analysis and Writing in April/May

There were a variety of steps taken to scrape analyze data and write about activity on Telegram through April and May. These are laid out below but note that each step of the process required a fair a significant amount of time to develop, as well as experimentation and trial and error to get processes to work.

- *March data scrape*. The team used the Telegram tools developed during monitoring on April 1 to scrape and archive all posts from the 72 channels and groups identified for monitoring. The project lead combined the posts into a single dataset and organized and then summarized findings into a table from which the conclusions about activity in the monitored universe could be drawn, but the dataset was far too large to manually review or analyze against developed frameworks.
- *Narrowing channel and group focus*. The team took two approaches to narrowing the volume of content it would focus on for further review. It first used a series of semi-automated and manual processes to select the 16 accounts based on their followings, activity levels, and generation of unique content (i.e., many of the accounts in the 72 often forwarded or reshared content from other accounts). Posts from these channels and groups were selected for further narrowing with a "harmful content" index developed for the project.
- *Utilizing a "harmful content" index*. The second approach to narrowing content for review was through development and use of an index of 51 keywords or phrases identified by the monitoring team as being frequently associated with posts containing disinformation, propaganda, hate speech, and speech that could lead to violence (primarily doxxing), and was designed with a focus on identifying the two latter types of content. The monitoring team drew from other similar indexes created by local organizations monitoring in Myanmar and added terms and phrases to the index that were considered relevant during content analysis. The scraping and analysis tool developed during monitoring was utilized to run the index and 2,433 posts were selected for manual review by my monitors.
- *Manual review and post coding with analysis frameworks.* The 2,433 posts selected for manual review were then organized into spreadsheets for monitor analysis. The spreadsheets contained the coding systems designed during framework development in March. Monitors were assigned a series of spreadsheets to review and code, and while doing so were also asked to flag posts they considered to represent the kinds of harmful content and themes so that they could be translated and incorporated in the report for context.
- *Identifying patterns of possible coordinated activity.* Several techniques were employed to try and better understand if the different channels and groups were coordinating to share the same or similar content and possibly run by the same entities. Among them, this included review of posts by using a Python script to determine frequency of the same

posts being shared, review of account handles for similar characteristics, review of the timing and frequency of posts, and extracting references to coordination made in posts by different channels and groups.

- *Analysis and writing.* After data was scraped, manually reviewed, and coded, the process of findings analysis and writing began. This was facilitated by the team meeting on multiple occasions to outline the report based on a list of what team members believed were the most important findings and evidence to support them. The process was not straight forward, iterative, and the report went through multiple reviews.

Key Takeaways/Lessons Learned

Like with most research and resulting products, especially in the nascent field of social media monitoring, arriving at this report was not a straightforward process. The whole process required trial and error, creativity, and the combination of tech and time-consuming manual data processing, all while keeping a focus on the utility and eventual impact of the work undertaken.

There were limitations to being able to answer the research questions asked at the outset, including for example "How much and what kind of harmful content is being shared?" and "Is there coordination happening between actors on the platform?

To the first question, high volumes of content being shared on the platform and resource and time constraints to reviewing, analyzing, and in some instances translating this content resulted in the team focusing on a subset of channels and groups and developing techniques and tools like the "harmful content" index to gain a better idea or estimate the volumes of different kinds of content on the platform versus arriving at a more exact figure. This underscores the importance of creating a research design and determining the accounts and dataset intended for review and methodology for getting useful answers. Also importantly, this underscored for the team the limitations of determining scale of problems on platforms without significant resources (and the assistance of platforms).

For the second research question, the inability to extract further channels and groups handles from existing channels and groups via Telegram's API led to the team using more manual approaches to identify channels and groups for monitoring, and possible coordination when more sophisticated approaches like social network analysis (SNA) could have been leveraged. This emphasizes the importance of companies making their data available and the continued exploration and development of tools that can further automate time consuming processes.

The "heavy lifts" in terms of time required were the manual analysis and review of content, and in some instances translation, that cannot be automated. This included the review of content to develop the analysis frameworks in March, the manual review and coding of posts pulled using automation, and finally the time it took to analyze and write up findings. Future efforts must actualize the "heavy lift" of monitoring and carefully balance being timely with the amount of work it takes. Positively, the initial "heavy lift" here has created a foundation for "quicker" future product.

Finally, additional tools can be developed to possibly to develop a more automated approach to identifying accounts (channels and groups) for monitoring, and to leverage social networking analysis (SNA), and natural language processing (NLP) to further automate some of the work done here!