# Study finds toxicity in the open-source community varies from other internet forums

June 28 2022, by Aaron Aupperlee



Credit: Pixabay/CC0 Public Domain

Trolls, haters, flamers and other ugly characters are, unfortunately, a fact of life across much of the internet. Their ugliness ruins social media networks and sites like Reddit and Wikipedia.

But toxic content looks different depending on the venue, and identifying online toxicity is a first step to getting rid of it.

A team of researchers from the Institute for Software Research (ISR) in Carnegie Mellon University's School of Computer Science recently collaborated with colleagues at Wesleyan University to take a first pass at understanding toxicity on open-source platforms like GitHub.

"You have to know what that toxicity looks like in order to design tools to handle it," said Courtney Miller, a Ph.D. student in the ISR and lead author on the paper. "And handling that toxicity can lead to healthier, more inclusive, more diverse and just better places in general."

To better understand what toxicity looked like in the open-source community, the team first gathered toxic content. They used a toxicity and politeness detector developed for another platform to scan nearly 28 million posts on GitHub made between March and May 2020. The team also searched these posts for "[code of conduct](#)"—a phrase often invoked when reacting to toxic content—and looked for locked or deleted issues, which can also be a sign of toxicity.

Through this curation process, the team developed a final dataset of 100 toxic posts. They then used this data to study the nature of the toxicity. Was it insulting, entitled, arrogant, trolling or unprofessional? Was it directed at the code itself, at people or someplace else entirely?

"Toxicity is different in open-source communities," Miller said. "It is more contextual, entitled, subtle and passive-aggressive."

Only about half the toxic posts the team identified contained obscenities. Others were from demanding users of the software. Some came from users who post a lot of issues on GitHub but contribute little else. Comments that started about a software's code turned personal. None of

the posts helped make the [open-source software](#) or the community better.

"Worst. App. Ever. Please make it not the worst app ever. Thanks," wrote one user in a post included in the dataset.

The team noticed a unique trend in the way people responded to toxicity on open-source platforms. Often, the project developer went out of their way to accommodate the user or fix the issues raised in the toxic content. This routinely resulted in frustration.

"They wanted to give the benefit of the doubt and create a solution," Miller said. "But this turned out to be rather taxing."

Reaction to the paper has been strong and positive, Miller said. Open-source developers and community members were excited this research was happening and that the behavior they had been dealing with for a long time was finally being recognized.

"We've been hearing from developers and [community members](#) for a really long time about the unfortunate and almost ingrained toxicity in open-source," Miller said. "Open-source communities are a little rough around the edges. They often have horrible diversity and retention, and it's important that we start to address and deal with the toxicity there to make it a more inclusive and better place."

Miller hopes the research creates a foundation for more and better work in this area. Her team stopped short of building a [toxicity](#) detector for the [open-source](#) community, but the groundwork has been laid.

"There's so much work to do in this space," Miller said. "I really hope people see this, expand on it and keep the ball rolling."

Joining Miller on the work were Daniel Klug, a systems scientist in the ISR; ISR faculty members Bogdan Vasilescu and Christian Kästner; and Sophie Cohen of Wesleyan University. The team's paper was presented at the ACM/IEEE International Conference on Software Engineering last month in Pittsburgh.

**More information:** Paper: [Did You Miss My Comment or What?" Understanding Toxicity in Open-Source Discussions](#)

Provided by Carnegie Mellon University