

Call for Public Comments

Submission by FORUM-ASIA

10th December 2025

The Asian Forum for Human Rights and Development (FORUM-ASIA) is a membership-based regional network of human rights organisations in Asia. Our collective strength draws from our extensive network of 86 member organisations across 23 countries, mainly in Asia.

We promote and protect human rights, civic space, and democracy in the region. We do this by empowering civil society and marginalised communities. FORUM-ASIA holds a consultative status with the UN Economic and Social Council as well as a consultative relationship with the ASEAN Intergovernmental Commission on Human Rights.

The human rights situation in Asia has been marred by shrinking civic space, repression of freedom of expression and the targeting of those who challenge the status quo. In online spaces, the spread of mis/disinformation and hate speech, and the harassment of human rights defenders has exponentially grown in the past few years. In this context, we strongly believe that companies such as Meta play a critical role in protecting people's fundamental freedoms online, and addressing any violations in a fair, equitable and transparent manner.

FORUM-ASIA would like to submit the following comments based on Meta's call for public comments for your consideration:

1. Factors to consider when adopting community-driven moderation, and how it could help or worsen misleading information
 - **Freedom of expression:** Will the system expand the public's ability to access contextual information, or will it disproportionately silence or bury views (especially minority or dissenting voices)?
 - **Non-discrimination and equality:** Could socially marginalized groups be excluded from contributor pools or targeted by notes?
 - **Safety, harm and retaliation:** Could contributors be exposed to state surveillance, reprisals, doxxing, or online/offline harm?
 - **Legal risks to contributors:** Are there criminal laws (e.g., sedition, defamation, hate speech, terrorism) or broad emergency powers that could criminalize note writing or contextualization?
 - **Likelihood of state capture:** Will governments try to manipulate contributor pools (e.g., coordinated volunteers, legal pressure on platforms)?

- **Local language coverage:** Could notes that don't reflect local languages/cultural framings be incomplete, ineffective or misleading?
 - **Representation across viewpoints, regions, genders, ethnicities, and expertise:** Is the contributor base broad enough to represent diverse viewpoints, or is it at risk of bias?
 - **Appeals and remediation:** Can content authors challenge notes? Is there an independent appeals path for violations?
 - **Stop/rollback criteria:** Are there clear thresholds (e.g., systemic harassment of contributors, sustained legal risk etc.) that justify pausing or reversing the launch if it is not working as intended?
2. The ability and limitations of consensus-based algorithms to function effectively in different political climates, especially where audiences are divided or information environments are weak
- **Potential for self-censorship:** In restrictive political climates, contributors may avoid participating out of fear, leading to misleading consensus and weakened correction capacity.
 - **Polarisation affecting consensus formation:** In highly divided political climates, cross-perspective agreement becomes harder to reach, meaning fewer notes will be displayed and harmful content may go uncorrected.
 - **Risk of majority or dominant-group capture:** In contexts where one political bloc or demographic dominates online spaces, "consensus" may simply reflect the views of the powerful, marginalizing minority voices.
 - **Susceptibility to coordinated manipulation:** Consensus mechanisms can be gamed by organised political actors or disinformation networks able to flood the system with aligned contributors.
 - **Recognising that consensus is not the same as accuracy:** Algorithms based on agreement may certify misleading or biased interpretations as "consensus" if coordinated networks or echo chambers reinforce one another.
 - **Reduced participation because of low civic trust:** In weak information environments where mistrust of institutions is high, fewer users may participate as contributors, skewing the diversity and reliability of consensus signals.
 - **Limited effectiveness in contexts with poor media literacy:** Where users struggle to distinguish reliable sources, consensus can normalize misconceptions rather than correct them.

3. Meta's human rights responsibilities when expanding, scaling back, or ending products and programs that address misleading information
 - **Duty to prevent harm:** Meta has a responsibility to assess and mitigate foreseeable risks to user rights - especially freedom of expression, safety, non-discrimination, and access to information - before expanding, reducing, or withdrawing tools that counter misleading content.
 - **Meaningful human rights due diligence:** Any expansion or rollback must be guided by robust, ongoing human rights impact assessments that consider local political contexts, risks to vulnerable groups, and potential unintended consequences. These assessments should be independent and accessible to the public to ensure transparency.
 - **Non-discrimination and equity across regions:** Meta must avoid creating unequal protections where users in some countries receive safeguards against misinformation while others - often in the Global Majority - are left exposed. Decisions to expand or limit coverage should not reinforce existing inequities in civic or information spaces.
 - **Transparency and access to information:** When scaling back or ending programmes, Meta has a responsibility to provide clear, accessible explanations about why a product is being changed, what risks were assessed, and what alternative safeguards will remain. Sudden withdrawals undermine user trust and can worsen harmful information flows.
 - **Accountability and access to remedy:** Users, human rights defenders, and affected communities should have channels to raise concerns, appeal adverse impacts, and receive timely responses. Ending or cutting back a product does not remove Meta's obligation to remedy harm if misinformation spikes or if communities are disproportionately affected.
 - **Protection of contributors:** If changes to a product increase risks to those who generate or rely on contextual information (e.g., Community Notes contributors, human rights defenders, journalists etc.), Meta must implement additional safeguards such as anonymity and emergency protocols, or reconsider the decision entirely.
4. The effectiveness of interventions beyond content removal, such as fact-checking, labelling, reduced distribution, added friction, and user-generated context, and how to avoid bias in these approaches
 - **Fact-checking may enhance accuracy but depends on local capacity:** Independent fact-checking can be a good way to correct information without censoring speech, but its effectiveness is limited if fact-checkers are biased, lack linguistic reach, face political pressure, fear reprisals, or cannot keep pace with high-volume misinformation.

- **Labelling and contextual warnings may help users make informed choices if bias is accounted for:** Clear, well-designed labels (e.g., “missing context,” “disputed claim” etc.) can significantly reduce belief in false content, especially when they explain why the content is misleading. However, labels must be applied consistently to avoid bias.
5. Country-level factors that influence the functioning of social media products and how these factors can be studied or measured
- **Strength of civic space and human rights protections:** Restrictions on expression, press freedom, or civil society influence user behaviour and the safety of contributors.
 - **Risk of state interference, surveillance, and censorship:** High surveillance or authoritarian contexts increase risks for users and contributors and can further distort platform behaviour.
 - **Political stability and level of polarisation:** Highly polarised or unstable political environments affect how users interpret information and how easily products can be weaponised.
 - **Media literacy and information ecosystem health:** Countries with weak independent media or low digital literacy are more vulnerable to mis/disinformation and manipulation.
 - **Prevalence of online harassment and coordinated abuse:** High levels of trolling, gendered harassment, or coordinated political attacks reduce participation in community-driven features and distort consensus.

FORUM-ASIA strongly believes that there needs to be comprehensive safeguards in place to prevent and address any human rights violations that may occur. We hope that these questions and suggestions will be taken into consideration when determining the expansion of the Community Notes programme.