

Tech Against Terrorism's Statement for Record to the Select Committee

"Examining Tech Platforms' Moderation Actions and Policies Related to the Attack on January 6 and Their Impact on Online Extremist Behaviour"

March 30, 2022

Tech Against Terrorism is an initiative supported by the United Nations Counter-Terrorism Executive Directorate (UN CTED) and was launched in April 2017. We work with the global tech industry to tackle terrorist use of the internet whilst respecting human rights. Tech Against Terrorism's support mechanisms for the tech industry span across bespoke training, mentorship, capacity-building, threat reports, software and product development, and operational support. Our team consists of open-source intelligence analysts, and counterterrorism and tech policy experts, and we closely monitor online terrorist and violent extremist spaces. Since our inception in 2017, we have engaged with and provided practical support to more than 400 global tech platforms. Within this engagement, Tech Against Terrorism has closely mentored 33 tech platforms to help them tackle terrorist use of their platforms whilst respecting human rights and freedom of speech.

Tech Against Terrorism has been asked to submit a written statement for the record to reflect our platform-specific expertise, especially within the realms of content moderation policies and their enforcement as they related to January 6. This submission of evidence highlights the online extremist behaviour around January 6 with open-source intelligence insights, as well as an in-depth analysis of tech platforms' content moderation policies before, during, and after the attack on January 6.

KEY TAKEAWAYS

Prior to January 6, 2021

- Prior to the attack on the Capitol on January 6, a range of extremist groups who supported the "Stop the Steal" movement, such as the Proud Boys, the Three Percenters and the Boogaloo Bois were able to operate relatively freely across the internet, including on mainstream platforms such as Facebook and Twitter.
- Policy-wise, most larger platforms had satisfactory content standards and prohibitions in place to counter election-related misinformation and calls to violence. Incitement to violence is a prohibition common to all platforms researched and most mainstream platforms prohibit misinformation.
- Enforcement-wise, correctly assessing what content constitutes misinformation and establishing effective thresholds for calls to violence is a challenge for all platforms.
- Larger and long-established platforms went beyond their usual Trust & Safety policies to roll out plans dedicated to ensuring the safety and integrity of the 2020 election. Throughout 2020, their approach gradually shifted from a focus on foreign interference to preventing misinformation and calls to violently disrupt the electoral process.

Aftermath of 6 January 2021: tech platform responses and adversarial shift

- A majority of the tech platforms responded by condemning the attack accompanied by swift content moderation actions. Larger tech platforms including Facebook, Twitter, and YouTube deplatformed then President Donald Trump and focused content moderation efforts on far-right groups, movements, and individuals associated with the attack.
- This resulted in a migration and scattering of far-right groups, movements, and individuals associated with the attack on the Capitol, across a variety of alternative platforms. Following the enforcement actions against far-right online activity by numerous tech platforms during and after January 6, Tech Against Terrorism observed a significant increase in the number of subscribers to multiple extreme far-right channels on Telegram. Tech Against Terrorism observed a concerted effort among the more longstanding far-right extremist Telegram community to “redpill” [radicalise and recruit] new arrivals to the platform.

Observations: tech platform mitigation approaches prior and after 6 January 2021

- Whereas large platforms were able to deploy the resources necessary to address issues specifically related to the 2020 election, solidify pre-existing policies, and deploy increased content moderation capabilities, smaller and newer platforms had to rely on their existing content standards. Further, while most mainstream platforms had prohibitions on misinformation and violent extremism prior to January 6, alt-tech platforms typically had more limited content standards and rarely acted beyond illegal content and direct calls to violence, with the exception of BitChute.
- Tech companies used different policy justifications for increased action. Tech Against Terrorism analysed 29 tech companies’ content standards and other publicly available information in the time leading up to, during, and following January 6. The most common policies invoked in the immediate aftermath of the January 6 events were incitement to violence, election misinformation, violent extremist or hate groups, and hate speech or hateful conduct.
- The bulk of the responses and actions from tech platforms on the January 6 events took place in the days after the events occurred and leading up to the inauguration later in the month. However, in a few cases, some actions and policies were slower to develop, and came about in the months following January 2021.

TECH AGAINST TERRORISM'S RECOMMENDATIONS

1. Tech companies should be more transparent¹ about their content moderation decision-making processes and their justification for policy and enforcement updates.² Tech Against Terrorism recommends that companies publish more information about the justification for changes to the content moderation process to improve overall transparency and accountability towards users.³
2. Tech companies should increase transparency about specific content moderation decisions to create clarity for users and the public about why certain decisions were taken and to support potential appeal processes. When tech platforms neglect to name specific policies which were violated when actioning accounts or content, this creates a gap in users' understanding of a platforms' terms and what is considered a violation of these terms. Naming the policies violated is also needed to facilitate users' right to appeal when their content is removed.⁴
3. The US government can, in order to facilitate content moderation decision-making for tech platforms, create more clarity around the legal basis for content removal. Defining what is lawful and unlawful online speech is the responsibility of democratically accountable governments, rather than private tech platforms. Therefore, we advise the US government to use designation as the legal grounding and guidance for tech platforms to counter domestic terrorist groups and related content. As it stands, tech platforms have no legal basis, nor requirements, to deplatform domestic terrorist or violent hate groups. The US government itself does not maintain a public list of which organisations are considered domestic terrorist organisations, and platforms that are taking the initiative to prohibit these groups have to rely on listing and information provided by civil society and expert organisations.⁵ Tech companies are expected to make decisions around groups that are not illegal by law, and the US government should provide rule of law-based guidance to the tech sector on countering domestic terrorism.
4. Any regulatory efforts seeking to counter terrorist and extremist content online should be drafted in a manner that respects the rule of law, human rights, and fundamental freedoms and considers the impact on smaller tech platforms. Regulatory efforts should avoid introducing measures that encourages over-zealous removal of legal speech, such as platform (and platform employee) liability and short removal deadlines.⁶

¹ This includes, for example, transparency on the ratio of human-to-machine review, tools used, and the guidelines used to implement policies in practice. We observed several instances in which policies were changed in the months following January 6 without context provided.

² Tech Against Terrorism supports tech platforms with increasing transparency around content moderation actions through the Mentorship Programme. We have additionally published [guidelines for tech platforms and governments](#), which seek to improve transparency and accountability from governments and the tech sector around online counterterrorism activities. The Guidelines serve as a starting point for increased transparency, and it is our aim that all governments and companies will report on the baseline set out in the Guidelines.

³ The importance of providing detailed information on the "behind the scenes" of content moderation was highlighted in a [recent article by the EFE](#). The article emphasises the need for companies to engage in "meaningful transparency" to be accountable to their users. Tech companies, for example, can consider clarifying the basics of moderation, such as the scale of the moderation team, the language capacity, the technologies used, and the human-machine moderation balance.

⁴ Explanation of moderation enforcement actions can occur on a case-by-case basis whereby tech companies explain to users which policies they violated when notifying them of actions taken on their account or content. This can equally be applied to a larger scale, such as in the aftermath of January 6, whereby tech companies publish statements on how they are strengthening their moderation enforcement in line with new or existing policies in order to respond to a specific event.

⁵ Many platforms operating or based in the US rely on the information provided by the Southern Poverty Law Center and the Anti-Defamation League to assess which groups to ban under their violent extremism policies.

⁶ Tech Against Terrorism's complete recommendations for policymakers on online regulation can be found in our [Online Regulation Series Handbook](#). We also commend [New Zealand](#) and [Ireland](#)'s regulatory framework for providing clear and rule of law-based guidance on harmful content removal.

ASSESSMENT: ONLINE EXTREMIST BEHAVIOUR AROUND JANUARY 6

During the two months between the November 3, 2020 election and January 6, 2021, the “Stop the Steal” movement gathered momentum and support from a range of extremist groups online including the Proud Boys, the Three Percenters, the Boogaloo Bois, as well as other radical pro-Trump activists and adherents to the QAnon conspiracy theory.⁷ During these months, these groups and individuals were active on a range of online platforms to discuss their views on the elections, plan offline events, and incite each other. Based on our monitoring and third-party press reporting, this took place on a wide range of platforms including Facebook and Twitter, as well as more fringe “alt-tech” spaces like Telegram, Parler, Gab.⁸

Following the removal of accounts operated by President Trump by Facebook, Twitter, and YouTube shortly after the January 6 insurrection, along with action against several far-right extremist and conspiracy theory-driven groups, there was a migration of such activity to a range of alternative platforms including messaging apps such as Telegram, alternative social media sites like Gab and Gettr, and video-sharing platforms including BitChute.⁹ The deplatforming of President Trump from mainstream social media inspired many of his ardent supporters to boycott those platforms and migrate to “free speech” alternatives. BitChute experienced a doubling of its web traffic in the week following January 6, 2021.¹⁰

Alt-tech platforms hosting extremist communities including Gab, Parler, were also the target of data scrapes, breaches or hacks in the months after January 6. We observed posts that indicated a feeling of paranoia within some affected far-right extremist communities, especially around infiltration from law enforcement.¹¹ In our assessment these security concerns have also contributed to an online fragmenting of far-right networks that preceded and were accelerated by the January 6 attack, across a greater number of alternative platforms. Tech Against Terrorism was closely monitoring the online activities of violent far-right networks across platforms around the events of January 6, 2021. In particular, we were tracking networks operating on Telegram Messenger, which has long been widely exploited by terrorist and violent extremist groups for communication and propaganda sharing. Following some of the enforcement actions against far-right online activity by numerous tech platforms during and after January 6, we observed a significant increase in the number of subscribers to multiple extreme far-right channels on Telegram, including ones that openly advocated for terrorism and other forms of political violence, and who regularly pushed conspiratorial narratives about the US election results. For the majority of these channels, the increase came following a period of relative stability in terms of their subscriber numbers.

⁷ Holt, Jared, [After the Insurrection: How domestic extremists adapted and evolved after the January 6 US Capitol attack](#), Atlantic Council

⁸ Criezis Meili, Galloway Brad (2021), [From MAGA to the Fringe: What was Happening Online Before the 6 January Insurrection and What Can We Do Now?](#) Global Network on Extremism & Technology.

⁹ Tech platforms’ moderation appeared to be focused primarily on groups, movements and individuals that were present during the attack at the Capitol. Several central figures lost access to their accounts on mainstream platforms such as Facebook and Twitter.⁹ For example, following their banning from mainstream social media platforms, data shows right-wing influencers including Proud Boys leader Enrique Tarrio, Stop the Steal founder Ali Alexander and former Trump Lawyer and conspiracist Sidney Powell began to establish themselves on alternative platforms such as Telegram, Gab, and Parler.

¹⁰ Zitser, Joshua, (2021), [Following Trump’s YouTube ban, it is feared his supporters are migrating to a ‘Wild West’ of video-sharing, mingling with far-right and neo-Nazi terror groups](#), Business Insider.

¹¹ Holt, Jared, [After the Insurrection: How domestic extremists adapted and evolved after the January 6 US Capitol attack](#), Atlantic Council

Coinciding with the migration to Telegram, we observed a concerted effort among the more longstanding far-right extremist Telegram community to “redpill” [radicalise and recruit] new arrivals to the platform. Referred to by some users and channels as “conservative conversion”, extreme far-right users suggested joining chat groups before contacting and befriending members in an attempt to radicalise their views to a more explicitly neo-Nazi perspective. This included frequent sharing of links in Trump-supporting channels to a 10-part anti-Semitic “documentary” series that defends Nazi Germany against an alleged international Jewish conspiracy. We saw this shared in channels including ones titled “Parler lifeboat” and “We The People Live” alongside subtle messaging encouraging the groups’ followers to watch the film.

SNAPSHOT: PLATFORM CONTENT MODERATION POLICIES BEFORE JANUARY 6

Two takeaways can be drawn from tech platform policy developments in relation to the presidential election throughout 2020:

1. Large tech platforms appear to have learned from the 2016 US election and initially focused their efforts to ensure the safety and integrity of the election on countering foreign interference and providing authoritative information on how and where to vote. Throughout the years, their efforts switched from focusing on foreign interference to domestic threat as their public stance on election-related policies increasingly tackled misinformation and calls to (violently) disrupt the election.
2. In contrast, smaller and newer platforms mostly relied on their existing content standards and made no specific announcements in relation to the presidential election throughout 2020. “Alt-tech” platforms relied on the free speech argument to take minimal actions.

Most tech platforms¹² had the necessary Community Guidelines and policies in place to ensure the safety of the 2020 election and related electoral process. Most leading tech companies operating in the US had prohibitions meant to counter misinformation, incitement to violence,¹³ violent extremism, and hate speech groups prior to the election in November 2020 and the January 6 attack on the Capitol.

Several platforms based or operating in the US had a prohibition on hate or violent extremist groups prior to January 6.¹⁴ These platforms opted to do so despite the absence of a terrorist designation process for “domestic” groups in the US, and therefore lack of legal obligations to counter their online presence. In the absence of a government designation list of domestic terrorist groups, tech platforms tend to rely on listing of violent extremist and hate groups provided by expert and activist organisations such as the Southern Poverty Law Center or the Anti-Defamation League, or work in consultation with counterterrorism experts such as Tech Against Terrorism to identify which non-designated groups represent a substantial threat to violence.

¹² The complete list of platforms reviewed to analysed the policy evolution prior to and after the January 6 attack can be found in Annex A.

¹³ “Incitement to violence” is the most commonly found prohibition in platforms’ Community Guidelines, in comparison to other types of content that platforms may have taken actions on to ensure the safety of the electoral process and presidential transitions – which typically include: misinformation, incitement to violence, violent extremism, and hate / speech groups. It is also the only category of speech to be unlawful in the US in comparison to the other types of content sanctioned around the election – incitement to violence is not covered by First Amendment protections, and incitement to violence or lawless actions is prohibited under different state and federal laws. To read more about this, see: [Inciting to Riot, Violence, or Insurrection](#).

¹⁴ Including but not limited to: Airbnb, Mailchimp, Pinterest, Eventbrite, Vimeo, BitChute, Meta.

Many of the more mainstream content-hosting platforms¹⁵¹⁶ However, most platforms did not have prohibitions related to conspiracy theories content until the second half of 2020, when platforms like Meta¹⁷ and YouTube¹⁸ began to announce they would take down conspiracy theory content that could be linked to offline violence.

In contrast to “mainstream” platforms which generally had policies in place to counter different types of misinformation, hate speech, and violent extremist / hate groups in place prior to January 6, alt-tech platforms typically take a different approach to content moderation. These platforms often have limited content prohibitions and only restrict explicitly illegal content, such as designated terrorist actors and incitement to violence, whilst rarely prohibiting misinformation and hate speech.¹⁹ Many of these platforms were created on a “free speech” argument, and often reject what they consider “censorship” by the larger social media platforms.

While most platforms had policies in place to allow for action against mis- and disinformation and violent extremism, enforcing these policies consistently is a challenge for all tech platforms. These types of content are often considered “grey area” content due to the lack of legal framework to base moderation policies and enforcement on, and require moderators to understand the context and nuance of online speech to be correctly assessed when gauging intent which presents a significant challenge for platforms to scale their moderation enforcement. This also means that it is more difficult for platforms to rely on automated moderation without the risk of removing a substantial amount of legal and non-harmful speech. Through Tech Against Terrorism’s work with tech platforms, we find that clear legal basis, including designation lists, are essential to ensure that platforms can make clear-cut and effective decisions with regard to content moderation.

Specific actions taken by platforms to safeguard the US 2020 Elections

Large tech platforms, such as Twitter, YouTube, Facebook, and Pinterest all announced measures to ensure the safety and integrity of the presidential election throughout 2020. These measures evolved throughout 2020 to increasingly address the issue of election-related misinformation and calls to disrupt the electoral process – including incitement to violence.

In early 2020, many of these measures focussed on countering foreign electoral interference and providing users with authoritative and neutral information on the electoral process (such as how and where to vote). In the early summer of 2020, efforts to counter election-related misinformation started being more frequently emphasised.²⁰ The focus on election-related misinformation further increased in the fall of 2020 and remained persistent until January 6.

¹⁵ Including but not limited to: Facebook, Youtube, Twitter, Vimeo, Pinterest

¹⁶ Tumblr, for instance, updated its Community Guidelines after 2016 to include an explicit provision on US “election integrity” – Tumblr’s Community Guidelines were last updated in [July 2020](#). Since 2016 Meta [had built up its Elections Operations Center](#), monitoring major elections globally, and built a “viral content review systems to flag posts that may be going viral”.

¹⁷ Meta (2020), [An Update to How We Address Movements and Organizations Tied to Violence](#).

¹⁸ YouTube (2020), [Managing harmful conspiracy theories on YouTube](#).

¹⁹ Unless mandated by law in certain jurisdictions. For instance, BitChute restricts content that constitutes incitement to hatred in the UK and EU to align with legal requirements. BitChute also goes beyond terrorist designation lists and prohibit violent extremism on its platform and, for instance, includes the Nordic Resistance Movement and the Order of Nine Angles in its prohibited organisation lists.

²⁰ It should be noted that platforms began to strengthen their efforts to counter electoral misinformation in the months after most tech platforms had to expand their prohibition on misinformation to counter the spread of Covid-19 related misinformation online and contain the associated risks of offline violence. See: Deverell Flora and Janin Maygane (2020), [Covid-19: far-right violent extremism and tech platforms’ response](#), Fondapol

In September and October 2020, certain platforms²¹ began to introduce prohibitions on misinformation related to the electoral process²² and its outcome. Google justified this in relation to the delay in electoral results to be expected because of postal voting. Around the same period, major platforms also began to prohibit calls to disrupt the electoral process – for instance calls to bring weapons to voting places or to intimidate election officials. For major platforms,²³ this appears to mark a shift in how they envisioned election integrity: as they continued to strengthen their efforts to counter election-related misinformation, they also increasingly began to focus on calls for disruption of the electoral process and presidential transition throughout the end of 2020 and early January 2021.

Very few platforms announced further moderation policy and enforcement changes between election night and January 6. For most platforms,²⁴ their last updates on election integrity efforts were published between September and November 2020, and no further announcements were made until January 6 and the following days. For instance, Facebook only publicly announced that it would remove “stop the steal” related content after January 6 and ahead of inauguration day.²⁵ A notable exception is Airbnb which cancelled bookings made by members of the Proud Boys for the D.C area in the days around January 6, 2021.²⁶

In parallel to electoral integrity efforts, large tech platforms also acted against US-based far-right violent extremist groups and movements throughout 2020. A notable effort at deplatforming online movements to safeguard the presidential election began in August 2020 when large platforms²⁷ began to crack down on QAnon supporters and conspiracy theorist movements in an effort to contain election-related misinformation and calls to disruption.

Large vs small platforms

The difference in how large and smaller and/or newer platforms prepared for the presidential election and its aftermath is likely to be explained by issues of resources and capacity rather than will. In our experience in directly supporting smaller and newer tech platforms as part of the Tech Against Terrorism Mentorship and Membership programmes²⁸, we find that the main challenge for smaller platforms in countering terrorist use of the internet is one of resources and understanding of the threat, not of will.

²¹ Including Meta, Pinterest, YouTube, and Twitter.

²² For example, claims of electoral fraud.

²³ Such as Meta, YouTube, and Twitter.

²⁴ Including Facebook, Twitter, Google / YouTube and Pinterest.

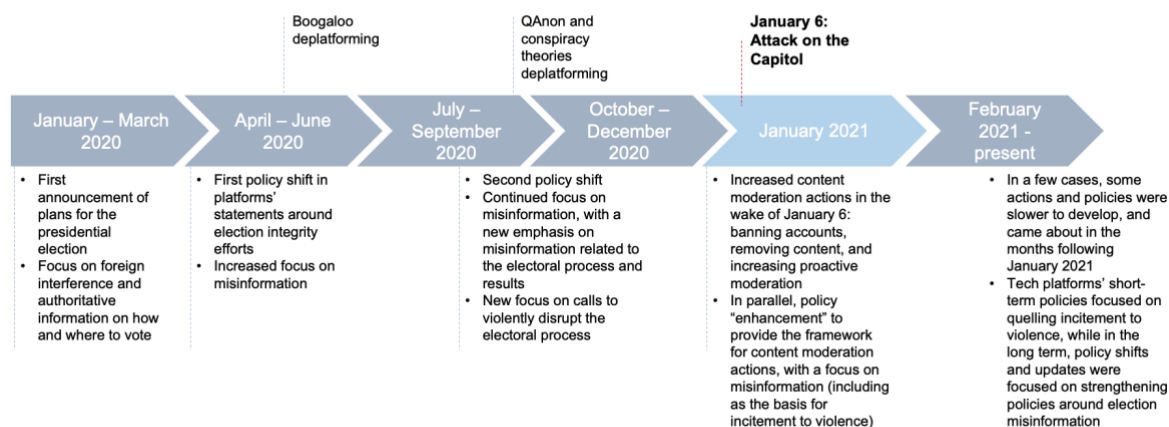
²⁵ Meta (2020), [Our Preparations Ahead of Inauguration Day](#).

²⁶ Beaujon, Andrew (2021), [Airbnb Says It Canceled Some Hate-Group Members' DC Reservations and Plans to Do More](#).

²⁷ Including Meta, YouTube, and Twitter. To read more on this, see: [Unwelcome On Facebook And Twitter, QAnon Followers Flock To Fringe Sites](#)

²⁸ To learn more about our mentorship programme please see: <https://ksp.techagainstterrorism.org/knowledgebase/tech-against-terrorism-mentorship-programme/>

Timeline: Policy Evolution January 2020 - Present



TECH PLATFORM CONTENT MODERATION AND POLICY ENFORCEMENT IN THE AFTERMATH OF JANUARY 6

In the immediate wake of the January 6 attack on the US Capitol, many tech platforms responded by removing or suspending accounts and content as well as by blocking the use of certain hashtags and phrases. Below we provide a summary of our observations in this regard.

Short term: Actions taken by tech platforms in January 2021²⁹

1. Moderation Actions

In the immediate wake of the attack, many tech platforms responded by releasing a public statement condemning the attack accompanied by swift content moderation actions. The different moderation actions taken by tech platforms on the day of the attack and in the immediate aftermath, include removal of content and channels, account suspensions and bans, suspension of ads, as well as blocking users, hashtags, payments, fundraising campaigns, and merchandise. From 29 tech platforms analysed by Tech Against Terrorism, 26 publicly recognised that they took action in some form of moderation against content and users affiliated with the attack on January 6 and in preparation of the inauguration later that month.

²⁹ For an overview of actions taken by different tech companies in the immediate days after the January 6 attack, please see First Draft's chart which documents how social media platforms responded to the January 6 events at the US Capitol, [here](#).

These actions would also include enhanced methods of increased proactive review and content moderation enforcement methods. For example, on January 12, 2021, Twitter released a statement³⁰ in which it announced that its teams were “continuing to aggressively deploy technology to surface potentially harmful Tweets for human review” to act on potentially violative content as quickly as possible. Twitter also stated that it blocked violating keywords from search and trends to prevent certain content from trending.³¹

As a result of increased moderation enforcement, some platforms reported an uptick in their content moderation actions in transparency reports for the first half of 2021. Pinterest’s transparency report, covering January to June 2021 exemplifies this.³² Pinterest accompanies this with a “recent trends” analysis, in which it highlights that as part of the company’s efforts to maintain a safe platform during the US presidential transition, there was a “relative increase in the number of boards and accounts deactivated for violation of [the Civic misinformation] policy in Q1 of 2021”. Similarly, in the transparency report’s section on “Graphic violence and threats” - which includes content “that shows the use of violence, threats and language that glorifies violence” - recent trends highlighted by Pinterest include mention of “a relative increase in the amount of content deactivated for violation of our graphic violence and threats policy during the US presidential transition in Q1 of 2021”.³³

Though a lot of the actions were alike and took place on similar grounding, such as election misinformation or promoting or glorifying violence, tech platforms varied in the types of policies that they explicitly used to justify their actions. The typical policies referenced with actions in the immediate aftermath of the January 6 events were incitement to violence, election misinformation, violent extremist or hate groups, and hate speech or hateful conduct. A few tech platforms did not specify exact policies but instead provided general reference to their Terms of Service or Community Guidelines as having been violated.

A. *Election Misinformation*

From an analysis of 29 tech platforms, five explicitly referred to policies on election or voting misinformation to justify content moderation actions following January 6. Examples include:

- TikTok clarified on January 7, 2021 that Trump’s speeches, where he reiterated claims of a fraudulent election, were being removed on the grounds that they violate the company’s misinformation policy.³⁴
- Video-sharing platform Vimeo announced on January 11, 2021 that it had updated its Acceptable Use Community Guidelines to make clear that it would remove “content containing false or misleading claims about the 2020 presidential election and its aftermath”. This included “claims that the election was stolen, fraudulent, or otherwise illegitimate, as well as claims that try to shift the blame for events of January 6 away from supporters of the current President”.³⁵

³⁰ Twitter, (2021), [An update following the riots in Washington, DC](#).

³¹ For operational security reasons these keywords are not public. Twitter prohibited keywords based on its Coordinated Harmful Activity, Civic Integrity, Hateful Conduct, Glorification of Violence, Violent Threats, and Sensitive Media policies. For more on this, see: [An update following the riots in Washington, DC](#).

³² Under the report’s “civic misinformation” section, Pinterest notes that it introduced its civic misinformation policy, which prohibits false or misleading content on Pinterest that impedes an election’s integrity or an individual, in 2020 as an addition to several existing Community Guidelines on misinformation and disinformation. Pinterest’s Transparency Report can be accessed [here](#).

³³ Pinterest’s Transparency Report can be accessed [here](#).

³⁴ Perez, Sarah (2021), [TikTok bans videos of Trump inciting mobk, blocks #stormthecapitol and other hashtags](#), TechCrunch.

³⁵ Vimeo, (2021), [Statement on videos promoting election disinformation and inciting violence](#).

B. Hate Speech or Hateful Conduct

- eBay stated that “any merchandise glorifying violence or hate will be removed from our marketplace, this includes QAnon merchandise.” This aligned with eBay’s seller policy which prohibits “listings that promote or glorify hatred, violence, or discrimination.”³⁶
- YouTube also removed channels of groups associated with January 6 events, such as those affiliated with Proud Boys and QAnon, for violating existing policies on hate, harassment, election integrity.³⁷

C. Violent Extremist or Violent Hate Groups and Organisations

Based on an analysis of 29 tech platforms, only two directly tied their content moderation actions to policies relating to violent extremist or violent hate groups and organisations, however, below are some examples:

- On January 13, Zello announced it had banned all militia-related channels over “evidence of Zello being misused by some individuals while storming the United States Capitol building.”³⁸
- On January 11, Airbnb reiterated and expanded upon its practices of banning individuals ahead of events associated with violent hate groups ahead of the far-right extremist gathering in Washington DC on January 6. Airbnb noted that, “...On an ongoing basis, Airbnb has removed people from the platform associated with violent hate groups in advance of specific events, including taking action ahead of the horrific attack on Congress, by cancelling reservations and removing accounts associated with hate group members, including Proud Boys.”³⁹

D. Incitement of Violence

Based on our analysis of the 29 tech platforms, it was found that tech platforms would most frequently refer to incitement of violence as the basis for which policies would be used to enforce content moderation around the January 6 events. Fourteen platforms explicitly referred to incitement of violence as the basis for their actions.

- Payment provider Stripe cited its Prohibited Businesses policy that restricts transactions with “high risk” business including one that “engages in, encourages, promotes or celebrates unlawful violence or physical harm to persons or property, or engages in, encourages, promotes or celebrates unlawful violence toward any group based on race, religion, disability, gender, sexual orientation, national origin, or any other immutable characteristic”.⁴⁰
- Twitter’s permanent suspension of Trump’s account was said to be “due to the risk of further incitement of violence” and grounded in Twitter’s glorification of violence policy.⁴¹

³⁶ Debter, Lauren (2021), [EBay Says It Will Pull QAnon Merchandise From Its Site](#), Forbes. For more on eBay’s policies, see [here](#).

³⁷ Ha, Anthony (2021), [YouTube will start penalizing channels that post election misinformation](#), TechCrunch. To read more on YouTube’s policies, see [here](#).

³⁸ Zello (2021), [Zello Takes Action Against Militias](#).

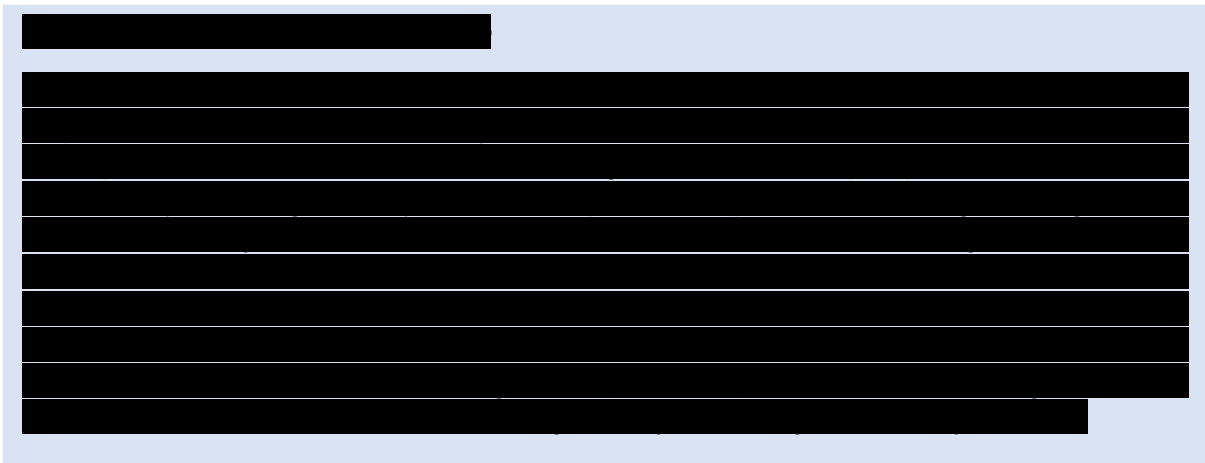
³⁹ Airbnb (2021), [Airbnb announces ‘Capitol Safety Plan’ for the Inauguration](#).

⁴⁰ Andriotis, AnnaMaria; Glazer, Emily; Rudegeair, Peter, (2021), [Stripe Stops Processing Payments for Trump Campaign Website](#), The Wall Street Journal. To read more on Stripe’s policy, see [here](#).

⁴¹ Twitter (2021), [Permanent Suspension of @realDonaldTrump](#). For more on Twitter’s policy, see [here](#).

- On January 8, Google Play suspended Parler’s app listings from the play store until it implemented “robust moderation for egregious content”. Google referenced policies that require “apps displaying user-generated content have moderation policies and enforcement that removes egregious content like posts that incite violence”, and additionally cited the “continued posting in the Parler app that seeks to incite ongoing violence in the US”.⁴²

Finally, of those analysed by Tech Against Terrorism, nine tech platforms refrained from providing specific policy references in their actions pertaining to the January 6 events. Rather than providing or naming specific policies, platforms chose to state that their actions were a consequence of violations of their Community Guidelines, Terms of Service, or Acceptable Use Policy generally. However, in many of these public statements where general Terms were provided as opposed to specific policies, companies still chose to ground their explanation of the violation and reasoning in similar themes as other platforms’ policies, including public calls to violence and election misinformation.



2. Policy “Enhancement”

In the short term and immediate aftermath of the events that took place on January 6, 2021, tech platforms introduced or solidified pre-existing policies to be able to increase and strengthen their content moderation capabilities. These updates by tech platforms would take form as either a blog post or statement re-iterating existing policies and prohibitions around misinformation and violence; include announcement of expansion of policies, such as Vimeo, Eventbrite, and YouTube’s updates on misinformation policy, or on their violent extremist policies, as updated by Zello; and finally, some platforms released a plan for inauguration day safety and outlined how they would be implementing policies and content moderation actions on certain behaviours and actors (see Annex B).

⁴² Matney, Lucas (2021), [Parler removed from Google Play Store suspensions reportedly looms](#), TechCrunch. For more information on Google’s policies, see [here](#).

Long Term Actions Taken by Tech Platforms Following January 2021

The bulk of the responses and actions from tech platforms on the January 6 events took place in January, days after the events occurred and leading up to the inauguration later in the month. However, in a few cases, some actions and policies were developed more slowly and came about in the months following January 2021 (see annex C).

A few tech platforms have updated relevant policies in the months following January 6, 2021, however, they do not explicitly mention any events, such as the attack on January 6, which caused the change or update in the policy. Based on two tech platforms in our analysis, we have noticed that elements of those policies seem to indicate that they were very likely drafted in response to the 6 January events and to claims that the 2020 US election was “stolen”. For example, the policies would include a prohibition on false and misleading content as well as about important events.

In 2022, Twitter convened an internal working group with members from various parts of the company to ensure the platform could enforce its rules and protect users around the one-year mark of January 6, 2021.⁴³ Also, in 2022 around the anniversary of the Capitol attack, YouTube clarified it has been monitoring trends around content and behaviour related to the Capitol attack and its anniversary. A spokesperson said YouTube had not detected an increase in content containing new conspiracy theories related to January 6 or the 2020 election that violates its policies and noted that YouTube’s “systems are actively pointing to high authority channels and limiting the spread of harmful misinformation for election-related topics.”⁴⁴

⁴³ Duffy, Clare (2022), [On the anniversary of January 6, social media platforms are on high alert](#), CNN.

⁴⁴ Duffy, Clare (2022), [On the anniversary of January 6, social media platforms are on high alert](#), CNN.

ANNEX.

- A. Tech Platforms analysed by Tech Against Terrorism** in terms of their policies and content moderation enforcement before, during, and after the attack on January 6. Though this list is not exhaustive of the research conducted by Tech Against Terrorism, given that we looked to additional platform’s content standards and publicly available information, the list below reflects the tech platforms which were more closely analysed.

Airbnb	Parler
Amazon	PayPal
Apple	Pinterest
Bumble	Reddit
Clubhouse	Snapchat
Discord	Stripe
DLive	Telegram
eBay	TikTok
Eventbrite	Twitch
Facebook	Twitter
Gab	Venmo
GoFundMe	Vimeo
Google Play	YouTube
Instagram	Zello
Mailchimp	

- B. Policy “Enhancement”:** examples of the short term and immediate aftermath of the events that took place on 6 January 2021, tech platforms introduced or solidified pre-existing policies to be able to increase and strengthen their content moderation capabilities directly following the January 6 events.

Eventbrite	Eventbrite published a statement following the US Capitol attack on January 6, 2021, titled “We stand in support of safe gathering” ⁴⁵ . In this blogpost, Eventbrite reminds users of existing prohibitions in its Community Guidelines, including regarding the promotion of violence, illegal activity and “misinformation that may result in harm”. In addition, however, Eventbrite importantly states that it will expand its misinformation policy to events “fuelling false claims about the 2020 election”. Further, Eventbrite concludes this blogpost by stating that they have a joint a statement from the Civil Alliance calling for a peaceful transition of power. ¹¹
Vimeo	On January 11, Vimeo published a blog post reiterating its commitment to safety “both for our users and broader society” and that content causing “real-world harm” is not allowed on the platform. ⁴⁶ In this blog post, Vimeo explains that it has updated its

⁴⁵ Eventbrite (2021), [We stand in support of safe gathering](#).

⁴⁶ Vimeo, (2021), [Statement on videos promoting election disinformation and inciting violence](#).

	<p>Community Guidelines to explicitly prohibit the sharing of content that constitutes voting misinformation. This restriction contains an explanation of what is considered to be voting misinformation, as well as a specific restriction on misinformation about the 2020 US elections: “In particular, content that claims that the 2020 presidential election was fraudulent or stolen or otherwise illegitimate will be removed.” Vimeo also states that, on the date of publication, it had already begun removing content in violation of its policy on election misinformation. Vimeo specifies that accounts that continue to upload violating content or “have demonstrated a propensity to do so” would be removed from the platform.⁴⁷</p>
Twitter	<p>On January 12, Twitter released a statement in which it noted that the company updated its Civic Integrity Policy after January 6 to increase enforcement action on misleading and false information surrounding the 2020 US presidential election as the basis for incitement to violence. The updated policy provides details about how Twitter actions on violations, including repeated sharing of Tweets that receive warning labels. Repeated violations of this policy can result in permanent suspension.⁴⁸ In this statement, Twitter also announced that it limited engagement by no longer allowing Tweets labelled for violations of Civic Integrity Policy to be replied to, liked, or retweeted.⁴⁹</p>
YouTube	<p>On January 7, YouTube introduced a policy where channels publishing “false claims” about the US election would be penalised with a strike which would temporarily suspend them from posting videos.⁵⁰ Before this, in December 2020, YouTube announced it would remove videos alleging fraud or election errors but there was a grace period so the videos would be removed without additional penalty. The grace period was supposed to end January 21, 2021, but it was extended. In February 2019 YouTube announced a strike system whereby users would first receive a warning for violative content, then a first strike if a user’s content doesn’t comply with Community Guidelines a second time. The first strike means users (for one week) cannot upload videos, live streams, stories; start a live stream; schedule a video; create a Premiere; add a trailer to a premiere or live stream; create thumbnails or community posts; create/edit/add other users to playlists; or add/remove playlists from the watch page using the save button.⁵¹ Therefore, this update means that users will receive a strike immediately instead of a warning. Therefore, not only would content be removed, but that user would be banned from posting (amongst using other YouTube features) for one week.⁵²</p>

⁴⁷ Vimeo, (2021), [Statement on videos promoting election disinformation and inciting violence](#).

⁴⁸ Twitter (2021), [An update following the riots in Washington, DC](#).

⁴⁹ Twitter (2021), [An update following the riots in Washington, DC](#).

⁵⁰ Ha, Anthony, (2021), [YouTube will start penalizing channels that post election misinformation](#), TechCrunch.

⁵¹ For more on YouTube’s policies, see [here](#).

⁵² For more on YouTube’s policies, see [here](#).

Zello	Zello published a statement on January 13, 2021, in which it outlined the company’s actions following the January 6 events, particularly related to militias. Zello mentioned a recent addition to their terms of service, which states that users agree not to use the service to “participate with or communicate to organizations whose principles specifically endorse or espouse violence”. Zello notes that, “while we used to impose limitations on the content alone, we have now extended the notion of abuse on the platform to include use by organizations whose principles or leaders specifically endorse or espouse violence. It is with that perspective that we have deleted 2,000+ channels associated with militias and other militarized social movements.” ⁵³
Airbnb	Ahead of the inauguration, on January 20, 2021, Airbnb announced a 7-step plan to prevent the use of its services by extremists and violent extremists and to protect hosts in the Washington D.C. area. ⁵⁴ As part of this plan Airbnb notably banned individuals identified to have been involved in criminal activity at the Capitol on January 6, including by banning the accounts of all individuals listed on the arrest logs of the D.C. Metro Police for that day, and by reviewing all bookings in the D.C. area ahead of the inauguration. ⁵⁵
Facebook	On January 11, 2021, Facebook published a statement of its Preparations Ahead of Inauguration Day which outlined the measures being taken to prevent the spread of misinformation and incitement to violence in the lead up to Inauguration Day. ⁵⁶ In line with Facebook’s existing Coordinating Harm Policy, ⁵⁷ Facebook banned all content and accounts related to or mentioning “stop the steal”. Facebook also announced a pause on all ads in the US about politics and elections. On January 15-16, Facebook updated its Preparations Ahead of Inauguration Day again to block the creation of new Facebook events happening in close proximity to the White House, US Capitol building, and any state capitol buildings through Inauguration Day. Additionally, Facebook’s operations centre would conduct secondary reviews of all Facebook events related to the inauguration and remove those which violate policies. Facebook would also continue to block event creation in the US by non-US based accounts and Pages. Finally, Facebook announced a ban on ads that promote weapon accessories and protective equipment in the US through Inauguration Day out of caution. In assessing the extent to which Facebook has responded to the events on January 6 on a more long-term basis, Facebook has maintained that its policies which were in place before the Capitol riots were comprehensive and have since only become more robust. ⁵⁸

⁵³ Zello, (2021), [Zello Takes Action Against Militias](#).

⁵⁴ Airbnb, (2021), [Airbnb announces ‘Capitol Safety Plan’ for the Inauguration](#).

⁵⁵ Airbnb, (2021), [Airbnb announces ‘Capitol Safety Plan’ for the Inauguration](#).

⁵⁶ Meta, (2021), [Our Preparations Ahead of Inauguration Day](#).

⁵⁷ For more on Meta’s Coordinating Harm Policy, see [here](#).

⁵⁸ Meta, (2021), [Our Preparations Ahead of Inauguration Day](#).

c. Policy “Enhancement”: Examples of long-term actions and policies which slowly developed and came about in the months following January 2021.

Facebook	<p>In assessing the extent to which Facebook has responded to the events on January 6 on a more long-term basis, Facebook has maintained that its policies that were in place before the Capitol riots were comprehensive and have since only become more robust. In September 2021, Meta noted in a blogpost that its Artificial Intelligence systems have improved at proactively removing problematic content such as hate speech. In its November 2021 Community Standards Enforcement Report, Meta said that across Facebook and Instagram, the prevalence of views of hate speech content versus other content had declined for the fourth consecutive quarter.⁵⁹ In October 2020, Facebook said that it was temporarily halting recommendations of political groups for US users in the run-up to the presidential election. In a call with investors on January 20, 2021, Mark Zuckerberg said the company would no longer recommend civic and political groups to users of the platform and announced this policy would be made permanent and global, calling the groups “breeding groups for hate” as the company was “continuing to fine-tune” its approach to countering extremism. This was never officially stated or reflected in Facebook’s policies, however, depicts an interesting policy announcement which was not materialised.⁶⁰</p>
Airbnb	<p>In March 2021, Airbnb published a blogpost to update users about its work to uphold its Community Standards. This post provides an overview of the evolution of Airbnb’s Community Standards and Trust & Safety practices to counter use of the services by terrorist and violent extremist groups, including actions taken to ensure the ban of violent extremists following the January 6, 2021 attack on the US Capitol and ahead of the January 20, 2021 US Presidential Inauguration – over 130 people have been banned due to their violent involvement in the January 6 event.⁶¹</p>

⁵⁹ Meta, (2021), [Community Standards Enforcement Report, Third Quarter 2021](#).

⁶⁰ Culliford, Elizabeth (2021), [Facebook says it will permanently stop recommending political groups to users](#), Reuters. Faife, Corin; Ng, Alfred, (2021), [After Repeatedly Promising Not to, Facebook Keeps Recommending Political Groups to Its Users](#), The Markup. BBC News (2021), [Facebook to Stop Recommending Civic and Political Groups](#).

⁶¹ Airbnb, (2021), [An update on our work to uphold our Community Standards](#).