**Pakistan Political Candidate Accused of Blasphemy**

**2024-031-FB-MR**

**Submission: Anonymous**

**Background:**

The February 2024 elections in Pakistan witnessed an innovative use of social media and artificial intelligence to reach and influence the vote of Pakistani citizens. One of Pakistan’s leading political parties, although disqualified by the Elections Commission Pakistan (ECP) at the time, used AI generated audio messages in their leader’s voice as part of their political campaign for the upcoming elections. Social media campaigning and discourse also brought forward an opportunity to spread both misinformation and disinformation through the use of deepfakes, generative AI content and altered/doctored audio/video clips. In addition, political party supporters used social media to garner mass support while also attempting to spread discontent against other candidates by indicating why they should not be elected. These posts ranged from showcasing the candidate’s or party’s incompetence, questioning their sexuality, passing misogynistic remarks to writing on sensitive topics such as personal religious beliefs and affiliations.

While the extent of freedom of expression in general remains to be a phenomenon with significant gray area, religion has historically been a sensitive topic in Pakistan. Blasphemy laws are the most crucial aspect of freedom of religious expression as penalties include severe outcomes such as death sentences. In addition to minority groups being under the public eye, journalists, political figures and civil society organizations remain at greater risk of being targeted. Previously, public figures have been accused of blasphemy by rivals, religious groups and general public alike resulting in arrests, violence, assassinations under the guise of vigilante/mob justice, and legal consequences. In 2011, the Governor of Punjab, [Salman Taseer](https://www.bbc.com/news/world-south-asia-12111831), was shot dead by his bodyguard, Mumtaz Qadri, as a result of his opposition to Pakistan’s blasphemy law. During a media interaction, Taseer had spoken in favor of pardoning Asia Bibi, a Christian woman accused of blasphemy. While [Qadri](https://www.dawn.com/news/1302289) was executed and hanged by the state, his act was lauded by conservative religious groups. Upon being hanged, he received a mass funeral and his grave was later turned into a shrine by his sympathizers. In 2023, as a result of a [blasphemy accusation](https://www.aljazeera.com/news/2023/8/23/pakistan-christians-in-fear-after-blasphemy-allegations-trigger-violence), 21 churches and 100s of houses belonging to the Christian community were burnt down in Jaranwala. Social media platforms were used to encourage and instigate violence against the accused individuals and their community. Additionally, in 2022, [Imran Khan](https://www.dawn.com/news/1753759) was accused of blasphemy by political rivals as a result of speaking in support of the Ahmadiyya community. Blasphemy laws and accusations being used against political rivals as a tool to discredit and malign them is alarming considering that the grave consequences of such accusations are not hidden from anyone. The weaponization of blasphemy allegations and laws to sideline political rivals is an issue that needs immediate attention as it puts electoral and political processes at greater risk of being altered and manipulated in terms of public opinion. In Addition, Blasphemy accusations pose a significant risk of individual safety as previous examples show cases of mob violence, killings and imprisonment without fair trial. In 2003, [Munawar Mohsin](https://go.gale.com/ps/i.do?id=GALE%7CA105114233&sid=sitemap&v=2.1&it=r&p=AONE&sw=w&userGroupName=edgewood_oscar&aty=ip) was subjected to life imprisonment for writing a supposedly blasphemous article.

**Meta’s Responsibilities and Policy Implications**

**Coordinating Harm and Promoting Crime Policy**

The Coordinating Harm and Promoting Crime Policy at Meta aims to safeguard at risk individuals through the removal of content that has the potential of inciting violence or subjecting them to persecution. Pakistan is a country where blasphemy allegations are a grave concern with the potential to result in serious offline harms, making this a vital policy. However, it is challenging to maintain a balancing act in terms of harmful content being removed without hampering freedom of expression particularly during timelines involving crucial events such as general elections 2024.

Taking this into account, it is important that Meta adopts a more nuanced approach in decision-making regarding content moderation by ensuring an in-depth understanding of not only the local realities but also the expected outcomes of blasphemy allegations. Meta needs to approach content related to blasphemy and blasphemy allegations through an intersectional lens and strategically remove content that clearly risks an individual’s safety by exposing them to violence and/or persecution without limiting important political discourse.

**Human Rights Responsibilities and Government Requests**

In Pakistan, strict blasphemy laws deepen the religious and sectarian divide. In such a situation, when Meta relies on local laws in terms of policy-making, it struggles to align with the fundamentals of the International human rights frameworks. Adhering to local laws in policy-making results in further legitimizing these laws despite being oppressive in nature while also increasingly marginalizing vulnerable groups/communities. This is a consequence of the blasphemy law frequently being used as an exploitative political tool to retain monopolistic power dynamics in society.

Meta has a responsibility to align its platform and community guidelines with the international human rights framework and other international frameworks such as the ICCPR, the Rabat Plan of Action and the UDHR. Upon receiving government requests for content removal, it is vital that Meta assesses the request through thorough and robust mechanisms to ascertain that the removal is not in violation of human rights by being lawful, necessary and proportionate. Meta needs to remain transparent and hold itself accountable under international frameworks to safeguard individual rights and retain public trust. In order to achieve this, Meta should conduct comprehensive human rights impact assessments through introducing human reviewers in greater capacity within the system. It is important to provide individual attention to each post instead of relying solely on content being flagged and removed by the algorithm to ensure thorough assessment and provide clear justifications for the removal of any content.

**Compliance with International Human Rights Frameworks:**

Examining Meta's content moderation policies through the lens of international human rights frameworks, particularly the International Covenant on Civil and Political Rights (ICCPR), it's essential to consider [**Article 19**](https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights), which upholds the right to freedom of expression. This encompasses the freedom to seek, receive, and impart information and ideas, albeit with certain responsibilities and limitations. These restrictions, as outlined in Article 19, must be lawful, necessary to respect the rights of others or protect national security, public order, public health, or morals. In the case of Pakistan, where strict blasphemy laws are in place, Meta must ensure its content moderation policies align with global legal standards. This requires transparency in the formulation and application of these policies to ensure they are clearly based on existing laws and understandable to users. Moreover, any restrictions should be proportionate to the intended protection, striking a balance between preserving freedom of expression and preventing harm.

[**Article 20(2)**](https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights)of the ICCPR prohibits the advocacy of religious hatred that leads to discrimination, hostility, or violence. Given the sensitive nature of blasphemy accusations in Pakistan, Meta must carefully navigate the removal of potentially inciteful content while upholding legitimate religious and political discourse. Achieving this balance necessitates a nuanced understanding of the socio-political context and thorough evaluation of content and intent.

[**The Rabat Plan of Action**](https://www.ohchr.org/en/freedom-of-expression) provides a structured approach to assessing the necessity and proportionality of content restrictions concerning incitement to discrimination, hostility, or violence. Meta can apply this framework by: contextualizing content, considering the speaker's identity and intent, analyzing the content itself, assessing its reach, and evaluating the likelihood of harm. This multifaceted approach will enable Meta to make informed decisions tailored to the specific dynamics of blasphemy-related content.

**UN Special Rapporteurs' resolutions** can help Meta gauge a more nuanced understanding of the need to protect vulnerable groups from hate speech and violence while safeguarding against the misuse of blasphemy laws for political suppression. Meta’s policies should be people centric where the priority should be protecting users from credible threats while ensuring blasphemy allegations are not weaponized to silence dissent. Transparency in Meta’s content moderation practices is cardinal to prevent the misuse of blasphemy laws and uphold the principles of freedom of expression.

Under the [**Universal Declaration of Human Rights (UDHR)**](https://www.un.org/en/about-us/universal-declaration-of-human-rights), Meta's content moderation strategy should respect freedom of expression while also protecting individuals from harm and reputational attacks. This entails allowing political discourse while intervening when necessary to prevent violence or discrimination, particularly in cases involving blasphemy.

As the digital landscape continues to evolve, Meta finds itself at a pivotal moment in its mission to cultivate a safe and inclusive online environment. The complexities surrounding content moderation present pressing challenges, particularly in regions like Pakistan where issues such as blasphemy accusations require careful navigation. It's evident that proactive steps are urgently needed to bolster content moderation and engagement. To address these challenges, Meta must embrace a comprehensive approach that integrates advanced AI technologies alongside human oversight, clear guidelines, and transparent practices. Prioritizing the delicate balance between freedom of expression and user safety is paramount in this endeavor. By doing so, Meta can fortify its content moderation infrastructure while steadfastly upholding fundamental human rights principles. Furthermore, fostering collaborative partnerships with local stakeholders and implementing robust community engagement initiatives will play a pivotal role in crafting context-sensitive strategies and nurturing mutual understanding. Below, we outline actionable recommendations for Meta to effectively navigate these challenges and ensure sustainable progress.

**Recommendations**

**Enhancing Content Moderation and Engagement**

* To bolster content moderation and engagement, it's imperative to fuse advanced AI capabilities with human oversight. We propose enhancing Meta's HERO system through sophisticated AI algorithms adept at discerning context-specific risks associated with blasphemy accusations. This involves integrating natural language processing (NLP) models trained on local languages and dialects to accurately detect nuanced blasphemous content and differentiate it from legitimate discourse.
* Moreover, human oversight by regional experts is paramount. By assembling a dedicated team of trusted partners, legal scholars, human rights activists, and local cultural analysts, Meta can ensure thorough reviews of flagged content. These experts would consider the socio-political context and potential repercussions of blasphemy accusations in Pakistan, thereby enhancing the accuracy and fairness of content moderation decisions.
* Real-Time Moderation capabilities are also crucial. By developing the capacity to swiftly identify and mitigate the spread of harmful content, particularly during critical periods such as elections or after high-profile blasphemy accusations, Meta can effectively curb the dissemination of damaging material.

**Clearer Guidelines and Transparency**

* Refinement of Content Moderation Guidelines is essential. Updating these guidelines to provide clear, specific criteria for handling blasphemy-related content, informed by international human rights standards and local legal requirements, will enhance consistency and fairness in moderation practices.
* Regular Transparency Reports are likewise imperative. By publishing detailed reports outlining actions taken on blasphemy-related content, including data on removals, reasons for removal, and appeals outcomes, Meta can foster transparency and accountability. Case studies illustrating decision-making processes would further elucidate moderation practices.
* User Education is pivotal. Conducting awareness campaigns to educate users about Meta's content policies, and the significance of reporting harmful content responsibly will promote responsible online behavior and contribute to a safer online community.

**Collaborative Frameworks**

* Strengthening relationships with trusted partners and acting on their escalations quickly. Establishing partnerships with local civil society organizations, human rights defenders, and legal experts is key. By collaborating with these stakeholders to develop context-sensitive approaches to content moderation and regularly consulting them to stay abreast of evolving political and legal landscapes, Meta can ensure that its moderation practices align with community needs and values.

**Intersectional Analysis**

* Comprehensive intersectional analyses are essential to understand the broader and localized impacts of content moderation decisions. By considering factors such as gender, ethnicity, religion, and political affiliation, Meta can ensure that its moderation practices are equitable and do not disproportionately affect any group.
* Developing and implementing risk mitigation strategies based on intersectional analyses will further prevent potential offline harm resulting from online content, particularly during politically sensitive times. Robust post-incident reviews will enable Meta to continually refine its approach.

**Human Rights and Due Diligence**

* Conducting thorough human rights impact assessments before removing content, especially in cases involving blasphemy accusations, is imperative. Evaluating the potential impact on freedom of expression, the right to security, and protection from persecution will help Meta uphold human rights principles.
* Enhanced review processes for content removal decisions, particularly for high-stakes blasphemy-related content, will ensure that decisions are well-informed and considerate of local contexts.

**Public Reasoning and Accountability**

* Providing clear and detailed explanations for content removal decisions in its transparency reports, particularly those related to blasphemy, is crucial.
* Strengthening accessible appeals and redress mechanisms will enable users to challenge content removal decisions transparently, ensuring that the process is fair and independent.

**Strengthening Meta’s Content Moderation Infrastructure**

* Establishing regional content moderation hubs staffed with experts familiar with local language, culture, and legal landscape will enhance Meta's ability to handle the complexities of blasphemy-related content.
* Continuous training for content moderators on local laws, cultural sensitivities, and socio-political contexts will ensure that moderators are well-equipped to make informed decisions.

For content related to blasphemy and blasphemy allegations, Meta’s aim should be to ensure online safety and privacy where freedom of thought and expression is safeguarded and offline safety against acts of violence is equally prioritized. In addition to assessing [political ads](https://www.brecorder.com/news/40265527), it is important that Meta adopts a more nuanced approach towards ensuring elections integrity, especially content on blasphemy and blasphemy accusations. This is important as the post in question was highlighted during the 2024 general elections. Meta, in terms of its policies, needs to find a balance ensuring sensitivity to regional circumstances particularly in areas where blasphemy laws and accusations dangerously aggravate socio political situations. In order to achieve balance, platforms need to engage and work with government bodies, trusted partners, civil society organizations, activists and marginalized communities to ascertain that freedom of expression is not endangered as a result of ensuring individual safety. For instance, while assessing government requests for the removal of content related to blasphemy, Meta should conduct due diligence to avoid putting human rights at risk. When working with government bodies, trusted partners, trusted partners and civil society organizations, Meta should ensure transparency in terms of reasoning behind removal of particular content. In the reasoning, it should be vital to indicate that the removal of the content is a result of ensuring both the individual’s safety and protecting the religious sentiments of particular religious groups/communities. On Meta’s part, such statements would serve as balancing acts necessary to ensure that freedom of expression is maintained without risking individual security. It is of utmost importance that Meta’s policies and decisions regarding such content are in accordance with the international human rights framework to avoid any and all human rights infringement.