MediaWell

News Item

Facebook is overhauling its hate speech algorithms | The Washington Post

By Elizabeth Dwoskin, Nitasha Tiku, and Heather Kelly
December 3, 2020

> Facebook is embarking on a major overhaul of its hate speech algorithms, reversing years of so-called “race-blind” policies.
>
> Those practices resulted in the company being more vigilant about removing slurs lobbed against White users while flagging and deleting innocuous posts by people of color on the platform.
>
> The overhaul, which is known as the WoW Project and is in its early stages, involves re-engineering Facebook’s automated moderation systems to get better at detecting and automatically deleting hateful language that is considered “the worst of the worst,” according to internal documents describing the project obtained by The Washington Post. The “worst of the worst” includes slurs directed at Blacks, Muslims, people of more than one race, the LGBTQ community and Jews, according to the documents.
>
> […]

Source: Facebook is overhauling its hate speech algorithms | The Washington Post