Received: 12 June 2023 | Accepted: 4 October 2023

DOI: 10.1002/poi3.374

RESEARCH ARTICLE

P&I Policy & Internet PSO WILEY

# "Highly nuanced policy is very difficult to apply at scale": Examining researcher account and content takedowns online

**Aaron Y. Zelin**[1,2]

[1]Department of Politics, Brandeis University, Waltham, Massachusetts, USA

[2]Counterterrorism and Intelligence Program, Washington Institute for Near East Policy, Washington, District of Columbia, USA

**Correspondence**

Aaron Y. Zelin, Department of Politics, Brandeis University, 415 South St., Waltham, MA, 02453-2728, USA.
Email: azelin@brandeis.edu

## Abstract

Since 2019, researchers examining, archiving, and collecting extremist and terrorist materials online have increasingly been taken offline. In part a consequence of the automation of content moderation by different technology companies and national governments calling for ever quicker takedowns. Based on an online survey of peers in the field, this research highlights that up to 60% of researchers surveyed have had either their accounts or content they have posted or stored online taken down from varying platforms. Beyond the quantitative data, this research also garnered qualitative answers about concerns individuals in the field had related to this problem set, namely, the lack of transparency on the part of the technology companies, hindering actual research and understanding of complicated and evolving issues related to different extremist and terrorist phenomena, undermining potential collaboration within the research field, and the potential of self-censorship online. An easy solution to this would be a whitelist, though there are inherent downsides related to this as well, especially between researchers at different levels in their careers, institutional affiliation or lack thereof, and inequalities between researchers from the West versus Global South. Either way, securitizing research in however form it evolves in the future will fundamentally hurt research.

**KEYWORDS**

censorship, cloud storage, extremism, messaging applications, research, social media

## INTRODUCTION

Since 2019, there has been greater evidence that there are broader consequences for researchers examining, archiving, and collecting extremist materials online. When social media companies began taking down extremist and terrorist content—most notably with the Islamic State first—it mainly focused on particular users within a network that were promoting such materials and ideas online (Berger & Perez, 2016). Later through the creation of a hash database that is now shared by the technology consortium the Global Internet Forum to Counterterrorism,[1] these platforms were able to target particular media items based solely on a digital signature or fingerprint.

With advancements in algorithms and artificial intelligence (AI), these tools have become far more hard-hitting, resulting in less discriminant practices in who particular platforms ban. Moreover, governments, who are worried about the spread of extremist and terrorist materials online, have begun enacting or pushing for laws that force companies to take down accounts and content in a swift and limited timeframe, usually within 24 h. Brian Fishman, former head of Dangerous Organizations at Facebook, highlighted in an interview with the Select Committee to Investigate the January 6th Attack on the US Capital that "highly nuanced policy is very difficult to apply at scale." (U.S. House of Representatives, 2022).

As a consequence, in recent years, you will see messages like this on Twitter: "@TwitterSupport, please reinstate @X's account. This person is a credible researcher working on X issue." Denoting fellow researchers pleading with a platform like Twitter to reinstate someone's account, even though it lacks transparency which rules were broken. This scenario and resulting ritual has played out among the research community many times now in recent years. To better understand the ramifications of these applied policies, this paper seeks to provide a more detailed understanding of how this has affected those that work to study and counter various extremist and terrorist actors in the real world and online.

It is important to highlight that account and content takedown is not the largest problem within academia or the world. Nor is it likely to create existential consequences for individuals in the West, at least, in the short-term. Though there are worries in nondemocratic states how this could create security dilemmas and undermine serious research for those that could provide insights that maybe someone from outside the country might not be able to grasp or understand since they are not from there. However, as the pace of restrictions related to freedom has eroded all over the world in the past decade and technology companies have been another avenue for repressive regimes to suppress anything critical or deemed sensitive to their rule even if it is legitimate academic work that sheds light on extremist phenomena within their own country settings. Beyond stifling academic research, this trend inhibits any accountability on the part of technology platforms, which has longer-term implications for those conducting primary source research online.

## BACKGROUND

The first sign that researchers would get enveloped in a takedown scheme occurred in late November 2019 when dozens, if not more, researchers were banned from Telegram when Europol in coordination with Telegram had a day of action against Islamic State propaganda (European Union Agency for Law Enforcement Cooperation, 2019). A steady stream of researchers have since been taken down from other platforms on and off again in subsequent years. Since then there has been nothing as large-scale, but rather random cases among individuals over time.

Unlike in the case of Telegram, which established a whitelist (a list of people and their accounts that is considered to be acceptable or trustworthy), no other larger companies have implemented such policies. When asking a former senior-level official at Facebook about this after the company blocked access to my WhatsApp account in early May 2021, the individual responded that "we are trying to get enforcement on WhatsApp better, but it will be blunt."[2] When this author followed up specifically about a whitelist, the individual would not directly respond.

Therefore, any redress is an ad-hoc process that does not follow any particular protocols. Furthermore, when researchers have been allowed back on WhatsApp or other platforms, many of the extremist groups they had been following remained online, illustrating gaps and issues with the algorithmic system itself. This has extended beyond the social media arena to also include other messaging applications and cloud-based storage providers.

## Literature

It is important to highlight that compared to 2015, it is much more difficult for extremists and terrorists to operate on the biggest online platforms even if the different platforms do not have a 100% success rate in takedowns. According to various research products from multiple researchers in the field, deplatforming and takedowns related to individuals, influencers, organizations, and content from the jihadi movement and the extreme far-right when tested shows that after concerted and sustained efforts, there are diminishing returns for those trying to radicalize individuals online to their cause (Mirrlees, 2021; Thomas & Wahedi, 2023).[3] However, in the case of the jihadi movement, Conway, Khawaja, Lakhani, Reffin, Robertson, and Weir argue that there is a difference in the level of deplatforming when comparing the Islamic State to other jihadi organizations (Conway et al., 2019). Highlighting discrepancies in how different companies might police their platforms.

These takedowns have since led many extremist and terrorist movements to smaller platforms, which although is not ideal is a tradeoff considering there is less exposure to the broader public. As Danny Klinenberg's research looking at the migration of far-right extremists to Bitchute, "deplatforming works in decreasing a content creator's overall views and revenue" after they have been taken off of YouTube (Klinenberg, 2023). Similarly, Richard Rogers shows when exploring "canceled extreme celebrities" that moving to the encrypted application Telegram "does not substitute" for the prior audience on mainstream platforms Facebook, Instagram, Twitter, and/or YouTube due to "audiences on the new platforms hav[ing] thinned." (Rogers, 2020).

Relatedly, as takedowns hit a broader extremist community online, according to Elizabeth Pearson, when exploring IS supporters online ("the baqiya family"),[4] responses to suspension enforce gender "norms that benefit the group: the shaming of men into battle and policing of women into modesty" (Pearson, 2018). Other researchers such as Myagkov, Shchekotin, Chudinov, and Goiko, have shown in both the jihadi community and extreme far-right on the Russian social media platform VKontakte when taken down they most often respond with "the mimicry of ideologically neutral content" as a way to try and overcome platform sanction in the future (Myagkov et al., 2020). Similarly, McMinimy, Winkler, Lokmanoglu, and Almahmoud have shown a shift by IS in its release of media products in the 6 months before the aforementioned November 2019 Europol takedown and 6 months after; particularly, with its "agenda-setting strategy" and "alter[ing] frames related to images showing opposing militaries and military outcome" (McMinimy et al., 2023).

It should be clarified, however, that this particular research does not specifically have to do with extremists and terrorists being taken offline. Rather, the second-order effects and how that has since hindered the research community. Thus, while the above literature is

important to the field, it is only relevant to this paper insofar as it is background to why researchers have since been caught up in the takedown dragnet. This research is not arguing against content moderation of extremists and terrorists in of itself. Instead, it seeks to highlight that as a consequence of these takedowns they have now begun to affect those that are doing legitimate research on these topics. Thus, a mechanism, such as a whitelist, even if it is not a perfect solution is better than the status quo.

## Government legal action

Beyond the policies implemented by different technology companies through their terms of service, this issue has been made even more complicated by policies pursued by governments around the world, most notably in the West (Council of Europe, 2017; Ramati, 2020). There is no doubt that many politicians are genuinely concerned about vulnerable individuals being exposed to extremist or terrorist materials. Yet the solutions that have been proposed or enacted to do so can actually act counter to the interests of those seeking to study these complicated issues. This in turn hurts a government's ability to combat those actually involved in extremist or terrorist activity since researchers can provide deep insights on a particular phenomenon, individual, or group that governments do not have the time to do when faced with many competing priorities.

Therefore, even if there is greater demand for speed and action when removing such content, there are potential third-order consequences that can inhibit the ultimate goal of understanding and deterring extremist and terrorist movements that should be considered in the creation of these processes. For example, Caleb Weiss, editor of *The Long War Journal*, noted that YouTube is "a large red flag for not only researchers, but all sorts of content creators. I'm on YouTube a lot, but one of the main things with YouTube right now is, they're using bots to flag videos for milliseconds of content that's questionable and accounts are getting strikes and users are making claims against that and YouTube has flat-out refused to do something about it."[5] This matters because it also takes down researcher work, news stories, and human rights archives. Regarding the latter, YouTube has removed videos showing atrocities committed by the Assad regime during the Syrian uprising and war, which takes away a potential key tool for evidence if the architects of its war crimes are ever brought in front of an international criminal court (Browne, 2017).

This issue isn't necessarily specific to extremist and terrorism-related issues too. For instance, governments rightfully want child pornography to be taken down right away. Yet, that is a far more black and white issue when determining what should be taken offline. Extremist and terrorist materials, however, are a bit more complicated given their political nature and the variety of views held by different individuals on what is deemed as extremist or terrorist content, especially beyond the West. Moreover, we lack a universal definition of what even a terrorist or extremist is—it often depends on different state actors.

Furthermore, even within the West itself, as researcher J.M. Berger has noted, extremists can become the mainstream, like the Nazi regime in Germany or the American slave regime. It is not always a fringe phenomenon (Berger, 2018). There are already a plethora of cases in Venezuela, Vietnam, Russia, Iran, Turkey, Thailand, India, Kenya, and Brazil whereby repressive governments have tried to quell speech deemed inappropriate even though it has nothing to do with extremism or terrorism (Salina, 2023). Plus, legal scholars of content moderation, such as Stanford Law School's evelyn douek, suggest that "giving platforms undeserved legitimacy dividends, allowing them to wrap themselves in the language of [International Human Rights Law] IHRL even as what is required by that body of norms remains indeterminate and contested." (Douek, 2021).

For example, the European Union enacted a regulation in April 2021 that calls for terrorist content to be takedown from different technology platforms within an hour of when it is identified. Meaning within 27 country jurisdictions, any technology company, whether as large as Facebook or as small as a new start-up, who might not even have a staff of more than a few people, must quickly take down content. Such rapid measures make it impossible, due to the sheer number of users depending on the particular technology platform, to actually evaluate, using human insight, if an account or piece of content is truly extremist or terrorist in some manner. As Rachel Griffin of Sciences Po Law School notes, "EU law incentivises the deletion of various broadly-defined types of illegal content, which is also likely to suppress large amounts of legal and harmless content. Evidence of how social media platforms moderate content suggests that this over-enforcement will disproportionately suppress marginalized users and nonmainstream viewpoints." (Griffin, 2022). Thus, automated algorithms and now what is perceived to be more sophisticated AI are tasked with this complex process.

In other spheres, we are already seeing a number of negative consequences of using AI to solve problems. For instance, in late May 2023, after firing its helpline human staff and replacing it with AI, the National Eating Disorder Association (NEDA) had to take down its AI chatbot called Tessa, due to the harmful responses it provided to potential users (Xiang, 2023). Plus, in other areas, AI such as ChatGPT or Bing Chat have been known to lie or make things up when individuals ask it for answers to specific questions (Stokel-Walker, 2023; Verma & Oremus, 2023). Put together, there are already limitations with more simple tasks like a generative AI chat or a more complicated issue such as an AI helpline. Grafting such AI solutions onto divisive, sometimes politicized, issues such as taking down online extremist and terrorist materials increases the likelihood of future failings. This also does not address how all of this relates to the rule of law as Macdonald, Correia, and Watkin have argued (Macdonald et al., 2019).

It should be highlighted though that there are downsides to only relying upon humans for content moderation too. For one, the scale of content is impossible for a staff of humans to cover on a minute, if not hourly or daily basis. More importantly, with extremist and terrorist materials there are mental health risks with constantly binging such horrific messaging and content.[6] Even more experienced researchers in the field are affected by this too, including myself as shown by Pearson, Whittaker, Baaken, Zeiger, Atamuradova, and Conway (Pearson et al., 2023). Thus, it is understandable that automated solutions appear to make the most and easiest sense. As a result, creating some form of whitelist procedure to overcome the large content moderation dragnet that researchers are increasingly being caught up in would be a measured solution to a complicated and sensitive problem.

## The securitization of research

In many ways, this research is a follow-on from this author's auto-ethnography related to the website Jihadology, which was changed after efforts by the UK government to have a password-protection functionality added due to alleged claims that since it had been an open website before April 2019, jihadis were exploiting it for their own ends (Zelin, 2021). Yet "three separate studies at varying times, on varying [social media] platforms [Telegram and Twitter], and examining different languages [Arabic and English] suggest that the concern over [Jihadology] within the jihadi online ecosystem is misplaced: more conjecture than based on empirical data." (Zelin, 2021, p. 235). While in some Western governments policies may be more transparent and citizens could influence policy, there are not many checks and balances or ways to affect much change from the outside at technology companies. Plus, governments have made this

effort that much more difficult without creating any serious regulatory regimes to govern these platforms. Even some companies like Facebook have called for greater regulation (Hutchinson, 2021). This is why this new study is so crucial so as to give an early look into how researchers are currently being affected by technology company policies and not on random suppositions.

Thus, this study wanted to try and have a better idea of how this issue is affecting researchers based on semistructured interviews with others in the field who this author knows have dealt with these problems as well as this author's own experiences. As Christian von Soest recently wrote in *Perspectives on Politics*, "qualitative expert interviews have an important role to play. This is particularly true for the analysis of complex decision-making processes, where there is a dearth of data." (von Soest, 2022). Therefore, this study combines one-on-one interviews with an online survey that was conducted from December 1 to December 25, 2022. This way, this study could garner both qualitative and quantitative data to try and gain some baseline statistics and understandung on this issue too. Those individuals that are quoted later in this paper as part of the interviews and survey have given this author approval to use their names on the record.

It is important to also point out that this is not only a problem for individuals conducting research on extremism or terrorism-related issues. It has also come up in the ad-transparency and misinformation space. For example, in August 2021, academics from New York University were banned from Facebook. According to Laura Edelson, who was the principal investigator on the study they were conducting, "Facebook is silencing us because our work often calls attention to problems on its platform. Worst of all, Facebook is using user privacy, a core belief that we have always put first in our work, as a pretext for doing this. If this episode demonstrates anything it is that Facebook should not have veto power over who is allowed to study them." (Vincent, 2021). More recently, Instagram banned and then unbanned The Real Facebook Oversight Board after it criticized Facebook's parent company Meta for its recent threat to cut off news access to news links in Canada and California (DeGeurin, 2023). Even though they were unbanned it illustrates the lack of transparency on the part of platforms and the potential to go after critics. These problems are a harbinger of likely greater issues that will affect researchers focusing on sensitive topics in the years to come if not resolved in an amicable manner.

## Ethical consideration

Before getting into some of the findings, there are important ethical considerations that should be brought up related to the conduct of researchers when working in this online space and how it relates to this study in particular. It is understandable that researchers would want to have some level of operational and personal security. Increasingly, according to the Association of Internet Researchers, there are death threats, potential real-world retaliation, and doxxing (Franzke et al., 2020). Thus, it is understandable that someone if they are only interested in collecting data and ideological materials would create a pseudonym to protect oneself. However, it is different if one is establishing a pseudonym and pretending to be a member of an extremist organization to regularly interact with subjects for research. It is hardly surprising that this latter type of account would be taken down due to its perception as being an extremist.

Thus, this study is not focused on the latter scenarios considering the potential inherent ethical issues with such a methodology, but also it would not truly provide us a greater understanding for those ethically conducting research in online spaces and how one might be affected in relation to unnecessary takedowns.

## Methodology

In addition to interviewing individuals in the field one-on-one over Zoom about this issue, a survey was also created to get a better sense of how researchers focusing on terrorism and extremism have been impacted by social media, messaging applications, and storage site companies taking down accounts and content that has allegedly violated their terms of service. The survey includes 18 prompts to respond to. Some were to confirm the identity is actually a real person or works in the field (name, institutional email address, employer), others were for demographic details (what field and career level?), and finally substantive questions related to the issue in both a qualitative and quantitative manner.[7] The next section will mainly focus on the quantitative data, while the qualitative responses from the survey are also included in the later analysis related to the section with those this author interviewed one-on-one since many of those questions were similar. A survey with quantitative data among peers in the field provides a deeper understanding of how widespread or limited the issue is. To make sure the survey was seen by as many people as possible, the survey was shared via this author's Twitter account and the website Jihadology on December 1, 2022, and promoted it online until December 25, 2022.[8] Over that time, there were 51 individuals that filled out the survey.

Of the 51 responses to the survey, there were seven among them that were extremists attempting to troll the results, mainly of the anti-Semitic variety as well as a couple of Islamic State supporters. For example, one of these extremist individuals named itself as the "KikeKiller," had its first account takedown on Hitler's birth date, and said the number of accounts that had been taken down was 1488.[9] This was why there were control questions so as to sift out any attempt to undermine the survey. Despite excluding such responses, there was a serious cohort of 44 individuals that can be drawn upon to get a sense of how some in the field have experienced the issue of account or content takedowns from social media, messaging applications, and cloud storage providers. Of course, it would have been ideal to have even more responses to the survey, but there are also a limited number of individuals that not only study extremism and terrorism, utilize primary sources in their research, and also have a public-facing presence online whereby they might also be swept up in takedown efforts by technology companies.

## Limitations

There are certain limitations to this study. For one, content policy norms change all of the time. For example, YouTube after cracking down on American election denialism about the 2020 election after August 2021, reversed course in June 2023, now allowing such content on its platform (Fischer, 2023). Moreover, following the takeover of Twitter by Elon Musk in October 2022, he dissolved its Trust and Safety Council, fired many staff on the trust and safety team and thus far as of June 2023, two heads of trust and safety at the company resigned as a consequence of Musk gutting large parts of the work that had been done over the prior 8 years (Alba & Wagner, 2023; NPR, 2022; PBS News Hour, 2023). Therefore, at least in the case of Twitter, it is plausible that it might be less likely for a researcher to be taken down now as it had in the past since its policing on the platform is less robust than it had been in the past. That being said, this study provides a useful snapshot of an issue that many researchers feel they have been caught up in due to the nature of their work. The following section will provide some quantitative data on those that responded to the survey about researcher takedowns.

## QUANTITATIVE FINDINGS

To get a better idea of the 44 individuals that responded to the survey, here are some demographic details provided. Most of the individuals that responded to the survey were in academia, likely in their mid-20s to late 40s, and primarily focused on jihadism in their research, but have also diversified in recent years to other topics like the extreme far-right.

*Field (some individuals choose multiple options)*

- Academia: 65.9%
- Journalism: 15.9%
- Nongovernmental Organization: 11.4%
- Think tank: 9.1%
- Government: 6.8%
- Independent: 4.6%

*Career-level (some individuals choose multiple options)*

- Graduate school (MA/PhD): 36.4%
- Early career (1–10 years): 31.9%
- Mid-career (10–30 years): 27.3%
- Late career (30–45+ years): 6.8%
- Post-doc: 4.8%
- Undergraduate school: 2.3%
- Retired individual: 2.3%

*Areas of focus (multiple choice)*

- Jihadism: 90.9%
- Far-right: 47.8%
- Incel/male supremacist: 27.3%
- Far-left: 22.7%
- Hindu: 4.5%
- Jewish: 4.5%
- Buddhist: 2.3%
- Ethnic: 2.3%

From this particular demographic of people that responded to the survey, more than half of them have been affected by the algorithms and AI that different platforms have implemented to takedown accounts on social media or content from cloud storage services. This would suggest that more often than not, individuals that work on extremist primary sources are likely to be affected by the scalability issue mentioned earlier. Therefore, research becomes a more arduous task for those working on sensitive topics, even when the understanding of such movements are of importance to the public interest -- whether to the general public or those working on these issues in government to safeguard the citizenry.

*Have you ever had one of your accounts or content you stored in the cloud taken down?*

- Yes: 56.8%
- No: 43.2%.

The top platform that individuals have been taken down from is Telegram, due to the aforementioned November 2019 wide-spread action taken against IS's online network there. Yet, when this occurred, Telegram was responsive to those that did work on Islamic State online networks, media, and ideology. If one emailed Telegram's recovery email address (recover@telegram.org) from one's own institutional email account and provided one's phone number attached to one's Telegram account, a short biography of oneself, and linked to proof of this online, Telegram would reinstate one's account. In this author's case, it took Telegram 6 h to respond and reinstate the account. Based on conversations with dozens of colleagues in the field, they had a similar experience. Likewise, as part of this policy, Telegram has established a whitelist for those using the platform for legitimate research purposes. Therefore, repeat takedowns have not been a problem. This proactive way of dealing with this dilemma highlights a potential smart way forward for other platforms, which do not have a standardized redress mechanism or for a way forward that the same situation does not repeat for a researcher.

*Of those who had been taken down, which platforms? (multiple choice)*

- Telegram: 31.8%
- YouTube: 18.2%
- Facebook: 15.9%
- Google Drive: 15.9%
- Twitter: 13.6%
- DropBox: 11.4%
- WhatsApp: 11.4%
- Instagram: 4.5%
- Discord: 2.3%
- Hoop: 2.3%
- Mastodon: 2.3%
- Pastebin: 2.3%
- RocketChat: 2.3%
- Vimeo: 2.3%
- VKontakte: 2.3%

The next section will dive deeper into broader questions related to these platforms, researchers interactions (or lack thereof) with them in how to resolve issues, consequences of current technology platform policies and how it affects ones research, and looking forward at how individuals in the field see how this issue is evolving, and what could alleviate the current problems that occur for those that have thus far been affected by these policies or others that worry about the broader consequences for academic freedoms and general moves toward censorial policies that undermine serious work on sensitive topics like extremist and terrorist groups and movements.

## QUALITATIVE DATA

Before analyzing how others have dealt with this problem, it is worthwhile to share an anecdote from my own experience with regard to a WhatsApp account previously used for research. It was taken down four times within a few months in 2021 for following jihadi groups. There was no interaction on my part with anyone, just observing and collecting potential data that could be used for research. The first time my account came back online it

happened within a couple of hours, the second within half a day, the third time 3 weeks later, and the fourth time 3 months later.

When my account returned from the third WhatsApp ban after a 3-week penalty, 24 active jihadi accounts were still operating that were previously being followed before that ban from groups like al-Qaeda in the Arabian Peninsula, al-Shabab, the Taliban, and Hayat Tahrir al-Sham.[10] The continued presence of these groups on WhatsApp illustrates the limitations of the current take down process. After that fourth and final ban, I decided to delete all the groups that were being followed since it was not worth it anymore. This all happened while there had been contact with senior-level individuals that worked on these issues at WhatsApp and Facebook. Imagine this happening to younger or independent researchers? It is likely they might never get their account back.

This was the case for Riccardo Valle, who focuses on jihadism in the Afghanistan and Pakistan region, who at the time of his accounts being taken down was not affiliated with any institution, but now is the Head of Research at The Khorasan Diary. In 2021, his Facebook and Instagram accounts were taken down. He said that he "tried to recover [his Facebook account] for one month, but they never wrote [him] back."[11] Likewise, Meili Criezis, a PhD candidate in the Department of Justice, Law, and Criminology at American University, noted in the survey that "WhatsApp has been terrible, so I just gave up on that."[12] It also affects more experienced researchers and journalists in the field like J.M. Berger, who is currently a Research Fellow with VOX-Pol, and said in the survey that YouTube took down his content six times even though there was "no terrorism intent" in his videos.[13] As a result, Berger "removed the content as it wasn't worth fighting over."[14] Again, signaling the limitations of the scalability issue that technology companies implement.

These issues as they become potentially broader could disincentivize individuals from even doing the work to better understand far-right extremist movements as Anna Meier, Assistant Professor in the School of Politics and International Relations at the University of Nottingham, notes, "I worry about younger researchers wanting to study the far right (and who have the technical know-how regarding newer online platforms that many older researchers do not) being turned off because of insurmountable legal hurdles. Our ethics review board at my institution is worried about students being arrested for possessing 'terrorist' content, and they have little information about how the law actually works in this space. I've done what I can, but bureaucratic processes are sticky and hard to change."[15] This thus undermines the ability to track and do research. Also highlighting the real consequences laws can have on research freedom.

Criezis further highlighted the downsides of the current policy trajectory of technology companies, "If we are unable to closely track extremists and collect their primary sources, I'm afraid this will obscure our ability to understand these groups/individuals/ideologies on a deeper level… Just because their accounts are deleted doesn't mean that they're gone and as we've seen with IS supporters in particular, they have fully adapted to deplatforming efforts." Moreover, Meier explained that, "the push to remove content rapidly, especially in Europe, gives the impression that governments are doing a lot, when actually they're not helping address the spread of far-right material among extremists and are making things more difficult for those working on counter-extremism."[16]

These issues don't only pertain to social media or encrypted messaging applications. These setbacks have also occurred on cloud-based storage websites. It is only likely to get more difficult for researchers due to a new policy Google Drive announced in mid-December 2021: "Google has announced a new policy for cloud storage service Drive, which will soon begin to restrict access to files deemed to be in violation of the company's policies." (Kahili, 2021). And while Google "explained in [its December 2021] blog post, there is a system to request a review of a decision if someone feels a file has been restricted unfairly, but it's

unclear how the process will be handled on Google's end and how long it might take." (Kahili, 2021).

Yet, according to Caleb Weiss, who is also a senior analyst at the Bridgeway Foundation, already in May 2021, "Google deleted [his] entire main account and everything associated with it - including all of [his] data sheets for work."[17] Even worse, Michael Loadenthal, Founder and Executive Director of the Prosecution Project, which tracks and provides an analysis of felony criminal cases involving illegal political violence occurring in the United States since 1990, had his entire Google Drive for the project "wiped [of] storage and [Google] refused to answer any emails." As a consequence, "nothing was recovered … it set our team back 5 months and caused some folks to quit the project in frustration."[18]

Similarly, Daniele Garofalo an independent security consultant, reiterated the time cost of takedowns and how it has affected him in doing his work publicly: "every time I post some content [online] I am terrified that [social media companies] will block me. It's stressful and frustrating, because I can't afford to lose all my research, my material, my contacts every time, as well as losing time to start over."[19] This has also occurred on other larger platforms like Dropbox. In October 2021, Elisabeth Kendall, who is the Mistress of Girton College at Cambridge University and does research on Yemen, complained that her account "was suddenly 'disabled' with no warning, no explanation, no communication, and no response to [her] query."[20]

This is why, Weiss, alongside his colleagues at Bridgeway Foundation have shifted toward Tresorit, a cloud storage service with end-to-end encryption.[21] However, there are downsides to migrating from larger more-known platforms, according to Weiss, "Tresorit is not as user-friendly as Google Drive or Google in general, [which are] so easy to do collaboration… with these other third-party apps, it's way more counterintuitive."[22]

Therefore, these trends related to content takedowns will affect collaboration on the larger platforms going forward. This is because most platforms cannot share files large in size nor do they have the ease of use and ability to share content. It also undermines community among researchers. The anonymous Twitter account @Switch_d who works in the information security industry worries about new joint efforts: "researchers [are] moving off [of] Twitter entirely, or housing their content elsewhere only accessible by subscription. Twitter has been a great medium for sharing this type of work. Collaborations can develop in a moment, and real and meaningful things are accomplished as a result, a great many of which benefit the Twitter community."[23]

It is important to remember that this does not only affect academics either, but also people working in fields related to human rights abuses, journalists and documentarians, law enforcement and prosecution/defense teams, as well as other government officials. For instance, on April 10, 2022, a documentarian reached out to this author about using some clips from some Islamic State videos for a forthcoming documentary the individual and their team were working on. This author sent the videos to them at 4:05 p.m. and noted in the email to "please download ASAP before Google takes them down." When this author woke up the next day, 11 emails had been received from Google saying that "One of your files violates Google Drive's Terms of Service," which were sent between 11:52 p.m. on April 10 through 8:41 a.m. on April 11. In the past, this has also happened even if the name of the file doesn't have anything to do with the group due to the hash database.

Likewise, in mid-May 2022, according to Seamus Hughes, Senior Research Faculty and Policy Associate at University of Nebraska's National Counterterrorism, Innovation, Technology, and Education Center, he had a prosecutor in the United States "ask [him] for a copy of the [Buffalo] manifesto. And both times [he] couldn't share it on Google Drive."[24] This is in reference to the May 14, 2022, racist shooting at a predominantly black supermarket in Buffalo, New York perpetrated by Payton S. Gendron, which killed 10 people

and wounded three (The Buffalo News, 2022).[25] Thus, Hughes incredulously said "so, you have government officials asking for stuff and you can't actually give it to them."[26] This is because the file was too large to share another way. It also makes researchers potentially perpetually online, as Hughes continued: "You gotta grab the manifesto within the first 10 min or else you're never gonna get it and that's frustrating if you want to have a weekend plan, you know?"[27]

The logical conclusion of all of these issues is that researchers will begin to self-censor more. The aforementioned Riccardo Valle confided that "we fear that if we publish something, the day after our account will be deleted. So for my part, for instance, I stop frequently publishing or talking about the Islamic State propaganda in Afghanistan because I fear that [Twitter] could suspend my account and I also stopped sharing any pictures of anything related to extremism, even with covering logos; I just don't publish anything. If we all set limits, our research on these platforms– eventually all our research in general will be hampered."[28] Such consequences further hinders the enterprise of educating others about different extremist or terrorist organizations, while also undermining greater chances of collaboration between different researchers in the field who may realize they have similar research interests.

## Future research

It would be worthwhile for someone more embedded within the far-right or other extremist research spaces to conduct a similar survey and see how the results are similar or different. This is because jihadis were the first extremist-types that the different technology companies went after and thus it is more likely that those working on issues related to the jihadi movement have been more affected by takedowns than those that research other extremist trends. Further, unlike the jihadis, which are more organized as centralized groups, the far-right scene is much more diffuse. Plus, due to a lot of the far-right originating in the United States, they are likely afforded more leeway due to free speech norms and potential immediate political consequences of going after individuals or movements that are adjacent or touching parts of a mainstream political party. This way, there could be a more comparable understanding if certain networks of individuals focusing on different extremist ideologies might be affected by taken downs more than others.

# POTENTIAL SOLUTIONS

During the one-on-one interviews and with the online survey, individuals working on extremism and terrorism-related issues were asked, "beyond the government regulating technology platforms, is there anything that can be done in the meantime that these companies can do to help alleviate this problem and stop it from getting worse?" Diana Bolsinger who is a Lecturer and the Graduate Director of the Intelligence and National Security Studies Master of Science program at the University of Texas El Paso, said that it is important to "foster awareness among platform management of the importance of academic research to resolving the problem; offer specific guidelines for differentiating academic versus true terrorist/supporter accounts."[29]

From a government perspective, Alyssa Potter of the New Jersey Office of Homeland Security, suggested that "the content is so important to our work and needs to be stored and shared appropriately for research purposes."[30] This would fall in line more with the Jihadology model that the UK government pursued. Putting the onus on researchers in

many circumstances even though they likely have the least amount of resources to pursue such a path.

In contrast, Gina Vale who is a Lecturer of Criminology at the University of Southampton, shifts the focus onto platforms and their duty to assist in the research enterprise: "access to archived data from platforms needs to be made available to researchers - with safeguards in place for secure handling and storage of data."[31] An out-of-the-box-solution was offered by the aforementioned @Switch_d, who put forward the idea that "perhaps some sort of ranking system can be applied to the reporting [of] accounts to help assess their legitimacy. This might include account creation date, level of activity, etc."[32] Thereby giving preference to those that have a history of good use and thus likely not being someone trying to abuse the terms of service for more nefarious reasons.

The most common response among those this author spoke with one-on-one and in response to the survey simply called for technology companies to enact a whitelist similar to what Telegram has already done. This seems like the most logical and easiest path to stop having researchers accounts being taken down that focus on extremist and terrorism issues. There is also already a potential infrastructure that could be used on an institutional level to do this: the aforementioned GIFCT. As noted in the introduction, the GIFCT currently has a hash-sharing database related to extremist and terrorist content that all 26 member companies as of June 2023 have access to. An anonymized shared whitelist directory that these companies have access to could cut down on takedowns against legitimate researchers. There could also be an application process for independent researchers to gain access to this directory as well. Though, as the aforementioned Meili Criezis explained, "I know that system would still favor those who are fortunate/privileged to have institutional affiliation but it's the only solution that comes to mind at the moment and, as with anything, it isn't without potential issues or risks. This problem really highlights inequalities and power imbalances in the research community as a whole!!"[33]

Even if such a policy was enacted it still wouldn't include every single technology company which extremists and terrorists exploit. Plus, the GIFCT, while doing events and conferences with stakeholders outside of the West, the member companies appear to all be based in the West, thus discluding the potential to resolve this dilemma of researcher takedowns on platforms that are from the Global South. Beyond those limitations, there would also be other potential concerns about providing data to a technology company or an NGO like GIFCT for a whitelist. Criezis also said that when she gave her information to Telegram to return to the platform, she still felt like it was a "risk even revealing [her] username and fake number" yet she sighed and said "however, there is no other option."[34]

On top of this, fundamentally institutionalizing so many aspects of research and in many ways securitizing it will likely also undermine creativity in the field. Would something like Jihadology still be possible nowadays? What might be lost in the future if the current trends get worse? What better ways can someone new to the field come up with for disseminating primary sources based on new technologies that no one is even thinking about now? Would that individuals' curiosities be thwarted by policies that do not actually fundamentally get at the reasons why individuals become enmeshed in extremist milieus online?

It is important to remember that social media, messaging applications, and cloud storage sites are tools. A means to an end. Not the actual motivating factor behind it. Of course, it is important to safeguard against large-scale dissemination on the biggest platforms such as what happened with the Islamic State on Twitter. Yet it is also important to think about what the eventual price might be if technology companies and pushed by countries in Europe and authoritarian regimes continue to suppress anything that might be deemed problematic. Without resolving the issue of research related to sensitive topics like extremism and terrorism, the future of research will be hurt and those interested in conducting it could be penalized, and likely the broader knowledge base within the field writ-large could be eroded.

## ACKNOWLEDGMENTS

Thank you to Sarah Cahn, J.M. Berger, Devorah Margolin, and the anonymous reviewers for providing feedback on earlier drafts of this paper.

## ORCID

*Aaron Y. Zelin* http://orcid.org/0009-0002-6130-0229

## ENDNOTES

[1] Explained here: https://gifct.org/?faqs=what-is-the-hash-sharing-database.

[2] Email conversation with a former senior-level Facebook official when my WhatsApp account had been taken down, May 2, 2021.

[3] See Berger and Perez: http://extremism.gwu.edu/sites/g/files/zaxdzs5746/files/downloads/JMB%20Diminishing%20Returns.pdf

[4] A reference to IS's slogan baqiya wa tatamadad (remaining and expanding) related to its Caliphate project.

[5] Interview with Caleb Weiss, Zoom, June 16, 2022.

[6] For resources on this see: https://www.voxpol.eu/researcher-welfare-2-wellbeing/.

[7] You can find all of the questions at the form here: https://docs.google.com/forms/d/e/1FAIpQLSfrWSgt1-y05ZFQdCGDm512VW0JvMQ3fPXNebll8X3Hc6nDYw/viewform.

[8] See: https://twitter.com/azelin/status/1598345929043632129 and https://jihadology.net/2022/12/01/please-fill-out-the-researcher-online-account-and-content-takedowns-survey.

[9] According to the Antidefamation League, "1488 is a combination of two popular white supremacist numeric symbols. The first symbol is 14, which is shorthand for the "14 Words" slogan: "We must secure the existence of our people and a future for white children." The second is 88, which stands for 'Heil Hitler' (H being the 8th letter of the alphabet). Together, the numbers form a general endorsement of white supremacy and its beliefs." For more on this see: https://www.adl.org/resources/hate-symbol/1488.

[10] Screenshots taken of my WhatsApp account message stream on April 22, 2021.

[11] Interview with Riccardo Valle, Zoom, June 22, 2022.

[12] Meili Criezis response to "Researcher Online Account and Content Takedowns" survey.

[13] J.M. Berger response to "Researcher Online Account and Content Takedowns" survey.

[14] Ibid.

[15] Anna Meier response to "Researcher Online Account and Content Takedowns" survey.

[16] Ibid.

[17] See: https://twitter.com/caleb_weiss7/status/1390682599345987589

[18] Michael Loadenthal response to "Researcher Online Account and Content Takedowns" survey.

[19] Interview with Daniele Garofalo, Zoom, June 19, 2022.

[20] See: https://twitter.com/Dr_E_Kendall/status/1447959812461957128

[21] Interview with Caleb Weiss, Zoom, June 16, 2022.

[22] Ibid.

[23] Interview with @Switch_d, Twitter Direct Message, June 16, 2022.

[24] Interview with Seamus Hughes, Zoom, June 20, 2022.

[25] "Complete coverage: 10 killed, 3 wounded in mass shooting at Buffalo supermarket," *The Buffalo News*, November 1, 2022, https://buffalonews.com/news/local/complete-coverage-10-killed-3-wounded-in-mass-shooting-at-buffalo-supermarket/collection_e8c7df32-d402-11ec-9ebc-e39ca6890844.html.

[26] Interview with Hughes.

[27] Ibid.

[28] Interview with Riccardo Valle, Zoom, June 22, 2022.

[29] Diana Bolsinger response to "Researcher Online Account and Content Takedowns" survey.

[30] Alyssa Potter response to "Researcher Online Account and Content Takedowns" survey.

[31] Gina Vale response to "Researcher Online Account and Content Takedowns" survey.

[32] Interview with @Switch_d, Twitter Direct Message, June 16, 2022.

[33] Meili Criezis response to the survey.

[34] Ibid.

## REFERENCES


Alba D., & Wagner, K. (2023). Twitter cuts more staff overseeing global content moderation. *Bloomberg*. https://www.bloomberg.com/news/articles/2023-01-07/elon-musk-cuts-more-twitter-staff-overseeing-content-moderation

Berger, J. M. (2018). *Extremism*. MIT Press.

Berger, J. M., & Perez, H. (2016, February). The Islamic state's diminishing returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters. Program on Extremism. https://extremism.gwu.edu/sites/g/files/zaxdzs5746/files/downloads/JMB%20Diminishing%20Returns.pdf

Browne, M. (2017). YouTube Removes Videos Showing Atrocities in Syria *New York Times*, August 22 2017, https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html

Conway, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D. (2019). Disrupting daesh: Measuring takedown of online terrorist material and its impacts. *Studies in Conflict & Terrorism*, *42*, 1–2. https://doi.org/10.1080/1057610X.2018.1513984

Council of Europe. (2017, December 20). *Comparative study on blocking, filtering and take-down of illegal internet content*. Council of Europe. https://edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html

DeGeurin, M. (2023). Instagram bans, unbans facebook critic, and it won't say why. *Gizmodo*. https://gizmodo.com/instagram-bans-the-real-facebook-oversight-board-1850506902

Douek, E. (2021). The limits of international law in content moderation. *UC Irvine Journal of International, Transnational, and Comparative Law*, *6*, 37. https://scholarship.law.uci.edu/ucijil/vol6/iss1/4/

European Union Agency for Law Enforcement Cooperation. (2019, November 22). Referral Action Day against Islamic State online terrorist propaganda. *Europol*. https://www.europol.europa.eu/media-press/newsroom/news/referral-action-day-against-islamic-state-online-terrorist-propaganda

Fischer, S. (2023). Scoop: YouTube reverses misinformation policy to allow U.S. election denialism. *Axios*. https://www.axios.com/2023/06/02/us-election-fraud-youtube-policy

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & The Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3*.0. https://aoir.org/reports/ethics3.pdf

Griffin, R. (2022). The sanitised platform. *Journal of Intellectual Property*, *Information Technology and E-Commerce Law*, *13*(1). https://hal-sciencespo.archives-ouvertes.fr/hal-03586779

Hutchinson, A. (2021). Facebook launches new video ad series calling for improved regulations for social media. *Social Media Today*. https://www.socialmediatoday.com/news/facebook-launches-new-video-ad-series-calling-for-improved-regulations-for/608269

Kahili, J. (2021). Google drive could soon start locking your files. *Tech Radar*. https://www.techradar.com/news/google-drive-could-soon-start-locking-your-personal-file

Klinenberg, D. (2023). Does deplatforming work? *Journal of Conflict Resolution*. Advance online publication. https://doi.org/10.1177/00220027231188909

Macdonald, S., Correia, S. G., & Watkin, A.-L. (2019). Regulating terrorist content on social media: Automation and the rule of law. *International Journal of Law in Context*, *15*(2). https://www.cambridge.org/core/journals/international-journal-of-law-in-context/article/regulating-terrorist-content-on-social-media-automation-and-the-rule-of-law

McMinimy, K., Winkler, C. K., Lokmanoglu, A. D., & Almahmoud, M. (2023). Censoring extremism: Influence of online restriction on official media products of ISIS. *Terrorism and Political Violence*, *35*(4), 971–987. https://doi.org/10.1080/09546553.2021.1988938

Mirrlees, T. (2021). GAFAM and hate content moderation: Deplatforming and deleting the alt-right. *Media and law: Between free speech and censorship, sociology of crime, law and deviance* (26, pp. 81–97). Emerald Publishing Limited. https://www.emerald.com/insight/content/doi/10.1108/S1521-613620210000026006/full/html

Myagkov, M., Shchekotin, E. V., Chudinov, S. I., & Goiko, V. L. (2020). A comparative analysis of right-wing radical and Islamist communities' strategies for survival in social networks (evidence from the Russian social network VKontakte). *Media, War & Conflict*, *13*(4), 425–447. https://doi.org/10.1177/1750635219846028

PBS Newshour. (2023). Twitter's head of trust and safety resigns after criticism from Elon Musk. *PBS Newshour*. https://www.pbs.org/newshour/nation/twitters-head-of-trust-and-safety-resigns-after-criticism-from-elon-musk

Pearson, E. (2018). Online as the new frontline: Affect, gender, and ISIS-take-down on social media. *Studies in Conflict & Terrorism*, *41*(11), 850–874. https://doi.org/10.1080/1057610X.2017.1352280

Pearson, E., Whittaker, J., Baaken, T., Zeiger, S., Atamuradova, F., & Conway, M. (2023). Online extremism and terrorism researchers' security, safety, and resilience: Findings from the field. *Vox-Pol*. https://www.voxpol.eu/download/report/Online-Extremism-and-Terrorism-Researchers-Security-Safety-Resilience.pdf

Ramati, N. (2020, March 4). The legal response of western democracies to online terrorism and extremism and its impact on the right to privacy and freedom of expression. *Vox-Pol*. https://www.voxpol.eu/download/vox-pol_publication/The-Legal-Response-of-Western-Democracies-to-Online-Terrorism-and-Extremism.pdf

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to telegram and alternative social media. *European Journal of Communication*, *35*(3), 213. https://doi.org/10.1177/0267323120922066

Salina, F. (2023, April 7). Like, tweet, & torment: Repression by proxy. *Human Rights Foundation*. https://hrf.org/like-tweet-torment-repression-by-proxy

Stokel-Walker, C. (2023). Fluent answers from AI search engines are more likely to be wrong. *New Scientist*. https://www.newscientist.com/article/2371097-fluent-answers-from-ai-search-engines-are-more-likely-to-be-wrong/

Thomas, D. R., & Wahedi, L. A. (2023). Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proceedings of the National Academy of Sciences United States of America*, *120*, 1. https://doi.org/10.1073/pnas.2214080120

The European Parliament and the Council of the European Union. (2021). Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, Document 32021R0784. *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2021/784/oj

The Buffalo News. (2022, November 1). *Complete coverage: 10 killed, 3 wounded in mass shooting at Buffalo supermarket*. https://buffalonews.com/news/local/complete-coverage-10-killed-3-wounded-in-mass-shooting-at-buffalo-supermarket/collection_e8c7df32-d402-11ec-9ebc-e39ca6890844.html

NPR. (2022). Twitter's safety chief quit: Here's why. *NPR*. https://www.npr.org/transcripts/1140431011

U.S. House of Representatives. (2022, April 26). *Interview of Brian Fishman. Select committee to investigate the January 6th attack on the U.S. capital*. U.S. House of Representatives. https://www.govinfo.gov/content/pkg/GPO-J6-TRANSCRIPT-CTRL0000071093/pdf/GPO-J6-TRANSCRIPT-CTRL0000071093.pdf

Verma, P., & Oremus, W. (2023). ChatGPT invented a sexual harassment scandal and named a real law prof as the accused. *Washington Post*. https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/

Vincent, J. (2021). Facebook bans academics who researched ad transparency and misinformation on Facebook. *The Verge*. https://www.theverge.com/2021/8/4/22609020/facebook-bans-academic-researchers-ad-transparency-misinformation-nyu-ad-observatory-plug-in

von Soest, C. (2022). Why do we speak to experts? Reviving the strength of the expert interview method. *Perspectives on Politics*. https://www.cambridge.org/core/journals/perspectives-on-politics/article/why-do-we-speak-to-experts-reviving-the-strength-of-the-expert-interview-method/45E710F27CEC6E739B015E10A161E140

Xiang, C. (2023). *Eating disorder helpline disables chatbot for 'harmful' responses after firing human staff*. Vice. https://www.vice.com/en/article/qjvk97/eating-disorder-helpline-disables-chatbot-for-harmful-responses-after-firing-human-staff

Zelin, A. Y. (2021). The case of jihadology and the securitization of academia. *Terrorism and Political Violence*, *33*(2), 225–241. https://doi.org/10.1080/09546553.2021.1880191

**How to cite this article:** Zelin, A. Y. (2023). "Highly nuanced policy is very difficult to apply at scale": Examining researcher account and content takedowns online. *Policy & Internet*, 1–16. https://doi.org/10.1002/poi3.374