U.S. Commission on International Religious Freedom Hearing

<u>Online Hate Speech and Disinformation</u>
<u>Targeting Religious Communities</u>

<u>Opening Remarks as prepared for delivery</u>

Tony Perkins<u>, USCIRF Vice Chair</u>

Thank you very much, Chair Manchin. I would like to join in welcoming you all to today's hearing.

To enforce the community standards that Chair Manchin outlined, social media companies rely on a combination of artificial intelligence, or AI, and human analysts to wade through content to identify and remove hateful statements or disinformation prohibited in community standards. Notably, disinformation is not always removed, but instead may be downgraded or corrected. As identifying hate speech involves a great deal of nuance, context, and linguistic expertise, relying on

machines to recognize it remains a challenge. Social media companies have also struggled with having enough content moderators who speak local dialects and have the expertise needed to proactively identify hate speech around the globe. A recent audit of Facebook noted that their content moderation efforts remain "too reactive and piecemeal," and hateful content continues to spill through the cracks. Critics of current content moderation efforts have urged social media companies to move away from their current "whack a mole" approach and develop early warning policies that preventively flag situations where violence and atrocities are likely to occur. There is also concern that the overreliance on content removal can lead to online censorship that restricts fundamental freedoms and drives extremist views.

Around the globe, the spread of false or misleading information through social media is causing real harm by catalyzing violence and brutality towards religious communities. For example, in India, where Facebook has more users than any other country globally, its WhatsApp platform is used to spread hate speech and false information against religious minorities. In Pakistan, Facebook is used to target, frame, and accuse individuals of blasphemy, leading to detention, disappearances, extrajudicial killings, mob gatherings and public lynchings.

Government-sponsored hate speech and disinformation is particularly perilous as it fosters a dangerous culture of hate and religious intolerance, where both online and offline abuses are condoned. The Russian Federation employs a sophisticated disinformation network that targets religious minorities with sensational allegations designed to create fear and animosity against them. Jehovah's Witnesses are depicted in state media as dangerous and subversive, often with ties to western interests. Government news programs accuse religious minorities of ties to revolutionaries in neighboring Ukraine and depict peaceful Muslim groups as terrorists. In May 2019, a close advisor to President Putin published an op-ed claiming Americans and Israelis were plotting with Ukraine's Jewish President Zelensky to deport ethnic Russians from Eastern Ukraine and replace them with Jews.

In Iran, the government uses social media to spread anti-Baha'i propaganda while systematically harassing and jailing members of that community on the basis of their faith. Iran's Supreme Leader, Ayatollah Khamenei regularly tweets antisemitic vitriol from his official Twitter account, while restricting Twitter access for his own citizens. Iran's security apparatus regularly uses Instagram and Telegram to threaten members of Iran's Sufi community and followers of spiritualist Mohammed Ali Taheri with physical harm.

I will now turn to Vice Chair Bhargava to further explain what is being done by social media companies in response to online hate speech.