

Wikimedia is writing new policies to fight Wikipedia harassment

Wikipedia plans to crack down on harassment and other “toxic” behavior with a new code of conduct. The Wikimedia Foundation Board of Trustees, which oversees Wikipedia among other projects, voted on Friday to adopt a more formal moderation process. The foundation will draft the details of that process by the end of 2020, and until then, it’s tasked with enforcing stopgap anti-harassment policies.

“Harassment, toxic behavior, and incivility in the Wikimedia movement are contrary to our shared values and detrimental to our vision and mission,” said the board in a statement. “The board does not believe we have made enough progress toward creating welcoming, inclusive, harassment-free spaces in which people can contribute productively and debate constructively.”

It will draft a “binding minimum” standard for behavior this year

The trustee board gave the Wikimedia Foundation four specific directives. It’s supposed to draft a “binding minimum set of standards” for behavior on its platforms, shaped by input from the community. It needs to “ban, sanction, or otherwise limit the access” of people who break that code, as well as create a review process that involves the community. And it must “significantly increase support for and collaboration with community functionaries” during moderation. Beyond those directives, the Wikimedia Foundation is also supposed to put more resources into its Trust and Safety team, including more staff and better training tools.

The trustee board says its goal is “developing sustainable practices and tools that eliminate harassment, toxicity, and incivility, promote inclusivity, cultivate respectful discourse, reduce harms to participants, protect the projects from disinformation and bad actors, and promote trust in our projects.”

Wikipedia’s volunteer community can be highly dedicated but intensely combative, launching edit wars over controversial topics and harshly enforcing editorial standards in a way that may drive away new users. The Wikimedia Foundation listed harassment as one factor behind its relative lack of female and gender-nonconforming editors, who have complained of being singled out for abuse. At the same time, the project grew out of a freewheeling community-focused ethos – and many users object to the kind of top-down enforcement you’d find on a commercial web platform.

Wikipedia has struggled with toxicity for years

These problems came to a head last year, when the Wikimedia Foundation suspended a respected but abrasive editor who other users accused of relentless harassment. The intervention bypassed Wikipedia’s normal community arbitration process, and several administrators resigned during the backlash that followed.

The board of trustees doesn’t mention that controversy, saying only that the vote “formalizes years’ of longstanding efforts by individual volunteers, Wikimedia affiliates, Foundation staff, and others to stop harassment and promote inclusivity on Wikimedia projects.” But on a discussion page, one editor cited the suspension to argue that the Wikimedia Foundation shouldn’t interfere with Wikipedia’s community moderation – while others said a formal code of conduct would have reduced the widespread confusion and hostility around it.

Amid all this, Wikipedia has become one of the internet’s most widely trusted platforms. YouTube, for instance, uses

Wikipedia pages to rebut conspiracy videos. That's raised the stakes and created a huge incentive for disinformation artists to target the site. Friday's vote suggests the Wikimedia Foundation will take a more active role in moderating the platform, even if we don't know exactly how.

<https://www.theverge.com/2020/5/25/21269482/wikimedia-foundation-anti-harassment-code-of-conduct-vote>