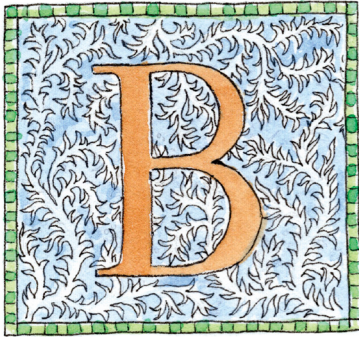


JONATHAN A. GREENBLATT

Complicit: Big Tech and Antisemitism



BEFORE Robert Bowers was a convicted mass murderer, he was a middle-aged truck driver who mostly kept to himself. In the 1990s, he became enthralled with talk radio. He was particularly drawn to the *Quinn in the Morning* show, whose host, Jim Quinn, ranted that Islam was the source of all our problems. Bowers listened, alone.

Years later, Bowers discovered the fringe online platform Gab. In this forum, he shared materials from the Christian Identity movement, a racist and antisemitic religious ideology popular in extreme right-wing circles. He fell deep into the rabbit hole of conspiratorial thinking. In particular, he adopted the “great replacement” theory, which posits that Jews are manipulating world events to bring more non-white people to Western countries to replace white people.

He posted comments including “Jews are the children of Satan”

and “Diversity means chasing down the last white man.” But this time, instead of being alone with his extreme beliefs, he found others to validate and encourage him.

Unlike in generations past, Bowers didn’t need to wear a white hood at Klan rallies in hidden forests when he made these claims. He could speak his mind at online “rallies” 24/7.

One morning, he reached a breaking point. He took out his phone, went to Gab, and wrote: “HIAS likes to bring invaders in that kill our people. I can’t sit by and watch my people get slaughtered. Screw your optics. I’m going in.” He drove to the Tree of Life Synagogue in Pittsburgh with multiple firearms and murdered 11 worshippers as they sat in the pews.



The same conspiracy theories that motivated Bowers still swirl around the muddy drains of the internet. In the nearly five years since the massacre, mass murderers have engaged similarly on Gab and other fringe platforms, trading information with echoes of replacement theory and posting their own manifestos. These men who became indoctrinated online went out and targeted marginalized communities from Buffalo to El Paso to Christchurch, New Zealand.

For those who already hate, online platforms like Gab mean they’re only a few clicks away from feeding the habit. The naïveté of those who are unaware of hateful ideologies can leave them vulnerable to radicalization.

My colleagues at the Anti-Defamation League report that anti-semitic content has become the norm rather than the exception on social-media services. It festers on every platform we monitor.

The very worst offenders can be found on niche services such as Telegram, Gab, and 4chan. These sites are populated by a small share

of social-media users, but a far larger share of the extremist community. The rhetoric on these sites demonstrates the depth and breadth of the challenges to addressing online antisemitism.

The founder of Gab, Andrew Torba, is a self-described Christian nationalist who claims that Jewish Americans have dual loyalty to the United States and Israel, that Jews are to blame for the crucifixion of Jesus, and that they control the U.S. government. Asked after the Pittsburgh massacre whether he would make any changes to the site's policies, Torba responded, "Absolutely not."

Some think these sites are essentially self-contained: magnets for extremists but nothing more. Our research shows that the influence of these outlets goes far beyond the platforms themselves. Then there are the large mainstream social-media companies, which in recent years have made solemn promises to do their utmost to remove hateful content on their own sites. But they aren't keeping them. Facebook and YouTube have reversed policies on curbing disinformation. Twitter was never the model for addressing intolerance even before Elon Musk acquired the platform; things have only gotten worse since then. A 2021 report from the Center for Countering Digital Hate found that the five leading social-media companies (Facebook, Twitter, Instagram, YouTube, and TikTok) failed to remove 84 percent of antisemitic posts—and those were just the ones that had been *flagged* by the tools these companies use to alert content moderators to problematic content.

In their defense, the big social-media companies say that, try as they might, it's impossible to police all the content on their sites, given the sheer volume of it. It's certainly true that there are obstacles in the way of combating online hate and antisemitism. Content moderation is a game of Whac-A-Mole. For example, several years ago, antisemites began using an "echo"—three parentheses bracketing a word—to refer to Jewish individuals. Content-moderation systems could take

down every instance of the echo, but that would also sweep up educational posts sharing what the echo means as well as posts by Jews using it to proudly self-identify in the face of hate. Automated content-moderation systems must be updated constantly to accommodate the shifting language and context of hate.

Another obstacle to blocking hateful online content is the tactic popular among social-media “influencers” who deliberately evade moderation by weaponizing talking points that incite *others* to harassment. The problem is particularly rampant on Twitter. Twitter, under Musk’s leadership, focuses on holding individual accounts responsible for harmful content. As a result, it frequently misses how influential accounts with large followings operate. When influential accounts become hubs of hate for other online users, antisemitic content proliferates. For example, when the far-right activist Ali Alexander tweeted about ADL, his followers replied with overt antisemitism. The platform is a case study in what the ADL calls stochastic harassment: a user “weaponizing talking points that incite others to harassment without being a harasser.”

Ultimately, companies lack adequate incentives to dedicate serious resources to overcoming such systemic obstacles. Firms face few consequences, financial or otherwise, for hosting and amplifying hate and harassment. Hateful content drives engagement, and engagement drives advertising revenue. Moreover, major social-media companies have gutted their trust-and-safety teams, despite claims to prioritize user safety.



Some critics of content moderation on social media insist that platforms provide a public service that advances free speech and that any curbs on the rights of people to say whatever they want violate

the spirit of the First Amendment. But the First Amendment doesn't apply to all forms of speech online—just as it doesn't apply to all forms of speech in the real world.

The exact moment when speech crosses the line from protected expression to harassment or threat is sometimes a nuanced one: There is a vast amount of speech that, while controversial and unpopular, is considered “awful, but lawful” and, therefore, safeguarded under the First Amendment. However, while some mistakenly contend that freedom of speech is an absolute right with no exceptions, there are several forms of speech that do not enjoy constitutional protection: true threats, incitement to imminent lawless action, defamation, speech integral to criminal conduct, and child pornography.

It is also essential to note that the First Amendment's restriction on abridging speech applies only to governmental actors. Although they are not legally mandated to do so, platforms can, and often do, implement robust policies against hateful speech and conduct in the same way that offline institutions always have. As private actors, social-media platforms are not bound by the First Amendment; in fact, courts have even understood platforms' moderation efforts to be protected speech in and of themselves.

If a person were to walk into a Starbucks and start yelling antisemitic epithets at the patrons, that person would surely be kicked out, because Starbucks, as a private company, can set rules for conduct in its stores. The same should hold true for the extremist who spews racist rhetoric or makes hateful antisemitic threats on privately owned online platforms.

Likewise, newspapers in America are shielded by their First Amendment right to criticize the government and public officials. But they aren't obliged to print anything. Editors select the forms of speech they want to platform and exclude the ones they don't—for instance, a letter from a white supremacist advocating a race war.

This is not censorship. It's a matter of maintaining editorial standards. The same must hold true for social-media companies that currently recommend and amplify content from white supremacists and other bigots.

These choices about what is and isn't permissible aren't always easy. For every neo-Nazi, there may be thousands more typing and posting opinions that, even if we disagree strongly with them, fall within the range of what ought to be considered acceptable speech. That's why it's essential for social-media companies to work with experts from civil society to parse nuance and understand how extremist behavior is changing and how evolving rhetoric affects targeted groups. Twitter used to be one model of this with their Trust and Safety Council, which Musk disbanded.

The key point is this: Freedom of speech does not mean *freedom of reach*. Social-media platforms are not obligated to provide a platform for bigots to spread hateful speech. They are certainly not obligated to amplify those messages using algorithms designed to generate interest among the like-minded. These are conscious choices made in the pursuit of the bottom line, not constitutional freedoms exercised for the greater good.

There is also a supposed question of cost: How can companies such as Facebook be expected to moderate the tidal wave of content being generated every hour on their platforms? The question is a little like asking why major automakers should be expected to produce safe and reliable cars given the many thousands they produce. Safety is part of the cost of doing business in every modern industry, and social-media companies — some of the most profitable firms in history — should hardly be exempt.

Nor should they be exempt from liability when things go wrong. When a large carmaker finds that its airbags are faulty, it is mandated by the government to recall its product and repair the prob-

lem. When a fast-food restaurant has a norovirus outbreak, it shuts down, updates its procedures, and pays a hefty fine. This is basic for public health. Social-media companies need to be held to the same commonsense standards.



If social-media companies are still unwilling to make changes, the advertising industry and nongovernmental organizations will need to once again step into the void. We've done this before, with the Stop Hate for Profit campaign launched with other NGOs in July 2020 to send a message that social-media companies need to be held accountable.

But awareness campaigns and public pressure won't be enough. Policymakers at the federal and state level must reshape their incentives to force behavior change. Two steps in particular could make a dramatic difference.

- The Communications Decency Act, which governs how social-media companies operate, was passed in 1996—before iPhones and Apple watches, Twitter and TikTok, and long before the age of artificial intelligence and synthetic media. As currently written, Section 230 of the Act provides platforms such as Facebook and Twitter with near-blanket immunity from liability for “user-generated content” published on their platforms, with few exceptions. In essence, unlike all other forms of media in our society, these companies are not liable even if they publish libel.

The existing law has become woefully insufficient to regulate tech companies and prevent platforms' ranking algorithms from recommending dangerous content. Section 230 must be updated to account for the reality that the platforms are exacerbating hate,

harassment, and extremism. Just as seat belts did not prevent people from reaching their destination, making social-media companies accountable for what they publish will reduce harm without sacrificing the connective power of the platforms.

- The regulatory toolbox for this issue should also include government-mandated transparency.

We know that social-media companies can't be trusted to regulate themselves. Companies should be required to disclose how they are actually enforcing content policies—the same kind of consumer protection we see in other industries. California signed an excellent model for this into law in 2022. California's A.B. 587 forces large social-media companies to publicly disclose their platform policies regarding online hate, racism, disinformation, extremism, harassment, and foreign political interference. It also mandates that they release data about their enforcement of those policies. The law doesn't require any content policies at all. Its premise is simple: We should know what social-media platforms' policies *actually* are and how well they're enforced. Federal legislation would be the most effective means to ensure transparency, but if Congress doesn't step in, more states could take action. At present, California, Florida, Texas, and New York have laws of this kind on the books.

American Jews and our friends and allies across the country are looking with alarm: The rate of antisemitic incidents rose nearly fivefold between 2013 and 2022, from 751 to 3,697. Nobody should think it's a coincidence that this dramatic increase coincided with the increasing ubiquity and influence of social media—a type of media that has for too long gotten away with maximizing its profits by minimizing its responsibilities.

It doesn't have to be this way. And we shouldn't have to wait for the next social-media-induced racist or antisemitic massacre before we act. *