**United Nations High Commissioner for Human Rights**

**Volker Türk**

**“Regulating online hate speech to avoid harm is not censorship.”**

Meta’s announcement that it is overhauling its policies on content moderation and fact-checking should alarm everyone who is concerned about freedom of expression. While these moves purport to protect free speech, they strike against it for many people and communities.

Meta announced it is recalibrating its automated content moderation to limit only so-called high-severity violations, such as those relating to terrorism. Taken together with policies adopted by other companies, including Telegram and X (formerly known as Twitter), this is likely to lead to far more abusive and hate-filled content on several of the world’s largest social media platforms.Some of this content will target marginalized communities, including LGBTIQ+ people, refugees, migrants, and minorities of all kinds. In doing so, it will drive away those communities from these platforms, limiting their visibility, isolating them further, and reducing their freedom of expression.

More broadly, these changes will cause harm far beyond specific people and groups. Freedom of expression requires not only that people should be able to express their views, but that they should also be able to seek and receive ideas and information. Poorly regulated social media platforms curtail this freedom in several ways. By silencing some people and communities, they restrict the range of information available to everyone. And by allowing the proliferation of lies and disinformation, they poison the information environment. They blur the line between fact and fiction, fracturing societies and eroding much needed public space for open debate, based on facts and basic common understanding.

The substantially increased volume of unmoderated content and hate speech that will result from these changes will be harmful at all times. But it could have particularly devastating consequences during conflicts, crises, and election campaigns, affecting hundreds of millions of people regardless of whether they are users of these platforms. I have just returned from Syria where I witnessed the toxicity of vengeful disinformation that leads to violence.

Indeed in 2018, Facebook itself recognized that it had not done enough to prevent its platform from being used to foment division and incite violence against the Rohingya in Myanmar. It acknowledged that it can and should do better. There are other similar examples. Election campaigns across the world have also shown that a lack of rights-based governance of social media environments can fray social cohesion and distort democratic decision-making. Over the past few years in election campaigns from Brazil, to Kenya, to Moldova, to Romania, there have been reports of disinformation and hateful content spreading on social media platforms. In such circumstances, it is crucial for companies to undertake human rights due diligence, and for States to ensure that debate spaces online and off remain free and open to all.

Content moderation is not easy, and it can be contentious. My Office has sounded the alarm in cases of over-enforcement, for example, when States have used blunt laws and policiesto silence dissenting voices and suppress unwelcome material in the public domain. And, to cite one example, several civil society organisations recently documented the suppression of material on Palestinian rights across social media platforms.

But carefully regulating online hate speech and moderating content to avoid real harm is not censorship. It is an essential plank of information integrity in the digital age – and the responsibility of social media platforms.

 The international community already has a framework to guide us through these issues: international human rights law. This agreed body of norms and standards is aimed precisely at protecting all freedoms, for everyone, while preventing incitement to hatred and violence. It is universal, it is dynamic, and it can adapt to emerging issues.

Human rights are not up for debate or redefinition. Our freedom of expression was hard won, through years of protests against censorship and oppression. We must be vigilant in safeguarding it. That means addressing incitement to hatred and violence head-on when it breaches the law, and protecting everyone’s right to access information, so that people can seek and receive ideas from a full range of diverse sources. The lessons of recent years are clear: platforms that are blind to violence and that ignore the threats faced by journalists and human rights defenders will inevitably fall short and undermine free speech.

Effective content governance must prioritize transparency, accountability, and provide for the ability to challenge content moderation decisions. It needs to consider context, nuance of local language, and who is controlling the content and its distribution. In short, it should take into account the broader information environment.

The effective governance of online content can only emerge from open, ongoing, and well-informed debates across society. My Office will continue to call and work for accountability in the digital space, in line with human rights law. Human rights must serve as our guiding compass to safeguard public discourse, build trust, and protect the dignity of all.