U.S. Senate Committee on Homeland Security and Governmental Affairs
Hearing on "Social Media Platforms and the Amplification of Domestic Extremism & Other Harmful Content"
Thursday, October 28, 2021

Karen Kornbluh Written Testimony

Chairman Peters, Ranking Member Portman, and distinguished Committee members, thank you for this opportunity to testify today on the relationship between social media platforms and the amplification of domestic extremism.

My name is Karen Kornbluh and I direct the Digital Innovation and Democracy Initiative at the German Marshall Fund of the U.S., where I am also a Senior Fellow. Previously, I served as United States Ambassador to the Organization for Economic Cooperation and Development.

The topic of this hearing is critical for the future of the country and our national security. In the words of Timothy Langan, Assistant Director of the FBI Counterterrorism Division, "The greatest terrorism threat…today is posed by lone actors or small cells who typically radicalize online and look to attack soft targets with easily accessible weapons."[1]

An internal Facebook test showed how these actors can be radicalized online. A fake Facebook account created for a fictional "Carol Smith," a 41-year old conservative mother from North Carolina, was recommended pages and groups related to QAnon within days of its creation and was recommended an account associated with the militia group Three Percenters within three weeks.[2]

This research and other documents released by the Facebook whistleblower underscore that design features of large social media platforms are creating a feedback loop that pushes some Americans toward violent extremist ideologies and are facilitating large-scale extremist organizing.

They also make clear that the companies' current strategy of after-the-fact, "whack-a-mole" take-downs is grossly insufficient to address this systemic vulnerability.

Until social media companies' incentives are changed, the problem of online radicalization and violent extremism will continue to grow.

Down the "rabbit hole" -- social media goes well beyond providing users tools to connect organically with others; it pulls users into rabbit holes and empowers small numbers of extremist recruiters to engineer algorithmic radicalization.

---

[1] Timothy Langan, "Confronting White Supremacy: Examining the Biden Administration's Counterterrorism Strategy," September 29, 2021.
[2] Ryan Mac and Sheera Frenkel, "Internal Alarm, Public Shrugs: Facebook's Employees Dissect Its Election Role," New York Times, October 22, 2021.

The documents released by Facebook whistleblower Frances Haugen reveal the risk that algorithms that rank and recommend content based on user engagement lead some users down information "rabbit holes" into increasingly narrow echo chambers where violent conspiracy theories thrive. People radicalized through these rabbit holes make up a small slice of total users, but at scale that means a great many users.

An August 2019 Facebook internal memo admitted very clearly that, "the mechanics of our platform are not neutral" and, in fact, that core product mechanics, including virality, recommendations, and optimizing for engagement are key to why hate and misinformation flourish on the platform.[3] Research shows this is true for other platforms as well.

Facebook groups are a key vector of recruitment. Internal research found 70% of Facebook political groups in the U.S. were rife with hate, bullying, harassment, misinformation and other rule violations, and that many of the most toxic civic groups were "growing really large, really fast."[4] Organizers benefited from a variety of tools to build groups and pages. For example:

- Facebook's own algorithms recommend extremist groups to users – to such an extent that they were responsible for a striking 64% of these groups' new members in 2016.[5] For example, Facebook directs users who "like" certain militia pages toward other militia groups.
- "Super inviters" or "invite whales" created invitation links that could be shared on or off Facebook and can easily coordinate their invitations.
- The platform provides group members recommendations of others to invite based on data about users' activities, allowing groups to draw from other conspiracy and militia groups.
- A report revealed the problem of bait-and-switch groups, in which pages that post about cute animals and other innocuous topics build tens of thousands of followers then sell the page to the highest bidder, at which point it becomes a vector for extremist content pushed at unwitting users.
- And groups that were punished for breaking the rules easily could evade take downs by re-establishing themselves with new names, from which they could continue viral recruitment strategies.

As an example of the way these tools can be used to grow groups, a Facebook internal report on the Stop the Steal movement revealed that 0.3% of group members were responsible for 30% of invitations to join.[6] Organizers sent hundreds of invitations to members of other groups, resulting in high membership overlap with Proud Boy and militia groups and fueling Stop the Steal Groups' meteoric growth rates.

---

[3] Mike Isaac, "Facebook Wrestles With the Features It Used to Define Social Networking," New York Times, October 25, 2021.
[4] Shannon Bond and Bobby Allyn, "How the 'Stop the Steal' movement outwitted Facebook ahead of the Jan. 6 insurrection," NPR, October 22, 2021.
[5] Jeff Horowitz and Deepa Seetharaman, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive," May 26, 2020.
[6] Ryan Mac, Craig Silverman, and Jane Lytvynenko, "Facebook Stopped Employees From Reading An Internal Report About Its Role In The Insurrection. You Can Read It Here," BuzzFeed, April 26, 2021.

Ads are another vector for radicalization. These can be targeted to small audiences based on detailed data gathered on users. And, as the Facebook whistleblower emphasized to a parliamentary select committee, ads on Facebook are priced "partially based on the likelihood that people like them, reshare them, do other things to interact with them — click through on a link" and therefore, "An ad that gets more engagement is a cheaper ad."[7]

Similar algorithmic radicalization is evident on other platforms: TikTok's recommendation algorithm also promotes content from QAnon, the Patriot Party, Oath Keepers, and Three Percenters.[8] After users interacted with trans-phobic videos on TikTok, the recommendation algorithm fed users videos with hate symbols, white supremacist and anti-Semitic content, as well as coded calls to violence.[9]

YouTube has 290 extremist channels, according to new research. When researchers showed an interest in militant movements, YouTube suggested videos to them with titles like "5 Steps to Organizing a Successful Militia" and "So You Want to Start a Militia?" The platform also recommended videos about weapons, ammunition, and tactical gear to what the researchers at the Tech Transparency Project call "the militia-curious viewer."[10]

This algorithmic radicalization has already resulted in extremist violence. Air Force Staff Sergeant Steven Carrillo in 2020 shot and killed a protective security officer and wounded his partner, before killing a sheriff's deputy. He had begun engaging with the extremist group Boogaloo Bois on Facebook and eventually was in direct contact with prominent Boogaloo members. He purchased a device that converts AR-15 rifles into fully automatic machine guns from a website that advertised to Boogaloo Facebook groups. As Carrillo was being pursued by police, he sent a WhatsApp message to members of the heavily armed Boogaloo militia faction he had recently joined, telling them to join him.[11]

Radicalization is a cross-platform phenomenon.

Extremists can organize on specialized, less moderated sites and use other platforms to radicalize others, silence critics, or swamp the news cycle. The magazine *Nature* found individuals move from mainstream platforms to less moderated ones like 4Chan or Telegram in a few clicks and then reintroduce fringe content to the original mainstream platform.[12]

---

[7] Isobel Asher Hamilton, "Facebook whistleblower Frances Haugen says it's cheaper to run 'hateful' ads on the platform than other kind of adverts," Business Insider, October 26, 2021.
[8] Olivia Little, "TikTok is prompting users to follow far-right extremist accounts," Media Matters for America, March 26, 2021.
[9] Olivia Little and Abbie Richards, "TikTok's algorithm leads users from transphobic videos to far-right rabbit holes," Media Matters for America, October 5, 2021.
[10] Tech Transparency Project, "YouTube's Filter Bubble Problem is Worse for Fox News Viewers," October 24, 2021.
[11] Gisela Pérez de Acha, Kathryn Hurd, and Ellie Lightfoot, "'I Felt Hate More Than Anything': How an Active Duty Airman Tried to Start a Civil War," PBS, April 13, 2021.
[12] N. Velásquez, R. Leahy, N.J Restrepo, et al., "Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms," Scientific Reports 11, 2021.

The social media companies' whack-a-mole approach of taking down individual pieces of violative content or accounts after damage is done fails to prevent algorithmic radicalization.

When people inside Facebook discussed a more systematic approach – one that would be content and viewpoint agnostic -- to restrict design features that amplify incendiary and divisive posts, the company rejected most of these ideas. Head of Facebook Health Kang Xing Jin proposed in 2019 that the company dial back on automated recommendations. Other proposals included dialing back algorithmic virality.[13] Another option was to enforce its rules prohibiting individuals from operating multiple accounts, since many of these accounts are purveyors of violent political activity, according to Facebook employees. [14]

The company rejected most of these ideas and largely left it to content moderation to play after the fact whack-a-mole. But even then, Facebook tied moderators' hands behind their backs. It maintains a whitelist that exempts VIP users with the largest footprints from the stated rules -- even though that meant the company was "not actually doing what we say we do publicly," according to an internal report.[15]

The resulting moderation process is catching only small percentages of violative content: only 3 to 5% of hate speech and 0.6% of content that depicts or incites violent content.[16] Even though Facebook says militia groups are banned, in reality, roughly 70% of the Facebook militia pages identified in a Tech Transparency Project report had the word "militia" in their name.[17]

The platforms are considering doing more to address this serious problem. Twitter recently released an internal report on its algorithms. [18] YouTube instituted strict enforcement of their rules against false election claims. [19] Facebook itself has launched a new project to examine the pathways to radicalization.

While Congress works toward comprehensive privacy legislation and various antitrust investigations proceed, targeted steps are needed now to limit algorithmic radicalization.

---

[13] Jeff Horwitz and Justin Scheck, "Facebook Increasingly Suppresses Political Movements It Deems Dangerous," Wall Street Journal, October 22, 2021.
[14] Julia Arciga and Susannah Luthi, "How Facebook users wield multiple accounts to spread toxic politics," Politico, October 25, 2021.
[15] Jeff Horwitz, "Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt," Wall Street Journal, September 13, 2021.
[16] Deepa Seetharaman, Jeff Horwitz and Justin Scheck, "Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts," Wall Street Journal, October 17, 2021.
[17] Tech Transparency Project, "Facebook's Militia Mess," March 24, 2021. For a detailed breakdown of Facebook's tiered process of addressing Dangerous Individuals and Organizations, see Facebook Transparency Center, "Dangerous Individuals and Organizations," last accessed October 26, 2021.
[18] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjera, and Moritz Hardt, "Algorithmic Amplification of Politics on Twitter," Twitter, 2020.
[19] YouTube, "Supporting the 2020 U.S. election," December 9, 2020; YouTube, "Updated Policy," Twitter, January 7, 2021.

First, a"black box flight data" recorder, or actionable transparency is needed. We shouldn't need a whistleblower to access data. The Federal Trade Commission should require more transparency, just as the National Transportation Safety Board gets access to data on airplane crashes or the Environmental Protection Agency releases data on pollution.

- It can require third-party audits of terms of service enforcement that are routine and publicly available.
- Researchers also need access to privacy-protected retrospective data.
- The bipartisan Honest Ads Act would provide the same transparency about ads as is required on broadcast but should be supplemented by Know Your Customer rules that prevent dark money or foreign actor ad funding. Platforms must have robust systems for archiving political advertisements that are searchable and sortable through an API.

Second, the industry must implement and regulators must enforce a digital code of conduct. Where regulators lack explicit authority or the First Amendment prohibits them from telling companies what to do, Congress or the FTC can drive platforms to clean up their acts with watchful oversight and enforcement. Such a code should include commitments such as:

- Eliminating design features, such as automatic group recommendations, that provide turn-key solutions for radicalizers.
- Using "circuit breakers" to prevent quick viral spread of radicalizing content, while human reviewers determine whether the content violates platform policies or poses a risk to public safety. [20]
- Best practices to avoid linking to platforms that consistently permit illegal and terrorist activity. This would prevent violent extremists from organizing on smaller platforms and then using the major tech platforms to spread their content—including so-called "manifestos" purporting to explain and justify acts of violence.

Any violation of the code could be enforced as an FTC Act consumer protection violation. Conditioning Section 230 immunity on companies following a robust code of conduct could provide further incentives to adopt a code.

It is essential that we act to protect our country from violent extremism and divisive hate. The Facebook papers provide a telling look in the rearview mirror. But the line between our offline and online interactions is disappearing. As we move to Internet 3.0, interacting through our avatars, it's essential that we build in protections against further radicalization and violent extremism.

[20] Ellen Goodman and Karen Kornbluh, "Social Media Platforms Need to Flatten the Curve of Dangerous Misinformation," Slate, August 21, 2020.