Guide 3: Protecting Against Malicious Bots and Advocating for Bot Governance

BOTS.NGO

Executive Summary

Civil society organizations are on the front lines of the bot governance challenge. They are targets of bot-amplified harassment and disinformation. They depend on the integrity of the information environments — social media, regulatory processes, public discourse — that bots can corrupt. And they have a distinctive role as representatives of the public interest in the policy debates about how bots should be regulated. Yet most NGOs are poorly prepared for all three dimensions: they lack the detection capacity to identify when bots are targeting them, the defensive protocols to respond effectively, and the policy voice to shape how governments and platforms govern automated activity.

This guide addresses the protective and advocacy dimensions of the bot challenge for NGOs. It covers: how to detect and respond to bot-amplified harassment and disinformation campaigns; how to protect regulatory and advocacy processes from bot-flooding; how to engage in policy debates about bot governance and platform accountability; and how to build organizational resilience against the automation-driven manipulation of public discourse that represents one of the most significant threats to democratic civil society in the current era.

The guide is grounded in the recognition that the bot threat is not static — it is evolving rapidly. As of 2026, AI-powered bots can generate human-like text, hold coherent conversations, and create realistic personas. Swarms of coordinated AI bots have been shown to measurably shift people's beliefs and political opinions through repeated exposure to particular narratives. State actors and foreign governments operate large-scale bot networks targeting civil society organizations, media, and democratic institutions. The threat is real, it is growing, and it requires systematic organizational response.

Key takeaway: Defense requires preparation, not reaction. NGOs that wait until they are targeted to develop detection, response, and advocacy capacity will find themselves overwhelmed. Systematic preparedness — audits, protocols, relationships, and policy engagement — is what determines whether an organization survives and thrives in a bot-saturated information environment.

Evidence Table: Key Findings, Strength, and NGO Implications

Key Finding	Evidence Strength	NGO Implications
Just 6% of Twitter accounts — likely bots — were responsible for 31% of low-credibility information spread during the 2016 U.S. election.	Strong — Indiana University research; widely replicated	A small number of coordinated bot accounts can have an outsized impact on information environments. Don't assume that volume of content or apparent consensus represents genuine public opinion.
State actors and foreign governments operate bot networks specifically targeting civil society and democratic institutions.	Strong — BBC (2026) reporting on Russian bot campaigns; Brennan Center; Stanford Internet Observatory	International civil society organizations and those working on politically sensitive issues face heightened state-actor bot risk. Develop specific threat assessments for high-risk contexts.
In 2017, bots flooded the FCC with over 1 million fake public comments, threatening to subvert the regulatory process.	Strong — Brennan Center documentation; government investigation	Regulatory comment processes your NGO depends on are vulnerable to bot-flooding. Advocate for verification requirements and document when flooding occurs.
Bot-amplified harassment campaigns specifically target women and people of color in civil society, making normal work impossible.	Strong — feminist technology literature; documented case studies	Build staff protection protocols — not just organizational defenses — that specifically address the heightened personal risk faced by staff from marginalized groups.
Swarms of coordinated AI bots can measurably shift people's beliefs and political opinions through synthetic consensus.	Emerging — The Conversation (2026); academic research	Your organization's campaigns operate in an environment where bot-generated synthetic consensus can undermine authentic public opinion formation. Monitor for and counter synthetic consensus on your key issues.
Platform transparency reports and coordinated inauthentic behavior (CIB) enforcement are the primary tools for exposing and disrupting bot networks, but are inconsistently applied.	Strong — platform policy research; civil society technology reports	Build relationships with platform trust-and-safety teams. Use formal reporting mechanisms. Maintain records of reports and responses to inform advocacy for stronger platform accountability.
Calls for blanket bans on bots are misguided and would eliminate valuable civil society uses of automation while leaving underlying problems (microtargeting, filter bubbles, human disinformation) unaddressed.	Strong — Policy Options (IRPP); SSIR; technology policy literature	Advocate for transparency, disclosure, and accountability — not blanket bans. Develop specific, targeted policy asks that address the harms without eliminating legitimate automation.

Step-by-Step Framework

Step 1: Conduct a Bot Threat Assessment for Your Organization

Not all NGOs face the same bot threats. An animal welfare organization running corporate campaigns faces different risks than a democracy organization working on voter registration, which faces different risks than a human rights organization operating in a politically repressive country. Begin with a structured threat assessment that maps: (1) What are the politically sensitive issues your organization works on, and who has a strong interest in silencing or discrediting your work? (2) Which staff members are most likely to be personally targeted in harassment campaigns? (3) Which digital platforms and processes does your advocacy depend on, and how vulnerable are they to bot manipulation? (4) Has your organization been targeted by bot activity in the past, and if so, what form did it take?

This threat assessment should be conducted by a small team including communications, campaigns, technology, and leadership staff. It should be documented and updated annually, and immediately after any significant bot incident. The output is a prioritized list of threats and a corresponding list of defensive and preparedness investments that your organization should make.

Step 2: Build Detection Capacity

Detection is the foundation of defense. If you cannot identify when bots are targeting you or distorting your information environment, you cannot respond effectively. Build detection capacity across the key risk areas identified in your threat assessment.

For social media, establish regular monitoring for: sudden spikes in mentions, follower counts, or hashtag activity (often a sign of bot amplification); coordinated posting patterns (many accounts posting the same message within minutes); new accounts created recently with minimal posting history but high activity; unusual geographic patterns in account origins; and content that appears to be machine-generated (formulaic, repetitive, missing the contextual specificity of human conversation). Tools like Botometer (for Twitter/X accounts), Social Blade (for unusual account growth patterns), and basic social media monitoring platforms can support this work.

For regulatory and advocacy processes, monitor for: sudden large volumes of comments appearing in short time windows; comments that are suspiciously similar to each other; comments from email addresses or accounts that do not correspond to real individuals; and patterns suggesting coordinated submissions from a single source. If you detect suspicious activity in a regulatory process, document it carefully and report it to the relevant agency and to organizations like the Brennan Center that track bot activity in democratic processes.

Step 3: Develop an Incident Response Protocol

When bot-amplified harassment or disinformation targeting your organization is detected, the quality of your response will depend entirely on whether you have a plan. Organizations that have to build their response from scratch while the attack is happening — while staff are being harassed, while false narratives are spreading, while leadership is trying to understand what is happening — consistently perform worse than organizations that prepared in advance.

Your incident response protocol should specify: (1) Who is notified when a bot incident is detected, and how (the communications director, the technology lead, senior leadership)? (2) Who has authority to make decisions about the organizational response? (3) What is the initial response to staff who are being targeted for personal harassment — who supports them, what resources are available, what decisions are they not expected to make alone? (4) What is the communications response — do you respond publicly, through which channels, with what message? (5) What reporting does your organization file with platforms, and who is responsible? (6) What documentation does your organization maintain, and for what purposes (legal, regulatory advocacy, research)?

The most critical principle in incident response is: do not engage with individual bot accounts. Engaging amplifies the bot campaign's reach and exhausts staff. Respond through your own authentic channels, report through platform mechanisms, and focus organizational energy on maintaining authentic engagement with real supporters rather than on fighting the bots directly.

Step 4: Protect Your Regulatory and Advocacy Processes

If your NGO's theory of change includes influencing regulatory agency decisions, legislative processes, or other formal public participation mechanisms, you need to actively protect the integrity of those processes. This means: (1) making your organization's regulatory submissions as distinctive, substantive, and specifically attributable as possible — the opposite of what bots produce; (2) documenting and reporting suspicious activity in comment processes to the relevant agency; (3) building coalitions with other civil society organizations to jointly flag and respond to bot-flooding; and (4) advocating explicitly for verification and transparency requirements in regulatory comment processes.

In the United States, the Brennan Center and other organizations are actively working on policy proposals to require verification of identities in high-stakes regulatory comment processes, to allow agencies to disregard mass-produced bot-generated comments, and to create liability for orchestrating bot campaigns in regulatory contexts. Engage with these policy processes and lend your organization's voice to the case for reform.

Step 5: Build Staff Protection Infrastructure

Bot-amplified harassment campaigns disproportionately target individual staff members — particularly those who are publicly visible, work on controversial issues, or belong to demographic groups that are regular targets of coordinated online harassment. The organizational response must include a personal support infrastructure for targeted staff, not only a communications strategy for the organization.

This infrastructure includes: a clear process for staff to report harassment and receive immediate organizational support; access to mental health resources for staff experiencing sustained harassment; technical assistance for staff who want to improve their personal account security (two-factor authentication, private browsing, account privacy settings); a policy on organizational support for staff who face legal threats as a result of harassment campaigns; and explicit organizational acknowledgment that being targeted is not the fault

of the targeted staff member. The most effective organizations treat harassment protection as a staff care issue, not only a communications issue.

Step 6: Engage in Bot Governance Policy Advocacy

Civil society organizations have both a stake in and a distinctive voice in the policy debates about how bots and automated activity should be governed. The current policy landscape is contested and in flux: some advocate for mandatory disclosure (all bots must be labeled as bots), others for platform liability, others for criminal penalties for orchestrating bot campaigns in democratic processes, and others for technical standards that make bot detection easier.

Develop specific, evidence-based policy positions on bot governance that reflect your organization's experience and interests. Engage with the organizations most active in this space — the Brennan Center, Stanford Internet Observatory, the Alliance for Securing Democracy, the Global Network Initiative, the Center for Democracy and Technology. Submit comments to regulatory proceedings on platform accountability and election integrity. Brief legislators and regulators on your organization's experience of bot-amplified harassment or disinformation. Partner with academic researchers who study bot activity to ensure that civil society experience informs the scholarly literature.

The most important policy ask for most NGOs is relatively simple: require bots to be labeled as bots. Transparency and disclosure requirements would not prevent all harmful bot activity, but they would make the deception at the heart of malicious bot campaigns much harder to sustain, and they would empower users and organizations to make informed decisions about the content they consume and share.

Step 7: Build Long-Term Organizational Resilience

Bot threats are not going to disappear. As AI technology advances, the sophistication of automated manipulation will continue to increase. The most durable protection for NGOs is not any specific defensive technology or protocol, but organizational resilience: deep authentic relationships with supporters and communities; a strong reputation for credibility and transparency; a diverse information environment that is not wholly dependent on bot-susceptible platforms; and an organizational culture that treats information hygiene, digital security, and ethical automation as core organizational competencies.

Invest in building direct relationships with your supporters — email lists, community spaces, events — that give you communication channels that do not depend on social media platforms. Develop your own content publishing capacity so that you are not entirely dependent on platforms that can be manipulated by bot activity. Build your credibility as a source of reliable information so that when bot campaigns try to spread false narratives about your organization, your supporters have a strong prior that makes them resistant. And develop an organizational identity that is explicitly grounded in authentic human connection — the most powerful antidote to automated manipulation.

Tools & Templates

Bot Threat Assessment Template: A structured document mapping: politically sensitive issues, likely adversaries, high-risk staff members, vulnerable digital platforms and processes, and past incidents. Reviewed annually.

Incident Response Protocol Template: A pre-written document covering: notification chain, decision authority, staff support procedures, communications response options, platform reporting steps, and documentation requirements. Reviewed and tested annually.

Social Media Monitoring Dashboard: A template for tracking key bot detection indicators across your most important social media platforms: mention volume, follower count changes, hashtag activity, and suspicious account characteristics. Updated weekly.

Regulatory Comment Protection Checklist: A checklist for all regulatory submissions: organizational letterhead and contact details, substantive and distinctive content, specific regulatory citations, human signatory identification, and submission confirmation documentation.

Staff Harassment Support Protocol: A written protocol specifying: how staff report harassment, who they contact, what organizational support is available (counseling, security advice, legal support), and how the organization communicates publicly about harassment of staff.

Platform Reporting Guide: A reference document with the specific reporting mechanisms for coordinated inauthentic behavior on each major platform (Facebook/Meta CIB reporting, Twitter/X synthetic manipulation reporting, YouTube coordinated gaming reporting), with links and step-by-step instructions.

Case Vignettes

Vignette 1: Russian Bot Campaign Against Civil Society Organizations (2024–2025)

BBC Monitoring analysis documented Russian-linked bot campaigns that used coordinated bot networks to spread false narratives, including campaigns that spread misinformation about a bed bug outbreak at the Paris Summer Olympics — a case study in how state-operated bot farms deploy manufactured content to undermine trust in democratic institutions and civil society. The campaigns used bot accounts that mimicked genuine public concern, amplified by networks of inauthentic accounts to create the appearance of a viral grassroots response.

For civil society organizations working on issues that intersect with geopolitical interests — human rights, democracy, election integrity, climate, animal welfare in industries with strong state backing — the risk of state-actor bot campaigns is real and documented. The case illustrates that the bot threat is not only from domestic political operatives: it is also from state actors with significant resources and sophisticated capabilities who view civil society organizations as legitimate targets.

Lessons learned: (1) State-actor bot campaigns are a real and documented threat to civil society, not a hypothetical. (2) The goal of these campaigns is often to undermine trust and create confusion rather than to advance a specific policy outcome — making them harder to refute than simple misinformation. (3) Relationships with journalists, researchers, and platform trust-and-safety teams who can help detect and document state-actor activity are among the most valuable assets an NGO can build.

Vignette 2: The FCC Bot Flood and the Regulatory Process Defense

In 2017, the Federal Communications Commission's public comment period on net neutrality was flooded with over 1 million bot-generated fake comments — a coordinated campaign by opponents of net neutrality to manufacture the appearance of public opposition. The Brennan Center and subsequent investigations documented the campaign in detail, but the damage to the integrity of the comment process was already done: regulators initially counted the bot-generated comments as authentic public input.

The case has been cited by scholars and policymakers as one of the clearest demonstrations that regulatory comment processes — a cornerstone of democratic governance and NGO advocacy strategy — are vulnerable to bot manipulation at scale. For animal rights NGOs, environmental NGOs, and any organization that depends on regulatory consultations to give civil society a voice, protecting the integrity of those processes is not an abstract policy interest — it is directly relevant to your ability to influence the decisions that determine outcomes for the animals and communities you serve.

Lessons learned: (1) Regulatory processes are not bot-proof, and cannot be assumed to accurately reflect public opinion when bot-flooding is possible. (2) Active documentation and reporting of suspicious activity in regulatory processes is a civic responsibility for NGOs, not optional engagement. (3) The case for verification requirements in high-stakes regulatory processes is clear and urgent — NGOs should be actively advocating for this.

Metrics & KPIs

Metric	What It Measures	How to Track
Bot threat assessment currency	Whether the assessment has been completed and updated within 12 months	Document version date; annual review calendar
Incidents detected and responded to	Number and nature of bot-related incidents	Incident log
Average incident response time	Speed of organizational response when an incident is detected	Incident log timestamps
Platform reports filed and outcomes	Engagement with platform trust-and-safety systems	Platform reporting log
Staff harassment support activations	Number of times the staff harassment protocol was activated	HR/communications log
Regulatory comment quality	Whether submissions are substantive, distinctive, and properly attributed	Submission review by legal/communications team

Metric	What It Measures	How to Track
Policy advocacy engagements on bot governance	Participation in policy processes related to bot governance	Activity log
Media literacy content produced	Number of materials educating constituents about bot risks	Content calendar

Risks & Mitigations

Risk: Incident response escalating bot campaigns rather than containing them. Responding publicly to each individual bot account or engaging in arguments with bot-amplified content can amplify the campaign's reach.
Mitigation: Train all staff on the principle of not engaging with individual bot accounts. Focus response energy on authentic channels — your own communications, your real supporters — rather than fighting bots directly.

Risk: Regulatory process engagement undermined by bot-flooding. If regulatory processes where your NGO submits comments are flooded with bot-generated fake submissions, your authentic voice may be drowned out or the process may be delegitimized.
Mitigation: Make submissions as distinctive and substantive as possible. Document and report suspicious activity. Advocate for verification requirements in the regulatory processes your NGO depends on.

Risk: State-actor bot campaigns targeting high-profile staff. Internationally operating NGOs and those working on geopolitically sensitive issues face heightened risk of state-actor bot campaigns that target individual staff members.
Mitigation: Develop specific threat assessments for high-risk staff. Provide individualized digital security support. Build relationships with organizations specializing in digital security for civil society (like Access Now or Front Line Defenders).

Risk: Bot governance policy advocacy positions being weaponized. If your organization advocates for specific bot governance policies (e.g., mandatory bot disclosure), opponents may use bots to misrepresent your positions or generate artificial controversy around your advocacy.
Mitigation: Document your policy positions clearly and publicly. Build relationships with journalists and researchers who can verify your positions if they are misrepresented.

Risk: Synthetic consensus shifting your organization's strategic decisions. If your organization monitors social media to track public opinion on your key issues, bot-generated synthetic consensus can make fringe positions appear mainstream — causing strategic miscalibration.
Mitigation: Use multiple methods to assess public opinion, not only social media monitoring. Conduct genuine constituent engagement — surveys, focus groups, community conversations — that cannot be bot-manipulated.

Implementation Checklist

- [ ] Complete bot threat assessment for your organization
- [ ] Develop and test incident response protocol
- [ ] Set up social media monitoring dashboard for bot detection indicators
- [ ] Build staff harassment support protocol and communicate to all staff
- [ ] Establish relationships with trust-and-safety contacts at key platforms
- [ ] Complete platform reporting guide for relevant platforms
- [ ] Review regulatory comment submissions for distinctiveness and attribution
- [ ] Develop organizational bot governance policy positions
- [ ] Engage in at least one policy process related to bot governance per year
- [ ] Schedule annual threat assessment review and incident response protocol test

Glossary

Coordinated inauthentic behavior (CIB): Organized manipulation of public discourse using fake or compromised accounts, often bot-driven. The primary enforcement standard used by major social media platforms.

Bot-amplified harassment: A form of online abuse in which bot networks are used to amplify and coordinate harassment campaigns targeting individuals or organizations, generating far more abuse than any organic human campaign could produce.

Synthetic consensus: The manufactured appearance of widespread agreement created by bot networks. Can shift genuine public opinion by making minority positions appear mainstream.

Bot farm: An organized network of computers or devices running bot software, often operated by state actors or political operatives for large-scale manipulation.

Astroturfing: Using bots or fake accounts to manufacture the appearance of genuine grassroots support for a position or cause.

Platform trust and safety: The internal teams at social media companies responsible for enforcing community standards and terms of service, including action on coordinated inauthentic behavior.

Botometer: A tool developed by Indiana University that analyzes social media accounts and scores their likelihood of being automated bots.

Disclosure requirement: A regulatory requirement mandating that automated accounts identify themselves as bots, not human users.

Digital resilience: An organization's capacity to maintain operations, communications, and advocacy effectiveness in the face of digital threats including bot campaigns, hacking, and disinformation.

Threat assessment: A structured analysis of the specific bot-related threats an organization faces, used to prioritize defensive investments and preparedness activities.

Coordinated inauthentic behavior (CIB) reporting: The formal reporting mechanism on major social media platforms for flagging networks of fake accounts engaged in coordinated manipulation.

Access Now: A civil society organization that provides digital security support to NGOs and human rights defenders facing advanced threats, including state-actor bot campaigns.

References

1. Brennan Center for Justice. (2023). Artificial Intelligence, Participatory Democracy, and Responsive Government. brennancenter.org
2. BBC World Service. (2026). How Bots Manipulate Social Media. bbc.com
3. The Conversation. (2026). Swarms of AI Bots Can Sway People's Beliefs — Threatening Democracy. theconversation.com
4. PMC / NIH. (2021). Bots and Misinformation Spread on Social Media. pmc.ncbi.nlm.nih.gov
5. Policy Options (IRPP). (2025). Minimize Harmfulness of Bots, But Don't Ban Them. policyoptions.irpp.org
6. Ford Foundation. (2026). Why You Should Care About Bots If You Care About Social Justice. fordfoundation.org
7. Stanford Social Innovation Review. (2018). Leveraging the Power of Bots for Civil Society. ssir.org
8. Alliance for Peacebuilding / Toda Peace Institute. (2018). Civil Society in the Age of Automation. toda.org
9. Tech Policy Press. (2025). Autocrats' Digital Advances Underscore the Need for Civil Society. techpolicy.press
10. Atmos Marketing. (2025). The Perils of AI Bots in Social Media: Navigating Risks and Ethics. atmos-marketing.com
11. Knight-Hennessy Scholars. (2026). How AI Threatens the Communities That Sustain Democracy. knight-hennessy.stanford.edu
12. PMC. (2020). Social Media Effects: Hijacking Democracy and Civility in Civic Engagement. pmc.ncbi.nlm.nih.gov