



Volume 14 Issue 1



RESEARCH
ARTICLE



OPEN
ACCESS



PEER
REVIEWED

Stop hate for profit: Evaluating the mobilisation of advertisers and the advertising industry to regulate content moderation on digital platforms

Steph Hill *University of Leicester*

DOI: <https://doi.org/10.14763/2025.1.1825>

Published: 31 March 2025

Received: 17 March 2024 **Accepted:** 11 October 2024

Funding: This research was funded in part by the Social Sciences and Humanities Research Council of Canada [752-2018-2077].

Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Licence: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>
Copyright remains with the author(s).

Citation: Hill, S. (2025). Stop hate for profit: Evaluating the mobilisation of advertisers and the advertising industry to regulate content moderation on digital platforms. *Internet Policy Review*, 14(1). <https://doi.org/10.14763/2025.1.1825>

Keywords: Platform governance, Content moderation, Advertisers, Activism

Abstract: This article compares the goals and outcomes of industry- and activist-led efforts that leverage advertisers to influence platform content moderation, and considers how these efforts fare as governments intervene more strongly in platform governance. It draws on contemporary documents and interviews to construct qualitative case studies of the Global Alliance for Responsible Media, a content moderation focused initiative of the World Federation of Advertisers, and the Stop Hate for Profit advertiser boycott of Facebook.

This paper is part of **Content moderation on digital platforms: beyond states and firms**, a special issue of *Internet Policy Review* guest-edited by Romain Badouard and Anne Bellon.

Introduction

Large social media platforms generate “substantially all” (Facebook Inc., 2023) of their revenue from advertisements. On social media, advertising interests are closely linked to content moderation practices, where filtering, removing, and recommending content “produc[es] a media commodity” (Gillespie, 2018, p. 43) meant to attract users and advertisers, and where advertising itself has been used to circulate propaganda and misinformation. Observers of the regulation of social media platforms have suggested that advertiser investment in norms of “civility and democracy” (Craig & Cunningham, 2019, p. 8) make them plausible allies in efforts to moderate extreme content online. However, examples of advertiser efforts to restrict funding for controversial content also suggest that half-hearted attempts to institutionalise advertiser-preferred content standards may have far-reaching consequences for user content, creator livelihoods, and the economics of journalism outlets (Aronczyk, 2020; Bishop, 2021; Parker, 2021; Willens, 2020; Yin & Sankin, 2021). Regardless, in a platformised online economy (Helmond, Nieborg, & van der Vlist, 2019), advertisers have sometimes been approached by civil society as productive actors with which to contest content moderation and monetisation, raising questions about legitimising private companies as appropriate governance intermediaries (Braun et al., 2019). This article compares the goals and outcomes of industry- and activist-led efforts that leverage advertisers to influence content moderation, and considers how these efforts fare as governments intervene more strongly in platform governance.

The Global Alliance for Responsible Media (GARM) and the Stop Hate for Profit (SHfP) campaign represent two efforts that leverage advertisers’ economic relationship with platforms to influence content moderation. The two initiatives—an alliance of major advertising associations, agencies, and large platforms in the case of GARM and an organised boycott of Facebook’s advertising services for the SHfP campaign—demonstrate some of the complex configurations of actors that come together around content moderation below the level of the state. These initiatives indicate the central place of the advertising industry in contests over platform content moderation as both a structure that influences content production and an active agent in public disputes over content. In the GARM case, the advertising industry partnered with large platforms directly to try to create consensus

around platform content moderation standards and to solve recurring issues with quality and controversy in online advertising. In the Stop Hate for Profit case, civil society and activist groups pressured and encouraged advertisers to boycott Facebook advertising services to pressure the platform to change its content moderation practices around hate speech and racism. While each existed in a specific context, the two approaches continue as voluntary advertising organisations, such as the Conscious Advertising Network, organise to influence how the industry addresses topics such as human rights and climate change misinformation, and activist groups such as Check My Ads, investigate how bad actors manipulate the adtech industry.

This article compares the actors and outcomes of the GARM and the Stop Hate for Profit campaign and assesses their implications for platform content regulation. It considers the fragmentation of the platform landscapes and efforts to support or contest forms of content moderation centred on the advertiser-platform relationship. It finds that boycotting advertisers' claims to fund Facebook were undermined by the scale of the platform's advertising network; that the GARM was used by Facebook as a shield from the more radical demands of the Stop Hate for Profit coalition; and that social justice values such as opposition to racism that were present in the Stop Hate for Profit campaign did not appear in the governance documents created by the GARM. Finally, it notes the absence of state actors, policymakers, and regulators in both cases, demonstrating how advertiser standards of brand safety are being institutionalised as content moderation standards by some groups even as those efforts are contested in the early stages of a regulatory turn (Flew, 2022) that is taking very different forms in Europe and the United States. The article concludes with a discussion of how policy is responding to platform content moderation and whether this regulation will disrupt the two strategies observed in Stop Hate for Profit and the GARM. It compares the aims of the Digital Services Act in the EU with lawsuits against the GARM for anti-competitive practice over their interventions in platform content moderation.

Literature review

The power of digital platforms, their dominance in online advertising, and their entanglement with nearly all business operations has prompted calls for heightened corporate responsibility in governing platform surveillance, data collection, and the attention economy (for instance, in Flyverbom et al., 2019). How ads are placed and who sees them are controlled by the platform, which retains a central place in the power structure (Helmond et al., 2019), placing advertisements using

proprietary algorithms and moderating content mostly in line with the normative values of US social and legal systems (Barrett & Kreiss, 2019; Klonick, 2017). However, the advertising industry has been singled out for its potential to influence platform practices because of the financial relationship between their industry and platform profits.

Critical political economists have theorised that advertisers' financial power over their publishing partners gives them a negative, censor-like power over media content through structural practices as well as direct interventions (Hardy, 2023). Online, scholars argue that the prestige and financial power of major advertisers are formative to the structure of many parts of the internet, including social media platforms (Gehl, 2014). Automatic placement of ads, combined with social media's struggles over content moderation has led to a series of public crises for online advertisers. In 2017, investigative journalist Alexi Mostrous revealed how the advertisements of prestigious brands were being placed automatically on websites and YouTube videos linked to terror groups and inappropriate sexual content (Mostrous, 2017). In response, advertisers such as Procter & Gamble and Pepsico boycotted YouTube, demanding guarantees of a brand safe environment, meaning one in which advertiser reputations are not damaged by content on the site (Cavale, 2020; Statt, 2017). To meet advertiser expectations of brand safety, YouTube changed its standards for sharing advertising revenue with partner channels, including adding stricter use of keyword blocking and higher bars before channels could earn advertising revenue. The resulting "adpocalypse" transformed monetisation for YouTube creators, creating new standards for assessing content suitability and fueling the "tiered governance" model that advantages some creators and content over others, likely affecting moderation and recommendation of content overall (Burgess & Green, 2018; Caplan & Gillespie, 2020; Hill, 2019; Kumar, 2019). The GARM institutionalised these standards of brand safety, and broader concepts such as brand suitability, to guide platforms on how to categorise content as compatible with major advertisers and incentivise platforms to restrict content that is not, with potentially widespread impacts on content and creators able to earn advertising money and content moderation practices more generally (Griffin, 2023).

Advertisers' perceived responsibility for content online is reinforced by activist groups such as Sleeping Giants, whose successful pressure on advertisers to abandon alt-right outlets such as Breitbart won them international accolades as well as scrutiny for their role in legitimising brands as governors of public discourse (Braun, Coakley, & West, 2019). Campaigns like Sleeping Giants leverage advertis-

ers' social reputation to incentivise them to withdraw advertising money from publications featuring hate speech, misogyny, and other perceived social ills. In effect, they practice a form of public shaming intended to have a regulatory effect on the profitability of the content, with perhaps knock on effects on its production. While campaigns like Sleeping Giants are novel in pressuring online advertisers directly, businesses have been targets of public campaigns, both socially progressive and conservative, over content at least as far back as television in the 1980's (Fahey, 1991; Turow, 1984). The sometimes cooperative, sometimes combative relationship between platform content moderation and advertiser interests continues in advertising industry-led initiatives like the GARM and activist-led campaigns like Stop Hate for Profit, and it is worth reconsidering these dynamics in the midst of a shifting regulatory environment that includes litigation against advertisers working together to influence content moderation, as well as legislative efforts to address content moderation and its effects directly.

Long considered "lawless" (Suzor, 2019), platforms are governed by overlapping commitments and structures, including voluntary agreements, codes of conduct, working groups, and other informal arrangements (Gorwa, 2019). Normative values and public opinion also contribute to this network of formal and informal governance mechanisms (Ananny & Gillespie, 2016; Barrett & Kreiss, 2019). Within this governance framework, platform terms of service nominally ban hate speech and other problematic content. However, content that violates platform terms of service, including hate speech, continues to spread widely (Giansiracusa, 2021) and platform structures support the circulation of violent and hateful content (Gerbaudo, 2018; Hopster, 2020; Mirrlees, 2021b), with key figures benefitting from dynamics between social media and legacy media to reach their audiences (McKelvey et al. 2021; Phillips, 2018). Persistent challenges with content moderation and revelations about platforms' limitations have motivated state-led regulatory responses, in what some scholars have described as a "regulatory turn" (Flew, 2022). In the US, Section 230 nominally shields platforms from liability and allows them to moderate on a discretionary basis, though recent antitrust lawsuits challenge platform power in other ways. In democratic contexts outside the US, platform regulation is changing rapidly with regulation in Europe, the passage of the Online Safety Act in the UK, and required payment for news content in Australia and Canada. Most notably for this article, the passage of the Digital Services Act in the EU in 2022 has created stronger legal requirements for liability and content moderation, as well as requirements for due diligence over illegal content and risk assessments for the largest platforms. While these state-led efforts clearly shift the platform landscape further from lawlessness, it is worth considering if and how a more reg-

ulated internet disrupts campaigns that leverage advertiser's financial power to influence content moderation below the level of the state.

Methods and data collection

This is an empirically grounded comparative account of two cases that share an understanding of advertisers' potentially powerful relationship with platform content moderation, but act on that understanding in different ways. The case studies of Stop Hate for Profit and the GARM presented for this article are constructed based on analysis of over 100 contemporary campaign documents, media coverage, and public interviews, complemented by eight interviews with advertisers, activists, and businesses associated with one or both campaigns. The materials were chosen intentionally to provide insight into the context of these groups' formation; the membership of the groups; their demands and intentions for content moderation, and their relationship to states and other potential regulatory actors. Documents collected included newsletters, social media posts, and press releases by advertisers that joined the Stop Hate for Profit boycott; public interviews and blog posts by the civil society groups that organised the boycott; Facebook's blog posts, US Security Exchange (SEC) filings, public interviews about the SHfP boycott; and publications and interviews with the GARM. For a complete list of documents examined, not all of which are cited in this article, see Appendix 1. Interview participants for this project included three companies that participated in the Stop Hate for Profit campaign; one of the civil society organisations that led the Stop Hate for Profit campaign; two prominent activists working with the advertising industry to defund hateful online speech; and two representatives of advertising industry associations familiar with the industry's relationship with Facebook and other social media (see the table of interview participants below). Semi-structured research interviews were conducted between January and August 2020 and participants were asked about the environment of corporate social advocacy, the role of communication technology, the motivations for corporate social advocacy and, where applicable, about their involvement in Stop Hate for Profit and the GARM.

TABLE 1: Interview participants

	PARTICIPANTS
1	representative, international advertising association
2	representative, national advertising association
3	owner, small public relations agency

PARTICIPANTS	
4	owner, crisis communications agency
5	communications, outdoor clothing brand
6	co-founder, advertising industry non-profit
7	founder, corporate accountability activist non-profit
8	director of strategy, consumer activism organisation

Both cases explored here focus on Facebook, to some degree. However, the GARM and its members worked with many of the largest advertising-supported social media platforms, including YouTube and Twitter (now X). The pressure campaign that Stop Hate for Profit ran against Facebook was modeled after global efforts led by Sleeping Giants affiliates to pressure advertisers to break ties with publications that feature hate speech. Efforts to use the financial ties between platforms and major advertisers to change content moderation and other issues online continue, including in voluntary industry groups such as the Conscious Advertising Network, and activists continue to expose ties between advertising money and objectionable content on many platforms. So, while the cases presented in this article are no longer active, the strategies that they represent continue to be relevant as examples of content moderation efforts beyond the state.

Case 1: Global Alliance for Responsible Media

The World Federation of Advertisers (WFA) announced the Global Alliance for Responsible Media—a coalition that included YouTube, Twitter (now X), Facebook, and Instagram alongside a collection of the world’s largest advertisers—in 2019 as a response to proliferating concerns over platform content and advertising placement. In September 2020, the GARM announced a “brand safety floor + suitability framework” that provided agreed-upon standards for defining content as “harmful and sensitive,” defining where ads should not appear, and labelling sensitive content on a scale of risk that advertisers could choose between (The Global Alliance for Responsible Media, 2020). This brand safety floor structured annual transparency reporting that platforms submitted to the GARM and has been described by researchers as part of “the ongoing expansion and institutionalisation of advertiser influence” over content moderation, with outsized consequences for the moderation of content from already marginalised groups (Griffin, 2023, p. 9).

For years before the GARM was formed, the advertising industry had been grappling with the reputational and value-for-money risks of their relationship with

platforms, including through limited boycotts by individual or small groups of advertisers. After shootings at two mosques in Christchurch were livestreamed on Facebook, New Zealand advertisers boycotted the platform and convinced the WFA to call for “all brands globally to hold social media platforms to account in light of recent failures to block dangerous and hateful content” (World Federation of Advertisers, 2020). The WFA subsequently launched the Global Alliance for Responsible Media (GARM) to facilitate more coordination between advertisers, agencies, and platforms. This alliance brought major advertisers and ad agencies together with platforms to find consensus on issues of responsible social media governance—including ad fraud and “strict brand safety protection” (World Federation of Advertisers, n.d.). The GARM and the WFA are led by companies with very established public reputations. As a representative of the WFA explained, “we represent companies that already have a very developed and sophisticated position on their role in society, CSR, those types of responsibility issues” (personal communication). The WFA had 127 corporate members and 55 national advertising association members. Their clients included historic leaders in the industry, such as P&G and Unilever, as well as newcomers, such as Airbnb. WFA member advertisers represented a small number of Facebook’s millions of advertising partners, “...of all the companies that are advertising on Facebook and YouTube—our members are a percentage of that, but they’re not quite as significant as they would be if you looked at TV channels” (personal communication, representative of international advertising association). The GARM represents the interests and viewpoints of established brands in their relationship with media, meaning that while the group includes very large and very well known brands they do not necessarily represent the majority of platforms’ earnings. For context, in a 2019 earnings call then-Facebook COO Sheryl Sandberg claimed that the top 100 advertisers on the platform represented less than 20% of their total revenue (Facebook Inc., 2019).

The GARM “was formed to identify specific collaborative actions, processes and protocols for protecting consumers and brands from safety issues” (World Federation of Advertisers, 2020) and close the gap between social media platforms’ hands off approach to content moderation and the social responsibility commitments of major advertisers. The GARM developed a Global Media Charter as well as the “Brand Safety Floor + Suitability Framework” to outline its priorities. The Global Media Charter’s eight principles outlined quality control issues for the online advertising industry, including user experiences of advertising, ad-fraud, and a tiered approach to brand safety risk. It emphasised addressing “walled garden issues” and third-party verification, referencing the inability of most advertisers to independently verify the numbers provided to them by private platforms. This em-

phasis echoes the widespread unease among advertisers about the ability of Facebook and other platforms to “grade its own homework” in terms of ad performance, viewership metrics, and content quality (personal communication; Monllos, 2020). The Brand Safety Floor + Suitability Framework addressed a range of topics, beginning with “adult & explicit sexual content” and including arms and ammunition, crime, death and injury, piracy, hate speech, obscenity, drugs, spam, terrorism, and “debated sensitive social issue” (Global Alliance for Responsible Media, 2020), in that order. The GARM’s Global Media Charter incentivised over-removal of content. For example, while informative sexual content was rated as “low-risk,” the suitability of platforms for advertising was ranked by the prevalence of “harmful” ad impressions and the effectiveness of platforms in removing “violative content” (World Federation of Advertisers, 2021), providing an incentive for platforms to err on the side of caution by removing sexual content of many kinds. The GARM existed to address issues of verification and value for money alongside an approach to content moderation that incentivised risk averse moderation likely to over-remove content that did not violate the standards directly.

National governments and multilateral institutions were absent from the discussion of the establishment of the GARM. Advertisers were aware of the lack of regulatory checks and balances and some identified it as a challenge to the legitimacy of the industry’s self-regulation: “[self-regulation] only works if we actually act responsibly and deal with any issues. [Platforms] talk about control but I don’t think there’s a high level of trust around that. And it’s not independent, it’s internal... there’s a vacuum there, which then says government regulators need to step in” (representative, national advertising association). The GARM’s later willingness to engage directly with EU regulatory processes by, for example, signing the EU Commission’s Code of Practice on Disinformation, may reflect this openness to state interventions to solve content moderation challenges. While it was active, there were doubts about whether platforms were fully complying with the GARM’s brand safety standards (Kaye, 2021). However, the implications of reinforcing content moderation in line with standards of brand safety was interpreted by scholars as potentially significant (Bishop, 2021; Griffin, 2023). Regardless, the GARM abruptly suspended its activities in 2024 after Elon Musk sued the organisation, claiming it was engaged in anti-competitive collusion over content (O’Reilly, 2024). It is unclear whether its efforts to formalise advertiser interests in platform content moderation will be revived.

Case 2: Stop Hate for Profit (activist-initiated effort at moderation via advertisers)

The year after the formation of the GARM, during the Black Lives Matter protests in the summer of 2020, the Stop Hate for Profit campaign recruited advertisers to pressure Facebook over perceived gaps in how its content moderation policies applied to hate groups and inadequate support for targets of hate and harassment on the platform. The Stop Hate for Profit advertiser boycott happened one month after George Floyd's murder by police in Minneapolis on 25 May 2020. The boycott targeted Facebook in part over differences between its moderation of inflammatory statements by Donald Trump and those of other platforms. Later investigations uncovered that Mark Zuckerberg had phoned Trump directly to discuss the content of his posts (Mac & Silverman, 2020). Facebook's willingness to bend its own rules on what was widely viewed as support for violence against protestors put them in conflict with some of the platform's advertisers that were facing pressure to prove that they understood the gravity of the situation and were taking meaningful action (personal communication, Stop Hate for Profit participant).

In this context, a coalition of American civil rights organisations, including the Anti-Defamation League, the NAACP, Color of Change, the National Hispanic Media Coalition and the online activist group Sleeping Giants organised the Stop Hate for Profit campaign. Stop Hate for Profit asked participants to suspend advertising on Facebook in July 2020. Stop Hate for Profit called on advertisers to suspend advertising on Facebook as a symbolic gesture against Facebook's role in enabling or amplifying "hate, bigotry, racism, antisemitism, and disinformation" (Stop Hate for Profit, n.d.). The group's nine-point list of demands focused on the themes of accountability, decency, and support for victims of hate and harassment. It emphasised civic solidarity with people and groups harmed by harassment, Facebook's lack of accountability to these same groups in its policies, and Facebook's failure to take responsibility for content on its platform. In contrast to the GARM's effort to align advertiser interests with platform practices, SHfP was much more narrowly focused on changing content moderation policies related to race and enforcement against hate speech.

At its height, the Stop Hate for Profit campaign included 1,100 advertisers, including 200 "major corporations" whose advertising value to Facebook was worth approximately USD \$7 billion (Levy, 2020). Speaking to the incentive for advertisers to intervene in controversies over advertising-supported content, representatives of the advertising industry noted "an external perception that advertisers should be the ones able to force change" (personal communication, national advertising

association). In particular, the investment that advertisers had made in social responsibility acted as an incentive. As a representative of a national advertising association communicated “the question becomes, if you are talking about brand purpose, CSR [corporate social responsibility], etc, what is your position around where your advertising may appear?” Moreover, “silence becomes a void that people can fill” (personal communication, small ad agency), opening companies up to accusations that they did not care about racial justice or other social issues if they were not publicly stating their support through participation in initiatives like SHfP. Many companies felt pressure internally from employees to demonstrate their commitment at a significant moment, particularly those that had been “late” in responding to the protests in May (personal communication, Stop Hate for Profit participant). Others had a history of support for issues of representation and diversity and felt that the campaign was “something doable” (personal communication, small ad agency) in terms of pressuring other small businesses and demonstrating leadership. The investments made by these companies in their reputations for social responsibility created internal incentives to push back against a platform framed as accommodating extremist rhetoric.

However, the campaign’s significance was soon subject to critical scrutiny. Facebook’s revenue was over USD \$21 billion over the quarter and did not dip while the boycott was active (Glenday, 2020). Some companies, such as Unilever, suspended advertising on the platform for all of 2020 independent of the SHfP campaign, citing the “toxic environment” online during the US presidential elections (McCarthy, 2020). Others, including P&G, whose year-long boycott of YouTube in 2017 was credited with an instrumental role in changes to the platform (Bergen, 2018), did not participate in the SHfP boycott (Cavale, 2020). Advertisers participating in the campaign felt a swift change in media interest from positive coverage of advertiser participation “to a kind of nitpicky, looking for gaps kind of thing” (personal communication, Stop Hate for Profit participant) that scrutinised the timing of the campaign, their use of other advertising tools linked to Facebook, and the amount advertising spending was being cut. Some industry activists saw the campaign as disingenuous. As one industry activist told me, “What I saw with Stop Hate for Profit was them bragging about having gotten companies to stop advertising. But those companies hadn’t really stopped advertising through Facebook” (personal communication, consumer activism organisation). Widely circulated comments such as Mark Zuckerberg expecting advertisers to be “back soon enough” (Heath, 2020) and Stop Hate for Profit’s characterisation of Facebook as offering “the same old defense of white supremacist, antisemitic, Islamophobic and other hateful groups on Facebook” seemed to reflect the frustration of the boycott’s

short-term demands. Observers deemed participation in the campaign for most companies as “at best, symbolic, and unlikely to harm Facebook financially” (Wodinsky, 2020, para. 7).

However, changes to content moderation in response to Stop Hate for Profit nominally included the removal of “content encouraging or calling for the harassment of others, which was a top concern of civil rights activists” and the removal of “more than 100,000 pieces of FB & IG content for violating voter interference policies” (Sharing Our Actions on Stopping Hate, 2020). Facebook’s responses to the boycott also echoed the language used by the campaign: “We do not profit from misinformation or hate, and we do not want this content on our platforms” (Mark Zuckerberg, in Facebook Inc’s Second Quarter Results Conference Call, 2020). In the end, the campaign claimed victory: “we forced an unprecedented public examination of Facebook’s deep harms to marginalized communities” (Statement from Stop Hate for Profit on Ad Pause Success, 2020). The campaign included the creation of a senior civil rights position at Facebook, the release of an unflattering civil rights audit, the creation of a team to study algorithmic bias, the promise of an independent audit, and the removal of extremist groups such as Boogaloo as evidence of the success of the initiative (Statement from Stop Hate for Profit on Ad Pause Success, 2020). Facebook did commit to changes in content moderation, such as the introduction of warning labels on contentious content (Bond, 2020), banning holocaust denial, and efforts to reduce the impacts of its hate speech policies on marginalised groups (Dwoskin et al., 2020; Wodinsky, 2020). The boycott, as contested and symbolic as it was, added to public criticisms of Facebook’s content moderation policies that motivated changes to the platform. It was also a reversal of decades of advertiser interventions that primarily restricted advertising on issues that served the self-interest of the advertisers, rather than addressing social issues (Hardy, 2023).

The GARM, and the commitments of the parties that signed onto it, became a touchstone for Facebook when disputing some of the claims of the Stop Hate for Profit campaign. Facebook’s VP of Global Marketing Solutions referenced commitment to GARM in their messages to advertisers and the GARM appeared in Facebook’s July 30 update to its public statement on the Stop Hate for Profit campaign, which addressed the campaign’s claims and demands in a point-by-point format (Sharing Our Actions on Stopping Hate, 2020; Fischer, 2020). At the time, Facebook’s CEO claimed that the platform was “not going to change our policies or approach on anything because of a threat to a small percent of our revenue” (Mark Zuckerberg, in Clayton, 2020, para. 4). Instead, it was making sure that those poli-

cies “support the advertising community through the Global Alliance for Responsible Media (GARM) and continue to engage with civil rights leaders about our policies and practices” (Sharing Our Actions on Stopping Hate, 2020, Stopping Hate section). In this case, Facebook used its membership in the GARM as evidence that it was aligned with advertiser standards of responsible content moderation without committing the platform to actually meet the demands of the Stop Hate for Profit boycott participants. National governments and multilateral institutions were absent from the discussions between Facebook and Stop Hate for Profit. While Facebook referred to multi-stakeholder agreements such as The Christchurch Call to Action, a set of voluntary commitments agreed to by governments and online service providers intended to address terrorism after the shootings in Christchurch, there was little suggestion of another credible authority outside of Facebook and its advertisers

Discussion

The cases of the GARM and the Stop Hate for Profit campaign demonstrate two different efforts to influence the scope and focus of social media content moderation through advertisers and the advertising industry. The first is advertiser activism, often pressured by civil society and the public, as in the Stop hate for Profit campaign. The other is the institutional approach—advertiser standards enacted by the industry, as in the GARM. These two approaches are analogous to the “instrumental” and “structural” understandings of advertiser influence over media in critical political economy, with one representing intentional interventions over content and the other the aggregate effects of the industry’s practices and preferences (Hardy, 2023). While the original Stop Hate for Profit campaign was relatively short-lived, working through advertisers and the advertising industry to incentivise responsible content moderation on platforms remains an active strategy for now. Adtech-focused watchdog Check My Ads declared the end of GARM as an opportunity to “reset” brand safety efforts from advertisers and marketers around fundamental reputational concerns, rather than abandon those efforts (Garcia, 2024). The institutional approach works to craft longer term commitments that can be supported by advertisers, instituted by platforms, and potentially supported by other institutions. This approach continues with groups such as the Conscious Advertising Network, a coalition of advertisers that sign onto shared “manifestos” on issues from ad fraud to climate misinformation that is hosted regularly by the United Nations (Hobbs, 2020). The power of these institutional connections may be diluted if changes to content moderation on major platforms continue to encourage advertisers to reduce their spending on those platforms (Ostwal, 2024).

Both the activist approach of individual advertisers and the institutional approach are highly relevant to considering content moderation beyond the state as they operate either as bilateral (advertiser-social media platform) or tri-lateral (advertiser-platform-civil society) governance negotiations, without a necessary role for state regulators and policymakers. They exemplify new complexities in how advertisers influence content moderation that complicate accounts that position advertisers as censors of the media, particularly media that serves marginalised groups (Hardy, 2023). Both strategies will be sorely tested by rapid changes in content moderation on large platforms, particularly on X and Meta which have made definitions of hateful conduct more permissive (Bradley & Joseph, 2025; Vranica & Haggin, 2025). The complexity around advertiser approaches to content moderation, and the questions for policy intervention, is most evident in the two main groups marginalised by the advertiser interventions in the GARM and SHfP. Some are marginalised for their participation in the kinds of misinformation and hate speech that the campaigns were created to combat, others were caught up by the vague definitions of risk employed by advertisers' standards and platforms' implementation.

Consequences, implications, and the regulatory turn

Two main groups oppose advertiser influence on platform content moderation: those who feel targeted by advertiser activism and those who feel marginalised by advertiser standards. Stop Hate for Profit drove an exodus of alt-right content creators from platforms like Facebook to alt-tech platforms, such as Gab, that market themselves in opposition to the content moderation of mainstream social media platforms, which they characterise as hostile to free speech and conservative viewpoints (Jasser, 2021). The content moderation policies enacted by social media platforms in response to advertiser activism resulted in the removal of hundreds of millions of pieces of hateful content on Facebook, YouTube, Twitter, and other social media sites, including many associated with alt-right groups and individuals (Mirrlees, 2021b; Renton, 2021). More generally, social media platforms, based on the policies that the GARM institutionalised, remove hate speech, violent rhetoric, misinformation, and insensitive treatment of debated topics. Increasingly, figures within the alt-right movement complain that "there's no economic model" (Steve Bannon in Embury-Dennis, 2019) based on advertising for sites such as Gab, Breitbart, Parler, or for alt-right celebrities, such as Alex Jones or Milo Yiannopoulos. As others have noted, "Big Tech's hate speech rules have thrown a monkey wrench into the social media machinery of some far Right's propagandist-entrepreneurs and denied them some platforms to make money, mobilize, and recruit in the digital mainstream" (Mirrlees, 2021a, p. 270). At the same time, conservative lawmakers have begun to push back against these efforts. In some cases, politicians accuse

the advertisers of collusion, as in the case of the chair of the US Judiciary Committee discussed below. In others, there is political pressure to exclude activist advertisers from policy processes, as in efforts to remove the Conscious Advertising Network from consultations with the United Kingdom Department of Media Culture and Sport (Young, 2023). These efforts reached a new level of prominence and effectiveness with Musk's takeover of Twitter (now renamed X). As advertisers left the platform over concerns about content and changes to policy, including significant reductions to the trust and safety team (Mac, 2024; Thomas, 2024), Musk publicly told them to "go fuck yourself" (Goode, 2023). Subsequently the company sued the GARM, the World Federation of Advertisers, and some of their most prominent supporters, accusing them of antitrust violations, continuing a pattern of lawsuits intended to intimidate advertisers and activists from taking action against the platform (Ortiz, 2024).

In a very different political context, many researchers argue that advertiser interventions further marginalise content created by and for women, sexual expression, and sexual and gender minorities, as well as journalism and activist content (Aronczyk, 2020; Bishop, 2021; Burgess & Green, 2018; Cotter, et al., 2021; Griffin, 2023; Mueller, 2015; Siapera & Viejo-Otero, 2021). For example, when Tumblr tried to make the platform more attractive to advertisers by banning adult content, it disrupted forums that had fostered LGBTQ communities, and attracted public contestation of the platform's governance practices by its users (Sybert, 2021). Tumblr's shifts in governance cost the platform much of its user base, but also destroyed what had been an important forum for marginalised sexual communities, including trans people and sex workers (Bronstein, 2020). Using keywords and algorithms in online content moderation increases the likelihood that non-target content will be removed or demonetised (see Urban et al., 2016 for a discussion of this trend in response to legal rulings). The GARM and its standards exemplifies these concerns (Griffin, 2023). That Facebook cited the GARM as a rebuttal to demands for solidarity with the aims of Black Lives Matter activists seemed to confirm the industry's institutional role in setting a risk-averse and relatively conservative approach to content moderation through their definitions of brand safety and suitability.

These are not equivalent claims, but do share a view of advertiser power over platforms. Both groups see the effects of advertiser interventions as at best incentivising the platforms against specific groups and their content and at worst practicing a form of "collateral censorship" (Balkin, 2009) where advertiser pressure, rather than legal liability, on platforms restricts discussion of certain topics. Reactions from groups that see themselves as disadvantaged by advertiser interventions in

content moderation include abandoning platforms perceived to be hostile, as in the case of some of the alt-right, and contestation of platform policies, as in the Stop Hate for Profit campaign and the Tumblr example. However, state-led regulatory developments, particularly in the US and EU, raise questions about whether advertiser influence over content might be constrained. The next section considers whether new regulations, in particular the EU's Digital Services Act, interrupt or diminish the role of advertisers as informal content regulators in social media environments. It also considers the rhetoric of antitrust used to challenge advertiser cooperation on content moderation, including the lawsuit filed by Elon Musk against the GARM that ended the organisation's activities.

Interventions by state and regional regulators

Neither the GARM nor the Stop Hate for Profit referred to the state as a solution to their content moderation challenges. The relative absence of states in the regulation of platforms is changing, however, with notable antitrust and anti-competition cases brought in the United States, the passage of online safety legislation in the United Kingdom, proposed online harms legislation in Canada, and, most notably, the passage of the Digital Services Act and Digital Markets Act in the European Union.

Advertisers and content moderation after gatekeeper legislation

The Digital Services Act

The Digital Services Act promises to significantly change how intermediary liability functions in the EU, incorporating considerations of fundamental rights into platform's moderation (Heldt, 2022). For this article, the question is really whether this legislation will disrupt the direct institutionalisation of advertiser interests in content moderation, or campaigns like SHfP that try to leverage that relationship. The DSA obligates platforms to take fundamental rights into account in the application of their terms and conditions (Art. 14(2)). Very large online platforms and search engines have additional obligations to assess risks related to potentially dangerous content, discrimination, and threats to fundamental rights under Article 34(2)(d). These obligations include consideration of content moderation systems and systems for placing advertisements. The fundamental rights including human dignity, respect for private and family life, personal data, freedom of expression and information, non-discrimination, the rights of the child, and consumer protection (Art. 34(1)(b)). Theoretically, these obligations could encourage platforms to

consider the fundamental rights of individuals alongside advertiser priorities in content moderation. They could also reinforce legal efforts charging platforms with discrimination based on monetisation policies and recommendation, strategies that have been unsuccessful so far (Henn, 2021; Stempel, 2023).

Elsewhere, the DSA and related legislation encourages platforms to work with advertisers to govern transparency in advertising placement, ad targeting, and the monetisation of misinformation. Where the DSA explicitly deals with advertisers and advertisements, it is primarily concerned with requiring transparency around who purchases advertisements and how data is used to target individuals (Articles 26, 39, 46). Much of this work is expected to take place through voluntary codes of conduct that stakeholders in the platform advertising chain will sign onto. Some potential codes of conduct, such as the EU Code of Practice on Disinformation, explicitly identify brand safety tools as important levers in preventing monetisation of disinformation and misinformation. The Code of Practice may soon be added to the DSA, but platform compliance with the Code is largely poor, adding to concerns that platform compliance with the DSA may be lacking (Mündges & Park, 2024). Platforms' history of inconsistent compliance, sometimes even with their own policies (Horwitz, 2021), suggests that the DSA may not significantly shift the balance of power in content moderation.

Griffin (2023) argues that the DSA/DMA is unlikely to disrupt advertiser impacts on content, given that the two pieces of legislation do not address the indirect effects of advertiser priorities or platform policies themselves, only their application. She adds that the Code of Practice on Disinformation is "effectively legitimising this [advertiser] censorship power and charging them with using it responsibly" (Griffin, 2023, p. 18). However, there is an argument to be made that the DSA's risk assessment provisions in Article 34 should cover advertiser impacts on content moderation and support legal challenges if advertiser standards can be shown to consistently impact freedom of expression, though this will take time. There may also be good reasons to distinguish between censorship, with its connotations of restrictions on predefined types of political speech, and commercial constraints on online platforms, which play a more complex role in algorithmic media, as well as working in multiple directions, politically speaking (Hardy, 2023). Content moderation in the name of safety and anti-discrimination has become politicised and subject to attack from policymakers and powerful individuals in the United States where, in contrast to the DSA's theorised effects, antitrust lawsuits have already forced the suspension of advertisers' formal content moderation efforts, including the GARM itself.

Antitrust, collusion, and backlash

In March 2023, the chairman of the US Committee on the Judiciary accused both the Global Alliance for Responsible Media (GARM) and the World Federation of Advertisers, of “potentially violating US antitrust law by coordinating their members’ efforts to demonetize and eliminate disfavored content online” (Jordan & Nadler, 2023). Jordan and Nadler’s accusations were used as the basis for a lawsuit filed by X in US federal court against the GARM, World Federation of Advertisers, and several of their highest-profile members for allegedly engaging in antitrust activity (Scanlon, 2024). In response, the WFA chose to suspend the GARM’s activities (Joseph & Scanlon, 2024). Observers doubt the legal merit of the case, and suggest that these cases are political (Elon Musk can’t force advertisers to spend, 2024) and continue a pattern of pushback against advertisers restricting ad placement over media content. Conservative media outlets have doxxed activists involved in *Sleeping Giants* (Greene, 2018). Earlier legal efforts by X/Twitter against activists and advertisers that were dismissed as Strategic Lawsuits Against Public Participation (SLAPP) meant to discourage reporting on the scale of hate speech and misinformation on X (Ortiz, 2024).

Participation in the GARM did not require advertisers to reduce advertising on X. Many jurisdictions, including the EU, protect horizontal cooperation between market actors when it serves a legitimate purpose, such as sustainability ends, and does not disrupt competition (Antonazzi, Kuiper, & Cramer, 2023). The GARM acted as another voluntary code of conduct, like many other voluntary agreements that govern platform content moderation beyond the state (Gorwa, 2019). The risks that it posed were not collusion to demonetise disfavoured content online, but of economic and public relations incentives that shored up existing platform content moderation trends that favour professional, advertiser-friendly content to the detriment of small producers, vulnerable groups, and educational and news services. By suing the GARM out of activity, X has not eliminated the indirect effects of an advertiser funded system, existing advertiser tools that platforms provide, the monetisation policies that came from codifying advertiser preferences under the GARM, or even strategies like Stop Hate for Profit that recruited individual advertisers. However, the case, and others like it, especially if supported by the second Trump administration, may have a chilling effect on efforts to challenge injustices in content moderation and discourage compliance with efforts to demonetise disinformation and hate speech. This is already visible in dramatic shifts to Meta’s content moderation policies (Vranica & Haggin, 2025). The use of legal intimidation may undercut efforts to engage in robust public deliberation over content moderation, including contestation of advertisers’ brand safety standards and their

effects on marginalised groups.

Conclusion: Zones of discretion

While the Digital Service Act may well increase platform transparency, it is not clear that it will disrupt the content moderation dynamics taking place between platforms and advertisers. As Nic Suzor and his collaborators in research on digital constitutionalism have pointed out, even with increased nation-level regulation of platforms, a considerable amount of platform content moderation will be left to the discretion of the platforms themselves (Suzor, 2020; Suzor & Gillett, 2022). These persistent “zones of discretion” (Suzor & Gillett, 2022, p.261) mean that self-regulation, informed by profit motives, will be with us as long as platforms exercise their “custodial” powers over online content (Gillespie, 2018) and as long as they are constructed around an economic model tied to advertising. Regulation that acknowledges the role of platforms’ economic model in issues of content moderation, as the Digital Services Act does, has the potential to push back against overreach by those standards when it affects fundamental rights. This approach is preferable to one that characterises efforts to demonetise misinformation and hate speech as anti-competitive collusion or censorship and potentially chills efforts to exercise content moderation in the name of safety and fairness.

References

- Ananny, M., & Gillespie, T. (2016). *Public platforms: Beyond the cycle of shocks and exceptions*. <http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/anannyGillespie-publicPlatforms-oi-i-submittedSept8.pdf>
- Antonazzi, L., Kuipers, P., & Cramer, T. (2023). The new EU regime for horizontal agreements: Overview of the main changes. In *Bird & Bird*. <https://www.twobirds.com/en/insights/2023/global/the-new-eu-regime-for-horizontal-agreements-overview-of-the-main-changes>
- Aronczyk, M. (2020). Brands and the pandemic: A cautionary tale. *Social Media + Society*, 6(3). <https://doi.org/10.1177/2056305120948236>
- Barrett, B., & Kreiss, D. (2019). Platform transience: Changes in Facebook’s policies, procedures, and affordances in global electoral politics. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1446>
- Bergen, M. (2018, April 20). *P&G ends its YouTube advertising boycott, but with a catch*. Bloomberg. <https://www.bloomberg.com/news/articles/2018-04-20/p-g-ends-its-youtube-advertising-boycott-but-with-a-catch>
- Bishop, S. (2021). Influencer management tools: Algorithmic cultures, brand safety, and bias. *Social Media + Society*, 7(1). <https://doi.org/10.1177/20563051211003066>

Bond, S. (2020, June 26). *In reversal, Facebook to label politicians' harmful posts as ad boycott grows*. NPR. <https://www.npr.org/2020/06/26/883941796/unilever-maker-of-dove-soap-is-latest-brand-to-boycott-facebook>

Bradley, S., & Joseph, S. (2025, January 8). *Meta follows Musk's lead on censorship – but ad industry keeps its distance from panic*. Digiday. <https://digiday.com/marketing/meta-follows-musks-lead-on-censorship-but-ad-industry-keeps-its-distance-from-panic/>

Braun, J. A., Coakley, J. D., & West, E. (2019). Activism, advertising, and far-right media: The case of sleeping giants. *Media and Communication*, 7(4), 68–79. <https://doi.org/10.17645/mac.v7i4.2280>

Bronstein, C. (2020). Pornography, trans visibility, and the demise of tumblr. *TSQ: Transgender Studies Quarterly*, 7(2), 240–254. <https://doi.org/10.1215/23289252-8143407>

Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*. John Wiley & Sons. <https://ayomenulisfisis.wordpress.com/wp-content/uploads/2019/02/youtube-online-video-and-participatory-culture.pdf>

Caplan, R., & Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305120936636>

Cavale, S. (2020, July 1). *P&G says will not disclose advertising decisions as Facebook ad boycott grows*. Reuters. <https://www.reuters.com/article/us-facebook-ads-boycott-p-n-g-idUSKBN242605>

Clayton, J. (2020, July 2). *Zuckerberg: Advertisers will be back to Facebook "soon enough"*. BBC. <https://www.bbc.com/news/technology-53262860>

Cotter, K., Medeiros, M., Pak, C., & Thorson, K. (2021). "Reach the right people": The politics of "interests" in Facebook's classification system for ad targeting. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211996046>

Craig, D., & Cunningham, S. (2019). *Social media entertainment: The new intersection of Hollywood and Silicon Valley*. NYU Press. <https://www.jstor.org/stable/j.ctv12fw938>

Dang, S., & Paul, K. (2020, July 2). *Facebook frustrates advertisers as boycott over hate speech kicks off*. Reuters. <https://uk.reuters.com/article/us-facebook-ads-boycott-idUKKBN2424GS>

Duffy, B. E. (2020). Algorithmic precarity in cultural work. *Communication and the Public*, 5(3–4), 103–107. <https://doi.org/10.1177/2057047320959855>

Dwoskin, E., Tiku, N., & Kelly, H. (2020, December 3). Facebook to start policing anti-black hate speech more aggressively than anti-white comments, documents show. *Washington Post*. <https://www.washingtonpost.com/technology/2020/12/03/facebook-hate-speech/>

Embury-Dennis, T. (2019, April 4). *Steve Bannon caught on video admitting Breitbart lost 90% of advertising revenue due to boycott*. The Independent. <https://www.independent.co.uk/news/world/americas/us-politics/steve-bannon-breitbart-boycott-advertising-sleeping-giants-trump-a8854381.html>

European Commission. (n.d.). *Digital services coordinators*. Shaping Europe's Digital Future. <https://digital-strategy.ec.europa.eu/en/policies/dsa-dscs>

Facebook Business. (2020, July 30). *Sharing our actions on stopping hate*. Meta. <https://www.facebook.com/business/news/sharing-actions-on-stopping-hate>

Facebook, I. (2019). *First quarter 2019 results conference call*. [https://s21.q4cdn.com/399680738/file_s/doc_financials/2019/Q1/Q1-'19-earnings-call-transcript-\(1\).pdf](https://s21.q4cdn.com/399680738/file_s/doc_financials/2019/Q1/Q1-'19-earnings-call-transcript-(1).pdf)

Fahey, P. M. (1991). Advocacy group boycotting of network television advertisers and its effects on programming content. *University of Pennsylvania Law Review*, 140(2), 647–709. <https://doi.org/10.2307/3312353>

Financial Times. (2024, August 7). Elon Musk can't force advertisers to spend. *Financial Times*. <https://www.ft.com/content/12a4e348-021a-4f88-a372-2b83c3cbb869>

Fischer, S. (2020, July 17). *Advertising giants agree to evaluate mutual definition of hate speech*. Axios. <https://www.axios.com/advertising-hate-speech-tech-companies-facebook-5d2b4ebf-e412-4075-ad1f-77aa2b7a07b3.html>

Flew, T. (2022). Beyond the paradox of trust and digital platforms: Populism and the reshaping of internet regulations. In T. Flew & F. R. Martin (Eds.), *Digital platform regulation* (pp. 281–309). Springer International Publishing. https://doi.org/10.1007/978-3-030-95220-4_14

Flyverbom, M., Deibert, R., & Matten, D. (2019). The governance of digital technology, big data, and the internet: New roles and responsibilities for business. *Business & Society*, 58(1), 3–19. <https://doi.org/10.1177/0007650317727540>

Garcia, A. (2024, August 12). *The end of GARM is a reset, not a setback*. Adexchange. <https://www.adexchange.com/data-driven-thinking/the-end-of-garm-is-a-reset-not-a-setback/>

Gehl, R. W. (2014). *Reverse engineering social media: Software, culture, and political economy in new media capitalism*. Temple University Press. https://www.researchgate.net/publication/261471949_Reverse_Engineering_Social_Media_Software_Culture_and_Political_Economy_in_New_Media_Capitalism

Gerbaudo, P. (2018). Social media and populism: An elective affinity? *Media, Culture & Society*, 40(5), 745–753. <https://doi.org/10.1177/0163443718772192>

Giansiracusa, N. (2021, October 15). *Facebook uses deceptive math to hide its hate speech problem*. Wired. <https://www.wired.com/story/facebooks-deceptive-math-when-it-comes-to-hate-speech/>

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. <https://doi.org/10.12987/9780300235029>

Glenday, J. (2020, October 30). *Facebook revenue soars despite July ad boycott*. The Drum. <https://www.thedrum.com/news/2020/10/30/facebook-revenue-soars-despite-july-ad-boycott>

Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1407>

Greene, T. (2018, July 17). *The daily caller's doxxing of sleeping giants was a dick move*. The Next Web. <https://thenextweb.com/news/the-daily-callers-doxxing-of-sleeping-giants-was-a-dick-move>

Griffin, R. (2023). From brand safety to suitability: Advertisers in platform governance. *Internet Policy Review*, 12(3). <https://doi.org/10.14763/2023.3.1716>

Hardy, J. (2023). Marketing communications and media: Commercial speech, censorship and control. In *The routledge companion to freedom of expression and censorship*. Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429262067-37/marketing-communications-media-jonathan-hardy?context=ubx&refId=744e664a-d91a-4d93-a34d-9542fe9bddd7>

Heath, A. (2020, July 1). *Zuckerberg tells Facebook staff he expects advertisers to return 'soon enough'*. The Information. <https://www.theinformation.com/articles/zuckerberg-tells-facebook-staff-he-expects-advertisers-to-return-soon-enough>

Heldt, A. P. (2022). EU Digital Services Act: The white hope of intermediary regulation. In T. Flew & F. R. Martin (Eds.), *Digital platform regulation* (pp. 69–84). Springer International Publishing. https://link.springer.com/10.1007/978-3-030-95220-4_4

Helmond, A., Nieborg, D. B., & Van Der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-as-infrastructure. *Internet Histories*, 3(2), 123–146. <https://doi.org/10.1080/24701475.2019.1593667>

Hill, S. (2020). Politics and corporate content: Situating corporate strategic communication between marketing and activism. *International Journal of Strategic Communication*, 14(5), 317–329. <https://doi.org/10.1080/1553118X.2020.1817029>

Hobbs, T. (2020, March 6). *'We can be catalysts for change': What 4 marketers learned from addressing the UN*. The Drum. <https://www.thedrum.com/news/2020/03/06/we-can-be-catalysts-change-what-4-marketers-learned-addressing-the-un>

Horwitz, J. (2021, September 13). Facebook says its rules apply to all. Company documents reveal a secret elite that's exempt. *Wall Street Journal*. <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353>

Jasser, G. (2021). Gab as an imitated counterpublic. In J. Bessant, M. Devries, & R. Watts (Eds.), *Rise of the far right: Technologies of recruitment and mobilization* (pp. 193–213). Rowman & Littlefield. <https://www.nomos-elibrary.de/en/10.5771/9781786614933/rise-of-the-far-right>

Jordan, J., & Nadler, J. L. (2023, March 22). *Chairman Jordan to Robert Rakowitz and Raja Rajamannar on potential anti-trust violations. 118th Congress*. <https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/2023-03-22-jdj-to-rakowitz-rajamannar-garm.pdf>

Joseph, S., & Scanlon, K. (2024, August 9). *Elon Musk's lawsuit shatters GARM, revealing industry's fracture and fear*. Digiday. <https://digiday.com/marketing/elon-musks-lawsuit-shatters-garm-revealing-industrys-fracture-and-fear/>

Karaganis, J., Schofield, B., & Urban, J. M. (2017). *Notice and takedown in everyday practice* (No. UC Berkeley Public Law Research Paper No. 2755628). UC Berkeley. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628

Kaye, K. (2021, May 24). *Getting Facebook, YouTube, TikTok, Twitter and others to independent GARM brand safety verification is a diplomatic dance*. Digiday. <https://digiday.com/marketing/as-facebook-commits-to-independent-garm-brand-safety-verification-getting-youtube-tiktok-twitter-and-others-on-board-is-a-diplomatic-dance/>

Kumar, S. (2019). The algorithmic dance: YouTube's adpocalypse and the gatekeeping of cultural content on digital platforms. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1417>

Levy, S. (2020, August 6). *Facebook has more to learn from the ad boycott*. Wired. <https://www.wired.com/story/rashad-robinson-facebook-ad-boycott/>

Mac, R., & Silverman, C. (2020, July 23). *"Hurting people at scale": Facebook's employees reckon with the social network they've built*. BuzzFeed. <https://www.buzzfeednews.com/article/ryanmac/facebook-employee-leaks-show-they-feel-betrayed>

McKelvey, F., Langlois, G., Coulter, N., & Elmer, G. (2022). Introduction: Connection issues. *Canadian Journal of Communication*, 47(1), 4-pp.

Mirrlees, T. (2021a). GAFAM and hate content moderation: Deplatforming and deleting the alt-right. *Media and Law: Between Free Speech and Censorship*, 26, 81–97.

Mirrlees, T. (2021b). 'Resisting' the far right in racial capitalism: Sources, possibilities and limits. In J. Bessant, M. Devries, & R. Watts (Eds.), *Rise of the far right: Technologies of recruitment and mobilization* (pp. 261–281). Rowman & Littlefield. https://www.researchgate.net/publication/367238116_%27Resisting%27_the_Far_Right_in_Racial_Capitalism_Sources_Possibilities_and_Limits

Monllos, K. (2020, July 9). 'We're letting Facebook grade their own homework': Here's how advertisers' desired changes differ from overall boycott. Digiday. <https://digiday.com/marketing/were-letting-facebook-grade-their-own-homework-heres-how-advertisers-desired-changes-differ-from-overall-boycott/>

Mostrous, A. (2017, February 9). Big brands fund terror through online adverts. *The Times*. <https://www.thetimes.com/business-money/technology/article/big-brands-fund-terror-knnxfgb98?region=global>

Mueller, M. L. (2015). Hyper-transparency and social control: Social media as magnets for regulation. *Telecommunications Policy*, 39(9), 804–810. <https://doi.org/10.1016/j.telpol.2015.05.001>

Mündges, S., & Park, K. (2024). But did they really? Platforms' compliance with the code of practice on disinformation in review. *Internet Policy Review*, 13(3). <https://doi.org/10.14763/2024.3.1786>

Ortiz, A. G. (2024, March 25). *Elon Musk vs. Center for Countering Digital Hate: Nonprofit wins dismissal of 'baseless and intimidating' lawsuit brought by world's richest man*. Counterhate. <https://counterhate.com/blog/elon-musk-vs-ccd-h-nonprofit-wins-dismissal-of-baseless-and-intimidatory-lawsuit/>

Ostwal, T. (2024, November 14). *Comcast, Disney, and IBM are among advertisers returning to X after ad freeze*. Adweek. <https://www.adweek.com/media/advertisers-returning-to-x/>

Parker, B. (2021, January 27). *The dangers of 'brand safety' ad technology*. The New Humanitarian. <https://www.thenewhumanitarian.org/analysis/2021/01/27/brand-safety-ad-tech-crisis-news>

Phillips, W. (2018, May 22). *The oxygen of amplification*. Data & Society. <https://datasociety.net/library/oxygen-of-amplification/>

Renton, D. (2021). *No free speech for fascists: Exploring 'no platform' in history, law and politics*. Routledge. <https://doi.org/10.4324/9781003153931>

Scanlon, K. (2024, August 6). *X files federal antitrust suit against GARM, WFA, CVS Health, Mars, Orsted, Unilever*. Digiday. <https://digiday.com/marketing/x-files-federal-antitrust-suit-against-garm-wfa-cvs-health-mars-orsted-unilever/>

Siapera, E., & Viejo-Otero, P. (2021). Governing hate: Facebook and digital racism. *Television & New Media*, 22(2), 112–130. <https://doi.org/10.1177/1527476420982232>

Statt, N. (2017, March 24). *YouTube is facing a full-scale advertising boycott over hate speech*. The Verge. <https://www.theverge.com/2017/3/24/15053990/google-youtube-advertising-boycott-hate-speech>

Stempel, J. (2023, August 17). *YouTube defeats racial bias lawsuit by black, hispanic content creators*. Reuters. <https://www.reuters.com/legal/youtube-defeats-racial-bias-lawsuit-by-black-hispanic-content-creators-2023-08-17/>

Stop hate for profit. (2022, January 29). #StopHateForProfit. Stop Hate for Profit. <https://www.stophateforprofit.org>

Suzor, N., & Gillett, R. (2022). Self-regulation and discretion. In T. Flew & F. R. Martin (Eds.), *Digital platform regulation* (pp. 259–279). Springer International Publishing. https://link.springer.com/10.1007/978-3-030-95220-4_13

Suzor, N. P. (2019). *Lawless: The secret rules that govern our digital lives*. Cambridge University Press. <https://doi.org/10.1017/9781108666428>

Sybert, J. (2022). The demise of #NSFW: Contested platform governance and Tumblr's 2018 adult content ban. *New Media & Society*, 24(10), 2311–2331. <https://doi.org/10.1177/1461444821996715>

The Wall Street Journal. (2021, October 1). The Facebook files. *The Wall Street Journal*. <https://www.wsj.com/articles/the-facebook-files-11631713039>

Turow, J. (1984). Pressure groups and television entertainment: A framework for analysis. In W. D. Rowland & B. Watkins (Eds.), *Interpreting television: Current research perspectives* (pp. 142–162). Sage Publications.

Vranica, S., & Haggin, P. (2025, January 25). Meta's free-speech shift made it clear to advertisers: 'Brand safety' is out of vogue. *The Wall Street Journal*. <https://www.wsj.com/business/media/meta-brand-safety-content-moderation-policy-changes-17308d9e>

Watson, I. (2018, December 14). *Conscious advertising network's UN address highlights need for ethical digital advertising*. The Drum. <https://www.thedrum.com/news/2018/12/14/conscious-advertising-networks-un-address-highlights-need-ethical-digital>

Willens, M. (2020, March 9). *Coronavirus climbs up keyword block lists, squeezing news publishers' programmatic revenues*. Digiday. <https://digiday.com/media/coronavirus-climbs-keyword-block-lists-squeezing-news-publishers-programmatic-revenues/>

Wodinsky, S. (2020, December 21). *Rest in peace, Facebook ads boycott (2020-2020)*. Gizmodo. <http://gizmodo.com/rest-in-peace-facebook-ads-boycott-2020-2020-1845911249>

World Federation of Advertisers. (n.d.). *Global media charter*. World Federation of Advertisers (WFA). <https://wfanet.org/leadership/global-media-charter>

World Federation of Advertisers. (2020a). *GARM Brand Safety Floor + Suitability Framework*. World Federation of Advertisers (WFA). <https://wfanet.org/l/library/download/urn:uuid:7d484745-41cd-4cce-a1b9-a1b4e30928ea/garm+brand+safety+floor+suitability+framework+23+sept.pdf>

World Federation of Advertisers. (2020b, March 3). *GARM: Small steps and big leaps in pursuit of online safety*. World Federation of Advertisers (WFA). <https://wfanet.org/knowledge/item/2020/03/03/GARM-Small-steps-and-big-leaps-in-pursuit-of-online-safety>

World Federation of Advertisers. (2021, April 22). *GARM aggregated measurement report – April 2021*. World Federation of Advertisers (WFA). <https://wfanet.org/knowledge/item/2021/04/22/GARM-Aggregated-Measurement-Report-April-2021>

World Federation of Advertisers. (2022). *GARM: 3 years of progress*. World Federation of Advertisers (WFA). <https://wfanet.org/leadership/garm/garm-resource-directory-%28weblog-detail-page%29/2022/06/20/GARM-3-Years-of-Progress>

Yin, L., & Sankin, A. (2021, April 9). *Google blocks advertisers from targeting Black Lives Matter YouTube videos*. The Markup. <https://themarkup.org/google-the-giant/2021/04/09/google-blocks-ad>

vertisers-from-targeting-black-lives-matter-youtube-videos

Young, G. (2023, September 7). *How conscious advertising network could redeem its status as an honest broker*. The Drum. <https://www.thedrum.com/opinion/2023/09/07/how-conscious-advertising-network-could-redeem-its-status-honest-broker>

Appendix 1: Documents collected for Stop Hate for Profit and Global Alliance for Responsible Media case study

Abril, D. (2020, July 23). These big companies have cut the most daily ad dollars during the Facebook boycott. *Fortune*. <https://fortune.com/2020/07/23/facebook-ads-boycott-big-companies-cuts-ad-spending-microsoft-samsung-starbucks-stop-hate-for-profit/>

All the companies quitting Facebook. (2020, June 29). *The New York Times*. <https://www.nytimes.com/2020/06/29/business/dealbook/facebook-boycott-ads.html>

An update on combating hate and dangerous organizations. (2020, May 12). *About Facebook*. <https://about.fb.com/news/2020/05/combating-hate-and-dangerous-organizations/>

Anti-Defamation League. (2020a, June 25). *An open letter to the companies that advertise on Facebook*. <https://www.adl.org/resources/letter/open-letter-companies-advertise-facebook>

Anti-Defamation League (Director). (2020b, June 26). *Stop Hate for Profit* [Video recording]. https://www.youtube.com/watch?v=STeH3uJDC_w

Anti-Defamation League (Director). (2020c, July 21). *Stop Hate for Profit: "Dear Mark"* [Video recording]. https://www.youtube.com/watch?v=AUT__Y-VXtl

ANZA calls for international solidarity. (2019, March 27). *WARC*. http://www.warc.com/newsandopinion/news/anza_calls_for_international_solidarity/41861

Ben & Jerry's. (2020, June 23). We're joining the #StopHateForProfit campaign. <https://www.Benjerry.Com>. <https://www.benjerry.com/about-us/media-center/stop-hate-for-profit>

Bergen, M. (2018, April 20). P&G ends its YouTube advertising boycott, but with a catch. *Bloomberg.Com*. <https://www.bloomberg.com/news/articles/2018-04-20/p-g-ends-its-youtube-advertising-boycott-but-with-a-catch>

Bickert, M. (2020). *Charting a way forward: Online content regulation*. Facebook.

https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf

Big outdoor brands join #StopHateForProfit campaign, boycott Facebook and Instagram ads. (2020, June 22). *TechCrunch*. Retrieved 10 September 2020, from <https://social.techcrunch.com/2020/06/22/facebook-instagram-rei-patagonia-north-face-brands-join-stop-hate-for-profit-boycott/>

Bond, S. (2020, June 26). In reversal, Facebook to label politicians' harmful posts as ad boycott Grows. *NPR*. <https://www.npr.org/2020/06/26/883941796/unilever-maker-of-dove-soap-is-latest-brand-to-boycott-facebook>

Bradley, S., & Joseph, S. (2025, January 8). Meta follows Musk's lead on censorship—But ad industry keeps its distance from panic. *Digiday*. <https://digiday.com/marketing/meta-follows-musks-lead-on-censorship-but-ad-industry-keeps-its-distance-from-panic/>

Carruthers, N. (2020, June 29). Diageo to stop advertising on social media. *The Spirits Business*. <https://www.thespiritsbusiness.com/2020/06/diageo-to-stop-advertising-on-social-media/>

Cavale, S. (2020, July 1). P&G says will not disclose advertising decisions as Facebook ad boycott grows. *Reuters*. <https://www.reuters.com/article/us-facebook-ads-boycott-p-n-g-idUSKBN242605>

Chairman Jordan Subpoenas GARM and WFA for documents and communications. (2023, May 5). *House Judiciary Committee Republicans*. <http://judiciary.house.gov/media/press-releases/chairman-jordan-subpoenas-garm-and-wfa-documents-and-communications>

Clayton, J. (2020, July 2). Zuckerberg: Advertisers will be back to Facebook 'soon enough'. *BBC News*. <https://www.bbc.com/news/technology-53262860>

Clegg, N. (2020a, May 6). Welcoming the Oversight Board. *Meta*. <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>

Clegg, N. (2020b, July 1). Facebook does not benefit from hate. *Meta*. <https://about.fb.com/news/2020/07/facebook-does-not-benefit-from-hate/>

Dang, S., & Paul, K. (2020, July 1). Facebook frustrates advertisers as boycott over hate speech kicks off. *Reuters*. <https://uk.reuters.com/article/us-facebook-ads-boycott-idUKKBN2424GS>

Dear Mark. (2020, July 27). *Stop Hate for Profit*. <https://www.stophateforprofit.org/dear-mark-video>

Desjardins Group. (2020, July 2). *Desjardins suspends posting and advertising on*

Facebook and Instagram. Retrieved 28 January 2025, from <https://www.newswire.ca/news-releases/desjardins-suspends-posting-and-advertising-on-facebook-and-instagram-823878509.html>

Disney quietly slashes Facebook spend, adding pressure on the platform. (2020, July 21). *WARC*. <http://www.warc.com/newsandopinion/news/disney-quietly-slashes-facebook-spend-adding-pressure-on-the-platform/en-gb/43876>

Dwoskin, E. (2020, June 28). Zuckerberg once wanted to sanction Trump. Then Facebook wrote rules that accommodated him. *Washington Post*. <https://www.washingtonpost.com/technology/2020/06/28/facebook-zuckerberg-trump-hate/>

Dwoskin, E., Tiku, N., & Kelly, H. (2020, December 3). Facebook to start policing anti-Black hate speech more aggressively than anti-White comments, documents show. *Washington Post*. <https://www.washingtonpost.com/technology/2020/12/03/facebook-hate-speech/>

Dwoskin, E., Zakrzewski, C., Menn, J., Verma, P., Siddiqui, F., Osaka, S., O'Donovan, C., Gurley, L. K., Tiku, N., & Kelly, H. (2020, July 8). Facebook's own civil rights auditors say its policy decisions are a 'tremendous setback'. *Washington Post*. <https://www.washingtonpost.com/technology/2020/07/08/facebook-civil-rights-audit/>

Elon Musk can't force advertisers to spend. (2024, August 7). *Financial Times*. <https://www.ft.com/content/12a4e348-021a-4f88-a372-2b83c3cbb869>

Embury-Dennis, T. (2019, April 4). Steve Bannon caught admitting Breitbart lost 90% of advertising revenue due to boycott. *The Independent*. <https://www.independent.co.uk/news/world/americas/us-politics/steve-bannon-breitbart-boycott-advertising-sleeping-giants-trump-a8854381.html>

Facebook, Inc. (2019). *First quarter 2019 results conference call*. [https://s21.q4cdn.com/399680738/files/doc_financials/2019/Q1/Q1-19-earnings-call-transcript-\(1\).pdf](https://s21.q4cdn.com/399680738/files/doc_financials/2019/Q1/Q1-19-earnings-call-transcript-(1).pdf)

Facebook, Inc. (2020). *Second Quarter 2020 Results Conference Call*. https://s21.q4cdn.com/399680738/files/doc_financials/2020/q2/FB-Q2%2720-Earnings-Transcript.pdf

Facebook joins other tech companies to support the Christchurch Call to Action. (2019, May 15). *Meta*. <https://about.fb.com/news/2019/05/christchurch-call-to-action/>

Fischer, S. (n.d.). *Advertising giants agree to evaluate mutual definition of hate speech*. *Axios*. Retrieved 21 July 2020, from <https://www.axios.com/advertising-hate-speech-tech-companies-facebook-5d2b4ebf-e412-4075-ad1f-77aa2b7a07b3.html>

Fischer, S. (2020, July 28). Scoop: Facebook boycotters lobby lawmakers on antitrust. *Axios*. <https://www.axios.com/2020/07/28/scoop-facebook-boycotters-lobby-lawmakers-on-antitrust>

Franklin, H., & Clement, C. (2020, July 7). Facebook advertisers revolt: It's time to #StopHateForProfit. *Free Press*. <https://www.freepress.net/our-response/expert-analysis/explainers/facebook-advertisers-revolt-its-time-stop-hate-for-profit>

Fried, I. (2020, June 2). Civil rights leaders blast Facebook after meeting with Zuckerberg. *Axios*. <https://www.axios.com/2020/06/02/civil-rights-leaders-blast-facebook-after-meeting-with-zuckerberg>

Ghaffary, S. (2020, June 3). Read the transcript of Mark Zuckerberg's tense meeting with Facebook employees. *Vox*. <https://www.vox.com/recode/2020/6/3/21279434/mark-zuckerberg-meeting-facebook-employees-transcript-trump-looting-shooting-post>

Glenday, J. (2020a, September 23). Facebook, YouTube and Twitter advance hate speech talks with brands. *The Drum*. <https://www.thedrum.com/news/2020/09/23/facebook-youtube-and-twitter-advance-hate-speech-talks-with-brands>

Glenday, J. (2020b, October 30). Facebook revenue soars despite July ad boycott. *The Drum*. <https://www.thedrum.com/news/2020/10/30/facebook-revenue-soars-despite-july-ad-boycott>

The Global Alliance for Responsible Media. (2020). *GARM Brand Safety Floor + Suitability Framework*. <https://wfanet.org/l/library/download/urn:uuid:7d484745-41cd-4cce-a1b9-a1b4e30928ea/garm+brand+safety+floor+suitability+framework+23+sept.pdf>

Goode, L. (2023, November 29). Elon Musk just told advertisers, 'Go fuck yourself'. *Wired*. Retrieved 18 December 2024, from <https://www.wired.com/story/elon-musk-x-advertisers-interview/>

Honda. (2020, June 4). Our perspectives: Philosophy without action is worthless. *Honda Corporate Social Responsibility*. <https://csr.honda.com/2020/06/04/our-perspectives-philosophy-without-action-is-worthless/>

Horwitz, J. (2021, September 13). Facebook says its rules apply to all. Company documents reveal a secret elite that's exempt. *WSJ*. <https://www.wsj.com/articles/facebook-files-x-check-zuckerberg-elite-rules-11631541353>

Howard, J. (2020, June 22). Tech CMOs: Together we can make Facebook stop hate. *Dashlane Blog*. <https://blog.dashlane.com/make-facebook-stop-hate/>

Hsu, T., & Friedman, G. (2020, June 26). CVS, Dunkin', Lego: The brands pulling ads from Facebook over hate speech. *The New York Times*. <https://www.nytimes.com/>

2020/06/26/business/media/Facebook-advertising-boycott.html

Hsu, T., & Lutz, E. (2020, August 1). More than 1,000 companies boycotted Facebook. Did it work? *The New York Times*. <https://www.nytimes.com/2020/08/01/business/media/facebook-boycott.html>

Introducing new brand safety controls for advertisers. (n.d.). *Meta for Business*. Retrieved 28 January 2025, from <https://en-gb.facebook.com/business/news/introducing-new-brand-safety-controls-for-advertisers>

Join Mozilla in telling major companies to #StopHateForProfit. (2020, July 2). *Mozilla Foundation*. <https://foundation.mozilla.org/en/blog/join-mozilla-telling-major-companies-stop-hate-for-profit/>

Joseph, K., & Scanlon, S. (2024, August 9). *Elon Musk's lawsuit shatters GARM, revealing industry's fracture and fear*. Digiday. <https://digiday.com/marketing/elon-musks-lawsuit-shatters-garm-revealing-industrys-fracture-and-fear/>

Joseph, S. (2020a, June 23). '180 degrees from the intent:' Why marketers are living in constant fear of the screenshot. *Digiday*. <https://digiday.com/media/180-degrees-from-the-intent-why-marketers-are-living-in-constant-fear-of-the-screenshot/>

Joseph, S. (2020b, August 3). As the Facebook boycott ends, brand advertisers are split on what happens next with their marketing budgets. *Digiday*. <http://digiday.com/media/as-the-facebook-boycott-ends-brand-advertisers-are-split-on-what-happens-next-with-their-marketing-budgets/>

Kaplan, J. (2025, January 7). More speech and fewer mistakes. *Meta*. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>

Karr, T. (2020, July 2). What pundits get wrong about the Facebook ad boycott. *Free Press*. <https://www.freepress.net/blog/what-pundits-get-wrong-about-facebook-ad-boycott>

Kaye, K. (2021, July 12). Facebook delays its brand safety audit a year after ad boycott raged. *Digiday*. <https://digiday.com/marketing/facebook-delays-its-brand-safety-audit-a-year-after-ad-boycott-raged/>

Kelly, M. (2020, June 26). Unilever will pull ads from Facebook, Instagram, and Twitter for the rest of the year. *The Verge*. <https://www.theverge.com/2020/6/26/21304619/unilever-facebook-instagram-twitter-ad-boycott-spending-dove-hellmans>

Klein, E. (2018, April 2). *Mark Zuckerberg on Facebook's hardest year, and what comes next*. Vox. <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge>

- Levy, S. (2020, August 6). Facebook has more to learn from the ad boycott. *Wired*. <https://www.wired.com/story/rashad-robinson-facebook-ad-boycott/>
- Liffreing, I. (2020, September 9). *Coke, McDonald's, Nike still boycotting Facebook, but it's no longer just about hate*. <https://adage.com/article/cmo-strategy/coke-mcdonalds-nike-still-boycotting-facebook-its-no-longer-just-about-hate/2278981>
- Mac, R., & Silverman, C. (2020, July 23). "Facebook is hurting people at scale": Mark Zuckerberg's employees reckon with the social network they've built. *BuzzFeed News*. <https://www.buzzfeednews.com/article/ryanmac/facebook-employee-leaks-show-they-feel-betrayed>
- Magalhães, J. C. (2024, December 4). Musk Reinvented Platform Power. We Should Take Heed. *Tech Policy Press*. <https://techpolicy.press/musk-reinvented-platform-power-we-should-take-heed>
- Mars, Incorporated. (2020, June 30). *Mars' Response to Hate Speech and Misinformation on Social Media*. <https://www.mars.com/news-and-stories/press-releases/mars-social-media-response>
- McCarthy, J. (n.d.). Conversation, not boycotts, is the way to fix social media says Unilever marketing boss. *The Drum*. Retrieved 11 July 2020, from <https://www.thedrum.com/news/2020/07/07/conversation-not-boycotts-the-way-fix-social-media-says-unilever-marketing-boss>
- Monllos, K. (2020, July 9). 'We're letting Facebook grade their own homework': Here's how advertisers' desired changes differ from overall boycott. *Digiday*. <https://digiday.com/marketing/were-letting-facebook-grade-their-own-homework-heres-how-advertisers-desired-changes-differ-from-overall-boycott/>
- Mostrous, A. (2017, February 9). Big brands fund terror through online adverts. *The Times*. <https://www.thetimes.co.uk/article/big-brands-fund-terror-knnxfgb98>
- Murphy, H., Thomas, D., & Platt, E. (2024, November 13). Advertisers set to return to X as they seek favour with Elon Musk and Donald Trump. *Financial Times*. <https://www.ft.com/content/34b6fc20-23f7-4e08-9ac4-ef05d5d66c13>
- Murphy, L. W. (2020). *Facebook's Civil Rights Audit—Final Report*.
- National Hispanic Media Coalition. (n.d.). *Non-profit SHFP letter to Facebook Oversight Board*. Retrieved 28 January 2025, from <https://www.nhmc.org/non-profit-shfp-letter-to-facebook-oversight-board/>
- News Release: CMA supports efforts to eliminate hatred on social platforms. (2020, July 2). *Canadian Marketing Association*. <https://thecma.ca/topic/media/2020/07/02/news-release-cma-supports-efforts-to-eliminate-hatred-on-social-platforms>

- Ostwal, T. (2024, November 14). *Comcast, Disney, and IBM are among advertisers returning to X after ad freeze*. <https://www.adweek.com/media/advertisers-returning-to-x/>
- Pash, C. (2019, March 27). The NZ advertising industry wants a global boycott of Facebook. *AdNews*. Retrieved 23 January 2020, from <https://www.adnews.com.au/news/the-nz-advertising-industry-wants-a-global-boycott-of-facebook>
- Patagonia joins the Stop Hate for Profit campaign. (2020, June 21). *Patagonia Works*. Retrieved 7 August 2020, from <http://www.patagoniaworks.com/press/2020/6/21/patagonia-joins-the-stop-hate-for-profit-campaign>
- Paul, K., & Dang, S. (2020, July 8). Facebook ad boycott organizers see ‘no commitment to action’ in Zuckerberg meeting. *Reuters*. <https://www.reuters.com/article/us-facebook-ads-boycott-idUSKBN2482W2>
- PUMA. (n.d.). We are proud to join the #StopHateforProfit boycott. We will stop all advertisements on Facebook and Instagram throughout July. *Twitter*. Retrieved 2 July 2020, from <https://twitter.com/PUMA/status/1277719215265525761>
- Recommended Next Steps. (n.d.). *Stop Hate for Profit*. Retrieved 16 July 2020, from <https://www.stophateforprofit.org/productrecommendations>
- Sandberg, S. (2020, July 8). Making progress on civil rights – but still a long way to go. *Meta*. <https://about.fb.com/news/2020/07/civil-rights-audit-report/>
- Sarang, V. (2020, August 11). Independent audit of community standards enforcement report metrics. *About Facebook*. <https://about.fb.com/news/2020/08/independent-audit-of-enforcement-report-metrics/>
- Sharing Our Actions on Stopping Hate. (n.d.). *Meta for Business*. Retrieved 28 January 2025, from <https://www.facebook.com/business/news/sharing-actions-on-stopping-hate>
- Sharing Our Actions on Stopping Hate. (2020, July 30). *Facebook for Business*. <https://www.facebook.com/business/news/sharing-actions-on-stopping-hate>
- Sloane, G. (2020, August 3). *Facebook’s Carolyn Everson opens up about the boycott, that Trump post, and actions vs. lip service*. <https://adage.com/article/digital/facebooks-carolyn-evererson-opens-about-boycott-trump-post-and-actions-vs-lip-service/2272076>
- Statement from Stop Hate for Profit on Ad Pause Success and #StopHateForProfit Campaign. (2020, July 30). *Stop Hate for Profit*. <https://www.stophateforprofit.org/july-30-statement>
- Statement from Stop Hate For Profit on Meeting with Facebook | NAACP*. (2020, July 7).

<https://naacp.org/articles/statement-stop-hate-profit-meeting-facebook>

Statt, N. (2017, March 24). *YouTube is facing a full-scale advertising boycott over hate speech*. The Verge. <https://www.theverge.com/2017/3/24/15053990/google-youtube-advertising-boycott-hate-speech>

Statt, N. (2020, June 25). Verizon is the biggest advertiser to join the Facebook ad boycott so far. *The Verge*. <https://www.theverge.com/2020/6/25/21303717/verizon-facebook-adl-ad-boycott-instagram-north-face-rei-ben-and-jerrys>

Stop Hate for Profit. (n.d.). *Color Of Change*. Retrieved 2 July 2020, from <https://colorofchange.org/stop-hate-for-profit/>

Stop Hate For Profit. (n.d.). *Anti-Defamation League*. Retrieved 29 January 2022, from <https://www.adl.org/stop-hate-for-profit-0>

Stop Hate for Profit. (n.d.-b). Stop Hate for Profit. Retrieved 29 January 2022, from <https://www.stophateforprofit.org>

Stop Hate for Profit update to advertisers. (n.d.). *Stop Hate for Profit*.

Times, F. (2020, June 25). Facebook executive admits to “trust deficit” on call with advertisers. *Ars Technica*. <https://arstechnica.com/tech-policy/2020/06/facebook-executive-admits-to-trust-deficit-on-call-with-advertisers/>

Vranica, S. (2020a, June 18). Ad agency encourages clients to join Facebook ad boycott. *Wall Street Journal*. <https://www.wsj.com/articles/ad-agency-encourages-clients-to-join-facebook-ad-boycott-11592517885>

Vranica, S. (2020b, July 18). Disney slashed ad spending on Facebook amid growing boycott. *Wall Street Journal*. <https://www.wsj.com/articles/disney-slashed-ad-spending-on-facebook-amid-growing-boycott-11595101729>

Vranica, S., & Haggin, P. (2025, January 26). Meta’s Free-Speech Shift Made It Clear to Advertisers: ‘Brand Safety’ Is Out of Vogue. *WSJ*. <https://www.wsj.com/business/media/meta-brand-safety-content-moderation-policy-changes-17308d9e>

We support the #StopHateForProfit movement. (2020, July 1). *Dogfish Head Craft Brewed Ales | Off Centered Stuff For Off Centered People*. <https://www.dogfish.com/blog/we-support-stophateforprofit-movement>

Wodinsky, S. (2020, June 26). The ‘Stop Hate For Profit’ Movement Isn’t Going to Stop Anything. *Gizmodo*. Retrieved 31 August 2020, from <https://gizmodo.com/the-stop-hate-for-profit-movement-isnt-going-to-stop-1844147197>

Wodinsky, S. (2020, December 21). Rest in peace, Facebook Ads Boycott (2020–2020). *Gizmodo*. <https://gizmodo.com/rest-in-peace-facebook-ads->

boycott-2020-2020-1845911249

Wong, Q. (2020, July 30). Why big brands are turning their backs on Facebook. *CNET*. Retrieved 31 August 2020, from <https://www.cnet.com/news/facebook-ad-boycott-how-big-businesses-hit-pause-on-hate/>

World Federation of Advertisers. (2019a, March 28). *WFA urges brands to hold platforms to account*. <https://wfanet.org/knowledge/item/2019/03/28/WFA-urges-brands-to-hold-platforms-to-account>

World Federation of Advertisers. (2019b, May 15). *WFA statement on the 'Christchurch Call to Action' in Paris*. <https://wfanet.org/knowledge/item/2019/05/15/WFA-statement-on-the-Christchurch-Call-to-Action-in-Paris>

World Federation of Advertisers. (2019c, June 18). *Global Alliance for Responsible Media launches to address digital safety*. <https://wfanet.org/knowledge/item/2019/06/18/Global-Alliance-for-Responsible-Media-launches-to-address-digital-safety>

World Federation of Advertisers. (2019d, June 18). *Global Alliance for Responsible Media launches to address digital safety*. <https://wfanet.org/knowledge/item/2019/06/18/Global-Alliance-for-Responsible-Media-launches-to-address-digital-safety>

World Federation of Advertisers. (2020a, January 22). *Marketing leaders take action on harmful online content*. <https://wfanet.org/knowledge/item/2020/01/22/Marketing-leaders-take-action-on-harmful-online-content>

World Federation of Advertisers. (2020b, March 3). *GARM: Small steps and big leaps in pursuit of online safety*. <https://wfanet.org/knowledge/item/2020/03/03/GARM-Small-steps-and-big-leaps-in-pursuit-of-online-safety>

World Federation of Advertisers. (2020c, June 22). *Advertisers demand transparency and clarity around platforms' content policies*. <https://wfanet.org/knowledge/item/2020/06/22/Advertisers-demand-transparency-and-clarity-around-platforms-content-policies>

World Federation of Advertisers. (2020d, July 1). *Nearly a third of advertisers pull back or consider pulling back from platforms*. <https://wfanet.org/knowledge/item/2020/07/01/Nearly-a-third-of-advertisers-pull-back-or-consider-pulling-back-from-platforms>

World Federation of Advertisers. (2021). *GARM Aggregated Measurement Report—April 2021*. World Federation of Advertisers. <https://wfanet.org/knowledge/item/2021/04/22/GARM-Aggregated-Measurement-Report-April-2021>

World Federation of Advertisers. (2022a). *GARM: 3 Years of Progress*. World Federation of Advertisers. <https://wfanet.org/leadership/garm/garm-resource-directory-%28weblog-detail-page%29/2022/06/20/GARM-3-Years-of-Progress>

World Federation of Advertisers. (2022b, June 21). *GARM announces guidelines on misinformation, standards on ad placements, and expansion to cover the metaverse*. <https://wfanet.org/knowledge/item/2022/06/21/GARM-announces-guidelines-on-misinformation-standards-on-ad-placements-and-expansion-to-cover-the-metaverse>

World Federation of Advertisers. (2023, April 25). *WFA issues new rallying cry for fairer, safer, more transparent and more sustainable media ecosystem for global advertisers, with publication of Global Media Charter 3.0*. <https://wfanet.org/knowledge/item/2023/04/25/WFA-issues-new-rallying-cry-for-fairer-safer-more-transparent-and-more-sustainable-media-ecosystem-for-global-advertisers-with-publication-of-Global-Media-Charter-30>

World Federation of Advertisers. (2024, August 9). *Statement on the Global Alliance for Responsible Media (GARM)*. <https://wfanet.org/leadership/garm/about-garm>

World Federation of Advertisers. (n.d.). *Global Media Charter*. Retrieved 30 July 2020, from <https://wfanet.org/leadership/global-media-charter>

Yin, L., & Sankin, A. (2021, April 9). Google blocks advertisers from targeting Black Lives Matter YouTube Videos. *The Markup*. <https://themarkup.org/google-the-giant/2021/04/09/google-blocks-advertisers-from-targeting-black-lives-matter-youtube-videos>

Young, G. (2023, September 7). How Conscious Advertising Network could redeem its status as an honest broker. *The Drum*. <https://www.thedrum.com/opinion/2023/09/07/how-conscious-advertising-network-could-redeem-its-status-honest-broker>

Zuckerberg, M. (2020, June 26). *An update from our company town hall*. Retrieved 10 September 2020, from <https://www.facebook.com/zuck/videos/an-update-from-our-company-town-hall/10112048862145471/>

Zuckerberg, M. (2020, July 30). Facebook post July 30. *Mark Zuckerberg Newsfeed*. Retrieved 12 August 2020, from <https://www.facebook.com/zuck/posts/10112048980882521>

Zuckerberg, M. (2019). Zuckerberg Facebook post about standing for voice and free expression. *Facebook*. https://epublications.marquette.edu/zuckerberg_files_transcripts/1022

Zuckerberg, M. (2020, June 5). *Facebook Post June 5*. <https://www.facebook.com/zuck/posts/i-just-shared-the-following-note-with-our-employees-and-i-want-to-share-it-with-/10111985969467901/>

Zuckerberg, Mark. (2020, May 29). *Facebook post May 29*. <https://www.facebook.com/zuck/posts/this-has-been-an-incredibly-tough-week-after-a-string-of-tough-weeks-the-killing/10111961824369871/>

Published by



ALEXANDER VON HUMBOLDT
INSTITUTE FOR INTERNET
AND SOCIETY

in cooperation with



CREATE



centre
— internet
et societe



R&I

IN3

Internet
interdisciplinary
Institute

Universitat Oberta de Catalunya



UNIVERSITY OF TARTU
Johan Skytte Institute of
Political Studies