MediaWell

News Item

Tackling online hate speech one emoji at a time | Oxford Internet Institute

By Hannah Rose Kirk
August 24, 2021

> New research from the Oxford Internet Institute (OII) exposes critical problems in how emoji-based online hate is tackled, with OII researchers uncovering critical weaknesses in how Artificial Intelligence (AI) systems detect hate speech involving emoji. Lead author of the study, OII DPhil researcher Hannah Rose Kirk explains more.
>
> Rise of online hate
>
> Social media platforms have opened up unprecedented channels of communication. While this greater communication brings about many benefits, it has also allowed for an expansion in the scope and virality of online harms. The volume of hate shared online far exceeds what human moderators can feasibly deal with, and AI content moderation systems have the potential to relieve the burden placed on these moderators. However, humans are creative in the way they express hate and the diversification in modalities of hate (such as the use of emoji) outpace what AI systems can understand.
>
> […]

Source: Tackling online hate speech one emoji at a time | Oxford Internet Institute