

TWITTER'S SCORECARD IN ADDRESSING VIOLENCE AND ABUSE AGAINST WOMEN ONLINE

CATEGORY	SUBCATEGORY	RECOMMENDATION	SCORE
TRANSPARENCY	Disaggregation	Improve the quality and effectiveness of transparency reports by disaggregating data along types of abuse, geographic region, and verified account status.	WORK IN PROGRESS
	Content Moderators	Increase transparency around the content moderation process by publishing data on the number of moderators employed, the types of trainings required, and the average time it takes for moderators to respond to reports.	NOT IMPLEMENTED
	Appeals	Increase transparency around the appeals process by publishing the volume of appeals received and outcomes of appeals.	NOT IMPLEMENTED
REPORTING MECHANISMS	Feature request	Develop more features to gather and incorporate feedback from users at every stage of the abuse reporting process, from the initial report to the decision.	WORK IN PROGRESS
	Appeals	Improve the appeals process by offering more guidance to users on how the process works and how decisions are made.	IMPLEMENTED
	Public campaign	Continue to educate users on the platform about the harms inflicted upon those who fall victim to abuse through public campaigns and other outreach efforts. This should include sending a notification/ message to users who are found to be in violation of Twitter's rules about the silencing impact and risk of mental health harms caused by sending violence and abuse to another user online.	WORK IN PROGRESS
ABUSE REPORT REVIEW PROCESS	Transparency	Provide clearer examples of what types of behavior rise to the level of violence and abuse and how Twitter assesses penalties for these different types of behavior.	WORK IN PROGRESS
	Automation	Automation should be used in content moderation only with strict safeguards, and always subject to human judgment. As such, Twitter should clearly report out on how it designs and implements automated processes to identify abuse.	NOT IMPLEMENTED
PRIVACY & SECURITY FEATURES	Feature request	Provide tools that make it easier for users to avoid violence and abuse on the platform, including shareable lists of abusive words and other features tailored to the specific types of abuse a user reports.	WORK IN PROGRESS
	Public campaign	Educate users on the platform about the privacy and security features available to them through public campaigns and other outreach channels and make the process for enabling these features as easy as possible.	WORK IN PROGRESS