Google rolls out AI tool to combat online trolls

February 24 2017

Google has developed a programming tool called Perspective to help editors trying to moderate discussions by filtering out abusive 'troll' comments

Google said it will begin offering media groups an artificial intelligence tool designed to stamp out incendiary comments on their websites.

The programming tool, called Perspective, aims to assist editors trying to moderate discussions by filtering out abusive "troll" comments, which Google says can stymie smart online discussions.

"Seventy-two percent of American internet users have witnessed harassment online and nearly half have personally experienced it," said Jared Cohen, president of Google's Jigsaw technology incubator.

"Almost a third self-censor what they post online for fear of retribution," he added in a blog post on Thursday titled "When computers learn to swear".

Perspective is an application programming interface (API), or set of methods for facilitating communication between systems and devices, that uses machine learning to rate how comments might be regarded by other users.

The system, which will be provided free to media groups including social media sites, is being tested by The Economist, The Guardian, The New York Times and Wikipedia.

Many news organisations have closed down their comments sections for lack of sufficient human resources to monitor the postings for abusive content.

"We hope we can help improve conversations online," Cohen said.

Google has been testing the tool since September with The New York Times, which wanted to find a way to maintain a "civil and thoughtful" atmosphere in reader comment sections.

Perspective's initial task is to spot toxic language in English, but Cohen said the goal was to build tools for other languages, and which could identify when comments are "unsubstantial or off-topic".

Twitter said earlier this month that it too would start rooting out hateful messages, which are often anonymous, by identifying the authors and

prohibiting them from opening new accounts, or hiding them from internet searches.

Last year, Google, Twitter, Facebook and Microsoft signed a "code of good conduct" with the European Commission, pledging to examine most abusive content signalled by users within 24 hours.

More information: jigsaw.google.com/projects/#perspective

© 2017 AFP

Citation: Google rolls out AI tool to combat online trolls (2017, February 24) retrieved 17 May 2024 from https://phys.org/news/2017-02-google-ai-tool-combat-online.html