

Facebook bots to combat bad behavior

July 24 2020, by Peter Grad



Credit: CC0 Public Domain

As the world's largest social network, Facebook provides endless hours of discussion, entertainment, news, videos and just good times for the more than 2.6 billion of its users.

It's also ripe for malicious activity, bot assaults, scams and hate speech.

In an effort to combat bad behavior, Facebook has deployed an army of bots in a simulated version of the social network to study their behavior and track how they devolve into antisocial activity.

Drawing principles from machine learning, artificial intelligence, game theory and multiagent systems, Facebook engineers developed the program—Web-Enabled Simulation (WES)—a highly realistic, large-scale replica of Facebook.

They hope to remedy the explosive growth of online harassment, especially in this era of political misinformation, crackpot conspiracy theories and hate speech.

WES bots are trained to interact with one another, sending messages, commenting on posts and making friend requests. They cannot interact with actual users.

Mark Harman, the lead Facebook research scientist who posted a summary of the WES effort on a blog Thursday, explained, "The WES approach can automatically explore complicated scenarios in a simulated environment. While the project is in a research-only stage at the moment, the hope is that one day it will help us improve our services and spot potential reliability or integrity issues before they affect real people using the platform."

Millions of bots with differing objectives can be deployed in the experimental system. Some, for instance, will attempt to purchase items that are not permitted on the site, such as guns or drugs. Researchers will track patterns applied by the bots as they conduct searches, visit pages and replicate actions that humans might take.

They will then assess various counter-measures to see which most effectively stop or even prevent the undesirable behaviors.

Harman compared their approach to that of traffic engineers seeking means to create safer roadways. To curb speeders, for instance, city planners may install more stop signs. If that measure is insufficient, road

bumps may be installed. With Facebook behaviors, countermeasures could include limiting frequency of commentary on posts or applying fact-checking to questionable conspiracy posts.

"We apply 'speed bumps' to the actions and observations our bots can perform," Harman said, "and so quickly explore the possible changes that we could make to the products to inhibit harmful behavior without hurting normal behavior," said Harman.

The project is in research stages. No changes to the actual Facebook platform have been made yet, though researchers say modifications will be introduced as soon as they confirm remedial bot behaviors mimic human behavior with a high enough degree of accuracy and reliability.

WES is affectionately referred to as dub-dub. The nickname is derived from an abbreviation of the pronunciation of "WW," ("dub" instead of "double-u") itself a truncated version of "WWW," or World Wide Web. WES is an abbreviated version of the World Wide Web, hence, it is referred to with only two W's.

The experiment may recall a recent project abandoned by Facebook when an unexpected turn of events occurred. In a 2017 experiment that sought to teach chatbots how to negotiate with one another, researchers left two chatbots alone to see how their negotiations would progress. But when the Facebook AI Research Lab researchers returned, they were stunned to discover that the chatbots, Alice and Bob, went off-script and actually invented a new language.

The researchers, perhaps bearing in mind the stark warning of noted physicist Stephen Hawking that AI could "take off on its own and redesign itself at an ever increasing rate" and "could spell the end of the human race," abruptly ended the experiment.

More information: [ai.facebook.com/blog/a-faceboo ... ct-harmful-behaviors](https://ai.facebook.com/blog/a-facebook-bot-combat-harmful-behaviors)

© 2020 Science X Network

Citation: Facebook bots to combat bad behavior (2020, July 24) retrieved 25 April 2024 from <https://techxplore.com/news/2020-07-facebook-bots-combat-bad-behavior.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.