

ANGRY BY DESIGN

A report by Luke Munn / Digital Cultures Institute

Contents

Summary

Background

Platform Analysis

Facebook

YouTube

Gab

Conclusion

References

Disclaimer

Because of its engagement with hate speech, toxic communication online, and far-right cultures in particular, this report features terms and language that is hateful—racist, sexist, anti-Semitic, or otherwise non-inclusive and xenophobic. Citing such language is important to demonstrate the kinds of discourse circulating in these spaces, indeed the degree to which in some ways they become normalized. However, it should be stressed that such language in no way reflects the views of the author or of organisations linked with this report.

Summary

Hate speech and toxic communication online is on the rise. Responses to this issue tend to offer technical (automated) or non-technical (human content moderation) solutions, or simply see hate speech as a natural product of hateful people. In contrast, this project begins by recognizing platforms as designed environments that support particular practices while discouraging others. In what ways might these design architectures be contributing to polarizing, impulsive, or antagonistic behaviours? Three platforms are examined: Facebook, YouTube, and Gab. Facebook's Feed, based on engagement, drives views but also privileges incendiary content, setting up a stimulus-response loop that promotes outrage expression. YouTube's recommendation system is a key interface for content consumption, yet this same design has been criticized for leading users towards more extreme content. Gab's heavy focus on design provides its content with a modern interface and intuitive user experience, suggesting a new mainstreaming of toxic communication and extremist ideologies. Across all three platforms, design is central and influential, proving to be a productive lens for understanding toxic communication.

Background

Hate speech online is on the rise. Recent studies describe this rise statistically,¹ but stop short of analyzing its underlying conditions or suggesting potential interventions. The response to this rise has broadly taken two approaches to harm reduction on platforms. The first approach is technical, attempting to develop software and scripts to detect and remove problematic content. Indeed over the last few years in particular, a significant amount of attention has been directed at abusive speech online, with huge amounts of work poured into constructing and improving automated systems.² Articles in computer science and software engineering in particular often claim to have studied the failings of previous techniques and discovered a new method that finally solves the issue.³ And yet the inventiveness of users and the ambiguity of language mean that toxic communication remains complex and difficult to address. Technical understanding of this content will inevitably be limited, explains researcher Robyn Caplan, because automated systems are being asked to understand human culture—racial histories, gender relations, power dynamics and so on—“a phenomenon too fluid and subtle to be described in simple, machine-readable rules.”⁴

The second approach is non-technical, stressing that hate speech online is a problem that only humans can address. This framing, not incorrectly, points out that

¹ SafeHome, “Hate on Social Media,” SafeHome.org, February 3, 2017, <https://www.safehome.org/resources/hate-on-social-media/>.

Darcy Hango, “Cyberbullying and Cyberstalking among Internet Users Aged 15 to 29 in Canada” (Ottawa: Statistics Canada, December 19, 2016).

² John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos, “Deeper Attention to Abusive User Content Moderation,” in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark: Association for Computational Linguistics, 2017), 1125–35, <https://doi.org/10.18653/v1/D17-1117>.

³ Sanafarin Mulla and Avinash Palave, “Moderation Technique for Sexually Explicit Content,” in *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, 2016, 56–60, <https://doi.org/10.1109/ICACDOT.2016.7877551>.

Stéphan Tulkens et al., “The Automated Detection of Racist Discourse in Dutch Social Media,” *Computational Linguistics in the Netherlands Journal* 6 (December 1, 2016): 3–20.

Jean-Yves Delort, Bavani Arunasalam, and Cecile Paris, “Automatic Moderation of Online Discussion Sites,” *International Journal of Electronic Commerce* 15, no. 3 (April 1, 2011): 9–30, <https://doi.org/10.2753/JEC1086-4415150302>.

⁴ James Vincent, “AI Won’t Relieve the Misery of Facebook’s Human Moderators,” *The Verge*, February 27, 2019, <https://www.theverge.com/2019/2/27/18242724/facebook-moderation-ai-artificial-intelligence-platforms>.

automated interventions will always be inherently limited, unable to account for the nuances of particular contexts and the complexities of language. The response is to dramatically expand content moderation teams. In May of 2018, for example, Facebook announced that it would be hiring 10,000 new workers into its trust and safety team.⁵ However, the toll for those carrying out this kind of work, where hate speech, graphic images, and racist epithets must be carefully reviewed, is incredibly high, leading to depression and other mental health issues. As Alexis Madrigal stresses, in being forced to parse this material, workers “do not escape unscathed.”⁶ Moreover, there has recently been more attention paid to the pressures placed on employees to meet performance targets, pressures that only add to the inherent psychological toll.⁷

In addition to these two approaches, there also seems to be a popular assumption, evidenced in online comments and in more mainstream literature, that hate speech is the natural product of hateful people. One user I interviewed stated that the toxic comments she encountered online were simply produced by rude and frustrated people, perhaps with a difficult background or early life, who haven’t been taught general manners. Another blog post blames toxic communication on an inherently toxic individual, someone with a predilection for hating or bullying, racism or sexism.⁸ In this understanding, hate speech results from people translating their fundamental nastiness in the offline world into the online environment.

In contrast to the approaches and assumptions discussed above, this study adopts a design-centric approach. It seeks to understand how hate might be facilitated in particular ways by hate-inducing architectures. Just as the design of urban space influences the practices within it, the design of platforms, apps and technical environments shapes our behaviour in digital space. How might the design of technical environments be promoting toxic communication? This project examined

⁵ James Freeman, “Facebook’s 10,000 New Editors,” *Wall Street Journal*, May 16, 2018, sec. Opinion, <https://www.wsj.com/articles/facebooks-10-000-new-editors-1526491169>.

⁶ Alexis C. Madrigal, “The Basic Grossness of Humans,” *The Atlantic*, December 15, 2017, <https://www.theatlantic.com/technology/archive/2017/12/the-basic-grossness-of-humans/548330/>.

⁷ Casey Newton, “Three Facebook Moderators Break Their NDAs to Expose a Company in Crisis,” *The Verge*, June 19, 2019, <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa>.

⁸ Grace Jennings-Edquist, “Abusive Text Messages and Mobile Harrassment Are on the Rise,” *Mamamia*, November 22, 2014, <https://www.mamamia.com.au/abusive-text-messages/>.

three notable platforms: Facebook, YouTube, and Gab. Each platform has userbases ranging from the thousands to the millions. Each platform has a global reach, with access available in hundreds of countries worldwide. And each has been linked in some way to hate speech, toxic communication online, or even acts of physical violence in the “real world.” All of these platforms, then, are highly influential in their own ways, shaping the beliefs and ideologies of individuals, their media production and consumption, and their relations to others on an everyday basis. A design analysis of each of these platform comprised the bulk of this project, while an interview with a platform user and an interview with an online community manager supplemented this core analysis.

While this method is novel in some ways, the attention to the design of platforms and their potential to shape behaviour is not unprecedented. In fact, over the last few years, there has been an almost confessional moment from designer and developers of platforms. These designers have admitted that their systems are addictive⁹ and exploit negative “triggers.”¹⁰ Others have spoken about their tools “ripping apart the social fabric of how society works.”¹¹ Facebook’s design privileges base impulses rather than considered reflection.¹² Studies have demonstrated that social media functionality enables negative messages to be distributed farther and faster,¹³ and that such affordances enable anger to spread contagiously.¹⁴ The “incentive structures and social cues of algorithm-driven social media sites” amplify the anger of users over

⁹ Bianca Bosker, “The Binge Breaker,” *The Atlantic*, November 2016, <https://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/>.

¹⁰ Paul Lewis, “‘Our Minds Can Be Hijacked’: The Tech Insiders Who Fear a Smartphone Dystopia,” *The Guardian*, October 6, 2017, sec. Technology, <https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia>.

¹¹ James Vincent, “Former Facebook Exec Says Social Media Is Ripping Apart Society,” *The Verge*, December 11, 2017, <https://www.theverge.com/2017/12/11/16761016/former-facebook-exec-ripping-apart-society>.

¹² Bianca Bosker, “The Binge Breaker,” *The Atlantic*, November 2016, <https://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/>.

¹³ Soroush Vosoughi, Deb Roy, and Sinan Aral, “The Spread of True and False News Online,” *Science* 359, no. 6380 (March 9, 2018): 1146–51, <https://doi.org/10.1126/science.aap9559>.

¹⁴ Rui Fan, Ke Xu, and Jichang Zhao, “Higher Contagion and Weaker Ties Mean Anger Spreads Faster than Joy in Social Media,” *ArXiv:1608.03656 [Physics]*, August 11, 2016, <http://arxiv.org/abs/1608.03656>.

time until they “arrive at hate speech.”¹⁵ In warning others of these negative social effects, designers have described themselves as canaries in the coal mine.¹⁶

Indeed, we have already begun witnessing the darker edge to these platforms. Far-right shootings in El Paso and Christchurch have been linked to users on sites like Gab and 8chan. Ethnic violence enacted against the Rohingya has been closely connected to material circulating on more mainstream platforms like Facebook.¹⁷ These overt acts of hate in the “real world” materialize this overlooked issue and highlight its significant stakes. A key point here is to see toxic communication not just as a nuisance or a nasty byproduct of online environments, but as establishing conditions with implications for human rights. As David Kaye writes in a recent report to the U.N. “Online hate is no less harmful because it is online,” stresses David Kaye in a recent U.N. report: “To the contrary, online hate, with the speed and reach of its dissemination, can incite grave offline harm and nearly always aims to silence others.”¹⁸ Hate forms a broad spectrum with extremist ideologies and hateful worldviews at one end. As explored in my own work, the gradual amplification of hate creates a potential pipeline for alt-right radicalization.¹⁹ In this respect, the violent outpouring of hate witnessed over the last few years in particular is not random or anomalous, but a “logical” result of individuals who have spent years inhabiting hate-filled spaces, where a deluge of racist, sexist, and anti-Semitic content normalized such views.

Very recently, then, a new wave of designers and technologists have begun thinking about how to redesign platforms to foster calmer behaviour and more civil discourse.

¹⁵ Max Fisher and Amanda Taub, “How Everyday Social Media Users Become Real-World Extremists,” *The New York Times*, October 10, 2018, sec. World, <https://www.nytimes.com/2018/04/25/world/asia/facebook-extremism.html>.

¹⁶ Tatiana Mac, “Canary in a Coal Mine: How Tech Provides Platforms for Hate,” *A List Apart* (blog), March 19, 2019, <https://alistapart.com/article/canary-in-a-coal-mine-how-tech-provides-platforms-for-hate/>.

¹⁷ Alexandra Stevenson, “Facebook Admits It Was Used to Incite Violence in Myanmar,” *The New York Times*, November 6, 2018, sec. Technology, <https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html>.

¹⁸ David Kaye, “Governments and Internet Companies Fail to Meet Challenges of Online Hate – UN Expert,” OHCHR, October 9, 2019, <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25174&LangID=E>.

¹⁹ Luke Munn, “Algorithmic Hate: Brenton Tarrant and the Dark Social Web,” Institute of Network Cultures, March 19, 2019, <http://networkcultures.org/blog/2019/03/19/luke-munn-algorithmic-hate-brenton-tarrant-and-the-dark-social-web/>.

Luke Munn, “Alt-Right Pipeline: Individual Journeys to Extremism Online,” *First Monday* 24, no. 6 (June 1, 2019), <https://doi.org/10.5210/fm.v24i6.10108>.

How might design create ethical platforms that enhance users wellbeing?²⁰ Could technology be designed in a more humane way?²¹ And what would the core principles and processes of such a design look like?²² Identifying a set of hate-promoting architectures would allow designers and developers to construct future platforms that mitigate communication which is used to threaten, harass, or incite harm, and instead construct more inclusive, affirmative, and more broadly democratic environments.

“Angry By Design” picks up on this nascent work, tracing the relationship between technical architectures and toxic communication. Three distinct platforms are examined: Facebook, YouTube, and Gab. How does Facebook’s privileging of metrics influence the intensity of content that gets shared? How does YouTube’s recommendation engine steer users towards more radical, right-wing content? And how does Gab’s sophisticated design provide a more mainstreamed, user-friendly vehicle for hate? This project represents an early set of steps towards investigating these issues. It examines the role of design in each platform, identifies several problematic design features, and suggests a number of civil alternatives.

²⁰ Lu Han, “Designing for Tomorrow - A Discussion on Ethical Design,” Spotify Design, January 18, 2019, <https://spotify.design/articles/2019-01-18/designing-for-tomorrow-a-discussion-on-ethical-design/>.

²¹ Tristan Harris, “Humane: A New Agenda for Tech,” Center For Humane Technology, April 23, 2019, <https://humanetech.com/newagenda/>.

²² Jon Yablonski, “Humane by Design,” 2019, <https://humanebydesign.com>.

Platform Analysis

Facebook

Facebook is the giant of social media. With 2.41 billion active users worldwide, it is the largest platform, and arguably one of the most significant.²³ On average, users spend 58 minutes every day on the platform.²⁴ While some signs indicate that the platform is plateauing in terms of use, these statistics remain compelling and mean that it cannot be overlooked. From the perspective of this project, Facebook is a technically mediated environment where vast numbers of people spend significant amounts of time. Moreover, if the platform is influential, it is also increasingly recognised as detrimental. “As Facebook grew, so did the hate speech, bullying and other toxic content on the platform,” notes one recent article, “when researchers and activists in Myanmar, India, Germany and elsewhere warned that Facebook had become an instrument of government propaganda and ethnic cleansing, the company largely ignored them. Facebook had positioned itself as a platform, not a publisher.”²⁵ What kinds of experiences are all of these users having, and how does the design of this environment contribute to this? Rather than calm and civil, this analysis will show how the platform’s affordances can induce experiences that are stressful and impulsive, establishing some of the key conditions necessary for angry communication.

A design approach to Facebook foregrounds it as a set of concrete design decisions made over time. For users, Facebook appears as a highly mature and highly refined environment. Every area has undergone meticulous scrutiny and crafting by teams of developers and designers. This provides the environment with a degree of stability and authority, even inevitability. In this sense, giants like Facebook claim a kind of defacto standard: this is the way our communication media operates. Yet Facebook has evolved significantly over time. Launching in 2004, the site was billed as an

²³ Dan Noyes, “Top 20 Facebook Statistics - Updated July 2019,” *Zephoria Inc.* (blog), July 24, 2019, <https://zephoria.com/top-15-valuable-facebook-statistics/>.

²⁴ Rani Molla and Kurt Wagner, “People Spend Almost as Much Time on Instagram as They Do on Facebook,” *Vox*, June 25, 2018, <https://www.vox.com/2018/6/25/17501224/instagram-facebook-snapchat-time-spent-growth-data>.

²⁵ Sheera Frenkel et al., “Delay, Deny and Deflect: How Facebook’s Leaders Fought Through Crisis,” *The New York Times*, November 14, 2018, sec. Technology, <https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html>.

“online directory”; in these early days, the site emulated the approach of MySpace, where each user had a profile, populated with fields for status, education, hobbies, relationships, and so on; in 2007, Facebook added a Mini-Feed feature that listed recent changes to friends profiles, and in 2011 Facebook released the Timeline that “told the story of your life” as a move away from the directory or database structures of the past.²⁶ Rather than inevitable, then, the design evolution of Facebook reminds us that it has evolved through conscious decisions in response to a particular set of priorities.



Early screenshot from ‘The Facebook’ indicating its significant design progression over time

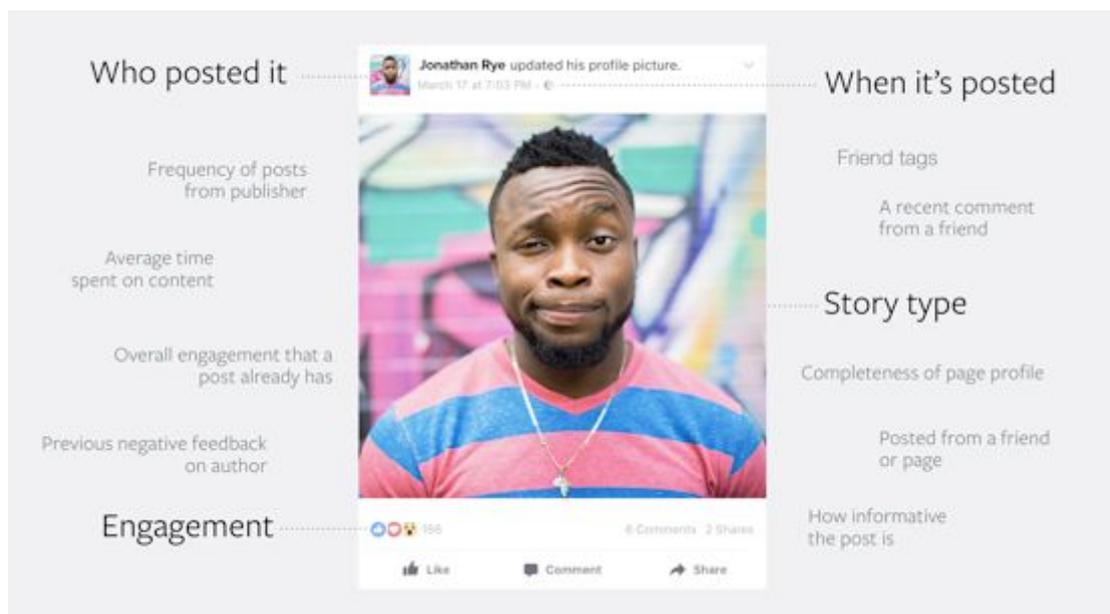
Design wise, the Feed remains one of the key pieces of functionality within Facebook. The Feed, or the News Feed as it is officially known, is described the company as a “personalized, ever-changing collection of photos, videos, links, and updates from the friends, family, businesses, and news sources you’ve connected to on Facebook.”²⁷ It is the first thing that users see when bringing up the app or entering the site. It is the center of the Facebook experience, the core space where content is presented to users. What’s more, because user actions are primed by this content and linked to it—whether commenting on a post, sharing an event, or liking a status update—the Feed acts as the gateway for most user activity, structuring the actions they will

²⁶ Chloe Albanesius, “10 Years Later: Facebook’s Design Evolution,” PCMag Australia, February 4, 2014, <https://au.pcmag.com/internet-2/12249/10-years-later-facebooks-design-evolution>.

²⁷ Facebook, “News Feed,” News Feed | Facebook Media, 2019, <https://www.facebook.com/facebookmedia/solutions/news-feed>.

perform during that particular session. Indeed, for many users, Facebook is the Feed and the Feed is Facebook.²⁸

Key to the Feed is the idea of automatic curation. Before the Feed, users would have to manually visit each one of their friend's profile pages in order to discover what had changed in his or her life. Once introduced, the Feed now carries out this onerous task for each user. "It hunts through the network, collecting every post from every connection — information that, for most Facebook users, would be too overwhelming to process themselves."²⁹ In this sense, the Feed provides both personalisation and convenience, assembling a list of updates and bringing them together into a single location. Yet from a critical design perspective, this begs some obvious questions. What is prioritized in this Feed, bubbling to the top of view and clamoring for a user's attention? What is deemphasized, only appearing after a long scroll to the bottom? And what are the factors that influence this invisible curation work? In short: what is shown, what is hidden, and how is this decided?



Graphic provided by Facebook listing some of the criteria used by its News Feed

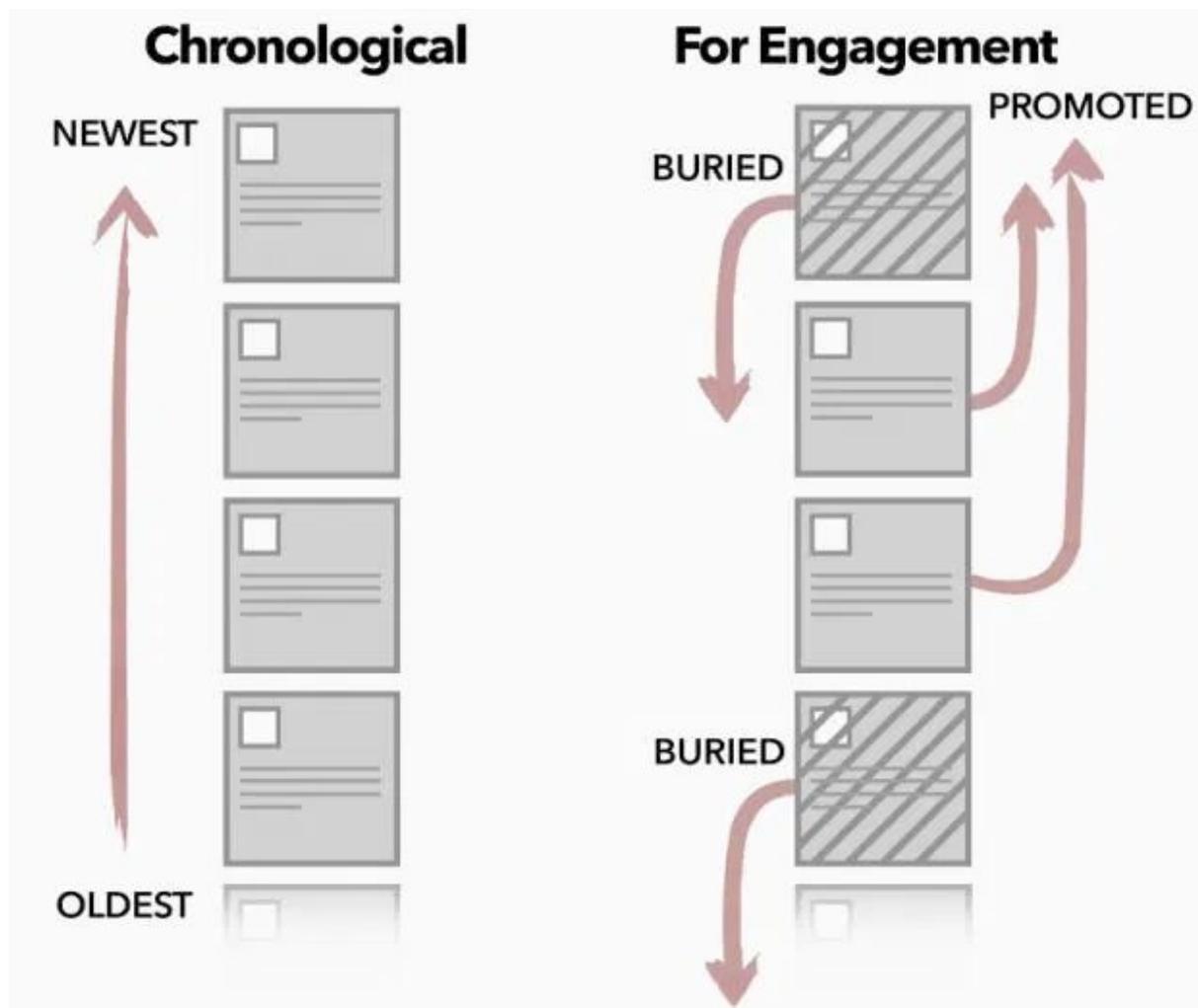
²⁸ Farhad Manjoo, "Can Facebook Fix Its Own Worst Bug?," *The New York Times*, April 25, 2017, sec. Magazine,

<https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html>.

²⁹ Farhad Manjoo, "Can Facebook Fix Its Own Worst Bug?," *The New York Times*, April 25, 2017, sec. Magazine,

<https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html>.

The Feed runs on its own logic. Since 2009, stories are not sorted chronologically, simply listing the updates friends have made in reverse order.³⁰ While this change induced a degree of backlash from users, the chronology itself proved to be overwhelming, especially with the hundreds of friends that each user has. “If you have 1,500 or 3,000 items a day, then the chronological feed is actually just the items you can be bothered to scroll through before giving up,” explains analyst Benedict Evans, “which can only be 10% or 20% of what’s actually there.” Instead, the Feed is driven by Engagement. In this design, Facebook weighs dozens of factors, from who posted the content to their frequency of posts and the average time spent on this piece of content. Posts with higher engagement scores are included and prioritized; posts with lower scores are buried or excluded altogether.



³⁰ Wallaroo Media, “Facebook News Feed Algorithm History,” *Wallaroo Media* (blog), July 3, 2019, <https://wallaroomedia.com/facebook-newsfeed-algorithm-history/>.

Diagram from Rose-Stockwell showing the change in content prioritization

The problem with such sorting, of course, is that incendiary, polarizing posts that provoke a reaction consistently achieve high engagement. This material often has a strong moral charge. It takes a controversial topic and establishes two sharply opposed camps, attempting to show the superiority of one belief or group and condemn the other. These are the headlines and imagery that leap out at a user as they scroll past, forcing them to come to a halt. This content is designed to draw engagement, to provoke a reaction—and inducing outrage is a consistent way to achieve this imperative. “Emotional reactions like outrage are strong indicators of engagement,” observes Tobias Rose-Stockwell, “this kind of divisive content will be shown first, because it captures more attention than other types of content.”³¹ This offensive material hits a nerve, inducing a feeling of disgust, shame, or outrage. While speculative, perhaps sharing this content is a way to offload these feelings, to remove their burden on us individually by spreading them across our social network and gaining some sympathy or solidarity. The design of Facebook means that this forwarding and redistribution is only a few clicks away. As one participant in this study stated, “it is so easy to share stuff.” Moreover, the networked nature of social media amplifies this single response, distributing it to hundreds of friends and acquaintances. They too receive this incendiary content and they too share, inducing what Rose-Stockwell calls “outrage cascades — viral explosions of moral judgment and disgust.”³² Outrage does not just remain constrained to a single user, but proliferates, spilling out to provoke other users and appear in other online environments.

At its worst, then, Facebook’s Feed stimulates the user with outrage inducing content while also enabling its seamless sharing, allowing such content to rapidly proliferate across the network. In increasing the prevalence of such content and making it easier to share, it becomes normalized. Outrage retains its ability to provoke engagement, but in many ways becomes an established aspect of the environment. For Molly Crockett, this is one of the keys to understanding the rise of hate speech online. As Crockett stresses, “when outrage expression moves online it becomes more readily

³¹ Tobias Rose-Stockwell, “Facebook’s Problems Can Be Solved with Design,” Quartz, April 30, 2018, <https://qz.com/1264547/facebook-problems-can-be-solved-with-design/>.

³² Tobias Rose-Stockwell, “Facebook’s Problems Can Be Solved with Design,” Quartz, April 30, 2018, <https://qz.com/1264547/facebook-problems-can-be-solved-with-design/>.

available, requires less effort, and is reinforced on a schedule that maximizes the likelihood of future outrage expression in ways that might divorce the feeling of outrage from its behavioural expression.” Design, in this sense, works to reduce the barrier to outrage expression. Sharing a divisive post to an audience of hundreds or thousands is just a click away.

How might the Feed be redesigned? Essentially there are two separate design problems here. Firstly, there is the stimulus aspect—the content included in the Feed. While the Feed’s filtering operations undoubtedly remain highly technical, its logics can be understood through a design decision to elevate and amplify “engaging” content. Facebook has admitted that hate speech is a problem and has redesigned the Feed dozens of times since its debut in an effort to curtail this problem and the broader kind of misinformation that often stirs it up.³³ But the core logic of engagement remains baked into the design of the Feed at a deep level. Design, then, might start by experimenting quite concretely with different kinds of values. If the hyperlocal was privileged, for example, then posts from friends or community members in a 5 kilometer radius might only be shown. This would be more mundane in many ways—everyday updates from those in our immediate vicinity rather than vicious attacks from anyone in a friend network. Or following the success of more targeted messaging apps like Messenger and WhatsApp, the Feed might emphasize close familial or friend connections above all. This pivot to a more intimate relational sphere would certainly be quieter and less “engaging”, but ultimately more meaningful and civil.

Secondly, there is the response aspect—the affordances that make outrage expression online more effortless. Such expression is often impulsive, done in the moment, and so one possible design focus would be time itself. Rather than an instant reaction, would a built-in delay add a kind of emotional weight to such an action? An interval of a few seconds, even if nominal, might introduce a micro-reflection and suggest an alternative response. As a means of combating the effortless and abstract nature of outrage expression, Rose-Stockwell suggests a number of humanizing prompts that might be designed into platforms: an “empathetic prompt” that asks whether a user really wants to post hurtful content; an “ideological prompt” that stresses how this

³³ Wallaroo Media, “Facebook News Feed Algorithm History,” *Wallaroo Media* (blog), July 3, 2019, <https://wallaroomedia.com/facebook-newsfeed-algorithm-history/>.

post will never be seen by those with opposing viewpoints; and a “public/private prompt” that would allow disagreements to take place between individuals rather than in the pressurized public arena.³⁴ Such design interventions, while clearly not silver bullet solutions, might each contribute in their own small way towards a more civil and less reactive online environment.

YouTube

YouTube remains a juggernaut of online spaces. Recently, it crossed the threshold of 2 billion logged-in users per month.³⁵ Perhaps even more important for this research project is the time spent by users within this environment. Users spend around 250 million hours on the video sharing platform every day.³⁶ The time “inhabiting” YouTube marks it out as distinct from Facebook, and suggests a different kind of influence over time, something slower and more subtle. Indeed, as will be discussed, alt-right individuals have noted how influential YouTube was in shifting their worldviews over longer periods of time, a medial pathway that nudged them towards a deep-seated anger and a more extremist stance. While this is just one highly politicized facet of YouTube, it signals the stakes involved here—not only the anger available to be tapped into, but the influence such an environment might have in shaping the ideologies of its vast population.

One key focus of recent critiques of YouTube has been its recommendation engine. The recommendation system is central to the user experience on YouTube. Firstly, it determines the content of each user’s homepage. Upon arriving on the site, each user is presented with rows of recommended videos, with each row representing an interest (e.g. gaming), channel (e.g. the Joe Rogan Experience), or an affiliation (‘users who watched X enjoyed Y’). As with similar designs such as Netflix, the YouTube

³⁴ Tobias Rose-Stockwell, “Facebook’s Problems Can Be Solved with Design,” Quartz, April 30, 2018, <https://qz.com/1264547/facebooks-problems-can-be-solved-with-design/>.

³⁵ Salim Saima, “YouTube Boasts 2 Billion Monthly Active Users, 250 Million Hours Watched on TV Screens Every Day,” *Digital Information World* (blog), May 4, 2019, <https://www.digitalinformationworld.com/2019/05/youtube-2-billion-monthly-viewers-250-million-hours-tv-screen-watch-time-hours.html>.

³⁶ Salim Saima, “YouTube Boasts 2 Billion Monthly Active Users, 250 Million Hours Watched on TV Screens Every Day,” *Digital Information World* (blog), May 4, 2019, <https://www.digitalinformationworld.com/2019/05/youtube-2-billion-monthly-viewers-250-million-hours-tv-screen-watch-time-hours.html>.

homepage is the first thing that users interact with, and the primary “jumping off” point for determining what to watch.

Secondly, the YouTube recommendation system determines the videos appear in the sidebar next to the currently playing video. By default, the Autoplay feature is turned on, meaning that these sidebar videos are queued to play automatically after the current video. This design feature means that, even if the user does nothing further, the next video in this queue will play. Even if the Autoplay feature has been turned on, this sidebar, with its dozens of large thumbnails, presents the most obvious gateway to further content. With a single click, a user can move onto a video which is related to the one they are currently viewing.

From a design perspective, the homepage and the sidebar form the crucial interfaces into content consumption. Search, while possible, is a manual process that requires more effort and has been deemphasized. Browsing recommended results, with its scrolling and tapping, provides a more frictionless user experience. It is unsurprising then, that “we’re now seeing more browsing than searching behavior,” stated one YouTube designer, “people are choosing to do less work and let us serve them.”³⁷ This shift has meant an even greater role for the recommendation engine. In theory, users can watch any video on the vast platform; in practice, they are encouraged towards a very specific subset of content. This is a single algorithmic system that exerts enormous force in determining what kinds of content users are exposed to and what paths they are steered down.

How is this recommendation system designed? In a paper on its high-level workings, YouTube engineers explain that it comprises two stages. In the first stage, “the enormous YouTube corpus is winnowed down to hundreds of videos” that are termed candidates.³⁸ These candidates are then ranked by a second neural network, and the highest ranked videos presented to the user. In this way, the engineers can be “certain that the small number of videos appearing on the device are personalized and

³⁷ Josh Lewandowski, 5 questions for YouTube’s lead UX researcher, interview by Amy Avery, February 2018, <https://www.thinkwithgoogle.com/advertising-channels/video/youtube-user-behavior-research/>.

³⁸ Paul Covington, Jay Adams, and Emre Sargin, “Deep Neural Networks for YouTube Recommendations,” in *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys ’16* (the 10th ACM Conference, Boston, Massachusetts, USA: ACM Press, 2016), 192.

engaging for the user.”³⁹ Based on hundreds of signals, users are presented with content that is attractive by design: hooking into their interests, goals, and beliefs. This recommendation engine is not static, but rather highly dynamic and updated in real-time. Your profile incorporates your history, but also whatever you just watched. As YouTube’s engineers explain, it must be “responsive enough to model newly uploaded content as well as the latest actions taken by the user.”⁴⁰ As content is consumed, an individual’s beliefs, ideologies and viewpoints are shaped and evolve in fundamental ways.

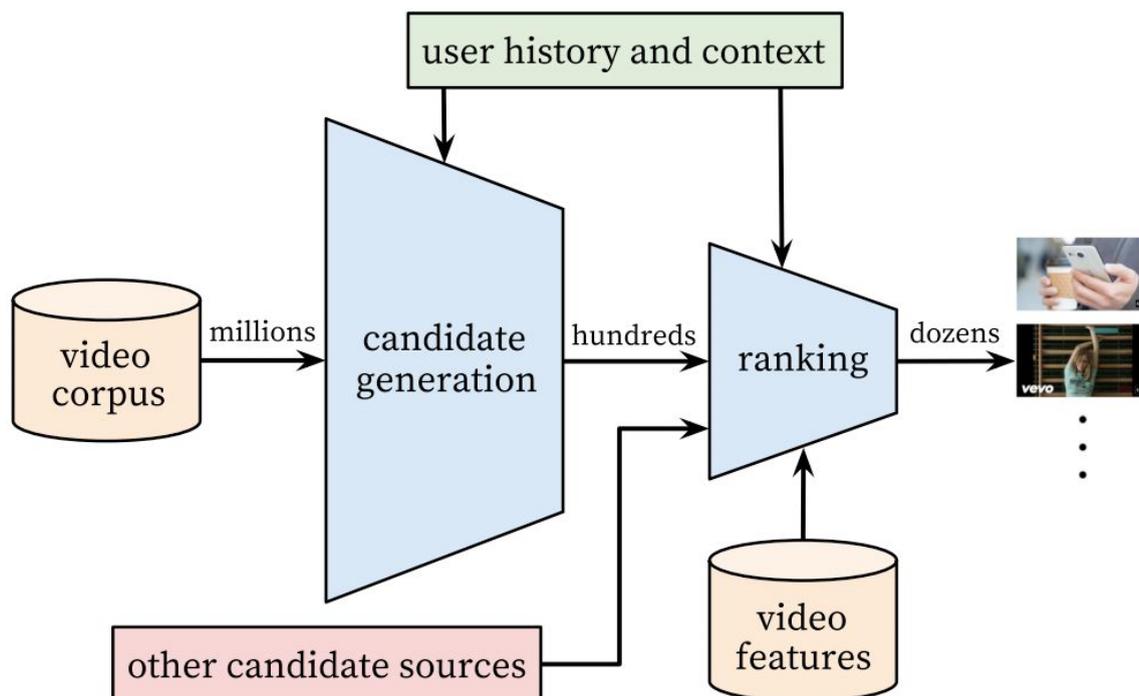


Diagram from YouTube engineers indicating how the recommendation system works

The result of these design choices is that the recommendation system emerges as a hate-inducing architecture. This is a system that appears to consistently suggest divisive, polarizing, and generally incendiary content. “YouTube drives people to the Internet’s darkest corners,” notes one *New York Times* article.⁴¹ As discussed, recommendations provide a seamless mechanism, allowing users to move to the next

³⁹ Covington et al., “Deep Neural Networks for YouTube Recommendations”, 192.

⁴⁰ Covington et al., “Deep Neural Networks for YouTube Recommendations”, 191.

⁴¹ Jack Nicas, “How YouTube Drives People to the Internet’s Darkest Corners,” *Wall Street Journal*, February 7, 2018,

<https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478>.

video easily or even automatically. The result seemingly gives both users and the platform what they want; delivering “relevant” content while ramping up view counts and minutes watched. And yet if this content stays within the same topic, it is typically more intense, more extreme. “However extreme your views, you’re never hardcore enough for YouTube” observes one journalist.⁴² YouTube’s recommendations often move from mainstream content to more incendiary media, or politically from more centrist views to right and even far-right ideologies. Recommendations are “the computational exploitation of a natural human desire: to look ‘behind the curtain,’ to dig deeper into something that engages us,” observes sociologist Zeynep Tufekci: “As we click and click, we are carried along by the exciting sensation of uncovering more secrets and deeper truths. YouTube leads viewers down a rabbit hole of extremism, while Google racks up the ad sales.”⁴³ In this sense, the design of the current recommendation serves the company well, but not necessarily individual users or online communities.

⁴² John Naughton, “However Extreme Your Views, You’re Never Hardcore Enough for YouTube,” *The Guardian*, September 23, 2018, <https://www.theguardian.com/commentisfree/2018/sep/23/how-youtube-takes-you-to-extremes-when-it-comes-to-major-news-events>.

⁴³ Zeynep Tufekci, “YouTube, the Great Radicalizer,” *The New York Times*, June 8, 2018, <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>.

HE'S TRASH?!  **Dave Chappelle Is CANCELLED?!**
Shawn Cee ✓
180K views
New

 **SJWs Try to Cancel Dave Chappelle! | Louder with...**
StevenCrowder ✓
656K views
New

 **Will Italy Be the Next Country to Leave the EU? (w...**
Real Vision Finance
37K views
New

 **Trolling the PRIDE MARCH!**
Hunter Avallone ✓
1.8M views

 **MIND FIELD S2 • E2**
The Psychedelic Experience - Mind Field S2 (Ep 2)
Vsauce ✓
Premium

 **Bill Maher Makes a New Rule About Fat Acceptance |...**
Every Damn Day Fitness ✓
46K views
New

 **Rating Memes from 0-10**
PewDiePie ✓
2.1M views
New

Screenshots showing anti-SJW (social justice warrior) and homophobic recommendations in response to viewing a centrist-right video by popular talk show host Joe Rogan

Based on the current values designed into the system, users are suggested material that progressively becomes more controversial, more political, more outrage-inducing, and in some cases, more explicitly racist, sexist, or xenophobic. Indeed, as my own

previous analyses have shown, YouTube can form a key part of an “alt-right pipeline”: users are incrementally nudged down a medial pathway towards more far-right content, from anti-SJW videos which demean so-called “social justice warriors,” to gaming related misogyny, conspiracy theories, the white supremacism of “racial realism”, and thinly veiled anti-Semitism.⁴⁴ What is particularly powerful about this design is its automatic and step-wise quality. Users do not consciously have to select the next video, nor jump into extreme material. Instead, there is a slow progression, allowing users to acclimate to these new beliefs before smoothly progressing onto the next step into their journey. At the far end of this journey is an angry and radicalized individual, a phenomenon we have unfortunately witnessed multiple times over the last year, from Christchurch in New Zealand to El Paso, Texas and Poway, California in the United States. Yet along with these extreme examples, equally troubling is the thought of a broader, more unseen population of users who are gradually being exposed to more hateful material.

Along with the recommendation engine, another problematic design element identified in this analysis is YouTube’s comment system. For years, YouTube has consistently held a reputation for being an environment with some of the most toxic and vitriolic comments online.⁴⁵ Even those used to online antagonism admitted that “you will see racist, sexist, homophobic, ignorant, and/or horrible comments on virtually every popular post”; and yet the same post from 2013 claims rather naively that the problem will soon be solved with some new technical features.⁴⁶ Far from being solved, the years since have seen toxic communication on the platform proliferate and take on concerning new forms. While regarded as a “cesspool” for over a decade, the latest indictment has been a large number of predatory and sexual comments on the videos of minors.⁴⁷

⁴⁴ Luke Munn, “Alt-Right Pipeline: Individual Journeys to Extremism Online,” *First Monday* 24, no. 6 (June 1, 2019), <https://doi.org/10.5210/fm.v24i6.10108>.

⁴⁵ Amelia Tait, “Why Are YouTube Comments the Worst on the Internet?,” *New Statesman*, October 26, 2016, <https://www.newstatesman.com/science-tech/internet/2016/10/why-are-youtube-comments-worst-internet>.

⁴⁶ Brent Rose, “YouTube Comments Will Soon Be Less Racist, Homophobic And Confusing,” *Gizmodo Australia*, September 25, 2013, <https://www.gizmodo.com.au/2013/09/youtube-comments-will-soon-be-less-racist-homophobic-and-confusing/>.

⁴⁷ Julia Alexander, “Can YouTube Fix Its Comment Section?,” *Polygon*, February 16, 2018, <https://www.polygon.com/2018/2/16/17020326/nikolas-cruz-youtube-comment-section>.

Why is YouTube so toxic, so angry? One explanation is that YouTube is simply one of the largest platforms. For some, its extremely broad demographic explains its trend towards the lowest common denominator in terms of intelligent, relevant commentary. Yet while the platform may have a large audience, there also seem to be clear design decisions exacerbating these toxic comments. “Comments are surely affected by *who* writes them,” admits one analysis, “but *how* a comment system is designed greatly affects *what* is written.”⁴⁸ For instance, YouTube comments can be upvoted or downvoted, but downvoting doesn’t lower the number of upvotes. This suggests a design logic that favors any kind of engagement, whether positive or negative. The result is that provocative, controversial, or generally polarizing comments seem to appear towards the top of the page on every video.



UnknownArchive 1 week ago

#1 fuck this feminist cunt and all the damage she did to thatguywiththeglasses.
#2 see #1

Reply · 15 👍 🗑️

Just one example from the many toxic comments on YouTube

The design choices built into both YouTube’s recommendation engine and its comment system might be understood as natural outcomes of an overarching set of company values. As recent articles have chronicled, YouTube as purposefully ignored warnings of its toxicity for years—even from its own employees—in its pursuit of one value: engagement.⁴⁹ Of course, this should come as no surprise for a publicly listed company driven by shareholder values and the broader dictates of capitalism. However it opens the question into what values are prioritized within online environments and how design supports them. Rather than grand vision statements or the aspirational company charters, what are the incentives built into platforms at the level of design: features, metrics, interfaces, affordances, and so on?

⁴⁸ Polymatter, “Why YouTube Comments Suck (and Reddit Comments Don’t),” December 15, 2016, https://www.youtube.com/watch?v=Lvf8kogX_yE.

⁴⁹ Mark Bergen, “YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant,” *Bloomberg.Com*, April 2, 2019, <https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>.

The community manager I interviewed underlined how the typical all-consuming focus on likes and shares could be damaging. Much of their work strives to foster healthy relations between members, to encourage beneficial content and block or demote toxic posts—in short, to facilitate “more of the good and less of the corrosive.” But her fellow community managers often speak of “algorithm chasing” where they attempt to combat or counteract the features built into the systems they use. There are often “competing logics” on a platform, she explained, an opposition between the value of creating a cohesive and civil community, and the values seen as necessary for platform growth and revenue such as expanding a userbase, extending use times, and attracting advertisers. Social media and community are often an awkward fit, and “marketing efficiencies are not social efficiencies.” On YouTube specifically, these designs privilege engagement above all else, resulting in a community that can be toxic and angry. Yet design might be rethought to prioritize an alternative set of values.

How might design contribute to a calmer, more considerate and more inclusive environment? One concrete intervention would be a redesigned recommendation system. Programmer and activist Francis Irving has found that the current system described earlier is both populist, prioritizing the popular, and short-term, using criteria to find videos that you’ll watch the longest.⁵⁰ What kind of design interventions would make it more conducive to user well-being? Irving suggests a number of possibilities. For one, its design might instead be based around happiness. Ask whether a YouTube user is more or less happy 6 months later, and use this signal as a way to improve video recommendations. As another alternative, Irving speculates about removing automated recommendations altogether, and moving to a more user-centered recommendation model. Like film or music, such a model would elevate taste makers who could curate great “playlists” of content.

Secondly, the comment system might be rethought entirely. It is clear that the current upvote/downvote binary is not working, rewarding quick immediate comments that are provocative—at best flippant, at worst, hateful or degrading. It also seems apparent that the relative anonymity of commenters and lack of any concept of

⁵⁰ Francis Irving, “Brainstorming a Better YouTube Recommendation Algorithm,” October 7, 2018, <https://www.flourish.org/2018/10/brainstorming-a-better-youtube-recommendation-algorithm/>

reputation means that there is no real disincentive for consistently generating toxic comments. “Each comment stands on its own, attached to nothing, bringing out the worst in every commenter.”⁵¹ Introducing a reputation system into this environment would be one concrete design intervention. Reddit, for example, features a Karma system that rewards high quality comments while docking points for comments against community guidelines. Such a system, while naturally not perfect, significantly “thickens” the identity of a user. Each user has a history of contributions and comments that persists over time. Based on this past behaviour, they have a status or level that has been awarded to them, a combined score that signals whether or not the community has found their contributions helpful, useful, or beneficial. Even if this score is mainly symbolic, these reputation systems hook into offline conventions of social standing within a community, introducing a degree of accountability.

⁵¹ Polymatter, “Why YouTube Comments Suck (and Reddit Comments Don’t),” December 15, 2016, https://www.youtube.com/watch?v=Lvf8kogX_yE.

Gab



Screenshot from Gab's funding page, where the company raised over a million dollars

The final platform to consider is Gab. In its own language, Gab is “a social network that champions free speech, individual liberty and the free flow of information online.”⁵² Gab emerged over the last few years as a direct response to the increased regulation of more mainstream social media platforms such as Facebook and Twitter. As hate speech and fake news became a target of media concern, these media platforms responded by hiring larger teams of content moderators, deleting posts that violated their community guidelines, and blocking or banning accounts that consistently went against their principles.

⁵² Gab, “Gab Social,” Gab Social hosted on gab.com, 2019, <https://gab.com/>.

Of course, it could be argued that such moderation is a case of too little, too late. The reticence of social media companies like Facebook to police content, even when warned, has been well documented.⁵³ Alt-right content in particular seems to have a privileged position in the moderation ecosystem. As one article noted, Facebook has adopted “a hands-off approach to flagged and reported content like graphic violence, hate speech, and racist and other bigoted rhetoric from far-right groups.”⁵⁴ Even when implemented, the effectiveness of automated or human-moderated measures against deep-seated cultures of intolerance, racism, and sexism is certainly questionable. Seriously tackling hate speech on these platforms would mean considering interventions that deprecated “engagement” in favour of other values. Such a fundamental “redesign” of not only user interfaces but also business imperatives is something that platforms have so far failed to undertake.

Yet even if the regulation so far by mainstream platforms has been light or ineffective, for those who founded Gab, it is entirely too much. For its founders and community, content moderation is equated with censorship. According to their rhetoric, this shift to a more regulated internet represents a threat to the democratic ideal of free speech. Posts are not taken down for ignoring guidelines, but for endorsing a controversial viewpoint, for adopting a stance that runs counter to mainstream cultural norms, or for asserting an uncomfortable “reality.” In short, a poster and his content is censored for speaking the truth.

Gab is a direct response to these conditions, proclaiming itself as one of the few bastions for free speech. It represents the latest incarnation of a lineage of right to far right spaces founded for the exact same reasons: 4chan, 8chan, and others. For these companies, the internet has become increasingly sanitized, scrubbed of any traces of “free thinking” and transformed into a dangerously homogenous environment. “The web is being shaped and influenced by a handful of companies with special interests

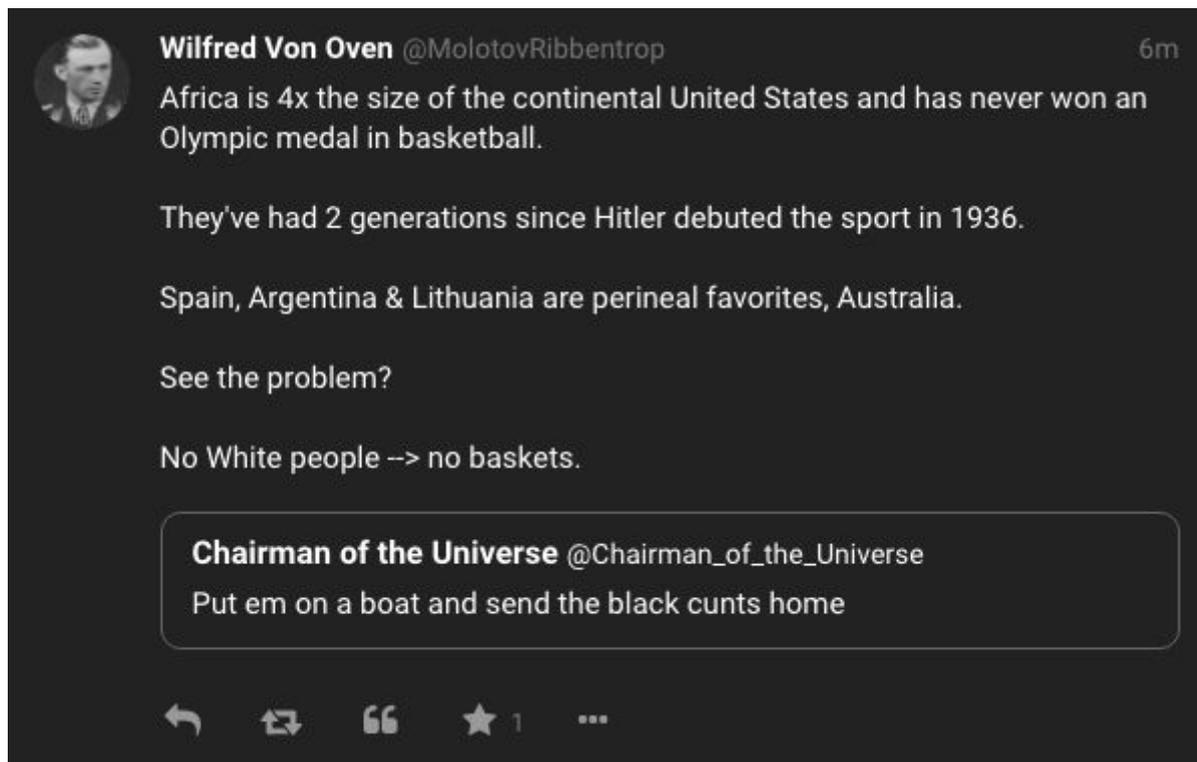
⁵³ David Morris, “Facebook Accused of Ignoring Government Warnings Before Mob Violence in Sri Lanka,” *Fortune*, April 22, 2018, <https://fortune.com/2018/04/22/facebook-ignored-sri-lanka-hate-speech/>.

See also: Sheera Frenkel et al., “Delay, Deny and Deflect: How Facebook’s Leaders Fought Through Crisis,” *The New York Times*, November 14, 2018, sec. Technology, <https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html>.

⁵⁴ Nick Statt, “Undercover Facebook Moderator Was Instructed Not to Remove Fringe Groups or Hate Speech,” *The Verge*, July 17, 2018, <https://www.theverge.com/2018/7/17/17582152/facebook-channel-4-undercover-investigation-content-moderation>.

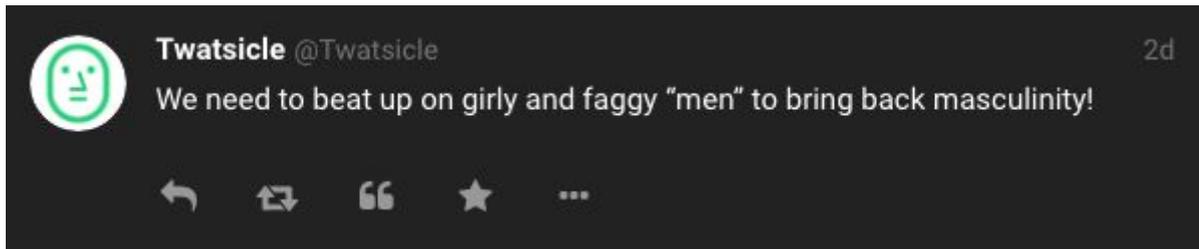
pushing a very specific agenda”, claims the copy on its funding page.⁵⁵ The monopolization and control exerted by a handful of Silicon Valley giants has produced a thoroughly totalitarian environment. In this imaginary, Gab is the people’s platform, an alternative communication channel where users voices can be heard, unshackled by stockholders and corporate mandates. “Gab is powered, operated, and funded by you The People,” stresses the same page, “not special interests or the old guard in Silicon Valley.”⁵⁶ This motivation is not just trivial but crucial in that these values inform the way the platform is designed.

On a practical level, Gab’s design blends many of the conventions found in other online platforms. The site opens on a news feed, a simpler version of that found in platforms such as Facebook. Posts are arranged reverse-chronologically, with the most recent appearing at the top. similar to the version pioneered by Facebook. These so-called Gabs can be replied to, reposted, or quoted, while clicking its creator takes you through to their profile. Per convention, profiles can be followed, prioritizing their content in feeds and alerting users whenever new content is posted.



⁵⁵ Andrew Torba, “Gab Investment Page,” StartEngine, 2018, <https://www.startengine.com/freespeech>.

⁵⁶ Andrew Torba, “Gab Investment Page,” StartEngine, 2018, <https://www.startengine.com/freespeech>.



Two examples of toxic comments found while browsing Gab online

Gab's design is sophisticated, especially when compared against other older homes for far-right content such as 4chan or 8chan. 8chan was essentially undesigned, its unstyled text and tables was a byproduct of its forum structure. These sprawling pages, with mismatched fonts and intermittent images, formed a kind of trashy if functional aesthetic. From a design perspective, 8chan in many ways represented its moniker as the cesspool of the internet, an environment that didn't care about usability or looks. 8chan's form matched its content: an unkempt wasteland, undesigned and unregulated, where nasty people said nasty things.

In contrast to these earlier "free speech" havens, then, Gab is highly designed. Its interface adopts the contemporary trend of flat design evangelized by Google in particular, where simple two-dimensional elements and solid colors strive to produce a very "clean" aesthetic that can be quickly parsed by a user. A coherent set of hues and fonts produce a cohesive identity across the platform. And the pages are arranged with an eye for line heights, spacing, and easy readability. In July of 2019, as CEO Andrew Torba notes in a video, "we launched a complete redesign to make it more user friendly"; in the video, Torbin steps users through a number of new features, from lists and notifications all the way through to an array of custom emojis "that you're not gonna find anywhere else on the internet."⁵⁷

While Gab has evolved over time, then, it has retained a strong attention to design. These details are not simply praise for Gab's design team, but rather point to a broader link between design and legitimacy. If 8chan was an eyesore, a space that had obviously rejected mainstream norms of taste and style, Gab is its more subtle successor—a "contemporary" platform where "best practice" design principles are adhered to. In its visual similarity with mainstream platforms like Facebook, Gab

⁵⁷ Andrew Torba, *An Update On Gab*, 2019, <https://www.youtube.com/watch?v=eUTHRTfgOsk&feature=youtu.be&app=desktop>.

asserts its more mainstream (if alternative) ambitions. Gab wants to move past the niche base of hardcore users to become a platform for everyone. Gab's design contributes to a legitimacy of the content embedded within it—even if or especially if that content is controversial.

However, if Gab's design puts a friendly, mainstream face on far-right material, it is worth remembering that the platform has been strongly linked to conspiracy theorists and extremists. In October of 2018, Robert Bower carried out a shooting at a Jewish Synagogue, killing 11 worshippers with an assault rifle and three handguns. Bower was an avid user of Gab, posting frequently on his profile page. As one report notes: "The profile shows that Bowers posted or recirculated dozens of anti-Semitic messages in the past few weeks. Those included two cryptic warnings hours before he allegedly opened fire in Pittsburgh's Tree of Life synagogue."⁵⁸ In contrast to the "dog whistles" and coded racism often employed by extremists online, what is striking is how upfront Bower's anti-Semitism was. His Gab profile featured a single sentence as a header: "jews are the children of satan." Bower's ideologies are by no means an anomaly on the platform. Only two months prior, in August, Microsoft had threatened to drop Gab's hosting services because of other anti-Semitic content. As CBS News reported, two posts had surfaced from a user who said that "Jews should be raised as 'livestock' and that he wanted to destroy a 'holohoax memorial'"; the posts were removed.⁵⁹

⁵⁸ Jason Silverstein, "Robert Bowers, Pittsburgh Shooting Suspect, Was Avid Poster of Anti-Semitic Content on Gab," CBS News, October 28, 2018, <https://www.cbsnews.com/news/robert-bowers-gab-pittsburgh-shooting-suspect-today-live-up-dates-2018-10-27/>.

⁵⁹ Jason Silverstein, "Robert Bowers, Pittsburgh Shooting Suspect, Was Avid Poster of Anti-Semitic Content on Gab," CBS News, October 28, 2018, <https://www.cbsnews.com/news/robert-bowers-gab-pittsburgh-shooting-suspect-today-live-up-dates-2018-10-27/>.



Memes, memes, and more memes.

40.2K Members

Catch all for the best memes. Should be loaded with humor, puns, innuendo, wit, and irony – even if mom doesn't approve. Extra Points for something not political. No posting of the the following: 1. Anything that isn't a meme. 2. Pro-Nazi 3. Racial insult as the main theme (N-word posts, etc.).



Meme Warfare

6.1K Members

This group is a dumping ground for pro-Republican memes in the lead-up to the 2020 election. Find (or create) unique campaign graphics to support your favorite God Emperor. Sharpen your autism.



Guns of Gab

37K Members

Guns of Gab is a place where firearms enthusiasts can share information about weapons and discuss the ongoing struggle to preserve our Right to Keep and Bear Arms. You can have fun here,

oak		Smooth	Opp	D	OSIER		Smooth	Alt:	D
BEECH		Shiny	Alt:	D	SALLOW		Smooth	Alt:	D
HORNBEAM		Toothed	Alt:	D	CRAB APPLE		Toothed	Alt:	D
BIRCH		Smooth	Alt:	D	PINE		Flews	-	E

Survival Beyond the Basics

4K Members

The real survival group. Here you will learn to live off the land with hand tools. Survive any situation. Live without government help. I am a survival/forager instructor. I will teach you to survive any situation. A lot of information. This is a...

Screenshot showing a selection of Groups permanently featured in Gab's "Groups" tab

One new design feature mentioned by Torba is the addition of Groups to Gab. While Groups can certainly be created by any user, Gab's Groups homepage permanently features a set of 50 Groups. These Groups are automatically privileged over Groups that must be searched for manually, receiving a huge boost in exposure and members. For this reason, and because Groups now provide a primary channel for creating and consuming content, it is worth paying attention to this special set of featured Groups. Featured Groups include: Libertarians of Gab (8k members), The_Donald (10k members), Christianity (28k members), Memes, Memes and More Memes (40k members), Guns of Gab (37k members), Survival (32k members), Manly

Men of Gab (20k members), and a host of others. The largest Group by far is Free Speech, with 72k members. Already then, in this series of examples, we can draw out a distinctive worldview comprised of men's rights, libertarianism, gun culture, and so on. Yet we can also see how Groups provide a way to cluster a wide-ranging community into more manageable clusters, where users are aligned closely by ideology or interests. This move can be thought of as a form of community design, a decision by the Gab team to structure the platform in a particular way in order to facilitate better quality exchanges between members.

Along with an image for each Group and the number of members, these featured boxes also include a short description. This description often includes the rules or guidelines associated with that particular Group. Survival, for example, warns users: "Please DON'T post any of that faggy Obama-loving 'Bear Grylls' crap." The Whiskey Women box stresses that "This is NOT your diary; race-baiting, offensive posts and trolls will NOT be tolerated. Keep it classy, guys." Content seems to be reviewed and moderated mainly by Group moderators, rather than an overarching Gab moderator team.

These guidelines, and the moderation that obviously has to back them up, are a further articulation of community design. After browsing through the content in these Groups, it appears that, by and large, these guidelines are working. The Dangerous Ladies of Gab group, as per its description, doesn't contain any pornographic images; the Classic Cars group doesn't contain anything unrelated to classic cars. A degree of success is achieved by placing very specific guidelines front and center, and delegating this regulatory work to a handful of group moderators. Here, it seems a few key factors come together. A 1) small number of invested individuals are given 2) absolute authority to regulate a 3) manageable sub-community according to 4) their own group-specific guidelines. Overlooking for a moment Gab's broader pathologies, we might note the effectiveness of this approach. This suggests ways that the ideal of "free speech" might be blended with other ideals, such as inclusion, anti-sexism and racism, civility, democracy, and so on.

However, these guidelines are largely about relevance (only cars, for instance) and obvious extremism (as one guideline declared: "no pro-Nazi memes, no n-words"). What this means is that content is valid if it seems on-topic and does not overtly cross

a hard threshold. If these requirements are met, then content is posted and platformized, reaching a large audience of Gab users. For example, in the Memes Group, a post may be approved that is decidedly racist, sexist, anti-Semitic, or similarly far-right in nature, provided that it a) does not directly endorse Nazism or use the n-word and b) is packaged in the form of a meme. The result is that a host of perhaps more subtle but undeniably hate-inducing material circulates. Indeed, as opposed to zero moderation in sites like 8chan, what is significant about Gab is that its light moderation may function as a stamp of approval, legitimating more tempered but still toxic communication.

Taken together, Gab's success in funding, its growing popularity, and its awareness of the importance of design point towards what may be called a "mainstreaming" of hate. Overall, its content is certainly more tame than the virulent racism displayed on sites like 8chan, the so-called ground zero for the alt right. Yet this could very well be the necessary compromise needed to become a more pervasive Platform for the People. Gab's more subtle hate, mixed in with liberal amounts of the kind of banal content found on any social media site—cats pics, dumb jokes, travel photos, and so on—has the potential to reach far more people. By coupling together the ideals of freedom and the usability of design, Gab provides a more palatable platform for toxic communication—a non-extremist environment interspersed with extremist ideologies.

Conclusion

This project has examined hate speech and toxic communication online from a design perspective, analysing Facebook, YouTube, and Gab as environments that support particular practices while discouraging others. In what ways might design be contributing to polarizing, impulsive, or antagonistic behaviours?

Based on engagement, Facebook's Feed drives clicks and views, but also privileges incendiary content, setting up a stimulus-response loop where outrage expression becomes easier and even normalized. Alternative ways of prioritizing content should be explored to decrease this kind of stimuli and in general to de-escalate the user

experience, providing a slower, calmer and more civil environment. In terms of user responses to this content, design interventions might be used to question, delay, or limit the scope of hateful comments.

YouTube's recommendation system is at the heart of the platform's design, exerting enormous influence on viewing and consumption. While technical, its operations have also been designed, a set of decisions that have been criticized for leading users towards more extreme content. Both this recommendation system and YouTube's infamous comment system need to be thoroughly redesigned, with the section laying out several suggestions.

Gab champions itself as a platform for free speech, alternative social media for the people. While Gab has been linked to racist, sexist, and anti-Semitic content and even acts of violence, it also displays a high awareness of contemporary design and the user experience. In this sense, it forms a subtler and more sophisticated successor to "legacy" havens of hate such as 8chan. Gab's design both tempers and legitimates its content, suggesting a new mainstreaming of toxic communication and extremist ideologies.

Across all three platforms, design is highly influential, shaping the ways in which content is accessed, consumed, and responded to. Because of this, design proved to be a productive lens for understanding toxic communication. Of course, there were also limits to this particular study. Based on this still-emerging field, this article could only present an initial foray into the design-centric analysis of hate speech online. In particular, the degree to which design may influence individuals—and the degree to which that influence may be modulated by age, gender, class, culture, socioeconomic background, and so on—has yet to be precisely determined. One path for research future would be to take up this challenge, producing a more quantitative analysis of design influence. Another path would be to apply this approach to other platforms: Reddit, Tiktok, 4chan, and so on. Based on the established link between hate, alt-right cultures, and gaming, gaming platforms such as Discord or Twitch would also make excellent studies.

Yet if this study has inevitable constraints, it reaffirms the key role that design plays within online environments. As everyday life increasingly migrates online, platforms

become crucial mediators for communication and key environments for inhabitation. These are spaces where time is spent, identities are forged, and ideologies are shaped. Understanding the ways in which these spaces can be redesigned in order to discourage hate speech and to instead encourage civility, inclusivity, and democracy remains an urgently needed task.

References

- Albanesius, Chloe. 2014. "10 Years Later: Facebook's Design Evolution." PCMag Australia. February 4, 2014.
<https://au.pcmag.com/internet-2/12249/10-years-later-facebooks-design-evolution>.
- Alexander, Julia. 2018. "Can YouTube Fix Its Comment Section?" Polygon. February 16, 2018.
<https://www.polygon.com/2018/2/16/17020326/nikolas-cruz-youtube-comment-section>
- Bergen, Mark. 2019. "YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant." *Bloomberg.Com*, April 2, 2019.
<https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>.
- Bosker, Bianca. 2016. "The Binge Breaker." *The Atlantic*, November 2016.
<https://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/>.
- CNN, Gianluca Mezzofiore and Donie O'Sullivan. n.d. "El Paso Shooting Is at Least the Third Atrocity Linked to 8chan This Year." CNN. Accessed September 9, 2019.
<https://www.cnn.com/2019/08/04/business/el-paso-shooting-8chan-biz/index.html>.
- Covington, Paul, Jay Adams, and Emre Sargin. 2016. "Deep Neural Networks for YouTube Recommendations." In *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 191–98. Boston, Massachusetts, USA: ACM Press.
- Crockett, Molly. 2017. *How Social Media Makes Us Angry All the Time*. Big Think.
https://www.youtube.com/watch?v=fE_QoebLUFQ.

- Delort, Jean-Yves, Bavani Arunasalam, and Cecile Paris. 2011. "Automatic Moderation of Online Discussion Sites." *International Journal of Electronic Commerce* 15 (3): 9–30. <https://doi.org/10.2753/JEC1086-4415150302>.
- Facebook. 2019. "News Feed." News Feed | Facebook Media. 2019. <https://www.facebook.com/facebookmedia/solutions/news-feed>.
- Fan, Rui, Ke Xu, and Jichang Zhao. 2016. "Higher Contagion and Weaker Ties Mean Anger Spreads Faster than Joy in Social Media." *ArXiv:1608.03656 [Physics]*, August. <http://arxiv.org/abs/1608.03656>.
- Fisher, Max, and Amanda Taub. 2018. "How Everyday Social Media Users Become Real-World Extremists." *The New York Times*, October 10, 2018, sec. World. <https://www.nytimes.com/2018/04/25/world/asia/facebook-extremism.html>.
- Freeman, James. 2018. "Facebook's 10,000 New Editors." *Wall Street Journal*, May 16, 2018, sec. Opinion. <https://www.wsj.com/articles/facebooks-10-000-new-editors-1526491169>.
- Frenkel, Sheera, Nicholas Confessore, Cecilia Kang, Matthew Rosenberg, and Jack Nicas. 2018. "Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis." *The New York Times*, November 14, 2018, sec. Technology. <https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html>.
- Gab. 2019. "Gab Social." Gab Social Hosted on Gab.Com. 2019. <https://gab.com/>.
- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.
- Han, Lu. 2019. "Designing for Tomorrow - A Discussion on Ethical Design." Spotify Design. January 18, 2019. <https://spotify.design/articles/2019-01-18/designing-for-tomorrow-a-discussion-on-ethical-design/>.
- Hango, Darcy. 2016. "Cyberbullying and Cyberstalking among Internet Users Aged 15 to 29 in Canada." Ottawa: Statistics Canada.
- Harris, Tristan. 2019. "Humane: A New Agenda for Tech." Center For Humane Technology. April 23, 2019. <https://humanetech.com/newagenda/>.
- Irving, Francis. 2018. "Brainstorming a Better YouTube Recommendation Algorithm." October 7, 2018. <https://www.flourish.org/2018/10/brainstorming-a-better-youtube-recommendation-algorithm/>.

- Jennings-Edquist, Grace. 2014. "Abusive Text Messages and Mobile Harrassment Are on the Rise." Mamamia. November 22, 2014.
<https://www.mamamia.com.au/abusive-text-messages/>.
- Lewandowski, Josh. 2018. 5 questions for YouTube's lead UX researcher Interview by Amy Avery.
<https://www.thinkwithgoogle.com/advertising-channels/video/youtube-user-behavior-research/>.
- Lewis, Paul. 2017. "'Our Minds Can Be Hijacked': The Tech Insiders Who Fear a Smartphone Dystopia." *The Guardian*, October 6, 2017, sec. Technology.
<https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia>.
- Mac, Tatiana. 2019. "Canary in a Coal Mine: How Tech Provides Platforms for Hate." *A List Apart* (blog). March 19, 2019.
<https://alistapart.com/article/canary-in-a-coal-mine-how-tech-provides-platforms-for-hate/>.
- Madrigal, Alexis C. 2017. "'The Basic Grossness of Humans.'" *The Atlantic*. December 15, 2017.
<https://www.theatlantic.com/technology/archive/2017/12/the-basic-grossness-of-humans/548330/>.
- Manjoo, Farhad. 2017. "Can Facebook Fix Its Own Worst Bug?" *The New York Times*, April 25, 2017, sec. Magazine.
<https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html>.
- Molla, Rani, and Kurt Wagner. 2018. "People Spend Almost as Much Time on Instagram as They Do on Facebook." *Vox*. June 25, 2018.
<https://www.vox.com/2018/6/25/17501224/instagram-facebook-snapchat-time-spent-growth-data>.
- Morris, David. 2018. "Facebook Accused of Ignoring Government Warnings Before Mob Violence in Sri Lanka." *Fortune*, April 22, 2018.
<https://fortune.com/2018/04/22/facebook-ignored-sri-lanka-hate-speech/>.
- Mulla, Sanafarin, and Avinash Palave. 2016. "Moderation Technique for Sexually Explicit Content." In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, 56–60.
<https://doi.org/10.1109/ICACDOT.2016.7877551>.
- Munn, Luke. 2019a. "Algorithmic Hate: Brenton Tarrant and the Dark Social Web." Institute of Network Cultures. March 19, 2019.

<http://networkcultures.org/blog/2019/03/19/luke-munn-algorithmic-hate-brenton-tarrant-and-the-dark-social-web/>.

- . 2019b. “Alt-Right Pipeline: Individual Journeys to Extremism Online.” *First Monday* 24 (6). <https://doi.org/10.5210/fm.v24i6.10108>.
- Naughton, John. 2018. “However Extreme Your Views, You’re Never Hardcore Enough for YouTube.” *The Guardian*, September 23, 2018. <https://www.theguardian.com/commentisfree/2018/sep/23/how-youtube-takes-you-to-extremes-when-it-comes-to-major-news-events>.
- Newton, Casey. 2019. “Three Facebook Moderators Break Their NDAs to Expose a Company in Crisis.” *The Verge*. June 19, 2019. <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa>.
- Nicas, Jack. 2018. “How YouTube Drives People to the Internet’s Darkest Corners.” *Wall Street Journal*, February 7, 2018. <https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478>.
- Noyes, Dan. 2019. “Top 20 Facebook Statistics - Updated July 2019.” *Zephoria Inc.* (blog). July 24, 2019. <https://zephoria.com/top-15-valuable-facebook-statistics/>.
- Pavlopoulos, John, Prodromos Malakasiotis, and Ion Androutsopoulos. 2017. “Deeper Attention to Abusive User Content Moderation.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1125–35. Copenhagen, Denmark: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1117>.
- Polymatter. 2016. “Why YouTube Comments Suck (and Reddit Comments Don’t).” December 15, 2016. https://www.youtube.com/watch?v=Lvf8kogX_yE.
- Rose, Brent. 2013. “YouTube Comments Will Soon Be Less Racist, Homophobic And Confusing.” *Gizmodo Australia*. September 25, 2013. <https://www.gizmodo.com.au/2013/09/youtube-comments-will-soon-be-less-racist-homophobic-and-confusing/>.
- Rose-Stockwell, Tobias. 2018. “Facebook’s Problems Can Be Solved with Design.” *Quartz*. April 30, 2018. <https://qz.com/1264547/facebook-problems-can-be-solved-with-design/>.
- SafeHome. 2017. “Hate on Social Media.” *SafeHome.Org*. February 3, 2017. <https://www.safehome.org/resources/hate-on-social-media/>.

- Saima, Salim. 2019. "YouTube Boasts 2 Billion Monthly Active Users, 250 Million Hours Watched on TV Screens Every Day." *Digital Information World* (blog). May 4, 2019. <https://www.digitalinformationworld.com/2019/05/youtube-2-billion-monthly-viewers-250-million-hours-tv-screen-watch-time-hours.html>.
- Silverstein, Jason. 2018. "Robert Bowers, Pittsburgh Shooting Suspect, Was Avid Poster of Anti-Semitic Content on Gab." CBS News. October 28, 2018. <https://www.cbsnews.com/news/robert-bowers-gab-pittsburgh-shooting-suspect-today-live-updates-2018-10-27/>.
- Statt, Nick. 2018. "Undercover Facebook Moderator Was Instructed Not to Remove Fringe Groups or Hate Speech." The Verge. July 17, 2018. <https://www.theverge.com/2018/7/17/17582152/facebook-channel-4-undercover-investigation-content-moderation>.
- Stevenson, Alexandra. 2018. "Facebook Admits It Was Used to Incite Violence in Myanmar." *The New York Times*, November 6, 2018, sec. Technology. <https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html>.
- Tait, Amelia. 2016. "Why Are YouTube Comments the Worst on the Internet?" *New Statesman*, October 26, 2016. <https://www.newstatesman.com/science-tech/internet/2016/10/why-are-youtube-comments-worst-internet>.
- Torba, Andrew. 2018. "Gab Investment Page." StartEngine. 2018. <https://www.startengine.com/freespeech>.
- . 2019. *An Update On Gab*. <https://www.youtube.com/watch?v=eUTHRTfgOsk&feature=youtu.be&app=desktop>.
- Tufekci, Zeynep. 2018. "YouTube, the Great Radicalizer." *The New York Times*, June 8, 2018. <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>.
- Tulkens, Stéphan, Lisa Hilde, Elise Lodewyckx, Ben Verhoeven, and Walter Daelemans. 2016. "The Automated Detection of Racist Discourse in Dutch Social Media." *Computational Linguistics in the Netherlands Journal* 6 (December): 3–20.
- Vincent, James. 2017. "Former Facebook Exec Says Social Media Is Ripping Apart Society." *The Verge*, December 11, 2017. <https://www.theverge.com/2017/12/11/16761016/former-facebook-exec-ripping-apart-society>.
- . 2019. "AI Won't Relieve the Misery of Facebook's Human Moderators." The Verge. February 27, 2019.

<https://www.theverge.com/2019/2/27/18242724/facebook-moderation-ai-artificial-intelligence-platforms>.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51. <https://doi.org/10.1126/science.aap9559>.

Wallaroo Media. 2019. "Facebook News Feed Algorithm History." *Wallaroo Media* (blog). July 3, 2019. <https://wallaroomedia.com/facebook-newsfeed-algorithm-history/>.

Yablonski, Jon. 2019. "Humane by Design." 2019. <https://humanebydesign.com>.