MediaWell

News Item

Content Moderation Tools to Stop Extremism | Lawfare

By Daniel Byman
September 29, 2022

> Technology companies are more active than ever in trying to stop terrorists, white supremacists, conspiracy theorists, and other hateful individuals, organizations, and movements from exploiting their platforms, but government and public pressure to do more is growing. If companies decide to act more aggressively, what can they do? Much of the debate centers around whether to remove offensive content or leave it up, ignoring the many options in between. This paper presents a range of options for technology companies, discussing how they work in practice, their advantages, and their limits and risks. It offers a primer on the many choices available and then discusses the numerous trade-offs and limits that affect the various approaches.
>
> Broadly speaking, the actions companies can take fall into three categories. First, they can remove content, deleting individual posts, deplatforming users or even entire communities, and otherwise simply removing offensive and dangerous content. Second, they can try to reshape distribution—reducing the visibility of offensive posts, downranking (or at least not promoting) certain types of content such as vaccine misinformation, and using warning labels—and otherwise try to reduce or limit engagement with certain material but allow it to stay on their platforms. Finally, companies can try to reshape the dialogue on their platforms, empowering moderators and users in ways that make offensive content less likely to spread.
>
> […]

Source: Content Moderation Tools to Stop Extremism – Lawfare