Briefing- meeting with UN/ EU on Commission on the Status of Women

Caste-hate speech
Addressing hate speech based on work and descent

The impact on women of caste-based hate speech in the internet, social media, and technology spheres

"As my report on hate speech at the UN Human Rights Council on 15 March 2021 indicates, the pandemic of hate in social media targeting minorities such as Dalits remains largely unconstrained – and social media owners continue to profit from hate and operate with relative impunity. The time has come for some kind of international regulatory covenant to deal with this global scourge. alarming trend on social media platforms – the de facto normalisation of caste-hate speech as a means to oppress and humiliate Dalits. (...) Caste-hate speech should be recognised as a distinctive form of hate speech that merits attention from the UN, the EU and INGOs. I would argue that it needs to be tackled in a global regulatory framework so that states cannot so easily avoid or evade their human rights obligations"

Dr Fernand de Varennes, UN Special Rapporteur on Minority Issues.

EU policies

In 2008, the EU adopted a Framework Decision on "Combating Racism and Xenophobia" that rejects and condemns all forms of racism and intolerance that are incompatible with the EU values and principles. The Framework prohibits any speech intended to incite violence, intimidation, hostility or discrimination against people on the grounds of "race, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status. The

Framework "advocates that racist and xenophobic hate speech and hate crime must constitute a criminal offence in all member states and be punishable by effective, proportionate and dissuasive penalties".

The EU anti-discrimination guidelines from 2019 also address caste-based discrimination under the term descent stating that "'descent' does not solely refer to 'race' and has a meaning and application which complements the other prohibited grounds of discrimination. The Committee also affirms that discrimination based on 'descent' includes discrimination against members of communities based on forms of social stratification such as caste and analogous systems of inherited status."

The European Union has consistently [recognised](#) caste-based discrimination and violence against Dalits in business, the economy and, most importantly, in everyday practices. But the EU has yet concrete policy-oriented action plans to address caste discrimination either within the EU or other caste-affected countries.

Caste-hate speech in the UN context

Discrimination based on work and descent is the United Nations term for discrimination based on caste or other inherited status. UN human rights bodies have made a number of references to caste-based hate speech. In 2002, the UN Committee on the Elimination of Racial Discrimination (CERD) published its General Recommendation no 29 on Descent-based Discrimination. The document contains specific references to the dissemination of hate speech, including through the mass media and the internet. The Committee recommends states to "take measures against any dissemination of ideas of caste superiority and inferiority or which attempt to justify violence, hatred or discrimination against descent-based communities." It also calls upon them to "take strict measures against any incitement to discrimination or violence against the communities, including through the Internet."

In 2009, the UN Human Rights Council published the draft UN Principles and Guidelines for the Effective Elimination of Discrimination based on Work and Descent – a comprehensive legal framework developed to eliminate caste discrimination. This framework calls on governments to "review or enact… hate speech laws to explicitly prohibit and punish… speech inciting discrimination, hatred, or violence based on work and descent ."

In 2017, the OHCHR published a ground-breaking Guidance Tool on Descent-based Discrimination which sets out practical ways for UN country teams and other stakeholders to combat caste-based discrimination. It specifically mentions hate speech as an early warning sign for predicting violence against descent-based communities. Recent representations made by International Dalit Solidarity Network led to the inclusion and recognition of caste-hate speech in the "Recommendations from the United Nations European and Asia Pacific Regional Forums on "Hate Speech", Social Media and Minorities".

Manifestations of everyday caste-hate speech

In caste-affected societies, discrimination of any form, including speech, is illegal and punishable under the protection of constitutional rights and other legal instruments. However, implementation of this legislation is poor, and caste discrimination, including caste-hate speech, is often carried out with impunity.

Recently, an increasing number of news reports and media analysis have covered caste-hate speech on the Internet and Social Media Platforms (SMPs). Several studies have started to focus on the ever-increasing fake news, disinformation, trolling and religious and caste-related hate speech at the regional level in South Asia. Recent research2 on the growing intolerance by

Hindu nationalists in India analysed how techie-turned ideologues, party workers online, political intelligence consultants and entrepreneurs are drivers of extreme speech, fake news, trolling and disinformation. A BytesforAll Pakistan study has examined various levels of online hate speech against religious minorities that included Muslim subgroups. At the international level, Equality Labs - an activist organisation based in the US - published a study on Facebook's approach to caste-hate speech.

When tech giant Cisco Systems faced allegations of discrimination against a Dalit employee, leading western newspapers and online portals, especially in the US and the UK, started to acknowledge and report caste-based discrimination in the West. In 2015, the Special Broadcasting Services (Australia) ran an in-depth analysis of rampant caste-based discrimination practiced amongst Australia's half-million-strong South Asian community. And in a recent article, a Melbourne-based academic of Indian origin noted that "caste goes where South Asians go".

A leading Indian matrimonial site shaadi.com was accused of perpetuating caste-based social prejudices with an algorithm that discriminated against its users on caste grounds. According to The Sunday Times, a Brahmin profile was designed not to generate potential Dalit matches. The machine only generated matrimonial matches from respective castes, and the search engine did not produce inter-caste matches. Shaadi.com has denied any caste-based bias because it would not view the personal settings by its users as discriminatory.

In 2021, a search for "Brahmins" and "Dalits" on Google and Getty images could show these caste groups' stereotypes on the Internet, with god-loving, ritual-observing, prosperous and successful Brahmins and polluted, unclean, victimised, helpless and angry Dalits. Such filtering is possible because of the combination of photographers' (data producers) social understanding of caste; the acknowledgement of such pictures by the Internet (data scientists producing insights); and the machines trained to create patterns that perpetually produce such biased results.

Tech corporations lack clear and transparent caste-sensitive policies which can often lead to frustration and increase mistrust between users and these platforms.

Women and caste-based hate speech

Dalit women, who have long been subjected to intersectional and multiple discrimination through the twin oppressions of caste and gender, have found a space on Twitter and Facebook. "I do believe that online space is refreshing and a space we never had earlier," said activist Beena Pallical. "There used to be limited regional media spaces, but we are now visible, and much of our anti-caste conversations are now happening on social media platforms." Beena pointed out that stories about victims of cow vigilantes only appeared in national and international media because of mobile phone footage. "Una, a small city with 60,000 people, where Dalits were flogged and assaulted in full public view for allegedly skinning a dead cow, would have never made it to international media without mobile phones and the Internet. It forced the local government to react to the event".

But in Sri Lanka, caste-hate speech is prevalent, especially for Tamil communities of both Indian and Sri Lankan origin, but also amongst Singhalese. Women politicians deemed to be of 'lower caste' origin are vulnerable to discrimination: "Women politicians from lower castes are heavily discriminated in general and face online caste-based slurs. These women are mainly targeted and abused by mentioning their lower caste identities," said Ms D. Vibooshi Balakrishnan, a digital security human rights trainer.

The Internet has also exposed Dalits, especially women, and other oppressed caste members to a vast number of abusers: "We are suddenly faced with an army of caste-speech abusers from nowhere. We realised physical distance is no longer a barrier to these abusers; this has put all our Dalit activists in a very vulnerable place" said Beena. "Alongside opportunities, digital spaces have also created an atmosphere of fear and intolerance, as there is more often than not outburst of obnoxious caste-hate speech".

Asha Kowtal, another Dalit activist, has commented that "in the beginning, we used to engage with abusers and casteist bullies, and it put us in a defensive mode and caused severe stress on our fellow activists and me. We, therefore, thought about this strategy and decided to disengage with these caste-hate speech abusers." A recent report by Equality Labs found that caste and other forms of hate speech "almost completely remains up or is reinstated by moderators on Facebook, an increasing number of minority user accounts are being banned or removed entirely".

One main reason for this, according to Saurabh from Dalit Women Fight Collective, is that Community Standards guidelines of all social media platforms are not yet localised for local languages and contextualised to address caste-hate concerns. "This situation encourages abusers to be bold and more vicious". Suman Saurabh, an activist at Dalit Women Fight, started to notice that casteist handles were replying to anti-caste tweets with an "OK Bhimer" - a new derogatory reference to cancel followers of the philosopher and Dalit leader 'Bhim'rao Ambedkar. "I was extremely triggered by this dismissive tweet". Suman Saurabh and Riya Singh, from the Dalit Women Fight Collective, said they had taken it up with Twitter, but received no response.

Constant caste abuse and inadequate measures by SMPs to hold abusers to account can leave Dalit online users with psychological stress and anxiety. Anti-caste handles and online spaces are committed to ensuring equality and dignity. However, many activists feel that SMPs are letting them down. This situation leads to the ghettoisation of Dalit users and followers. As a result, some Dalit women choose to anonymise their identities so as not to give ammunition to casteist trolls. Social media can also expose Dalit activists to online abuse. In some instances, casteist handles join together and descend on Dalits until the latter give up and disappear. "We are very aware that we are constantly watched and monitored by trolls and haters. Frankly, we are scared of this (level of) exposure, especially when the political climate is increasingly anti-Dalit", said Mohini Bala, another Dalit Women Fight member. She has stopped sharing pictures and family details and now restricts her sharing to mundane work-related updates.

Other Dalit activists felt SMPs should be more caste-sensitive to support Dalit visibility and muchneeded anti-caste conversations. According to a woman activist, "I am popular in social media and attract lots of traffic to my page, but I choose not to upload my picture and name out of fear of being bullied by dominant caste abusers. At the moment, SMPs do not allow me a verified account unless I upload my real face and picture. I am willing to share this information privately, but SMPs tend to insist I upload and de-anonymise myself".

Ways forward- towards a caste-sensitive internet

Global policymakers should consider caste a protected characteristic related to hate speech policies; caste should be recognised and included in all international covenants related to human rights and hate speech. Action plans to mitigate hate speech in everyday conversations increasingly mediated by and saturated on digital platforms should be established.

The social media platforms Google, YouTube, Facebook and Twitter have acknowledged and included caste into the list of protected characteristics under their hate speech policies. This has happened in response to anti-caste activists and civil society organisations and is an important step in the right direction. According to Ms Fadzai Madzingira, Public Policy

Manager Content, Facebook, the platform has been having a series of internal meetings with stakeholders to acknowledge and mitigate caste hate speech on the platform.

A quick search on these platforms, however, reveals that caste-hate speech content is familiar, if not necessarily much viewed. Facebook adopts a "prevalence metric" method: a machine learning process that randomly selects and checks the most viewed (occurrence) content. By the company's admission, this method is very likely to leave out less viewed content, yet the consequence can be severe.

Principles behind search engines and machine learning need to be objective and fair. Reimagining the caste-sensitive Internet requires tech corporations to involve and empower Dalits and other oppressed communities to be part of fixing errors in these codes and principles. Tech corporations, online platforms and internet-related rights-based non-profits are unlikely to back Dalit and anti-caste causes.

Dalit activists across the globe strongly feel that SMPs should be a lot more caste-sensitive. Online activists want tech corporations and SMPs to engage with Dalit groups, as caste-hate speech is layered and locally contextualised. One possible way forward is to make policy manuals and decision-making processes more transparent about what qualifies and constitutes caste-hate speech. Such sharing might enable anti-caste activists to share their perspectives with SMPs to foster a caste-sensitive environment.

As part of its efforts to understand caste-related online safety concerns, Twitter India has started to work with the National Campaign on Dalit Human Rights (NCDHR) and Social Media Matters to connect more structurally and comprehensively with members of oppressed castes. These responses are encouraging, but many of those interviewed still believe that these platforms could and should invest in more resources to be caste-sensitive at organisational (diversity and inclusion) and policy levels to mitigate caste-hate speech. Building a caste-sensitive internet is not just about acknowledging caste in policy documents, it requires structural rethinking of the Internet.

Online platforms rely heavily on their community to identify hate content. Both Twitter and Facebook, for example, have mechanisms for its users to report hate content. Content moderators review this content and decide which action to take – ranging from force-delete to removing the handle permanently. Both Facebook and Twitter, in principle, have explicitly included "caste" as a protected characteristic in their hate speech policy. However, without sensitisation, understanding, and proper implementation, the policy is ineffectual.

Also, status as most viewed depends on various factors such as breaking news, geography, demography and influencers. This could mean that content by a Dalit user may not be qualified to be highly prevalent by the global standard. For instance, targeted caste-hate abuse is a common tactic, and it does not require hundreds of views or re-posts to dehumanise an individual or a group.

Dalits who could not find their own spaces and voices in the mainstream media space found alternative spaces for political participation. This change has led to the organic growth of independent Dalit media outlets producing their own content. For example, Dalit Camera – a YouTube channel modestly started with a still camera in 2011 – is now an influential anti-caste media outlet with over 69,000 subscribers and thousands of videos. Without funding from international media and development agencies it addresses the gap that the "mainstream" media continues to manufacture. Dalit Camera – a caste-hate speech free space – is now recognised as a credible news feed on Dalits and other oppressed groups by mainstream media in India. These anti-caste online media outlets also hold casteist practices of mainstream media to account.

However, sustained and critical reporting of caste and its effects is still rare in western media.

High profile media, communication, IT and development organisations in South Asia are also known to ignore the caste issue even though it is a powerful barrier to inclusive participation and development. However, the proliferation of mobile phones, access to social media platforms such as Facebook, Twitter, TikTok and Instagram, and affordable access to data plans have significantly influenced mobilisation of Dalit and anti-caste activist movements in the last two decades.

Recommendations

General Policies
At the global level
- Acknowledge and recognise that caste is a protected characteristic in all international covenants related to hate-speech and human rights.
- Recognise caste-hate speech as a distinctive form of hate speech that needs to be addressed by international platforms such as the UN, EU and INGOs in all their relevant policies and programs.
- As caste-hate speech sees no borders online, all countries, where South Asian communities have a presence, should recognise caste as a protected characteristic in Hate Crime laws and Equal Employment Opportunities Act.
- States affected by caste should also work closely with Dalit and anti-caste activists to build support for the law enforcement, judiciary, and other relevant stakeholders on caste-hate speech.

International Media, Technology-related Development agencies
- Recognise caste as a barrier to all aspects of social development.
- Support independent monitoring agencies, academic researchers, policymakers and NGOs to come together to produce research and engage in advocacy efforts to monitor and mitigate caste-hate speech.
- Collaborate with Dalits and other oppressed caste media professionals when developing training manuals, media, and development-related research frameworks to ensure caste-sensitivity.

Digital Spaces and Tech Corporations
Digital tech corporations should be sensitive to growing online caste-hate speech and ensure that measures to address this are incorporated into corporate human rights due diligence processes and requirements.

Fact-checking and content moderation
- Localise content moderation assignments;

- Work with Dalits and other minorities and, where relevant, seek their advice on policies and moderation practices related to caste-related contents.
- Train fact-checkers and content moderators about caste-hate speech.
- Develop training modules, with guidance from Dalit and anti-caste activists, on caste-hate speech.
- Recruit sensitive and local human moderators to resolve caste-hate speech-related matters.

Anti-caste diversity, representation and inclusion by tech corporations
- Commit to caste diversity of the workforce – especially Dalits and other minorities, at the managerial, policy, moderation and tech levels.
- Recruit anti-caste activists, anti-caste communication experts as coders, data scientists and AI experts on caste-related issues.
- Commit to translate content policies such as user conduct policies into local languages in South Asia.

Data and transparency
- Strive actively to disaggregate data on the incidences of and actions taken against caste hate speech and caste-supremacist related data along with profiles of triggering agencies, groups and individuals.
- Acknowledge that these disaggregated data are of public interest value and establish a framework to share the data with researchers and policymakers for a further and independent analysis.
- Commit to an ethical stand not to monetise data that promotes caste-practices, casteism, and actively removing such data.

Anti-caste machine learning practices o Create caste-sensitive algorithmic fairness towards building a caste-sensitive internet by working with Dalits and anti-caste activists.
- Improve the process of transparency when handling, deciding case-hate speech and when users appeal against companies' decisions.
- Ensure Dalits and other oppressed minorities are involved in analysing caste-hate speech's context and content.