



# Why did we do this study?



There is a growing awareness of the ways in which social media platforms have engineered a fundamental shift in the traditionally held notions of the public sphere. With their massive user bases across the globe, social media platforms today constitute a vital site for public communication and discourse. It is a site where fundamental freedoms such as the right to free speech are exercised, and curtailed. The normalisation of violence against women in the online public sphere is a pressing issue that demands our urgent attention.

Profitable Provocations: A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life

# Why did we do this study?

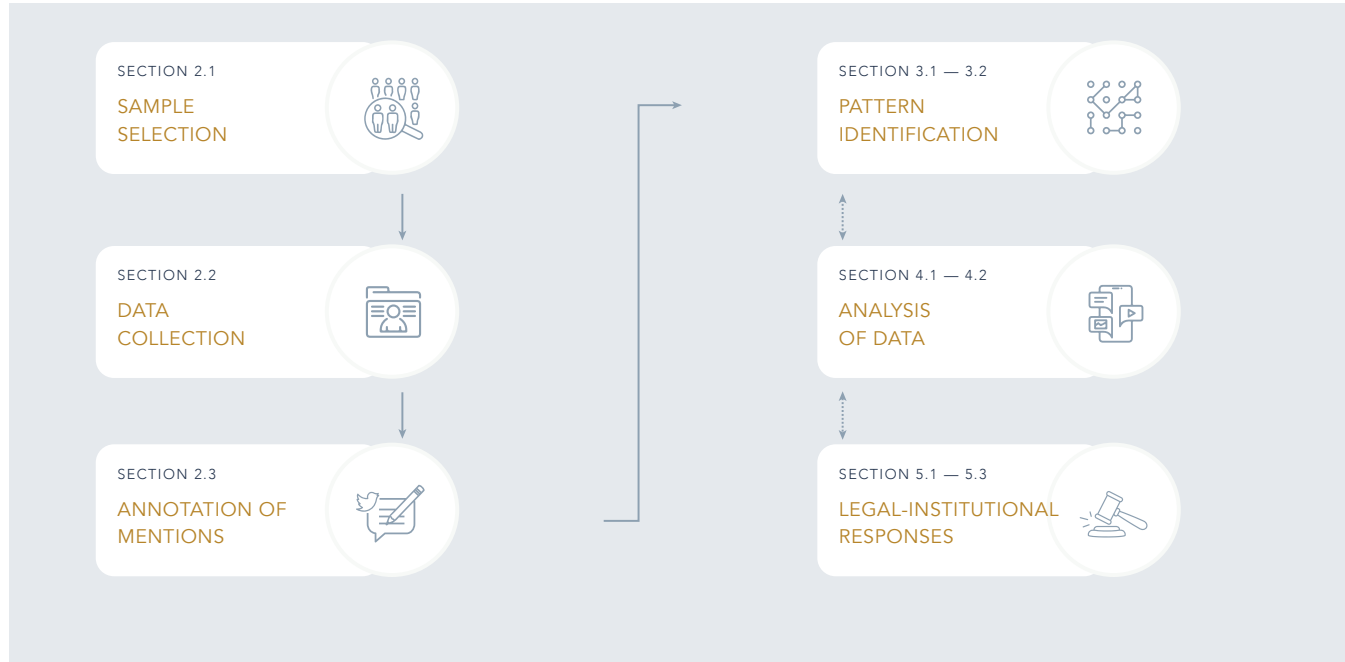


To combat the normalisation of such pervasive forms of gender-based violence, we undertook a systematic study of misogynistic speech on Twitter directed at 20 women in public-political life.

Objectives of the study:

- A mixed-methods exploration of online hate to capture not only the scale and incidence of violence but also the **nature of online hate and recurrent patterns of abuse**
- By bringing to light these forms of abuse that run rampant on social media platforms, **to propose a regulatory framework that is able to contend with the unprecedented challenges of viral hate in the online public sphere**

# Research Process



Profitable Provocations: A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life

# Methodology

20 women were selected as the sample group. This included women involved in formal party politics as well as vocal political commentators.

**20** Research Subjects



**7** Women MP/MLAs

**6** Women in Politics

**4** Political Commentators

**3** Political Figures with No Twitter Handle

# Annotation Process

Using the credentials provided by the Twitter developer platform, mentions directed at the women were collected from their public profiles for a period of one week between 26 November to 3 December 2020.

A set of annotation guidelines, developed through inductive coding, was then used to annotate the data. The 19 codes were then further condensed into 7 categories to identify more general patterns in the data.

1. Religious hate speech	Direct religious hate speech	Religious stereotyping	Indirect religious hate speech	
2. Abuse	Generic abuse	Asexual objectification	Dehumanising insult	
3. Misogyny	Derailing	Exercising dominance	Sexist slur	Stereotyping
4. Othering	Delegitimising by othering			
5. Sexualised references	Sexualised slur	Sexual harassment	Sexual objectification	Over-familiarity
6. Threat	Intimidation	Threats of violence		
7. Caste hate speech	Casteist slur	Casteist hate speech		

# Findings



## Broad patterns on the nature of misogyny



### Pervasiveness Of Misogynistic Speech

All women in our sample regardless of their political or ideological standpoints received some amount of abuse on the platform. No one was entirely spared.

### “Light-hearted” Trolling

A majority of abusive messages directed at the women were of a supposedly milder variety of tongue-in-cheek jokes and remarks.



### Herd Aggression

Trolls tended to strategically target certain women and certain posts to exploit the affordances of virality and the algorithmic amplification of content.



### Intersectional Violence

Muslim women, political dissenters, and political commentators received an overwhelming majority of abusive messages.



# The overarching subtext of Brahminical patriarchy





# Analysis & Recommendations

## Destabilising the concept of (online) violence

1. How does the online public sphere reproduce the unequal gender relations of the offline world?
2. How do the affordances of social media platforms produce new kinds of vulnerabilities for women?

Sahana Udupa's conception of "*gaali*" to capture the interlocking practices of insult, comedy, shame, and abuse that unfold in a blurred arena of online speech.

Raminder Kaur and William Mazzarella's work on censorship to understand how tactics of incitement and provocation are used by trolls to hijack the attention of users and gain control of the public narrative.

C. Thi Nguyen on how the design mechanics of social media platforms play a defining role in shaping online discourse

# Platform as critical actors in modern-day political formations



## A Crisis of Publicness

Platform architecture and protocols enable the creation of public communicative registers that routinise censure and abuse against 'erring' or 'transgressing' women.

## Platform Populism

The networked dynamics of platform sociality, together with the utter impunity that perpetrators of majoritarian and gendered violence enjoy, enable extreme ideas to be translated into populist ones.



## Under-resourced Languages

ML tools for detecting abusive speech in non-dominant languages are much less developed, hence allow the majority of such speech to escape the scrutiny of automated content moderation systems.



## Contextual Abusive Language

Special characters, alternative spellings, wordplay and rhyme are used to evade automated detection methods, making abuse difficult to detect and monitor.



# Legal- institutional responses

## PLATFORM ACCOUNTABILITY

- Platforms, as powerful actors in the political-economy, to be held to rigorous legal standards.
- Independent regulatory authority necessary for oversight of platform governance.
- Independence of regulatory authority crucial to guard against dangers of State excess.

## BRAKES ON AMPLIFICATION

- Need for wider recognition of the harms of virality in the online public sphere.
- Content posted by users with greater reach and visibility to be moderated more strictly.
- Platforms to invest in proactive monitoring tools to pre-emptively arrest the viral spread of violent content.

## GRIEVANCE REDRESSAL

- Process of filing grievances and appeals to be made easy to access and disposed of expeditiously.
- Need for platforms to be more responsive to grievances, especially in the category of abuse and harassment.
- User grievances to be utilised by platforms as key resources to identify problematic content.



## TRANSPARENCY

- Transparency reporting to extend beyond only content takedown/ reinstatement decisions, to include disclosures about algorithmic recommendation/ranking systems.
- Transparency reports to present disaggregated data for specific policy areas such as abuse and harassment, bullying, and nudity.
- Reporting formats to be consistent and standardised across platforms.

## CONTEXT SPECIFICITY

- Presence of local human moderators a non-negotiable to ensure attention to context.
- A baseline of context specificity in platform terms of service an urgent imperative.
- ML tools to be developed to detect problematic speech in regional languages.

## NORMATIVE BENCHMARKING

- Need to establish a minimum level of agreement across stakeholders regarding what constitutes gender-based violence.
- Need for express recognition of different modalities of gender-based violence in law.
- Free speech doctrine to be re-examined to emphasise positive freedoms in the online public sphere.