INTERVENTIONS FOR ONLINE HARASSMENT OF JOURNALISTS

MOLLY CINNAMON
Suddenly I was getting 100 tweets a minute. I’d open my phone at my kids’ playgroup and see a picture of myself in a gas chamber.”

This is what Bethany Mandel, a conservative Jewish journalist, faced after publishing an article about former President Donald Trump’s anti-Semitic following. In response, the police asked, “Why don’t you just stop writing things on the internet that make people upset?”

This unsympathetic reaction is precisely what the perpetrators of online harassment hope for. Online harassment of journalists aims to shape the broader civic conversation by limiting what can be said without destructive consequences. As more journalists are driven out of the public online sphere—or limit the scope or content of their coverage—due to harassment, one of the pillars of a free-thinking democracy is at stake: the freedom of the press.

This white paper aims to help address the problem of online harassment of journalists by identifying priority areas of technological, financial, or policy investment. The underlying research is rooted in workshops and breakout sessions with Berkman Klein Center for Internet and Society (BKC) affiliates who are experts in areas of technology, news media, law, and nonprofits. This white paper lays out these findings, identifying opportunities for cross-industry efforts that will multiply the protection of a single journalist into a network of press resiliency.

I. LIFESPAN OF JOURNALIST HARASSMENT ONLINE

Dogpiling – a coordinated effort to attack a target through “a barrage of threats, slurs, insults, and other abusive tactics” – intends to silence the individual, punishing them for what they have published. This phenomenon is all too common: 73% of women journalists have experienced online harassment. In the face of such risks, 38% of women journalists have reduced their presence online, and 20% reported withdrawing offline altogether, self-censoring out of fear of online threats turning into offline violence. These numbers do not even capture the totality of the problem. To determine how to disrupt dogpiling, the lifespan of harassment online must be understood: from the release of a journalist’s work, to the inciting harassment incident, to the dogpiling itself, to the resulting chilling effect.

Figure 1: Lifespan of journalist harassment online.

Today, when a journalist’s work is released, it often hits social media in tandem with publishing. The inciting incident occurs when trolls are catalyzed to harass the journalist, most often incited by a public figure (such as Tucker Carlson or Donald Trump) directing negative, vitriolic attention to the journalist or their work on traditional or social media. Trolls use anonymous forums like Kiwi Farms or Doxbins to coordinate a dogpile, initiating a constant flow of harassing, threatening, and violent messages that utilize and dox personal information online about the journalist, their friends, and their family. Sometimes the online violence morphs into physical incidents, such as swatting or stalking. The chilling effect sets in when journalists do not publish again or limit the scope of their discourse due to the lasting psychological or reputational harm once the dogpile has died down. This is exactly what the dogpile participants want: to mold civic discourse, determining what can be said without reprisal.

II. TODAY’S LANDSCAPE: SOLUTIONS & BLOCKERS

Any new interventions to address the problem of online harassment of journalists should capitalize upon the existing valuable work done by nonprofit organizations, newsrooms, and technology companies. This section maps those solutions against the harassment lifespan, illustrating areas for growth due to existing limitations of policy, law, and economic realities.

---


2 Id.


4 Id. at 12.

i. Pre-Inciting Incident: Limiting online exposure

Once a dogpile is underway, it is nearly impossible to stop. Thus, the most effective existing solutions are preventative. Digital safety trainings offered by nonprofits, well-funded newsrooms, and technology companies enable journalists to limit their online exposure and thus minimize fodder for harassment and doxing. For example, the nonprofits the International Women’s Media Foundation (IWMF) and the Committee to Protect Journalists (CPJ) offer trainings for journalists to protect their data online and ensure device and communications security. CPJ’s Digital Safety Kit even includes an editor’s checklist to protect staff and freelancers against online abuse. As a technical solution to this problem, the browser extension Privacy Party prompts social media users on how to lock down their accounts to maximize social media safety and reduce the content that can be weaponized against a journalist.

While journalists can be guided to manually scrub their social media profiles, they lack control over the personal information accumulated by data brokers, including their email addresses and phone numbers. Content removal services, such as DeleteMe, PrivacyPros, and Optery, enable the continuous deletion of this data at a price ranging from roughly $50 to $300 a year, depending on their thoroughness. Although these services are not a panacea—content removal requests are not enforced by law—they reduce a journalist’s potential attack surface.

Blocker: Uneven Deployment

Digital safety trainings and content removal services are valuable only if they are used. Unfortunately, many newsrooms and journalists are unaware of preventative resources, even if free. Often, a journalist only becomes aware of the amount of their online exposure because they have been harassed. Once dogpiling occurs, reducing information online is significantly less effective—one an address or phone number is circulated on social media or forums, the journalist has nearly no control over its removal. In the face of the risk of online abuse and a lack of understanding about effective solutions, some newsrooms simply tell journalists to avoid discussing controversial topics online altogether. But in doing

<table>
<thead>
<tr>
<th>Point in Lifespan</th>
<th>Current Solution</th>
<th>Solution provided by</th>
<th>Solution implemented by</th>
<th>Blockers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Inciting Incident</td>
<td>Limiting online exposure—social media lockdown</td>
<td>Nonprofits, newsrooms, technology companies</td>
<td>Journalists</td>
<td>Lack of awareness, often not actually completed before the inciting incident</td>
</tr>
<tr>
<td>Pre-Inciting Incident</td>
<td>Limiting online exposure—data broker content removal</td>
<td>Technology companies, newsrooms</td>
<td>Journalists, newsrooms</td>
<td>Cost, uneven deployment</td>
</tr>
<tr>
<td>Dogpiling</td>
<td>Guidance during the dogpile: digital security &amp; logging</td>
<td>Nonprofits</td>
<td>Journalists</td>
<td>Lack of awareness, lack of access to digital expert</td>
</tr>
<tr>
<td>Dogpiling</td>
<td>Physical security &amp; online security</td>
<td>Newsrooms, Nonprofits</td>
<td>Newsrooms, Technology Companies</td>
<td>Cost, often done too late, unsteady reliance on platforms</td>
</tr>
<tr>
<td>Dogpiling</td>
<td>Escalation Channels</td>
<td>Technology Companies</td>
<td>Technology Companies</td>
<td>Fickle, high reliance on platforms</td>
</tr>
<tr>
<td>Chilling Effect</td>
<td>Communities for mental health recovery</td>
<td>Nonprofits</td>
<td>Journalists</td>
<td>Lack of awareness, may be too late</td>
</tr>
</tbody>
</table>

11 The Chilling, supra n.8 at 14.
so, the press does the dogpilers’ work for them, undercutting its own freedom out of fear of harassment.

ii. During Dogpiling: Guidance & Efforts for Safety

Once dogpiling starts, well-funded newsrooms and nonprofits offer a patchwork of mitigating tactics. **Digital security and logging** efforts can help journalists respond to an ongoing attack and seek assistance from law enforcement, security experts, and platforms. For example, AccessNow’s Digital Security Helpline provides incident response-based guidance to journalists on how to manage privacy and cybersecurity concerns in the wake of a dogpile.12 Nonprofits RaReNet (Rapid Response Network) and CiviCERT have created the Digital First Aid Kit, guiding journalists on how to log evidence of attacks for technical and legal help.13

Well-funded newsrooms have some resources at this stage of harassment. When online harassment crosses over into in-person threats or actual violence, some newsrooms may offer **physical security** resources, including relocating a journalist during a period of in-person harassment after being doxed. They may also use informal **escalation channels** to contact Trust and Safety teams at social media platforms to expedite a response to their takedown requests of harassing content. Facebook, TikTok, and the Twitter (as it was known before Elon Musk’s takeover)14 use a “Trusted Partners” system to engage with civil society organizations such as AccessNow. These relationships can enable informal escalation channels to raise urgent content removal requests.

**BLOCKER: Cost & Fickle Relationships with Social Media Platforms**

These tools have mental and economic costs. Digital security and logging efforts, while highly valuable, still require journalists to be the first responders to their own harassment—receiving the vitriol, blocking trolls, and managing their online presence.15 As for security solutions sustained by news outlets, those costs are likely out of scope for smaller shops. And even in large, well-established newsrooms, not all employees are full-time; many are contributors or freelancers who would not get the benefits of such protections.

Changing relationships with platforms have created uncertainty about solutions that rely on their cooperation. Given recent Congressional and political scrutiny of social media platforms and content removal,16 platforms are cautious about opening up informal escalation channels that could be criticized as perpetuating “censorship.” Even in the absence of such scrutiny, purely relationship-based escalation channels are innately vulnerable; if those contacts leave, are fired, or laid off, or even just reassigned within the company, those channels are closed. Additionally, some social media platforms have made interoperability with their technology unreliable; when Twitter shut off free access to its API in April 2023,17 many critical technologies mitigating harassment on the site became defunct. For example, the plug-in Block Party (by the same creators as Privacy Party) relied on Twitter’s API to enable users to shield themselves from harassment by mass-blocking users or filtering out certain content. Without reliance on APIs, much platform-enhancing technology to protect journalists from dogpiling cannot be built.

**iii. Preventing the Chilling Effect: Community-Building**

To preserve journalists’ mental health and prevent them from leaving the industry, nonprofits suggest forming communities to share the emotional burden with others similarly affected.18 Becaus...
Yet, for a journalist hoping to obtain affirmative recourse, the United States’ strong free speech protections, combined with the anonymous nature of most online accounts, make it difficult for journalists to go on the offensive against a dogpile. Litigation costs against hard-to-find and likely judgment-proof defendants comprise a barrier to entry for individual journalists and many news outlets, and the difficulty of winning defamation suits—especially when journalists themselves can be considered public figures—may mean there is no vindication at the end of the process. Platforms too are shielded from nearly all liability, even for actionable content, by Section 230. Thus, unless this legal gridlock changes, near-term efforts to reduce journalist harassment online must come from sources other than the law.

III. PROPOSED INTERVENTIONS

Informed by subject matter experts, the proposed interventions aim to halt progression in the online harassment lifecycle, ultimately subverting the chilling effect.

i. Trust & Safety Co-op

A Trust and Safety Co-op, consisting of Trust & Safety (T&S) professionals from various platforms, could enable a structured mechanism for anticipating, identifying, and halting dogpiling across the Internet. Because harassment rarely stays on one platform, T&S teams should harness their collective knowledge to inform awareness of ongoing or future harassment. Information sharing may be as simple as publishing alerts of dogpiling underway or as complex as sharing analytical tooling to predict when a dogpile is likely to occur. In response, T&S teams can apply their respective platforms’ policies to the content. But by working from a collective knowledge base, T&S teams can lower the risk of dogpiling on multiple platforms.

ii. Journalist Co-op

A Journalist Co-op, consisting of freelance journalists and journalists from small, under-resourced news outlets, would enable the pooling of resources to achieve the same protections from harassment as those at major, well-resourced news outlets.

Proactive online security is the most critical defensive tactic in the face of online harassment. This co-op would help every single member journalist lock down their social media profiles and enact strong cybersecurity protections, ideally before publishing their first story. For example, this co-op could form a mutually beneficial partnership with Privacy Party, the free browser extension that automatically locks down users’ social media settings: in ensuring that as many journalists as possible use this tool, the overall attack surface for online harassment would be minimized, and Privacy Party could expand their user base.

Full benefits of membership in the co-op should be gated based upon whether the journalist has taken free prescribed protective measures. Upon full membership, the co-op could fund a journalist’s subscription to data broker content removal services, connect them with digital experts, and provide them with funds for physical security in the event of severe doxing.

iii. Tool for Automated Harassment Reporting

A tool to automatically compile digital evidence of online harassment would enable journalists to more easily obtain aid from platform T&S teams, digital security experts, and law enforcement. The tool would sweep journalists’ multiple social media
accounts for harassing posts or messages, then compile and share the collected content into databases or reports, depending on the preferred mode of input for each receiving party. To prevent the need for manual review, targets of harassment could tag a few posts as indicative of the harassment, and artificial intelligence and natural language processing tools could be used to infer other posts within the same dogpile. If the journalist was willing to provide personal information like phone number and address, the tool could flag doxing content. An automated tool not only makes the process of reporting for platform, cybersecurity, and legal help easier, but it reduces the risk of a journalist slipping into a chilling effect because of the psychological burden of reviewing their own harassment.

NEED: TECHNICAL & LEGAL RESEARCH

Any technical intervention built on top of platforms, including an automated reporting tool, must overcome the lack of API access to the platforms. In the face of this obstacle, some developers have pivoted to building browser extensions to create middleware-like functionality for their users. A browser extension executes with the authority of the user and simply scripts on top of the site the user visits. For example, after Block Party was shut down due to Twitter rescinding free API access, founder Tracy Chou pivoted to creating Privacy Party as a browser extension.19 While scripting does not enable the same automated and mass data pulls as APIs, browser extensions are a valuable work-around to existing limitations. Still, more research is needed to expand the possibilities of how independent safety tools can be built without reliance on platforms. For example, can browser extensions be built to run on mobile, or even run tasks in the background (with user permission)? Could phone apps be written to access the local copy of social media content other apps load onto the device? Solving these technical problems would unlock a range of new tooling to mitigate online harassment.

But a new technological workaround creates new legal concerns. Certain end-user license agreements (EULAs) may prevent scripting over certain apps for certain purposes. To build browser extensions and scripting services that can reliably scale, more technical and legal research is needed to determine the limits of those technologies.

Near-term interventions: level up existing solutions

---

Developing an accurate, apolitical data observatory of online harassment of journalists requires buy-in from many journalists across political lines. And the stakes are high: if personal data from the research is leaked, harassment against these journalists (and the researchers who compiled the data) could begin with a renewed vigor.

**NEED: BUY-IN & TECHNICAL RESEARCH**

This research requires collaboration between journalists, academics, technologists, and data scientists. Developing an accurate, apolitical data observatory of online harassment of journalists requires buy-in from many journalists across political lines. And the stakes are high: if personal data from the research is leaked, harassment against these journalists (and the researchers who compiled the data) could begin with a renewed vigor. Additionally, such research would butt up against aforementioned limitations on API access. API workarounds, such as browser extensions, are a prerequisite to obtaining data inputs for research.

i. Enable Middleware to Unlock Third-Party Development

Cost-effective API access is critical to building technology that combats harassment. For example, tooling that is interoperable with social media platforms can automatically lock down a journalist’s online presence, enforce the strictest cybersecurity controls, and create content filters to minimize exposure to dogpiles. Such automated tooling would increase the broad implementation and efficacy of privacy and cybersecurity controls many times over. Simply restoring API access at Twitter would revive valuable tools such as Block Party.

Without government regulation, platforms do not have a strong market incentive to provide affordable API access. Change will only come with regulation, and those changes are slow moving. Federally, the agent of change may be antitrust regulation, breaking apart companies or at least requiring interoperability. In 2021, the FTC sued Facebook for anti-competitive practices, in part for closing off API access to third-party developers producing tooling similar to Facebook or connecting to other social media platforms. The viability of this claim and resulting damages remain to be seen; the case is ongoing. In states, the momentum of privacy legislation should be used to fuel interest in legislation for open APIs. The New York State Senate has led the way, proposing a bill requiring social media applications to maintain an open API accessible to third-party platforms. Funding is needed to continue fighting this fight, even if piecemeal at the state level.

ii. Pass Privacy Legislation to Reduce Online Exposure

Today, it is nearly impossible for an American online to track all of the public sites on which their data is listed. Even for someone with every proactive privacy protection in place, personal information can still be in circulation online, enabling doxing and stalking. In the absence of federal legislation,

---


"DEVELOPING AN ACCURATE, APOLITICAL DATA OBSERVATORY OF ONLINE HARASSMENT OF JOURNALISTS REQUIRES BUY-IN FROM MANY JOURNALISTS ACROSS POLITICAL LINES. AND THE STAKES ARE HIGH: IF PERSONAL DATA FROM THE RESEARCH IS LEAKED, HARASSMENT AGAINST THESE JOURNALISTS (AND THE RESEARCHERS WHO COMPILED THE DATA) COULD BEGIN WITH A RENEWED VIGOR."
State privacy laws must continue to expand consumers’ data rights and limit the reach of data brokers. For example, California passed the Delete Act at the end of 2023, putting legal teeth into the removal requests made on data broker sites. To put collective pressure on the industry to conform with privacy-preserving standards, more states need to pass laws of this nature. Still, much personal information used to fuel doxing and in-person harassment, such as residential addresses, can be found in public records, including property transactions and voter registrations. States and the federal government alike need to look more closely at the privacy and security concerns generated by online access to public records while balancing transparency interests.

**iii. Expand Harassment Laws to Encompass Online Threats**

State laws against harassment need to be evaluated and amended to ensure they provide protection for those experiencing online threats and related in-person stalking, both symptoms of dogpiling. Across all states, restraining orders are most commonly granted in cases of domestic and intimate partner violence, and judges are hesitant to grant orders that may restrict freedom of speech. For these reasons, journalists experiencing harassment online may have a difficult time convincing judges to extend their interpretation of harassment laws to credible threats and abuse from an online-only, non-interpersonal relationship. State laws must be tweaked to ensure they respond to an era in which online communication is a major medium for harassment.

**IV. CONCLUSION**

In today’s age, journalists face an impossible choice: publish groundbreaking truths on socially-critical issues and face seemingly endless harassment and threats, or forgo publishing altogether, saving themselves and their family from becoming the targets of online and offline violence. The interventions proposed in this white paper aim to eliminate this choice, or at least make it significantly easier, by undermining the effectiveness of the dogpile. If successful, journalists could publish freely, and freedom of the press would be protected. Critically, the impact of these proposed interventions would benefit more than just journalists. The co-ops, tools, and legal changes recommended in this white paper would give all individuals online tools to proactively and reactively respond to online harassment. By elevating existing solutions and harnessing collective power across industries, we can disrupt the dogpile and empower speech without reprisal—for journalists and anyone online.

---

23 California Senate Bill No. 362 (2023).

---

<table>
<thead>
<tr>
<th>Point in Lifespan</th>
<th>Solution</th>
<th>Goal</th>
<th>Solution Impacts</th>
<th>Resource Need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Inciting Incident</td>
<td>Regulations to force API access</td>
<td>Enable research &amp; technical solutions to automatically minimize online presence and enforce strong cybersecurity controls</td>
<td>Platforms, technologists, researchers</td>
<td>Policy change, technical skills, organization</td>
</tr>
<tr>
<td>Pre-Inciting Incident</td>
<td>Privacy legislation</td>
<td>Reduce the amount of uncontrollable personal information online</td>
<td>Journalists</td>
<td>Legal change</td>
</tr>
<tr>
<td>Dogpiling</td>
<td>Harassment law reform</td>
<td>Amend state laws for protective orders to ensure they provide recourse in dogpiling</td>
<td>Journalists</td>
<td>Legal research, legal change</td>
</tr>
</tbody>
</table>

Long-term interventions: shape policy & law to unlock new solutions
Online harassment of journalists poses a grave risk to press freedoms: retaliating in response to a disfavored article or post, trolls coordinate abuse in the hopes of silencing the journalist altogether. To protect journalists from this outcome, this white paper provides recommendations for interventions to interrupt the lifespan of harassment online. These recommendations are informed by a literature review, review of existing solutions, and interviews with stakeholders in areas of technology, news media, and nonprofits.

Recommended interventions that can be deployed near-term
- Trust & Safety Co-op: a community of T&S professionals empowered to share predictive analytics and reports of ongoing harassment, helping stop online harassment earlier
- Journalist Co-op: a community of journalists centered around resource-sharing, ensuring every journalist is equipped with tools to respond to harassment proactively or reactively
- Automated Harassment Reporting: a tool to automatically collate instances of harassment across platforms, allowing journalists to more easily secure aid from trust and safety teams, digital security experts, and law enforcement
- Public Reporting Dashboard: live metrics of journalist online harassment to apply public pressure to social media platforms, prompting their attention on this issue

Recommended interventions requiring longer-term investment
- Open API Access: policy change to force platform interoperability for third-party developers, enabling them to build tools to mitigate online harassment
- Privacy Legislation: legislation to enable journalists to exercise control over their personal information online, such as wiping their information from data brokers
- Expanded Harassment Laws: policy recommendations to ensure that state protective orders cover dogpiling-spurred harassment, not just domestic violence situations

These recommendations provide ready-to-deploy means to protect journalists from online harassment, ultimately providing a path forward to protecting freedom of the press.