

1 Chapter 9: Taking Action

1 Chapter 9: Taking Action	1
1.1 Reporting	1
1.1.1 Reporting inside the League	1
1.1.2 Reporting to law enforcement from the League	2
1.1.3 Reporting to platforms	2
1.2 Direct Action	2
1.3 References	2
1.3.1 Platform trust and safety policies	2

The point of real-time disinformation tracking is to be able to do something about it. Our basic actions include:

- Direct action.
- Asking someone directly connected to us to take action
- Reporting to someone not directly connected to us, so they can investigate and decide whether to take action.

Direct action: there are many small things that a team could do to disrupt a disinformation incident. These include:

- Flooding a disinformation hashtag or group with alternative information (be careful with this because if the original intent was confusion, you might be adding to it) etc

Asking someone connected to us to take action

- Reporting a suspicious domain to registrars. If we do this, it's on us to gather information to help them - e.g. screenshots of selling bleach 'cures' etc etc

Reporting to someone not directly connected

- This is most likely with the large social media platforms. We're going to find bots and botnets; we won't be able to remove them ourselves, we will be able to report them to platforms. It'll help if we have that reporting mechanism set up ahead of time.

1.1 Reporting

1.1.1 Reporting inside the League

If you know which organisation you need, use the /list_orgs and /list_contacts [org] slack command to find the person you need. More generally, look at the channels guide in the League handbook to see the right channel to report an incident or component to.

1.1.2 Reporting to law enforcement from the League

You can open an LE escalation ticket using the /lenew command

1.1.3 Reporting to platforms

Reporting to social media

- Reddit: <https://www.reddit.com/r/redditsecurity/>
- Twitter: [report-twitter-impersonation](#) and [twitter-rules](#)
- Facebook: [How to Report Things on Facebook](#)
- LinkedIn: [Reporting Inaccurate Information on Another Member's Profile](#)
- Instagram: <https://help.instagram.com/1735798276553028>
- YouTube: <https://support.google.com/youtube/answer/2802027>
- Google: is going to take some digging [Avoid and report Google scams - Google Help](#)

Pinterest

- Fast: <https://help.pinterest.com/en/article/report-something-on-pinterest>
- Slower: report on https://help.pinterest.com/en/contact?page=about_you_page - you'll need a Pinterest account to do this from.
 - Choice is porn, violence, hate speech, self harm, harassment/ exposed private information, spam; currently going with either hate speech, violence or harassment as appropriate.
 - Has an image filesize limit of 2MB
- community guidelines are <https://policy.pinterest.com/en-gb/community-guidelines>

1.2 Direct Action

1.3 References

1.3.1 Platform trust and safety policies

Pinterest community guidelines:

- “Hateful activities. Pinterest isn’t a place for hateful content or the people and groups that promote hateful activities. We limit the distribution of or remove such content and accounts, including:
 - Slurs or negative stereotypes, caricatures and generalisations
 - Support for hate groups and people promoting hateful activities, prejudice and conspiracy theories
 - Condoning or trivialising violence because of a victim’s membership in a vulnerable or protected group
 - Support for white supremacy, limiting women’s rights and other discriminatory ideas
 - Hate-based conspiracy theories and misinformation, such as Holocaust denial
 - Denial of an individual’s gender identity or sexual orientation, and support for conversion therapy and related programmes
 - Attacks on individuals including public figures based on their membership in a vulnerable or protected group
 - Mocking or attacking the beliefs, sacred symbols, movements or institutions of the protected or vulnerable groups identified below
- Protected and vulnerable groups include: people grouped together based on their actual or perceived race, colour, caste, ethnicity, immigration status, national origin, religion or faith, sex or gender identity, sexual orientation, disability, or medical condition. It also includes people who are grouped together based on lower socio-economic status, age, weight or size, pregnancy or ex-military status.
- Misinformation. Pinterest isn’t a place for misinformation, disinformation or mal-information. We remove or limit distribution of false or misleading content that may harm Pinners’ or the public’s well-being, safety or trust, including:
 - Medically unsupported health claims that risk public health and safety, including the promotion of false cures, anti-vaccination advice, or misinformation about public health or safety emergencies
 - False or misleading content about individuals or protected groups that promotes fear, hate or prejudice
 - False or misleading content that encourages turning individuals, groups of people, places or organisations into targets of harassment or physical violence
 - Conspiracy theories
 - False or misleading content that impedes an election’s integrity or an individual’s or group’s civic participation, including registering to vote, voting and being counted in a census
 - Content that originates from disinformation campaigns
 - Factual information that’s published or deliberately modified to erode trust or inflict harm, such as changing or omitting context, date or time
 - Fabricated or meaningfully manipulated visual or audio content that erodes trust or causes harm
- Harassment and criticism. Pinterest isn’t a place to insult, hurt or antagonise individuals or groups of people. There are good reasons to express criticism, but we may limit the

distribution of or remove insulting content to keep Pinterest a positive, inspiring place; this includes:

- Manipulated images intended to degrade or shame
- Shaming people for their bodies or assumed sexual or romantic history
- Sexual remarks about people's bodies and solicitations or offers of sexual acts
- Criticism involving name-calling, profanity and other insulting language or imagery
- Mocking someone for experiencing sadness, grief, loss or outrage
- We've also put together [some resources \(opens in a new window\)](#) for you to protect yourself."