

What happens when the US decides to designate a group as a terrorist organization? Insights from Mnemonic

Dia Kayyali
18 February 2022

As one of his last acts in office, the [Trump administration designated](#) the Houthis as a terrorist organization on January 10, 2021. When Biden took office, his administration [revoked that designation](#) effective February 16, 2021. They did so in recognition of the fact that “the designations could have a devastating impact on Yemenis’ access to basic commodities like food and fuel.”

In recent months the Houthis (officially known as Ansar Allah) have carried out [attacks on infrastructure](#) in the United Arab Emirates and Saudi Arabia. With urging from the UAE, the Biden administration has been re-considering the issue of the Houthis’ designation. While [news reports indicate](#) that he is “putting the plan on hold for now,” the issue isn’t gone. Based on our research of how this impacted human rights-related content last year, we want to flag the impacts of such a designation.

Though currently no single civil society organisation is capable of keeping track of how much human rights documentation is removed from social media platforms, we can get an idea of the scope through monitoring the availability of content that we have archived. By comparing what links still work against what we have archived, we are able to see what percentage of our collection is no longer available, as well as look at our stored copies to see what kind of content has been removed. We’ve seen [spikes in takedowns](#) before, often seemingly in response to new policies or new tools. This is, in fact, why we started working on content moderation and digital rights policy issues.

Our advocacy focuses on general content moderation issues, but there are a few categories of tools and policies that we are particularly concerned with. In our experience, and based on our monitoring, policies that address [moderation of content posted by organizations labelled “terrorist or extremist”](#), and the use of automation have the greatest impact on human rights documentation. Large swaths of what could constitute evidence are lost in a beat. These tools and policies are based, to a certain extent, on the US and UN’s lists of designated terrorist organizations. This question of adding the Houthis to one or both of those lists is a perfect example of how political decisions can immediately impact content moderation, which in turn can have irreparable damage to valuable documentation on all kinds of violations.



[Our investigations](#) are completely non-partisan and focus on human rights violations by all parties. We have focused on everything from attacks against journalists to attacks on hospitals. As with all of our investigations, much of that content comes from social media platforms. The source of the content, or who is posting it, is irrelevant once it has been verified.

The fact is that Houthi media channels post a significant amount of content that ends up in our collections, from claims about specific airstrikes to footage of bombings, which includes evidence of crimes. Many of our published investigations and data sets link to tweets that no longer exist publicly. If the US decides to re-designate the Houthis, even more content could come down, even if that content does not violate any laws or the platforms' terms of service.

Platforms have, by and large, stated publicly that they believe they are legally obligated to remove nearly all content related to organizations on the US's international terrorist organization list, or in relation to sanctions. Their interpretation of these obligations goes [far beyond the scope of existing legal precedent](#). To remove this content, they don't just employ human moderators. They also use the name of the organization, specified terms (like names of leaders or strongholds) and any logo or other branding as "qualifiers" for machine learning processes. They also use other training data, none of which is made public or done in collaboration with civil society. These processes, which are often particularly poorly designed, then accidentally sweep up content improperly. We have evidence of how poorly this can go in Arabic-speaking countries [from the Facebook papers](#). Some of the papers describe how content from Yemen is reviewed by content moderators in Morocco who do not speak the same dialect. Worse, those moderators are even labeling training data for machine learning, baking in their mistakes.

To try to get a sense of how changes in the Houthi designation could impact content, we reviewed our data on content availability for our Yemeni Archive from the time that the Department of State officially announced it was adding the Houthis to the list - as well as data from earlier, when that policy discussion was still happening. What we found was fascinating.

As of November 30 2020 we had 271,698 tweets archived. 79,812 of those were no longer available on Twitter, or 29% of the collection. What was interesting is that on October 1 2020 only 25,966 of that collection was unavailable, or 9% of the collection. In those two months, 53,846 tweets became unavailable. That's highly abnormal. By comparison, it's more common for us to see a few thousand tweets become available per month.



This massive takedown seems to have been a response to the potential designation of Houthis as a terrorist organization. As early as September 26 it [became public](#) that the Trump administration was [considering designating Houthis](#) as an international terrorist organization. We believe that Twitter removed accounts in anticipation of the change. We saw many Houthi accounts removed, but Twitter's efforts also swept up a number of non-Houthi twitter accounts. Some of the accounts not affiliated with Houthis spoke with Twitter about their suspensions and were reinstated. This resulted in 2,936 Tweets that Yemeni Archive has preserved being made publicly available again.

We reached out to Twitter for comment and they were able to share the following: "Our top priority is keeping people safe, and we remain vigilant about coordinated activity on our service. These accounts were suspended under our Platform Manipulation and Spam Policy, which prohibits a range of activity, including coordinated activity that attempts to artificially influence conversations through the use of multiple accounts, fake accounts and automation."

While we appreciate Twitter's response, we believe that conversation about designating Houthis made Twitter pay closer attention. It's important to point out that the types of activity prohibited under the [Platform Manipulation](#) policy are activities that arguably many expert institutional Twitter users have engaged in. For example, the policy prohibits "overlapping accounts", "mutually interacting accounts," and "posting identical or substantially similar Tweets or hashtags from multiple accounts you operate." These are policies that some NGOs have probably accidentally violated just by taking part in coordinated campaigning on Twitter. That's why we suspect that the conversation about designating Houthis made Twitter pay closer attention. However, even if it was only incidental, these mass takedowns are a good example of the potential impact new US, UN, or other legal designations could have on content from conflict areas.

The impact of this decision on the [already dire situation](#) in Yemen cannot be understated. Already, "more than five million people in Yemen are on the brink of famine, and 50,000 others were living in famine-like conditions" and COVID and other diseases are ravaging the population. At the same time hostilities between the Houthis and Saudi Arabia are escalating rapidly. To make matters worse, infrastructure attacks are [shutting down access to the Internet](#) for days at a time.

Human rights investigations are needed more than ever. In October, despite opposition from civil society organisations and several states, the UN Human Rights Council [voted against](#) continuing the mandate of the [Group of Eminent Experts on Yemen](#). In [a statement](#) they "call[ed] upon the international community to take specific initiatives at the international level or in third States, as appropriate, in pursuance of accountability."



The pursuit of accountability without a UN mandate puts an immense burden on civil society organisations like [Mwatana for Human Rights](#) and Yemen Archive. Without a UN mandate or other viable accountability mechanisms, social media platforms have no real reason to preserve content from Yemen that could be needed in courtrooms in the future. That's why the Yemeni Archive is rushing to preserve as much potential evidence as we can. But it's not enough. We know that there is content we won't find that could be essential documentation at risk of removal should this designation go through, and social media platforms fail to respond by preserving content.

Currently, platforms appear to hold on to content from 180 days to a year. In the time that content is removed to the time it is deleted permanently, there are no systems in place for working with civil society to ensure evidence is preserved and verified. Trying to get platforms to preserve all content that might be related to Yemen is a clear overreach. It's violative of individual privacy and security rights, amongst other issues, and these rights are at the heart of ethical and legal issues with preserving content. We don't want to create massive databases that could be vulnerable to government surveillance or other forms of misuse. As an organization, Mnemonic continues to work with other organizations, academics, and human rights defenders to try to solve this issue.

But this is a special situation that requires an immediate response. That's why we call on platforms to commit to preserving content and metadata from official Houthi channels, as well as "branded" Houthi content shared elsewhere, for two years, in the hope that a new UN mechanism or other accountability processes will be initiated in that time. We are comfortable doing this because this is not content posted by private individuals. It is not content from private accounts. There are still privacy concerns if privately identifiable information is included, and since this sometimes depict crowds or individuals that is possible. But weighed against the value of this content, this would be an appropriate temporary solution.

Ultimately, the push to remove so-called "terrorist and violent extremist content" is not going away. More and more governments and companies are signing up to the [Christchurch Call](#) and the [Global Internet Forum to Counter Terrorism](#). More companies are using automation to detect content, and lawmakers are introducing and passing [legislation to mandate rapid removal](#). We at Mnemonic have been considering how to respond to this paradigm for several years now. We will keep searching for solutions that balance accountability for war crimes against individual privacy and security. In this case, the balance is in favor of preservation.

