Reddit says it's 'impossible' to consistently enforce liberal speech rules 4

A leaked discussion reveals the difficult problem of training AI to detect and define slurs

By Shoshana Wodinsky

SHARE

A private chat thread released this weekend reveals that banning hate speech on Reddit is apparently too difficult a job to even attempt.

The news, <u>first spotted by the *Huffington Post*</u>, comes from a leaked conversation between Reddit CEO Steve Huffman and Reddit user Zachary Swanson, known as "whatllmyusernamebe" on the site. Swanson asked Huffman to reconsider his permissive stance on hate speech on the site, pointing out that <u>Reddit's rules</u> <u>already ban "violent" speech</u>, which the site defines as "content that encourages, glorifies, incites, or calls for violence or physical harm against an individual or a group of people."

In response, Huffman argued that hate speech is "difficult to define," adding that enforcing a total ban on hate speech is "a nearly impossible precedent to uphold," and "impossible to enforce consistently." After speaking with Huffman, Swanson posted screenshots of his conversation in a pair of subreddits, including r/stopadvertising, which is dedicated to curbing hate speech on the site. After he did, Swanson received an email saying he was suspended from the site for a week on charges of harassment.

""THERE'S A REASON WHY IT'S NOT REALLY DONE.""

