Fight Fire with Fire: Hacktivists' Take on Social Media Misinformation

Filipo Sharevski and Benjamin Kessell, ***DePaul University***

https://www.usenix.org/conference/soups2023/presentation/sharevski

This paper is included in the Proceedings of the Nineteenth Symposium on Usable Privacy and Security.

August 7–8, 2023 • Anaheim, CA, USA

978-1-939133-36-6

Open access to the Proceedings of the Nineteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Fight Fire with Fire: Hacktivists' Take on Social Media Misinformation

Filipo Sharevski
DePaul University

Benjamin Kessell
DePaul University

Abstract

In this study, we interviewed 22 prominent hacktivists to learn their take on the increased proliferation of misinformation on social media. We found that none of them welcomes the nefarious appropriation of trolling and memes for the purpose of political (counter)argumentation and dissemination of propaganda. True to the original *hacker* ethos, misinformation is seen as a threat to the democratic vision of the Internet, and as such, it must be confronted head on with tried hacktivism methods: deplatforming the "misinformers" and doxing their funding and recruitment. The majority of the hacktivists we interviewed recommended interventions for promoting misinformation literacy in addition to targeted hacking campaigns. We discuss the implications of these findings relative to the emergent recasting of hacktivism as a defense of a constructive and factual social media discourse.

1 Introduction

Steven Levy's portrayal of the hacker culture in his 1984 book *Hackers* largely remains the most influential reference to the public's general view of hackers [45, 67]. Recasting them as Robin Hood-style activists committed to a democratic vision of the Internet [101], Levy asserts that the hacker ethos embodies several sacrosanct postulates to the public good, notably that: (i) *all information should be free*, and (ii) *authority should be mistrusted and decentralization promoted* [67].

Later-day Internet hackers shifted towards an ideology oriented around autonomy in cyberspace. In this view, the Internet is seen as a politicized, public, information sharing space and as a valuable weapon against the neoliberal elites, who they see as responsible for economic and social disarray [40]. In other words, hacktivists took a front against the "*neoliberalism*" or the sociopolitical right-of-center positioning of individualized, market-based competition as the preferred governing principle for shaping human action in all areas of life – including the Internet – both at the individual and collective, societal levels [122]. Turning Internet activism into a form of socio-political resistance online [60] enabled a functional selection of issues that no longer necessitated lengthy preparations [77]. This, in turn, resulted in almost instant convergence and coordination of activities in response to the issues of interest. These campaigns in turn generated significant public visibility via coverage by mass media (e.g. television, newspapers, magazines, and radio) [49].

The Internet activism bifurcated to online campaigns concerned with the protection of the Internet as a relatively unregulated and unowned space (e.g. Anonymous, WikiLeaks, Snowden [23, 118, 120]) and online campaigns concerned with the protection of human rights and the environment (e.g. the Occupy movement, Arab Spring, Pirate Party [61, 84]). The former activism – or *hacktivism* – is often anonymous, performed in secret, and operates with a kind of impunity thus far afforded by networking technologies [121]. The latter activism – or *hashtag activism* – is usually public, openly leverages the Internet for political mobilization, operates primarily on the streets, and is subject to the dangers of crowd violence, harassment, and arbitrary arrest [104].

Hashtag activism historically utilized various technologies like petition websites (e.g. MoveOn.org for organizing political protests) or e-mail communication (e.g. Tea Party's campaign to reduce government spending and taxation) [18], but the advent of social media sites like Twitter, Facebook, and YouTube dramatically accelerated the self-organization and participation in the sociopolitical struggle (e.g. the #BlackLivesMatter and #SchoolStrike4Climate movements [37]). For hashtag activism there is a historical and ongoing essential dependence on social media [58]. The relationship between hacktivism and social media, however, is more complicated.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2023.
August 6–8, 2023, Anaheim, CA, USA

Hacktivists, in contrast, have hacked various technologies to defacing websites [102], broken into systems to "leak" private documents and "dox" individuals [118, 123], and have overwhelmed systems with traffic to cause a Denial-of-Service (DOS) [85]. Hacktivists' foray in social media mirrors these actions as campaigns were undertaken for hijacking/defacement of social media accounts (e.g., Anonymous's #OpKKK campaign [134]), doxing individuals on Twitter (e.g. the students of Covington High School [72]), and DoS Twitter topics (e.g. #IranTalks campaign [90]). But hacktivists also hacked the social media affordances for content amplification (e.g. StayWokeBot [39, 106]), early instances of trolling (e.g. Rickrolls [105]), and sharing memes (e.g. Lol Cats on 4chan [23]).

Despite the intuitive versatility of social media for such subversive operations, hacktivism became largely inactive on the mainstream platforms following some high profile run-ins of leading hacktivists with the legal authorities [55, 130]. The apparent absence of hacktivism created a vacuum where no one actively challenged the elites, defended freedom of expression, and appended the vision of democratic social media participation. It took little time, unfortunately, for this vacuum to be appropriated by state-sponsored actors hijacking the hacking playbook for actions aimed not just against the neoliberal elites but the entire social order [35]. Bot-enabled amplification aided political trolling and sharing of memes during the Brexit campaign in the UK [26] and the 2016 elections in the US [11]. The crucial difference in these instances was that the amplified memes and trolling were not pranks but damaging fake news, emotionally-charged memes, and conspiracy theories that instead of unifying the social media crowds for a cause, divided them in opposition camps that were pitted against each other [115].

In response to such a large-scale disruption on the social media landscape, one would have plausibly expected for hacktivists to retaliate, confront, expose, or counter-hack the state-sponsored "trolls" [141]. Misinformation, back to the Levy's depiction of hacker's ethics [67], runs counter the first postulate (i) *all information should be free* because it undermines the basic utility of information as a public good (i.e. truth and facts do not dwindle in supply as more people "consume" them and truth and facts are available to all people in a society) [34]. Misinformation also runs counter the second postulate (ii) *authority should be mistrusted and decentralization promoted* because it is promulgated by a state-sponsored "shadow authority," as evidence confirms in the aftermath of the Brexit and the 2016 US elections [50, 75, 140]. Surprisingly, the hacktivists never struck back [12], though they clearly possessed the capabilities to do so, as evidenced in the Anonymous's #OpISIS campaign, for instance, where the collective flagged about 101,000 Twitter accounts attributed to the Islamic-State [51].

The absence of response to misinformation on social media by the hacktivist community seemed quite perplexing and, in our opinion, worthy of in-depth inquiry with active "hackers" that still operate in the spirit of the Levy's code of ethics [67]. Through personal connections and snowball sampling, we identified 22 prominent hacktivists and conducted hour-long interviews with each of them to learn their take on the misinformation ecosystem, on responses to falsehoods on social media, and on the way misinformation impacts and shapes the hacktivists' agenda in the future. We found a consensus among the hacktivists against the present forms of misinformation as an ammunition for political counter(argumentation) and external propaganda. They recommended actions to deplatform, dox, and expose every "misinformer" that is believed to pollute the social media discourse, and suggested ways to improve the general misinformation literacy among users in addition to these targeted operations.

To situate our study in the intersection between the hacktivist counter-culture and the rise of misinformation on platforms, we review the interplay between Internet activism, social media, and false information in Section 2. We look in the broader context of misinformation in Section 3 to highlight the pressing need of (hack)tivism action to reclaim the social media space true to Levy's vision of Internet as an information exchange to the public good. In Section 4 we outline our research design and methodology. Sections 5, 6, and 7 expand on our findings and we discuss the implications of the hackers' disposition to social media misinformation in Section 8. Finally, Section 9 concludes the paper.

2 Internet Activism and Social Media

2.1 Hashtag Activism

Online social media activism – or *slacktivism*, *clicktivism* – emerged on popular platforms as a repertoire of low-risk, low-cost expressive activities for advocacy groups' agenda setting and political participation [103]. Social media users participated in petitions, changed personal avatars, added picture filters in support of a cause, and simply "liked" posts as an act of participation [43]. Slacktivists quickly realized they could use virality as a distinctive social media affordance to their advantage and move to use hashtags as the main drivers of mobilization, raising awareness, and demanding sociopolitical change. The practice of *hashtag activism* was instrumental for the success of social movements like #metoo, #takeaknee, and #BlackLivesMatter, allowing for visibility, expression of solidarity, and statement of victimhood [119]. This success, in turn, inspired a plethora of other movements advocating for health, human rights, social justice, and environmental issues across all social media platforms as a trend that remains active and prominent across online public discourse [54].

The advent of the hashtag activism, however noble, had to deal with the obvious threat of *hashtag hijacking*, or the appropriation of viral hashtags as a vehicle to inject contrary perspectives into the discourse [132]. This "hack" against

Internet activism is not just adding noise or attempting to result in a DoS, but is also used to disseminate hateful narratives and dilute the campaign itself (e.g. the hijacking of the #metoo hashtag [71]). Another similar threat is *hashtag co-opting*, or the contentious co-opting of the rhetoric of popular social movements (e.g. #HeterosexualPrideDay campaign co-opting the language of the mainstream LGBT movement [8]). Equally threatening is *counter hashtagging*, which concocts similar hashtags to garner opposition to well-established movements (e.g. #BlueLivesMatter countermovement to police reform in reaction to #BlackLivesMatter [63]). These antagonistic appropriations of social media virality enable political extremism to creep in the public discourse and embroil users in an emotionally-charged participation [99].

In an age of emerging social media polarization, it was a matter of time before fake news, offensive memes, and conspiracy theories would be weaponized against hashtag activism (e.g. the proliferation of fake news in the #Gunreformnow vs #NRA Twitter battle [20]). What was initially expected to remain on the fringes of the mainstream hashtag activism [36], quickly turned into information disorder on a mass scale. Today hashtag hijacking and co-opting develops *in parallel* with activism campaigns, feeding from and perpetuating an ecosystem of false and unverified information. This emotionally-charged participation has manifested within a global health panic (e.g. #FlattenTheCurve hashtag hijacking for COVID-19 misinformation [29]) and moral panic (e.g. the QAnon's co-opting of #SaveTheChildren hashtag [87]) in addition to the already growing political panic [89].

2.2 Hacktivism

Hacktivism was a term that "Omega," a member of the Texas-based computer-hacking group *Cult of the Dead Cow* (cDc) coined in 1996 in an email to the cDc listserv [78]. Characterized with the increasingly political ethos of hacking-for-cause, hacktivists primarily leveraged technology to advance human rights and protect the free flow of information in campaigns against the UK, US, and Chinese governments, as well as the UN [92]. In as much as hackers individually roamed the Internet, socialization was increasingly desired as many of them needed to establish a strong hacktivist network. Hacktivists' penchant for humorous memes (LOLCats) and gag hyperlinks (Rickrolls) [95] attracted an army of hackers to Christopher Poole's `4chan.org` social media website, setting the stage for the notorious hacktivist collective Anonymous [78].

While these hacktivists never displayed a predictable trajectory in their cyberoperations and political program [23], they narrowly utilized social media for self-promotion – announcing operations with an #Op prefixed hashtags [12] – and furthering relationships with other Internet activists. Anonymous cried foul on Twitter when WikiLeaks puts millions of its documents behind a pay wall [42], but also launched operation #Ferguson which doxed the St. Louis County police chief daughter's information in response to the shooting of the black teenager Michael Brown [10]. Hacktivists, in solidarity to the Arab spring uprisings, sent a care package composed of security tools and tactical advice though downplayed the touted "Twitter Revolution" [23].

True to their credo for utilizing Internet technologies against oppression, including social media, hacktivists launched the #OpKKK in support of #BlackLivesMatter protesters in Ferguson, Missouri to "unhood approximately 1000 Ku Klux Klan members" by gaining unauthorized access to a KKK Twitter account [134]. After a several years hiatus, perhaps due to arrests of some of the leading Anonymous hacktivists, the group resurfaced during the 2020 #BlackLivesMatter protests in response to the killing of George Floyd [56]. This time, in addition to leaking a 269 gigabyte trove of confidential police data (dubbed *BlueLeaks* [66]), the hacktivists launched social bot operations to amplify the online support for #BLM and criticize police actions.

Anonymous-affiliated hacktivists also utilized Internet technologies in the context of cyberwarfare. For example, the #OpIsis operation, which collated and published lists of tens of thousands of Twitter accounts that purportedly belonged to members of ISIS or its sympathizers, was launched in response to the terrorist attacks in France in 2015 [80]. Here, in addition to the identification efforts, hacktivists also waged a meme war and called for a "Troll ISIS Day" to provoke and disrupt ISIS-supported social media [79]. In early 2022 the Anonymous group took to Twitter to declare a "cyber war" to Russia in response to the Ukrainian invasion, launching DoS attacks against Russian's Federal Security Service's website and hacking Russian streaming services to broadcast war videos from Ukraine [108].

3 Internet Activism and Misinformation

3.1 Grassroots Misinformation Operations

Hacktivists, perhaps inadvertently, authored or gave popularity to the most utilized primitives for creating, propagating, amplifying, and disseminating misinformation - *trolling* and *memes*. This negative externality is unfortunate as trolling and memes were initially used by Anonymous against what they perceived a "misinformation campaign" by the Church of Scientology [78]. The "anon" members on `4chan.org` practically *hijacked* the term "troll" – initially meaning provoking others for mutual enjoyment – to abusing others for members' own enjoyment by posting upsetting or shocking content (usually on the /b/ channel of `4chan.org` [23]), harassing users (e.g. mocking funeral websites [14]), and spreading rumors [64]. What Anonymous did for the "lulz" (a brand of enjoyment etymologically derived from laughing-out-loud (lol)), nonetheless, showed the ease with which one could exploit the Internet technologies to be impolite, aggressive, disruptive, and manipulative to users' emotional states [23].

Trolling initially came in textual format as comments to posts, bulletin boards, and websites "deindividualized" people's lived experience for the "lulz" [14]. Gradually, hacktivists popularized a multimedia format of trolling or "memes," where textual commentary is superimposed over well-known imagery, typically representing different forms of power, such as political leaders, the police, and celebrities [79]. Memes, perhaps, were the actual rite of passage to true hacktivism – moving away from the early LOLCats – as they seek to deconstruct the power represented, contest censorship, and provide political commentary [91]. Memes as content were put to hacktivist use *en masse* in operations like "Troll ISIS day," where Anonymous proliferated memes with rubber-duck heads or rainbow stripes to ridicule ISIS propaganda imagery and disinformation narratives on Twitter [79]. Spread together with satirizing hashtags (e.g. #Daeshbags), the trolling memes achieved a cultural virality that brought hacktivism into the mainstream discourse online [96]. What the hacktivists did with the memes nonetheless, showed the ease with which anyone could disrupt, challenge, reimagine, and appropriate new political contexts by harnessing the virality and visibility of content spread on social media [88].

3.2 Mainstream Misinformation Operations

The hacktivists' playbook of trolling and meme dissent, though initially targeted *against* misinformation, was skillfully appropriated for the purposes of crafting and disseminating misinformation from 2014 onward, coinciding with the period of hacktivist inactivity [12]. This playbook alone was at first insufficient to achieve widespread political disruption, as it necessitated a support network of many accounts to gain traction. But the "appropriators" – privy to prior campaigns of disinformation and with the support of nation-state governments [117] – did not need to look further than the "sock puppet" accounts which were already utilized for spreading political falsehoods (e.g., Martha Coackey's "twitter bomb" disinformation campaign [89]). Having all the ingredients necessary to exploit the virality of social media and users' familiarity with emotionally-charged discourse, the "appropriators" established *troll farms* in the lead up to the UK's Brexit campaign and the 2016 US elections [75, 141].

The "army" behind the troll farms were particularly clever to integrate their social bots with "sock puppet" accounts that imitated ordinary users to systematically micro-target different audiences, foster antagonism, and undermine trust in information intermediaries [7]. Playing both sides in the emotionally-charged discourse already unfolding on social media, the troll farms posed as authentic, culturally competent personas (e.g. the so-called "Jenna Abrams" account [136]), and as vocal supporters of hashtag activism (counter) movements (e.g. BlackToLive in #BlackLivesMatter and SouthLoneStar in #BlueLivesMatter [124]). They also appropriated hashtag hijacking (e.g., #elections2016 and #ImVotingBecause tagging of quotes about Donald Trump and against Hillary Clinton [4]), hashtag co-opting (e.g. #BlackGunsMatter and #syrianlivesmatter [31]), and counter hashtagging (e.g. #NoDAPL against the Dakota Access Pipeline [47]). The troll farms even had the audacity to impersonate Anonymous themselves (e.g. the @_anonymous_news impersonation of the "Your Anonymous News" twitter account [22]).

The "meme game" of the troll farms was equally sophisticated and added to the initial success of their operations [86]. Trolls tested the waters around war-related memes regarding the opposition/support of the conflict in Syria [31], capitalized on both meme trolling and Internet activism to spread political memes through their fabricated Blacktivist social media accounts and co-opted Wikileaks in exploiting the leak of sensitive documents from the Democratic National Committee (DNC) [73]. Memes were also used to amplify conspiracies (e.g. QAnon, Pizzagate, and the murder of Seth Rich [138]), Texas secessionism (e.g. if Brexit why not #Texit [52]), and direct attacks (e.g. crooked Hillary [48]).

While the initial campaigns of the troll farms have been tracked, exposed, and brought into attention [31, 48], social media discourse has not recovered from this watershed period of meme and trolling appropriation for the purposes of conducting large-scale information operations [115]. Worse, the troll farm brand of political dissent was adopted by populist accounts keen on disseminating misinformation beyond just politics [53]. The trolling pandemonium spilled out of control with the COVID-19 pandemic as rumors, conspiracy theories, fake news, and out-of-context spins plagued the social media by hijacking the dominant hashtags like #COVID19, #coronavirus or #DoctorsSpeakUp [15], co-opting hashtags like #plandemic [62] and counter hash tagging with hashtags like #COVIDIOT [114]. Memes were distributed in conjunction with deepfake videos on platforms like YouTube [100] and TikTok [9] as well as blatant fake news on alt-platforms like Gab [21] to effectively reach a self-perpetuating bedlam of misinformation Internet counter-activism.

4 Hacktivism and Misinformation

In a radical state of ravaging misinformation campaigns on social media with no end in sight, one could wonder what the original activists on the Internet have to say in response. The unravelling of falsehoods is clearly a serious threat to the democratic vision of the Internet [101], as misinformation facilitated the rise of non-democratic communities contesting even factual knowledge and science (e.g. anti-vaxers, climate change deniers, etc. [133]). Hacktivists, as we have seen in Section 2, have fiercely opposed early misinformation campaigns in the past, but their means to do so were later "hijacked" for the purposes of mass misinformation production (i.e."disinformation" when spread with *intent* to deceive). One could attribute the paucity of hacktivists' involvement in the passing of the techno-liberal order of the Internet as the

rise of partisan-divided trust in facts and the politicization of science were already underway [38], but that alone is not a sufficient showstopper for action.

Regardless of any new Internet order, there is a reasonable expectation that one should still act upon the Levy's sacrosanct postulates [67], even if operating within an ecosystem polluted with misinformation. In addition to the public good arguments, misinformation is in conflict with the first (i) *all information should be free* postulate as it creates "information disorder" that, by the token of catalyzing polarization and emotionally-charged participation online, gives even more power to the neoliberal elites for perpetuating the economic and social (media) disarray [27]. Misinformation also conflicts with the second postulate (ii) *authority should be mistrusted, and decentralization promoted* as it stands in the way of independent truth discovery and dissemination online [69]. Should the new brand of reprehensible misinformation, therefore, be on the top of the hacktivists' agenda already?

4.1 Research Questions

To explore the gap in response to mass misinformation, we invited prominent hacktivists to address these questions:

- **RQ1:** How do contemporary hacktivists conceptualize the social media misinformation ecosystem?
- **RQ2:** What actions do hacktivists deem appropriate in response to misinformation on social media?
- **RQ3:** In what directions do the hacktivists see the misinformation ecosystem evolving toward in the future?

4.2 Sample

Our study was approved by the Institutional Review Board (IRB) of our institution before we invited, through personal contacts, and snowball sampling the hacktivists for a virtual interview session during 2022 and early 2023 with open-ended questions, listed in the Appendix. We sampled a population who were 18 years or older, from the United States, and that is an active contributor in the hacktivist community. As "active contributors" in the hacktivists space, our participants stated they are concerned with challenging online far-right extremism, help tracking criminals, and uncovering foreign countries' information operations. All of them were active in hacktivism prior to 2014 (and we conducted Open Source Intelligence (OSINT) investigations to verify that there is no evidence of them engaging in harmful or criminal activities in the past). The participants in our sample identified themselves using Levy's hacker ethos [67] and maintain presence on mainstream social media (Twitter, Facebook) and chat-based communities (Discord, Matrix). We used Zoom to conduct the interviews and allowed participants to choose whether to share a video feed or not. Every interview was recorded, stored in a secure server, and manually transcribed. We communicated with each interviewee to obtain final approval prior to starting the qualitative analysis.

Overall, our final sample contained 22 participants, all of which agreed to participate voluntarily. Gender demographics are given in Table 1. We made a deliberate attempt to produce a sample that is not a male-only or male-dominated, as previous studies indicate that the hacktivist community is imbalanced in regards gender [126]. Participants identities were not anonymous to us as researchers, but we deliberately suppress identifying statistics and potentially identifiable information in the reporting of our results to preserve participant anonymity to the general population, as a condition for their participation. In some cases, we used a direct censoring of names in citing participants' responses. We allowed the participants to skip any question they were uncomfortable answering. Each interview took around an hour to complete.

Table 1: Sample Demographic Distribution

Gender		
Female	**Male**	**Non-Binary**
8 (36.4%)	13 (59.1%)	1 (4.5%)

4.3 Methods and Instrumentation

To ensure validity to the task of conceptualizing misinformation, we introduced the participants to the generalized definition of social media misinformation from [135]. This also helped avoid confusion between past trolling and memes "for the lulz" and present information operations involving rumors, conspiracy theories, hoaxes, and clickbait. The hacktivists in our sample were invited to speak about their profiles, activity, and agendas online, before we asked their take on misinformation on social media. The qualitative responses were coded and categorized in respect to the following seven themes: a) antecedents to misinformation; b) mental models of misinformation; c) countering misinformation through leaking, doxing, and deplatforming; d) anti-misinformation operations (referred to as "ops") ; e) counter-misinformation tactics; f) misinformation literacy; and g) misinformation hacktivism.

Two independent researchers analyzed the interview transcriptions, achieving a strong level of inter-coder agreement (Cohen's $\kappa = .82$). We utilized a thematic analysis methodology to identify the aforementioned themes that naturally emerged from the responses in our sample. These themes were summarized to describe the conceptualization of, response to, and evolution of misinformation in the view of the contemporary hacktivists we sampled. In reporting the results, we prioritized verbatim quotation of participants' answers (emphasized and quoted in "*italics*") but omit reference to participant number in the sample to preserve their anonymity.

4.4 Hacktivists' Profiles

The hacktivists in our sample, true to the original ethos, represent the voice for advocacy and contemporary policy discussion. While they did not disclose their current operations, several of them hinted they are involved in tracking the rise of the far-right extremism, cybercriminals, as well as the information warfare part of the Ukraine invasion. A few of the hacktivists reported an agenda which comprised of leaking documents from companies and nation-state agencies as manifestation of their information freedom advocacy. Few of the hacktivists explicitly mentioned they still create and disseminate memes and participate in the "old school" trolling. And several of the hacktivists did actual *hacking* as in analyzing security problems (e.g. ransomware) and providing free tools for helping ordinary Internet users fend off related threats.

The majority of the hacktivists noted they have been active for a long time, being brought into the world of computers in childhood or early adolescence. Some of them resorted to hacktivism as a way to protect themselves against online bullies and some of them in response to state-sponsored offensive operations online, notably campaigns attributed to China and Russia. Several of them started with hacking operating systems to enable unrestricted access to games and/or bypass parental controls. While most of the participants in our sample cited curiosity as their driver to enter the "hacktivist conglomerate" and keep on hacking, there were many who voiced a strong support for cybersecurity education activism.

5 Misinformation Conceptualization

Evidence shows that social media users use multiple models to conceptualize misinformation – not just the traditional model that narrowly focuses on the fallacious nature of the information [113]. Beyond just fake news, misinformation is equally conceptualized as form of *political (counter)argumentation* where facts do selectively appear in alternative narratives relative to political and ideological contexts, often taken *out-of-context* with speculative intentions. Misinformation is also seen as *external propaganda* that includes *manufactured* facts and factoids disseminated and amplified online with the intention to create division. Given the radical transformation of the trolling and memes over time, our first research question aimed to learn the hacktivists' take on these competing conceptualizations amongst ordinary social media users.

5.1 Antecedents to Misinformation

The participants in our sample agree that trolling and meme dissemination has been hijacked for nefarious purposes, pointing out that they are not surprised about the current misinformation proliferation on the Internet. One participant summarized this evolution through first account experience:

> *I remember using sock puppet accounts way back in the early 2000s running forum raids as a* [redacted]*, specifically to run/post misinformation on other forums online. It was mostly for laughs, but we were massive monsters in those days. The only real major difference is these days is that the sock puppets are automated and put in action for keeping people tribalistic and resistant to opposing views.*

The use of "*sock puppets for running forum raids in the old days of hacktivism*," unfortunately, was not a serious enough threat for social media companies to implement "*strict policies of who and how can participate in the public discourses early on*" and counter to their business model of "*monetizing every possible engagement on their platforms,*" in the view of our participants and true to their innate resistance against the neoliberal appropriation of Internet freedoms.

Mainstream social media companies were accused of being the direct enablers of the "information disorder" as their models of engagement pushed "*less educational content the more an issue was important and demanded action*." This disorder played in the hands of the neoliberal elites and media outlets run by "*billionaires detached from reality to gain further control over public spaces*" as one of the participants put it. In the view of our participants, misinformation "*has always been there*" and pointed to the combination of "*self-proclamation of expertise online, cultivating followers, and playing on confirmation bias*" as the recipe the very hacktivists showed it works well in seeding misinformation:

> "*For example, look at the* [redacted]*. This person said they were a founding member of Anonymous and lots of people believed them. The person has spoken at conferences about it and even got jobs because of it. Literally dig slightly into that and it's clear that no one in the Anonymous community can vouch for the person and there's no evidence of them being linked. So, people are just too lazy to check stuff out because this person is kinda selling a story that fits with what they think so it must be true*."

5.2 Mental Models of Misinformation

The predominant mental model of misinformation amongst the hacktivists in our sample was *political (counter)argumentation*, where misinformation is disseminated for the sake of furthering a political argument or agenda [113]. In the original version of trolling and meme sharing the misinformation was seen as an alternative expression of disagreement, revolt, or ridicule without any context, but contemporary trolling and memes enter into the political context as content ready-made for the expression of political attitudes [94]. Despite fact checking being widely

available (and even suggested to users when content is moderated on social media [112]), the political appropriation of misinformation thrives because "*people won't fact check things and perpetuate them as long as these things align with their political ideology*." The reason why most social media users "*fall for misinformation*," in the view of our participants, is "*plain ignorance and stubbornness to hear anything contrary to their own political opinions*." One participant offered the following genesis of the misinformation problem:

> "*I think that people have learned that spreading disinformation through social media, Twitter for example, it's one of the best ways to get a word out. Twitter readers won't fact check things, especially if it aligns along with a political ideology people are passionate about so this word gets effectively to them. They'll believe whatever you tell them, and I think this is because there's a serious lack of, at least in the US, critical thinking education in schools.*"

In the view of the majority participants,"*both sides of the political spectrum spread misinformation and it further enables polarization*." While they acknowledge that "*the misinformation on social media is often identified with right-wing opinions*," participants recognize that "*we overuse the terms misinformation and disinformation to describe anything that is not a leftist opinion or fact.*" They point to the misinformation "stickiness" where the repeated exposure to speculative and false statements make them appear truthful [68], becoming the main theme of social media discourse. For example, one participant pointed out the Hunter Biden laptop saga [46]:

> "*It's usually the outrageous political claims that attract a lot of attention and people want a proof of concept, right? For example, take the bold claims aligned with the political message behind the Biden's laptop. Maybe there was a laptop but it's been politically disinfoed [sic] to death, to the point that the laptop leaks are irrelevant and can't be trusted as an evidence. These politicized things require a deeper dive into the actual truth as bold claims require bold evidence, but that's often missing so disinformation naturally creeps in.*"

Misinformation as political counter(argumentation) conflicts with the *all information should be free* postulate, which in turn forces mainstream social media platforms to "*restrict the flow of information.*" Misinformation, in the view of one of the participants, should not be restricted because "*people are entitled to see both sides of a proverbial political coin so the platforms must allow them to do so, otherwise by only showing heads or tails people will speculate about what's on the other side and assume the worst.*" The restriction of information on platforms conflicts with the *mistrust of authority and promote decentralization* hacker postulate because it allows "*the elites to define what constitutes 'truth' alone*," according to one participant. It also forces "*people to become rather tribalistic and a priori suspicious of people with different views.*" The "political tribalism" on social media [3], in turn, makes it "*easier to demonize people with different opinions and political attitudes and avoid scrutinizing the like-minded ones,*" playing directly in the hands of the "misinformers."

As for the "misinformers", our participants identified the state-sponsored "appropriators" that hijacked the original hacktivist playbook to spread *external propaganda* on social media. That other countries promulgated disinformation was not a news to the hacktivists (e.g. "*Russia has always been really good at it*"), but instead what surprised them was the "*audacity and the sophistication*" in utilizing trolling and memes on such a massive scale [140]. Reflecting on this shift in online operations, one participant believes that "*disinfo operations and hacking our intellectual property is all these other countries are left with because they can't beat US militarily or economically*." Not necessarily neoliberal, but nonetheless authoritarian, the elites behind the external propaganda in equal degree conflict with the *mistrust of authority and promote decentralization* hacker postulate because they are behind a "*blatant effort to control the social media turf and the mass of population spending their time there*", per one of the participants. The external propaganda nature of disinformation also conflicts with the *all information should be free* hacker postulate in the view of the hackers in our sample because "*overshadows and complicated an access to other more factual or useful information*."

6 Active Countering of Misinformation

Literature on misinformation focuses on helping the social media users discern falsehoods with strategies for "prebunking" (i.e. forewarning and preemptive refutation of the falsehoods [70]) or "debunking" (i.e. providing users verifiable corrections of the falsehoods from credible sources to break the illusion of truth [33, 97]). Algorithmic moderation tools are also available to mainstream social media platforms (the alternative ones do not deem misinformation as a problem [111]). These tools leverage natural language processing, image analysis, or metadata to detect trolling and memes [52,53,128]. Platforms have the option for "soft" moderation (by either obscuring trolling and memes with warnings covers or attaching warning labels [112,131]) and "hard" moderation (removing or suspending misinformer accounts [65]). None of these solutions, however fends off troll farms and meme disseminators effectively. Our second research question, therefore, sought to query hacktivists for their thoughts on countering this development.

6.1 Leaking, Doxing, and Deplatforming

The suspension of user acounts by social media platforms for breach of their code of conduct is referred to as "deplatforming" [2]. In the context of hacktivism, it takes a broader meaning, as hacktivists do investigative work that entails leaking and doxing but also confrontation with the misinformers that, in their subjective view, contradict the vision of a democratic Internet. For example, hacktivists did a massive API scraping of the alt-platform Parler to leak data that tied users to the Capitol Riots and the QAnon conspiracy [98], which in turn resulted in a massive account deplatforming on Twitter [17]. These activities spurred operations to confront and expose the QAnon conspirators on social media (e.g. @QAnonAnonymous [24]), amongst which some of our hacktivists have a direct role in "*dismantling the Qanon infrastructure*."

This deplatforming targeted political misinformation campaigns where our hacktivists "*compiled and leaked dossiers on individuals spreading hateful propaganda and those who seek to sow the seeds of violence*" on social media. These operations were targeted both on "*individual spreaders, nation-states, even companies with murky records*." Several mentioned their direct operations for exposing disinformation relative to the "*Ukrainian conflict*," praising the work of the Ukrainian IT Army outfit for dispelling the myth that Ukraine is committing genocide against Russians in the Donbas region [25]. Hacktivists were dedicated to "*doxing companies and governmental agencies in response to the political meddling in the US internal affairs from places like Russia, Iran, and China*." Misinformation "*sanctioned by the governments*" was targeted by the hacktivists in attempts to deplatform prominent "*disinformation front agents on social media, like* [redacted]*, for example*."

Leaks and doxing were equally utilized for misinformation beyond political counter(argumentation) and external propaganda. One of the participants has dedicated considerable time on exposing cryptocurrency scammers on social media and elsewhere, deeming the feeling of it as "*better than sex*." Another pushed back against criminal misinformation by doxing "*bullies, liars, and fraudsters*" and one "*anti-cancel culture in case of minors*" hacktivist noted that they "*successfully deplatformed major participants in hate campaigns and stalking of minors*" on social media. Another focused on leaking personal details about predators on social media who spread misinformation to cover their sexual harassment and cyberstalking towards women:

> *I've called it sometimes when I notice it on Twitter or elsewhere. I've exposed threat actors after tracing their activity and positively identifying them. Unfortunately, this is somewhat of a Bushido violation amongst fellow hackers but I am not concerned with such things. Some of these clowns have it coming to them. There was one person who went by the handle* [redacted] *who had been sexually harassing women, cyberstalking them, creating several sock puppet accounts, and just generally being a real nuisance in the community. Well, I doxed that person's real name on Twitter but I didn't post their address. This person thought he had cleaned up his tracks online being an 'infosec' professional but he underestimated someone like myself with a technical OSINT background. I easily found their information in an old resume on the Way Back Time Machine internet archive and posted their name. Some fellow 'infosec' pros didn't appreciate that I did so but honestly, the person had it coming and I don't regret it. I was careful about what I shared so no physical harm came to the them.*

6.2 Anti-Misinformation Operations

The hacktivists in our sample engaged in misinformation saturation operations, true to the their commitment to fight misinformation with more information. One of the hacktivists stated that it is "*expected from the hacktivist community to combat misinformation in such a way*" and noted that "*it is the sole reason they maintain a Twitter account.*" Another one seconded this posture noting that "*it is frustrating to see misinformation from others and other creators but that is the main reason I continue to post on TikTok.*" In the words of one participant, "*there is more ideological aspect of it when I am fighting disinformation*," directly invoking the mission of the true hacktivists to become reflexively "*loud and determined*" to speak true information in response to the "*general assholery of misinformation on internet.*"

Partaking in operation #NAFO (North Atlantic Fellas Organization) dedicated to countering Russian propaganda and disinformation in Ukraine by weaponizing memes [107], our participants materialized a combination of saturation and doxing to "*curtail misinformers' ability to gain followers.*" They extended their work to counter "*extremists and fascists and their toxic conspiracy theories*" by disrupting their funding and deplatforming prominent followers, true to the spirit of the "Antifa" hacktivist counterculture [137]. In a similar vein, one of the hacktivists proclaimed that they "*greatly contributed in the #OpJane operation*." #OpJane is the latest operation launched by Anonymous against Texas for enacting the anti-abortion Bill 8 that allows "abortion bounty" for anyone who anyone who reports abortion in the state of Texas [41]. Interestingly, in the announcement of the operation, Anonymous calls for "fighting misinformation with enough plausible and difficult to disprove misinformation" to make any data these bounty hunters gather as useless [6].

7 Misinformation Evolution

As there is virtually no cost to disseminating misinformation [89], it is unlikely that the online discourse will rid itself

of the alternative narrative plague any time soon. Whether this gloomy prediction will eventually materialize [81], or whether new technologies will improve the public's ability to judge the quality and veracity of content [5], remains an open question. Because hacktivists are nonetheless stakeholders in resolving this issue, our third research question aimed to learn their thoughts about how online spaces will fare with trolling, memes, and falsehoods in the near future.

7.1 Counter-Misinformation Tactics

The hacktivists in our sample unanimously posit that "*it is hard for social media platforms to keep up with removing it, so people stepping in to help is going to be of critical importance*" for preserving a healthy discourse. The mobilization for "*justice and truth as a cause*" is important not just for curbing misinformation but also in "*reclaiming information back from the political hold.*" To help "*expose misinformation charlatans,*" hacktivists call for maintaining a code of conduct where "*no leak, doxing, or exposure action should cause anyone else harm (physical, reputation, mental).*" To begin with, one participants reckons we should do the following:

> "*If one is a disinformation actor and they're acting aggressively I feel like you have to respond in a similar measure, in this case. I identify what their weaknesses are, what is it that's going to trigger them? Trying to get their accounts to get shut down, trying to get them to react in a way that will expose them. That's something I think is fair, as long as you're doing it [via] legal means. Getting open-source information about the individuals, exposing them, I think that's totally fair. Disinformation actors always try to be anonymous, of course, but what is the intent of that? Being anonymous allows you to act with impunity to do these really nasty things? Whereas all of a sudden if the tables are reversed and a disinformation actor is exposed, now I feel like we're teaching them a lesson. I guess it is sort of vigilantism, but in certain cases it's warranted. And one thing we got to do is got stop treating disinformation as freedom of speech. It's one thing to think you can say what you want, but that shouldn't shelter you from the consequences.*"

As misinformers usually use the anonymous cloak to legitimize their aggressive actions on social media, the next step is to "*identify what their weakness are and what triggers them - deplatforming or provocation?*" If the misinformers are unresponsive spreaders, then "*exposing, doxing, and putting their real faces through OSINT*" is seen as justified, not just on mainstream social media but also alt-platforms, forums and everywhere on Internet. If they itch for a provocation, then "*orchestrated saturation*" might work better with "*shitposts, absurd trolling, and ridiculing memes*" in the view of our participants. Here, it is vitally important to *a priori* distance from a "*political whataboutery*" and avoid "*coming across as censorship, disagreement, canceling that only could cause argument or dismissal.*"

Some of the hacktivists were on the opinion that "*doxing is not hacking anymore per se because you can get stuff with a credit card and documents could be easily faked nowadays.*" One possible tactic, proposed by one of them, was to "*find exploits, vulnerabilities in their platforms and step-by-step expose misinformers' amateurish way of doing trolling, using bots, and feeding think tanks to get a credibility behind their propaganda.*" Another tactic was "*doxing for the purpose of having advertisers pull from supporting known misinformer influencers, like for example in the case of* [redacted]." Proposing a hybrid style of hacktivist tactics, one participant suggested "*a latent, yet coordinated psychological warfare where psychologists rip apart these people, conduct serious OSINT to find incriminating leaks on them, and even pay for billboards and radio ads to publicly shame them.*" Along these lines, another participant even suggested leveraging all available tactics, "*targeting them with a social engineering attack and compromise a piece of their core infrastructure, be that their servers, Internet access, or bot credentials.*"

7.2 Misinformation Literacy

Hacktivists in our sample echo the sentiment regarding social media users' susceptibility to false information found in scientific literature: laziness to check facts [93], resistance to corrections [59], allegiance [125], and ignorance [19]. As people that resort to action, hacktivists feel the obligation to propose ways to address this susceptibility. In the view of one participant, "*misinformation needs to be seen as something everyone is being watched for, and not just one group of people on the left or the right.*" A "misinformation social contract" [142] necessitates interventions such as "*a critical thinking curricula in schools,*" "*teaching hacking operational security skills as social responsibility and rise to action,*" and "*forcing professional communication norms on platforms.*"

As our participants have little direct control over these interventions, they frequently proposed the development of "*truth-spreading bots for a 'standoff' with misinformation-spreading bots*" as something that could complement the practice of leaks, doxing, and exposure. They recognized that these "truth-spreading bots" must help ordinary users to better find facts, as information literacy is the single most effective tool in dispelling falsehoods [57]. Hacktivists reiterate that platforms do have to let "*misinformation to float on social media and make bots visible, so they gets overwhelmed with factual information*" in order to demonstrate to ordinary users how to do it themselves.

Regardless of whether these stances are realistic or not, the participants in our sample believed that the current approach to improving misinformation literacy is ineffective because

it does not signal an "*unbiased attitude*" to the social media users in the wrong. Instead of an educational and respectable tone, "*rather a 'cancel culture' infused or a 'your opinion is wrong' tone*" plagues any attempt to help people to navigate and locate factual information. Rejection of misinformation, as a result of misinformation literacy, must come as an agreement that "*scientific facts do not have political properties, even if the social media platforms inherently do.*"

7.3 Misinformation Hacktivism

The participants in our sample acknowledge that orchestrated *misinformation hacktivism*, barring individual instances of operations against misinformers, is largely absent from social media. For the hacktivists to assume misinformation as a worthy cause for action, the conflict between the past "hacking for political causes" and [60] future "hacking against using falsehoods in furthering political causes" [24] must be resolved. Though this conflict is complex and evolving, several of the participants worried that it could nevertheless create a "*division between the hacktivists on political lines.*"

As a relative threat to the misinformation activism, one participant mentioned the hijacking of the hacktivists image for self-promotion, e.g. "*some like to portrait themselves as woke gods of the web with zero fuck-ups*." Another threat is the temptation of using misinformation against misinformation, as in the #OpJane campaign. While this strategy is true to the "fight-fire-with-fire" approach, it might backfire in circumstances where abiding to the hacktivist ethic comes secondary to expressing social and political angst on social media [83]. On top of this, one could argue that this conflict *per se* might be hard to resolve in the case of external propaganda, because even if the hacktivists are "hacking for the homeland," they are nonetheless doing it on political terms [28].

8 Discussion

8.1 Implications

The new brand of misinformation, our findings show, draws the attention of the hacktivists, who find the hijacking of discourse for political and propagandistic purposes reprehensible. The "fight-fire-with-fire" response – leaks, doxing, and deplatforming – though individually employed by some of the participants in our sample, has yet to be orchestrated and tested against serious disinformation outfits that, unfortunately, are still prominent on social media [50]. Early evidence from the Ukrainian IT Army's work against Russian wartime propaganda suggests that these new orchestration tactics bring a degree of success [25].

The hacktivists' resolve to go after the misinformers would certainly have implications for the content/user moderation on social media, user participation, and the future of Internet activism overall. Moderating users and content on social media was, and remains, the default response of mainstream platforms to political and public health misinformation [112]. On the other hand, alternative platforms like Gab, Gettr, and Parler, which are seen as breeding grounds for the misinformation [139], have not and currently do not employ the same content and user moderation [111]. While content/user moderation incites a migration from mainstream platforms to the alt-platforms [139], it remains to be seen whether deplatforming will have the same effect. Mainstream social media has had a mixed response to leaks and doxing in the past (e.g. allowing WikiLeaks [118] but barring the Hunter Biden laptop leaks [30]), which adds uncertainty to if and how the hacktivists' "fight-fire-with-fire" approach will be allowed, moderated, or perhaps even forced to migrate entirely outside of the social media space.

Trolling and memes might still maintain popularity amongst misinformers, but, the latest modes of social media participation like short videos on TikTok open new "fronts" for both the misinformers and the hacktivists. TikTok has increasingly been tested as the next "battlefield" of alternative narratives with evidence of health and abortion misinformation [9, 116] and individual engagement by at least one of participants. Recalling that hacktivists' #OpJane was waged in response to the abortion ban laws in Texas and called for "misinformation-against-(mis)information" [41], it is yet to be seen how leaks, doxing, and deplatforming will interact with meme-ified videos and trolling. The company behind TikTok claims it does moderate health and abortion misinformation [129], but evidence shows that this is lax and largely ineffective [16], adding an additional incentive for the adoption of this platform by disinformation campaigns.

TikTok is also poised to become the next platform for Internet activism where the hashtag activism. Here the hashtag activism is combined with videos expanding the developing news narratives, such as the coverage of the Black Lives Matter movement and the Capitol riot [74]. TikTok presents content not just from viral hashtags but also their variations (e.g. #abotion but also #abôrtion [116]) so the threat of hashtag hijacking, co-opting, and counter hash tagging will inevitably materialize here too. This particular affordance will likely facilitate the further weaponization of deepfakes in the near future, as they have already appeared on TikTok in misinformation videos about the COVID-19 pandemic [110]. All of these developments would certainly necessitate a dynamic adaptation in the way doxing, leaking, and deplatforming are performed in order to not just avoid the disintegration of Internet activism and hacktivism, but prevent another paucity in action like the one which brought state-sponsored misinformation *en masse* on social media in the first place [44].

8.2 Ethical Considerations

The purpose of our study was not to generalize to a population, but rather to explore the contemporary hacktivists' relation-

ship with misinformation in depth. To protect individual's privacy, and to avoid speculations, we omitted the names and some of the procedures mentioned during the interviews. We are careful with our study not to infringe upon the hacktivist's sensibilities nor to cause any retaliation with our findings. Though our analysis and interpretation of the findings is positioned on impartiality, the overall study suggests the need for a stronger ethical contextualization of the techniques expressed by the hacktivists, the harms done by this community in the past, and the risk of future harms (intentional or accidental) to individuals and their close ones.

Certain hacktivists have used the techniques discussed in this study to harass, dox and cause harm to intended targets and innocent bystanders alike. For example, the GamerGate scandal emerged as a result of doxing and harassment targeting female gamers – including rape and death threats on a daily basis – by hacktivists that perceived feminism as a threat to traditional values of video games [1]. Similarly, proponents of the #metoo movement have been targets of abusive comments, doxing, and trolling [82]. Even academics and journalists have experienced targeted harassment regardless of the impartiality in their inquiry and "good faith" reporting on the activities associated with the hacktivist communities [32].

These instances clearly contradict Levy's postulate that *computers can change your life for the better* [67] and also undermine the common hacktivist value of defending human rights from any oppression, especially in a sociopolitical, right-of-center context [76]. The ethical justification behind traditional hacktivism as civil disobedience is predicated, and still is, on the premise that "no damage is done to persons" during any hacktivist operation or action [127]. The above actions clearly violate this principle and, as such, lose credibility from a civil disobedience perspective (it is worth pointing out that participants in our study explicitly called for a code of conduct where "*no leak, doxing, or exposure action should cause anyone else harm (physical, reputation, mental)*" towards addressing this issue).

Ethics violations by a hacktivist calls into question the hacktivist's credibility to speak on misinformation, and entails a degree of wariness for future consideration of the proposed anti-misinformation approach. The fact that all the participants in our study, to the best of our knowledge, had no involvement in the aforementioned violations or other unethical activities, cannot alone verify the credibility of the results. Equally, the claims that several participants provided in taking actions *against* sexual harassment and cyberstalking towards women cannot be seen as a vote of confidence that applies to everyone in the sample or the hacktivist community. Our participants or other hacktivists do not always get things right and can be assumed to have the expertise to cover their tracks online [121]. Therefore, the findings reported in our paper do not grant any exemption nor condone engaging in morally dubious or illegal acts.

Our results alone serve neither as an approval nor a call for any hacktivist action. We make no claims to be able to identify and agree with hacktivists on all the "bad actors" in the misinformation space, nor even on a reliable definition of what content constitutes "misinformation" itself [13]. We maintain the caveat that any action – misinformation hacktivism or otherwise – must be morally justified separately. We do, however, identify with the many of the ideas and suggestions posed by the hacktivists in our study – particularly when in conformity with the guidelines put forth in Levy's hacker ethos [67] and when they are in support of a democratic vision of the Internet [45]. We do also support the idea of "fighting fire with fire" identified in our findings inasmuch as it seeks to address the power mismatch that has arisen in the context of regulation, lax policy, and perverse incentive structures on social media platforms – factors which have contributed to the current state of affairs where misinformation is a prominent component of everyday discourse [109]. Again, this is not to legitimize hacktivist actions across the board nor to herald them as the sole defenders but simply to highlight the pressing need for examples of organized resistance against misinformation and well-resourced troll farms online.

8.3 Limitations

Our research was limited in its scope to US-based hacktivists, so we exercise caution not to generalize the results across the entire Internet activist community. Hacktivist operations are often at the center of debates regarding the dimensions of civil disobedience, political participation, legality, and the ethical use of Internet technologies [109]. Our results seek neither to approve or disapprove of these operations, but rather share in-depth accounts of the perspective of this unique and engaged Internet minority. Even with such a relatively small sample, it is clear that hacktivism encompasses a wide variety of perspectives. That being said, we acknowledge the limitations of our sampling and that the individuals in our sample present a limited subset of views and experiences.

We are aware that our results only reflect the current, limited understanding of misinformation informed by the forms currently prominent on social media. Therefore, we are careful to avoid any predictive use of our results in future misinformation campaigns. We also do not know if, when or how the activists in our sample have used the proposed counter-misinformation tools, tactics, and procedures. Our results do not do not offer a blanket justification for the frivolous use of them across any online space. We note that this study reported on the evolving experience of dealing with misinformation by hacktivists and might miss some important aspects of meting out truth on social media. We advise caution in this, as we see our work as a synergistic line of scientific inquiry that addresses an important gap in voicing the opinions of those that actually introduced the means for mass producing of misinformation online in the first place.

8.4 Future Work

Our future research will continue to to trace out the ways in which the hacktivist community engages with misinformation. We plan to expand our work beyond the US and working with hacktivists worldwide, as misinformation influences geopolitical affairs across the globe. We are set to further explore the intersection of and interactions between hashtag activism and hacktivism targeting online misinformation, as synergy between the two have emerged, such as in the case of the #NAFO campaign on Twitter. Here, we would devote much attention to the new misinformation "battlefield" of short-form video platforms such as TikTok. It would be useful to study the emergent circumstances in which misinformation hacktivism mobilizes and empowers ordinary users to join future "Troll [target] Day" operations and and to catalog their experiences with such participation. Of equal importance would be to further study the use of "misinformation-against-(mis)information" as in the case of #OpJane to learn both the useful and harmful aspects of this approach.

9 Conclusion

Reflecting the communitarian ideals of free information and disobedience to authority, the hacktivists in our study showed a determination for a radical response against the reprehensible act of spreading falsehoods on social media. As misinformation is consequential to the trolling and memes of the early days of hacktivism, it is reassuring to observe that the contemporary hacktivists are outwardly against such a nefarious appropriation of their aesthetics. It is encouraging to reveal that hacktivists also advocate for general misinformation literacy as a strategic asset against an undemocratic Internet. These findings, we hope, will empower ordinary users in counteracting misinformation towards the vision of a democratic Internet.

References

[1] Sarah A. Aghazadeh, Alison Burns, Jun Chu, Hazel Feigenblatt, Elizabeth Laribee, Lucy Maynard, Amy L. M. Meyers, Jessica L. O'Brien, and Leah Rufus. *GamerGate: A Case Study in Online Harassment*, pages 179–207. Springer International Publishing, Cham, 2018.

[2] Shiza Ali, Mohammad Hammas Saeed, Esraa Aldreabi, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. Understanding the effect of deplatforming on social networks. In *13th ACM Web Science Conference 2021*, WebSci '21, page 187–195, New York, NY, USA, 2021. Association for Computing Machinery.

[3] Jennifer Allen, Cameron Martel, and David G Rand. Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in twitter's birdwatch crowdsourced fact-checking program. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[4] Omar Alonso, Vasileios Kandylas, Serge-Eric Tremblay, Jake M Hofman, and Siddhartha Sen. What's happening and what happened: Searching the social web. In *Proceedings of the 2017 ACM on Web Science Conference*, pages 191–200, 2017.

[5] Janna Anderson and Lee Rainie. The future of truth and misinformation online. 2017. https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/.

[6] Anonymous. Operation jane initiated. we're totally going to mess with texas. #anonymous, 2021. https://twitter.com/YourAnonNews/status/1433926829396668429.

[7] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. Acting the part: Examining information operations within# blacklivesmatter discourse. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–27, 2018.

[8] JP Armstrong. Twitter as a channel for frame diffusion? hashtag activism and the virality of# heterosexualprideday. *Rise of the Far Right: Technologies of Recruitment and Mobilization*, page 87, 2021.

[9] Corey H. Basch, Zoe Meleo-Erwin, Joseph Fera, Christie Jaime, and Charles E. Basch. A global pandemic in the time of viral memes: Covid-19 vaccine misinformation and disinformation on tiktok. *Human Vaccines & Immunotherapeutics*, 17(8):2373–2377, 2021.

[10] Ross W. Bellaby. An ethical framework for hacking operations. *Ethical Theory and Moral Practice*, 24(1):231–255, 2021.

[11] Yochai Benkler, Robert Faris, and Hal Roberts. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018.

[12] Davide Beraldo. Unfolding #anonymous on twitter: The networks behind the mask. *First Monday*, 27(1), 2023/01/20 2022. https://firstmonday.org/ojs/index.php/fm/article/view/11723.

[13] Sven Bernecker, Amy K Flowerree, and Thomas Grundmann. *The epistemology of fake news*. Oxford University Press, 2021.

[14] Jonathan Bishop. Trolling for the lulz?: using media theory to understand transgressive humor and other internet trolling in online communities. In *Transforming politics and policy in the digital age*, pages 155–172. IGI Global, 2014.

[15] Amanda S. Bradshaw. #doctorsspeakup: Exploration of hashtag hijacking by anti-vaccine advocates and the influence of scientific counterpublics on twitter. *Health Communication*, 0(0):1–11, 2022.

[16] Jack Brewster, Lorenzo Arvanitis, Valerie Pavilonis, and Macrina Wang. Beware the 'new google:' tiktok's search engine pumps toxic misinformation to its young users, 2022. https://www.newsguardtech.com/misinformation-monitor/september-2022/.

[17] David Bromell. *Deplatforming and Democratic Legitimacy*, pages 81–109. Springer International Publishing, Champaigne, 2022.

[18] Victoria Carty. *Social Movements and New Technology*. Taylor and Francis, 2018.

[19] Jin-Hee Cho, Scott Rager, John O'Donovan, Sibel Adali, and Benjamin D. Horne. Uncertainty-based false information propagation in social networks. *Trans. Soc. Comput.*, 2(2), jun 2019.

[20] Miyoung Chong. Discovering fake news embedded in the opposing hashtag activism networks on twitter: #gunreformnow vs. #nra. *Open Information Science*, 3(1):137–153, 2019.

[21] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The covid-19 social media infodemic. *Scientific reports*, 10(1):1–10, 2020.

[22] E Gabriella Coleman. Logics and legacy of anonymous. *Second International Handbook of Internet Research*, pages 145–166, 2020.

[23] Gabriella Coleman. *Hacker, hoaxer, whistleblower, spy: The many faces of Anonymous*. Verso books, 2014.

[24] Christopher T Conner and Nicholas MacMurray. The perfect storm: A subcultural analysis of the q-anon movement. *Critical Sociology*, 48(6):1049–1071, 2022.

[25] Ellen Cornelius. Anonymous Hacktivism: Flying the Flag of Feminist Ethics for the Ukraine IT Army. 2022.

[26] John Corner. Fake news, post-truth and media–political change, 2017.

[27] Brian Creech. Fake news and the discursive construction of technology companies' social power. *Media, Culture & Society*, 42(6):952–968, 2020.

[28] Michael Dahan. Hacking for the homeland: Patriotic hackers versus hacktivists. In *Proceedings of the 8th International Conference on Information Warfare and Security (Iciw-2013)*, pages 51–57, 2013.

[29] Philipp Darius and Fabian Stephany. How the far-right polarises twitter: 'hashjacking' as a disinformation strategy in times of covid-19. In Rosa Maria Benito, Chantal Cherifi, Hocine Cherifi, Esteban Moro, Luis M. Rocha, and Marta Sales-Pardo, editors, *Complex Networks & Their Applications X*, pages 100–111, Cham, 2022. Springer International Publishing.

[30] Glenn Diesen. *Conclusion: Anti-Russian Propaganda of a West in Relative Decline*, pages 255–258. Springer Nature Singapore, Singapore, 2022.

[31] Renee DiResta, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. The tactics & tropes of the internet research agency. 2019.

[32] Periwinkle Doerfler, Andrea Forte, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, and Damon McCoy. "i'm a professor, which isn't usually a dangerous job": Internet-facilitated harassment and its impact on researchers. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–32, 2021.

[33] Ullrich K. H. Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga, and Michelle A. Amazeen. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29, 2022.

[34] Abbas Ehsanfar and Mo Mansouri. Incentivizing the dissemination of truth versus fake news in social networks. In *2017 12th System of Systems Engineering Conference (SoSE)*, pages 1–6, 2017.

[35] Luca Follis and Adam Fish. *3 When to Hack*, pages 73–111. 2020.

[36] Deen Freelon, Alice Marwick, and Daniel Kreiss. False equivalencies: Online activism from left to right. *Science*, 369(6508):1197–1201, 2020.

[37] Deen Freelon, Charlton D McIlwain, and Meredith Clark. Beyond the hashtags:# ferguson,# blacklivesmatter, and the online struggle for offline justice. *Center for Media & Social Impact, American University, Forthcoming*, 2016.

[38] Gordon Gauchat. Politicization of science in the public sphere: A study of public trust in the united states, 1974 to 2010. *American sociological review*, 77(2):167–187, 2012.

[39] Jordana J. George and Dorothy E. Leidner. From clicktivism to hacktivism: Understanding digital activism. *Information and Organization*, 29(3):100249, 2019.

[40] Paolo Gerbaudo. From cyber-autonomism to cyber-populism: An ideological history of digital activism. *tripleC: Communication, Capitalism & Critique*, 15(2):477–489, May 2017.

[41] Claire Goforth. 'anonymous' hackers have a message for texas abortion 'snitch' sites: We're coming for you, 2021. https://www.dailydot.com/debug/anonymous-hactivists-texas-abortion-ban-operation-jane/.

[42] Atilla Hallsby. Psychoanalysis against wikileaks: resisting the demand for transparency. *Review of Communication*, 20(1):69–86, 2020.

[43] Max Halupka. Clicktivism: A systematic heuristic. *Policy & Internet*, 6(2):115–132, 2014.

[44] Jason Hannan. Trolling ourselves to death? social media and post-truth politics. *European Journal of Communication*, 33(2):214–226, 2018.

[45] Masayuki Hatta. Cowboys and the eternal september transfiguration of hacker aesthetics. *Annals of Business Administrative Science*, page 0210923a, 2021.

[46] Nolan Higdon, Emil Marmol, and Mickey Huff. Returning to neoliberal normalcy: Analysis of legacy news media's coverage of the biden presidency's first hundred days. In *The Future of the Presidency, Journalism, and Democracy*, pages 255–273. Routledge, 2022.

[47] Matthew Hindman and Vlad Barash. Disinformation, 'fake news' and influence campaigns on twitter. 2018.

[48] Philip N Howard, Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. The IRA, social media and political polarization in the United States, 2012-2018. 2019.

[49] Laura Illia. Passage to cyberactivism: how dynamics of activism change. *Journal of public affairs.*, 3(4), 2003-11.

[50] Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. Still out there: Modeling and identifying russian troll accounts on twitter. In *12th ACM Conference on Web Science*, WebSci '20, page 1–10, New York, NY, USA, 2020. Association for Computing Machinery.

[51] Leanna Ireland. We are all (not) anonymous: Individual- and country-level correlates of support for and opposition to hacktivism. *New Media & Society*, 0(0):14614448221122252, 0.

[52] Peter Jachim, Filipo Sharevski, and Emma Pieroni. Trollhunter2020: Real-time detection of trolling narratives on twitter during the 2020 us elections. In *Proceedings of the 2021 ACM workshop on security and privacy analytics*, pages 55–65, 2021.

[53] Peter Jachim, Filipo Sharevski, and Paige Treebridge. Trollhunter [evader]: Automated detection [evasion] of twitter trolls during the covid-19 pandemic. In *New Security Paradigms Workshop 2020*, NSPW '20, page 59–75, New York, NY, USA, 2021. Association for Computing Machinery.

[54] S.J. Jackson, M. Bailey, B.F. Welles, and G. Lauren. *#HashtagActivism: Networks of Race and Gender Justice*. MIT Press, 2020.

[55] Keenan Jones, Jason R. C. Nurse, and Shujun Li. Behind the mask: A computational study of anonymous' presence on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):327–338, May 2020.

[56] Keenan Jones, Jason R.C. Nurse, and Shujun Li. Out of the shadows: Analyzing anonymous' twitter resurgence during the 2020 black lives matter protests. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):417–428, May 2022.

[57] S Mo Jones-Jang, Tara Mortensen, and Jingjing Liu. Does media literacy help identification of fake news? information literacy helps, but other literacies don't. *American behavioral scientist*, 65(2):371–388, 2021.

[58] Andreas Jungherr, Gonzalo Rivero, and Daniel Gayo-Avello. *Retooling Politics: How Digital Media Are Shaping Democracy*. Cambridge University Press, 2020.

[59] Ben Kaiser, Jerry Wei, Eli Lucherini, Kevin Lee, J Nathan Matias, and Jonathan R Mayer. Adapting security warnings to counter online disinformation. In *USENIX Security Symposium*, pages 1163–1180, 2021.

[60] Vasileios Karagiannopoulos. *A Short History of Hacktivism: Its Past and Present and What Can We Learn from It*, pages 63–86. Springer International Publishing, Cham, 2021.

[61] Athina Karatzogianni. *Firebrand waves of digital activism 1994-2014: The rise and spread of hacktivism and cyberconflict*. Springer, 2015.

[62] Matthew D Kearney, Shawn C Chiang, and Philip M Massey. The twitter origins and evolution of the covid-19 "plandemic" conspiracy theory. *Harvard Kennedy School Misinformation Review*, 1(3), 2020.

[63] Vance D. Keyes and Latocia Keyes. Dynamics of an american countermovement: Blue lives matter. *Sociology Compass*, 16(9):e13024, 2022.

[64] Allison Klempka and Arielle Stimson. Anonymous communication on the internet and trolling. *Concordia Journal of Communication Research*, 1(1):2, 2014.

[65] Deepak Kumar, Patrick Gage Kelley, Sunny Consolvo, Joshua Mason, Elie Bursztein, Zakir Durumeric, Kurt Thomas, and Michael Bailey. Designing toxic content classification for a diversity of perspectives. In *SOUPS@ USENIX Security Symposium*, pages 299–318, 2021.

[66] Micah Lee. *Hack of 251 Law Enforcement Webiste Exposes Personal Data of 700,000 Cops*. 2020. https://theintercept.com/2020/07/15/blueleaks-anonymous-ddos-law-enforcement-hack/.

[67] Steven Levy. *Hackers: Heroes of the Computer Revolution - 25th Anniversary Edition*. O'Reilly Media, Inc., 1st edition, 2010.

[68] Stephan Lewandowsky, John Cook, Ullrich Ecker, Dolores Albarracin, Michelle Amazeen, Panayiota Kendou, Doug Lombardi, E Newman, Gordon Pennycook, Ethan Porter, et al. *The Debunking Handbook 2020*. 2020.

[69] Stephan Lewandowsky, Ullrich K.H. Ecker, and John Cook. Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4):353–369, 2017.

[70] Stephan Lewandowsky and Sander van der Linden. Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2):348–384, 2021.

[71] Simon Lindgren. Movement mobilization in the age of hashtag activism: Examining the challenge of noise, hate, and disengagement in the #metoo campaign. *Policy & Internet*, 11(4):418–438, 2019.

[72] Alexander J Lindvall. Political hacktivism: doxing & the first amendment. *Creighton L. Rev.*, 53:1, 2019.

[73] Darren L. Linvill and Patrick L. Warren. Troll factories: Manufacturing specialized disinformation on twitter. *Political Communication*, 37(4):447–467, 2020.

[74] Ioana Literat, Lillian Boxman-Shabtai, and Neta Kligler-Vilenchik. Protesting the protest paradigm: Tiktok as a space for media criticism. *The International Journal of Press/Politics*, 0(0):19401612221117481, 0.

[75] Clare Llewellyn, Laura Cram, Adrian Favero, and Robin L. Hill. Russian troll hunting in a brexit twitter archive. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, JCDL '18, page 361–362, New York, NY, USA, 2018. Association for Computing Machinery.

[76] Mark Manion and Abby Goodrum. Terrorism or civil disobedience: toward a hacktivist ethic. *Acm Sigcas Computers and Society*, 30(2):14–19, 2000.

[77] Martha McCaughey and Michael D Ayers. *Cyberactivism: Online activism in theory and practice*. Psychology Press, 2003.

[78] Ty McCormick. Anthropology of an idea hacktivism. *Foreign Policy*, (200):24–25, May/Jun 2013.

[79] Ally McCrow-Young and Mette Mortensen. Countering spectacles of fear: Anonymous' meme 'war' against isis. *European Journal of Cultural Studies*, 24(4):832–849, 2021.

[80] Virginia McGovern and Francis Fortin. The anonymous collective: Operations and gender differences. *Women & Criminal Justice*, 30(2):91–105, 2020.

[81] Lee McIntyre. *Post-truth*. MIT Press, 2018.

[82] Kaitlynn Mendes, Jessica Ringrose, and Jessalynn Keller. # metoo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies*, 25(2):236–246, 2018.

[83] Paul Mihailidis and Samantha Viotty. Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society. *American behavioral scientist*, 61(4):441–454, 2017.

[84] Stefania Milan. *Social movements and their technologies: Wiring social change*. Springer, 2013.

[85] Stefania Milan. Hacktivism as a radical media practice. In *The Routledge companion to alternative and community media*, pages 550–560. Routledge, 2015.

[86] Ryan M Milner. *The world made meme: Public conversations and participatory media*. MIT Press, 2018.

[87] Rachel E. Moran and Stephen Prochaska. Misinformation or activism?: analyzing networked moral panic through an exploration of #savethechildren. *Information, Communication & Society*, 0(0):1–21, 2022.

[88] Mette Mortensen and Christina Neumayer. The playful politics of memes. *Information, Communication & Society*, 24(16):2367–2377, 2021.

[89] Eni Mustafaraj and Panagiotis Takis Metaxas. The fake news spreading plague: Was it preventable? In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, page 235–239, New York, NY, USA, 2017. Association for Computing Machinery.

[90] Mahdi M. Najafabadi and Robert J. Domanski. Hacktivism and distributed hashtag spoiling on twitter: Tales of the #irantalks. *First Monday*, 23(4), Apr. 2018. https://firstmonday.org/ojs/index.php/fm/article/view/8378.

[91] Asaf Nissenbaum and Limor Shifman. Internet memes as contested cultural capital: The case of 4chan's/b/board. *New media & society*, 19(4):483–501, 2017.

[92] Taylor Owen. *Disruptive power: The crisis of the state in the digital age*. Oxford Studies in Digital Politics, 2015.

[93] Gordon Pennycook and David G. Rand. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188:39–50, 2019. The Cognitive Science of Political Thought.

[94] Gordon Pennycook and David G. Rand. The psychology of fake news. *Trends in Cognitive Sciences*, 25(5):388–402, 2021.

[95] W. Phillips and R.M. Milner. *The Ambivalent Internet: Mischief, Oddity, and Antagonism Online*. Polity Press, 2017.

[96] Whitney Phillips. The house that fox built: Anonymous, spectacle, and cycles of amplification. *Television & New Media*, 14(6):494–509, 2013.

[97] Man pui Sally Chan, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11):1531–1546, 2017.

[98] David Redding, Jian Ang, and Suman Bhunia. A case study of massive api scrapping: Parler data breach after the capitol riot. In *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2022.

[99] Eugenia Ha Rim Rho and Melissa Mazmanian. Political hashtags & the lost art of democratic discourse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[100] Daniel R´ochert, Gautam Kishore Shahi, German Neubaum, Björn Ross, and Stefan Stieglitz. The networked context of covid-19 misinformation: Informational homogeneity on youtube at the beginning of the pandemic. *Online Social Networks and Media*, 26:100164, 2021.

[101] Mark Rolfe. *Hacker: Creating the Narrative of the Digital Robin Hood*, pages 135–164. Springer Singapore, 2016.

[102] Marco Romagna. *Hacktivism: Conceptualization, Techniques, and Historical View*, pages 743–769. Springer International Publishing, Cham, 2020.

[103] Dana Rotman, Sarah Vieweg, Sarita Yardi, Ed Chi, Jenny Preece, Ben Shneiderman, Peter Pirolli, and Tom Glaisyer. From slacktivism to activism: Participatory culture in the age of social media. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, page 819–822, New York, NY, USA, 2011. Association for Computing Machinery.

[104] Rodrigo Sandoval-Almazan and J. Ramon Gil-Garcia. Towards cyberactivism 2.0? understanding the use of social media and other information technologies for political activism and social movements. *Government Information Quarterly*, 31(3):365–378, 2014.

[105] Madelyn R Sanfilippo, Shengnan Yang, and Pnina Fichman. Managing online trolling: From deviant to social and political trolls. In *50th Annual Hawaii International Conference on System Sciences, HICSS 2017*, pages 1802–1811. IEEE Computer Society, 2017.

[106] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, page 813–822, New York, NY, USA, 2016. Association for Computing Machinery.

[107] Mark Scott. The shit-posting, twitter-trolling, dog-deploying social media army taking on putin one meme at a time, 2022.

[108] Dimitrios Serpanos and Theodoros Komninos. The cyberwarfare in ukraine. *Computer*, 55(7):88–91, 2022.

[109] Philip Serracino-Inglott. Is it ok to be an anonymous? *Ethics & Global Politics*, 6(4):22527, 2013.

[110] Lanyu Shang, Ziyi Kou, Yang Zhang, and Dong Wang. A multimodal misinformation detector for covid-19 short videos on tiktok. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 899–908, 2021.

[111] Filipo Sharevski, Amy Devine, Peter Jachim, and Emma Pieroni. "Gettr-ing" User Insights from the Social Network Gettr, 2022. https://truthandtrustonline.com/wp-content/uploads/2022/10/TTO_2022_proceedings.pdf.

[112] Filipo Sharevski, Amy Devine, Peter Jachim, and Emma Pieroni. Meaningful context, a red flag, or both? preferences for enhanced misinformation warnings among us twitter users. In *Proceedings of the 2022 European Symposium on Usable Security*, EuroUSEC '22, page 189–201, New York, NY, USA, 2022. Association for Computing Machinery. https://doi.org/10.1145/3549015.3555671.

[113] Filipo Sharevski, Amy Devine, Emma Pieroni, and Peter Jachim. Folk models of misinformation on social media. In *Network and distributed system security symposium*, 2023.

[114] Filipo Sharevski, Alice Huff, Peter Jachim, and Emma Pieroni. (mis)perceptions and engagement on twitter: Covid-19 vaccine rumors on efficacy and mass immunization effort. *International Journal of Information Management Data Insights*, 2(1):100059, 2022.

[115] Filipo Sharevski, Peter Jachim, Emma Pieroni, and Nate Jachim. Voxpop: An experimental social media platform for calibrated (mis)information discourse. In *New Security Paradigms Workshop*, NSPW '21, page 88–107, New York, NY, USA, 2021. Association for Computing Machinery.

[116] Filipo Sharevski, Jennifer Vander Loop, Peter Jachim, Amy Devine, and Emma Pieroni. Abortion misinformation on tiktok: Rampant content, lax moderation, and vivid user experiences. *arXiv preprint arXiv:2301.05128*, 2023.

[117] Filipo Sharevski, Paige Treebridge, Peter Jachim, Audrey Li, Adam Babin, and Jessica Westbrook. Socially engineering a polarizing discourse on facebook through malware-induced misperception. *International Journal of Human–Computer Interaction*, 38(17):1621–1637, 2022.

[118] Micah L Sifry. *WikiLeaks and the Age of Transparency*. OR Books, 2011.

[119] Ellen Simpson. Integrated & alone: The use of hashtags in twitter social activism. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18, page 237–240, New York, NY, USA, 2018. Association for Computing Machinery.

[120] Edward Snowden. *Permanent record*. Pan Macmillan, 2019.

[121] Tom Sorell. Human Rights and Hacktivism: The Cases of Wikileaks and Anonymous. *Journal of Human Rights Practice*, 7(3):391–410, 09 2015.

[122] Simon Springer, Kean Birch, and Julie MacLeavy. *The handbook of neoliberalism*, volume 12. Routledge New York, 2016.

[123] Kevin F Steinmetz. Hacking and hacktivism. *Shades of Deviance: A Primer on Crime, Deviance and Social Harm*, 19, 2022.

[124] Leo G Stewart, Ahmer Arif, and Kate Starbird. Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web*, volume 70, 2018.

[125] Briony Swire-Thompson, Ullrich KH Ecker, Stephan Lewandowsky, and Adam J Berinsky. They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political psychology*, 41(1):21–34, 2020.

[126] Leonie Maria Tanczer. Hacktivism and the male-only stereotype. *New Media & Society*, 18(8):1599–1615, 2016.

[127] Paul Taylor. Hacktivism: in search of lost ethics? In *Crime and the Internet*, pages 71–85. Routledge, 2003.

[128] William Theisen, Joel Brogan, Pamela Bilo Thomas, Daniel Moreira, Pascal Phoa, Tim Weninger, and Walter Scheirer. Automatic discovery of political meme genres with diverse appearances. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):714–726, May 2021.

[129] TikTok. Tiktok safety, 2022. https://www.tiktok.com/safety/en-us/topics/.

[130] Justus Uitermark. Complex contention: analyzing power dynamics within anonymous. *Social Movement Studies*, 16(4):403–417, 2017.

[131] Anthony Vance, David Eargle, Jeffrey L. Jenkins, C. Brock Kirwan, and Bonnie Brinton Anderson. The Fog of Warnings: How Non-essential Notifications Blur with Security Warnings. In *Fifteenth Symposium*

on Usable Privacy and Security (SOUPS 2019), Santa Clara, CA, August 2019. USENIX Association.

[132] Courtland VanDam and Pang-Ning Tan. Detecting hashtag hijacking from twitter. In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16, page 370–371, New York, NY, USA, 2016. Association for Computing Machinery.

[133] Silvio Waisbord. Truth is what happens to news. *Journalism Studies*, 19(13):1866–1878, 2018.

[134] Jared M Wright, Kaitlin Kelly-Thompson, S Laurel Weldon, Dan Goldwasser, Rachel L Einwohner, Valeria Sinclair-Chapman, and Fernando Tormos-Aponte. Drive-by solidarity: Conceptualizing the temporal relationship between# blacklivesmatter and anonymous's# opkkk. *Contention*, 10(2):25–55, 2022.

[135] Liang Wu, Fred Morstatter, Kathleen M. Carley, and Huan Liu. Misinformation in social media: Definition, manipulation, and detection. *SIGKDD Explor. Newsl.*, 21(2):80–90, nov 2019.

[136] Yiping Xia, Josephine Lukito, Yini Zhang, Chris Wells, Sang Jung Kim, and Chau Tong. Disinformation, performed: self-presentation of a russian ira account on twitter. *Information, Communication & Society*, 22(11):1646–1664, 2019.

[137] Weiai Wayne Xu. Mapping connective actions in the global alt-right and antifa counterpublics. *International Journal of Communication*, 14:22, 2020.

[138] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. What is gab: A bastion of free speech or an alt-right echo chamber. In *Companion Proceedings of the Web Conference 2018*, pages 1007–1014, 2018.

[139] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. On the origins of memes by means of fringe web communities. In *Proceedings of the Internet Measurement Conference 2018*, IMC '18, page 188–202, New York, NY, USA, 2018. Association for Computing Machinery.

[140] Savvas Zannettou, Tristan Caulfield, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. Characterizing the use of images in state-sponsored information warfare operations by russian trolls on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):774–785, May 2020.

[141] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. Who let the trolls out? towards understanding state-sponsored trolls. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, pages 353–362, New York, NY, USA, 2019. Association for Computing Machinery.

[142] Melissa Zimdars and Kembrew McLeod. *Fake news: understanding media and misinformation in the digital age*. MIT Press, 2020.

Appendix

1. How do you describe your niche, role, activity, or agenda you have online?
2. What brought you to hacking, OSINT, cyber-threat intelligence, and any operations you have taken so far?
3. Have you faced any obstacles, challenges, repercussions because of your activity?
4. Has the obstacles, challenges, repercussions affected your commitment, motivation, and vision of your actions and in what way?
5. What is your take on the increased misinformation proliferation online?
6. Have you ever engaged or considered engaging in utilizing your actions in exposing disinformation campaigns? What was the disinformation about, in what capacity you participated, and what were the outcomes you were attempting to achieve?
7. What do you think the tools, tactics, and procedures undertaken in a hypothetical *misinformation hacktivism* operation might entail?
8. What in your opinion, is the way to continue evolving this work and in what shape and form?
9. Is there anything else that you would like to add or say that is relevant to the questions we have asked so far?
10. If you would like to share some demographic information, please do - we don't require it but it will help us better contextualize your effort and story.