

# Trust and Safety

## Disinformation at the Wikimedia Foundation:

### Who we are & what we do



"Wikipedia Mini Globe Handheld" by Lane Hartwell, is licensed under [CC BY-SA 3.0](#)

Thank you for reading this!

This document will help you to:

1. Understand the role of the Wikimedia Foundation.
2. Understand the ways in which the Trust and Safety unit of the Wikimedia Foundation works to reduce the exposure of harm to volunteer and reader communities.
3. Learn more about disinformation and what the Foundation is doing to support the communities in tackling disinformation.



"Community" by Jasmina El Bouamraoui and Karabo Poppy Moletsane, is licensed under [CC BY-SA 4.0](#)

## Trust and Safety Team

Trust and Safety (T&S) Team is a unit of the Foundation that focused on keeping the Wikimedia community safe. The unit has several workflows aimed to reduce the exposure of harm to volunteer and reader communities; administers the office action policy; supports communities in addressing misinformation and disinformation in the projects; and develops tools to support the global community in its own self-governance.

The Trust and Safety team works across several departments of the Foundation such as legal, talent and culture, research, public policy, product, learning and evaluation, and engineering to solve problems and develop innovative ways to combat harmful content. We do this through identifying, building and – as appropriate – staffing processes which keep our users safe. We also do this through designing, developing, and executing strategies that integrate legal, product, engineering, research, and learning & evaluation to proactively mitigate risk, as well as manage the overall safety of our online and offline communities.

Among other measures Trust and Safety receives and handles reports of major safety issues on Wikimedia projects, including suicide threats, threats of violence, and child pornography. The team also administers the policies regarding Wikimedia Foundation bans of users from the projects and from Foundation-funded or supported events. T&S also works with other Foundation teams to address concerns about user privacy and freedom that do not necessarily rise to the level of bans.

As a part of the Foundation's commitment to respect and foster community autonomy, the Trust & Safety team does not handle general community or community-member disputes that may be addressed through community processes, nor does it serve as an appeal avenue for community-made policies and decisions. While we are happy to assist community members in need of help, many times that help will consist of assisting the people to find the right community venue to solve their problem.

Similarly, the Foundation does not create or curate content on the sites, and Trust & Safety is not directly involved with monitoring for or removing disinformation. The community has many processes for managing content, and we are here to support them in the methods detailed below.

Trust and Safety works to provide compassionate, credible, and comprehensive Trust and Safety services to the Foundation and the volunteer communities and affiliates it supports.



"Friendship!" by Jasmina El Bouamraoui and Karabo Poppy Moletsane, is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

## Trust and Safety Disinformation Team

The Trust and Safety Disinformation Team works to ensure that the Wikimedia contributor communities are properly supported in combating disinformation and help them in their work to preserve the accuracy of the information resources they create and curate. Disinformation, commonly defined as sharing false or misleading information with the intent to deceive, is an increasing challenge across the world. This has also impacted multiple Wikimedia projects.

In order to empower our volunteers, T&S Disinformation Team offers support in different forms, depending on the needs of the different language communities. The support could be through:

1. Partnering with specific communities in order to learn about their current needs or problems in regard to disinformation threats.
2. Assisting the communities in identifying and, on request, investigating possible disinformation threats.
3. Sharing intelligence with the community volunteer functionaries who participate in our election integrity, disinformation evaluation (Croatian language Wikipedia review), and other cooperation projects with interested local communities.
4. The Trust & Safety Disinformation Team can help coordinate with WMF Communications on press issues if the need arises for a community during disinformation attacks on their wiki.
5. Providing language support to the community volunteers in handling disinformation threats as needed, for example quick Farsi or Hindi language evaluations of sensitive material for functionaries of other wikis who have requested such for their own investigations.
6. Providing disinformation-focused training to the communities.
7. Whenever the community feels it is becoming overwhelmed with disinformation threats against their physical safety or their content integrity, the Foundation through the Trust & Safety Disinformation Team can be directed by WMF Legal to issue office actions on accounts that pose a threat to the community.

## Role of Wikimedia

As a Foundation, we thank you for contributing to Wikipedia, Wikidata, and other Wikimedia Foundation knowledge projects. It is through your efforts, and many others like you, that we host the largest collection of free, collaborative knowledge in human history. Our community of Wikimedians makes it possible for hundreds of millions of people worldwide to access free knowledge. As a Foundation, our role is to empower and engage people around the world to join this community in collecting and developing educational content.

**If you have any questions or concerns, or would like to learn more about the work of our Trust & Safety teams, you can reach out to us at:**

**[ca@wikimedia.org](mailto:ca@wikimedia.org).**



"MediaWiki" by Jasmina El Bouamraoui and Karabo Poppy Moletsane, is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)