

Gender Justice through Ethical Algorithms

Developing an AI-driven digital support system for tackling online misogyny in India

Hameeda Syed

Supervised by: Dr. Maria Rún Bjarnadóttir

GRÓ GEST, University of Iceland

May 2024



Final assignment submitted in partial fulfilment of the requirements for post-graduate diploma in International Gender Studies at the GRÓ Gender Equality Studies and Training Programme.

© Hameeda Syed

Reykjavík, Iceland, 2024

Table of Contents:

Table of Contents:	3
List of Figures:	5
List of Tables:	6
List of acronyms:	7
Acknowledgements:	8
List of Definitions:	9
Executive Summary:	11
Foreword:	13
Global Overview:	16
AI and Global South:	19
The Indian Context:	23
Social hierarchies:	23
Religion:	23
Caste	25
Gender:	26
Law:	30
Praxis of action-research framework: Dignity in Difference	34
About the organization:	34
Action Research:	34
Research Design:	34
Discussion:	38
On experience of harassment:	39
On emotions and regulation:	40
On BridgeGPT:	41
Recommendations:	42
Conclusion:	42
Project Framework:	43
Project Rationale:	43
Project Goal:	45
Project Objectives:	45
Target group:	45
Theory of Change:	45
Theoretical Framework:	46
Digital Anthropology:	46
Citizen Science:	46
Data Science:	47
Algorithmic Failure:	47

Design Justice and Dialogic Partnership:	48
Contextual Integrity:	48
Project Outcomes and Outputs:	49
Project Design and Principles:	49
Project Principles and SDGs:	52
1. SDG 5 - Gender Equality:	52
2. SDG 9 - Industry, Innovation and Infrastructure:	52
3. SDG 10 - Reduced Inequalities:	53
4. SDG 16 - Peace, Justice and Strong Institutions:	53
Project Schedule:	54
Implementation Plan:	56
Capacity to implement:	56
Sustainability through stakeholder analysis:	57
Partnerships:	58
Nippon Foundation:	59
Counsel to Secure Justice:	59
WISCOMP:	59
Khabar Lahariya:	59
One Future Collective:	59
Budget:	59
Risk Management:	61
Evaluation and Impact Assessment:	62
Reporting, Communication and Advocacy:	64
Conclusion: What can Digital care look like?	66
Bibliography:	67
Appendices:	73
Appendix A: Interview Questions:	73
Appendix B: Project Logframe:	74

List of Figures:

Figure 1: Screenshots of the comments I received on [X \(formerly Twitter\)](#) and [YouTube](#) as a visible Muslim journalist in 2020

Figure 2: A quote on women's rights by Andrew Tate, a controversial online figure

Figure 3: Layers of online misogyny, depicted in an Intimate Partner Violence (IPV) case between US celebrities Johnny Depp and Amber Heard

Figure 4: A visual of a country's structural requirements for innovation in AI

Figure 5: The words mean "prostitute" in Hindi language. Automated moderation tools online could flag the first word more than the rest below

Figure 6: During COVID-19, cartoon depictions like above were published in national dailies after a Muslim organization organized a large gathering prior to the lockdown implementation

Figure 7: A snapshot of an Instagram post by transnational South Asian nonprofit, Equality Labs on reclaiming historically marginalized Dalit identities

Figure 8: A Venn Diagram showing how different forms of prejudice interact with multiple cultural factors to build unique manifestation

Figure 9: A snapshot of the Sull Deals app, auctioning the photos of 80-100 notable Muslim women personalities

Figure 9: A screengrab of a music video by an interfaith rapper.

Figure 10.1: A visual framework for the theoretical framework for the action-research

Figure 10.2: A visual framework for the action-research process throughout the GEST Fellowship

Figure 11.1: Part 1 of a snapshot of Bridge's response towards a survivor facing online misogyny

Figure 11.2: Part 2 of a snapshot of Bridge's response towards a survivor facing online misogyny

Figure 12: A visual Theoretical Framework used for the DSE.

Figure 13: A visual Project Design building the foundation of the DSE, taken from Google, 2021.

Figure 14: A visual mock-up of the DSE. (Dignity in Difference, 2024)

List of Tables:

Table 1.1: Laws related to Indian Penal Code on online misogyny.

Table 1.2: Laws related to Indian Penal Code on online misogyny

Table 1.3: Laws related to Information Technology Act on online misogyny

Table 1.4: Laws related to Information Technology Act and Code of Criminal Procedure on online misogyny and stakeholder accountability

Table 2: A short mapping of existing tools to counter online gender-based violence in the Majority World

Table 3: Project Timeline for the DSE from June-December 2024

Table 4: Stakeholders analysis mapping of DSE

Table 5: Total costs for the DSE

Table 6: Risk analysis and mitigation for DSE

Table 7: Monitoring and Evaluation system for DSE

Table 8: Reporting process for the DSE by Dignity in Difference

Table 9: Interview questions used during action-research phase for DSE

Table 10: Project Logframe for the DSE

List of acronyms:

IPV - Intimate Partner Violence

AI - Artificial Intelligence

ML - Machine Learning

NLP - Natural Language Processing

IBM - International Business Machines corporation

BJP - Bharatiya Janata Party

COVID-19 - Coronavirus 19

BBC - British Broadcasting Corporation

DBA - Dalit, Bahujan, Adivasi

IPC - Indian Penal Code

ITA - Information Technology Act

GRO-GEST - GRO Gender Equality Studies and Training

ChatGPT - Chat Generative Pre-Trained Transformer

DALL-E - A name mashup of a famous artist (Salvador Dali) and a cartoon robot (WALL-E)

GPT-4 - Generative Pre-Trained Transformer 4

DSE - Digital Support Ecosystem

DiD - Dignity in Difference

SDG- Sustainable Development Goal

UNESCO - United Nations Educational, Scientific and Cultural Organization

UN - United Nations

UNAOC - United Nations Alliance Of Civilizations

UNOCT - United Nations Office of Counter-Terrorism

WISCOMP - Women in Security, Conflict Management & Peace

Acknowledgements:

“Don't just learn, experience. Don't just read, absorb. Don't just change, transform. Don't just relate, advocate. Don't just promise, prove. Don't just criticize, encourage. Don't just think, ponder. Don't just take, give. Don't just see, feel. Don't just dream, do. Don't just hear, listen. Don't just talk, act. Don't just tell, show. Don't just exist, live.”

— Roy T. Bennett, *The Light in the Heart*

Every learning journey has its inspirations, and mine come from the immense wealth of knowledge I have gained from the careful curation of GEST professors, lectures, assignments, and cultural trips in Iceland.

I have warm gratitude and respect for the Director of the GRO-GEST program, Dr. Irma Erlingsdóttir, the Operations Manager, Guðrún Eysteinsdóttir; the Project Managers, Anna Guðrún Aradóttir, Thomas Brorsen Smidt and Védís Ólafsdóttir, and Dr. Giti Chandra, the Research Specialist and our academic coordinator for choosing my application and building a safe and thoughtful space for learning and training on Gender Studies. Many times, I have questioned the limits of my knowledge and challenged myself to be curious, only because of how much epistemic humility you taught me.

I am touched by the Final Assignment Module Coordinator, Margrét Ann Thors, for their gentle, and constructive nudges towards making this assignment a reality. Their presence was stable, encouraging and kind— and gave me the strength of discipline when my mood swings were like Icelandic weather.

I have immense respect for my assignment supervisor, Dr. Maria Rún Bjarnadóttir, who patiently alleviated the recurrence of imposter syndrome, and allowed me to see myself the way she sees me— as a thinker and researcher with lived experience.

I could not have applied for the program if my mentor hadn't seen any potential in me. I have constant gratitude to the head of the organization, Dignity in Difference, Himanshu Panday, for believing that my imagination of a gender just utopia in the Global South isn't a fantasy.

I have love, cheerfulness and gratitude to all 22 fellows in my cohort, the Student Paper group and the numerous friends, strangers and acquaintances in the Student Housing, in the University, and in Iceland, who saw me, acknowledged me, listened to my tiny ramblings, and allowed me glimpses of their universe. Who are we if not relational human beings that learn from each other?

List of Definitions:

Misogyny: A deep-seated prejudice or hatred against women ([Moloney et. al, 2018](#)).

Online Misogyny: Online manifestations of prejudice against women through hatred, harassment and targeted behavior ([Etherington, 2015](#)).

Sexism: A form of prejudice that leads to hierarching thinking about the fundamental roles of men and women in society. ([European Institute for Gender Equality, accessed 2024](#))

Benevolent sexism: Patronizing attitudes, mostly by men, to reinforce women's subordination in status and power ([Leaper et al, 2024](#))

Hostile sexism: Overly negative attitudes usually directed to women who disrupt traditional gender expectations ([Daniels et.al, 2011](#)).

Algorithms: Step-by-step instructions to perform specific objectives or tasks. The tasks can differ from analyzing data to making predictions ([Datacamp, 2023](#)).

Moderation: Managing and regulation of online content through removal of harmful posts. It has two types: automated and human moderation ([Gongane et. al, 2022](#)).

Automated moderation: Automated moderation uses machine learning (ML) systems to compare new content with a checklist of identified undesirable content. This creates a cycle of recognition based on existing patterns ([Ofcom, 2019](#)).

Human moderation: Establishes the standards of learning for automated systems' use through manual content removal under a minute ([Thornham, 2020](#)).

Artificial Intelligence (AI): Broadly refers to machine learning (ML), natural language processing (NLP), computer vision, and others ([Mozilla Foundation, 2022](#)).

Machine Learning (ML): It uses data and algorithms to help AI accurately imitate how humans learn. It is a branch of AI and computer science. ([IBM, accessed 2024](#)).

Natural Language Processing (NLP): Help enable computers and other systems to generate, translate and understand large volumes of speech and text. ([IBM, accessed 2024](#)).

Data labeling: Identifying raw data as one component in helping ML models make accurate predictions ([IBM, accessed 2024](#)).

Caste: A form of social hierarchy present in India creating rigid standards of one's status, behavior, profession and personal relationships from birth and passed down through descent ([Wikipedia, accessed 2024](#)).

Gendered cyberhate / Gendered hate speech / Gendered cyberbullying: Public discrimination through hate speech targeted towards gender minorities, particularly women, online (the cyber space) ([INACH, accessed 2024](#)).

Generative AI: Technology that generates responses to prompts or questions. Known through 'brand names' like ChatGPT, and DALL-E, it is able to generate text, audios, images and videos through a "simple input" ([UNESCO, 2023](#)).

Data: Information converted in a form suitable for digital processing ([Vaughan, accessed 2024](#)).

Dataset: Similar groups of data stored together for specific analysis. For eg: gender-based violence dataset in India ([Databricks, accessed 2024](#)).

Data crawling: Collecting large amounts of data from multiple sources, like websites ([Senkrondata, 2023](#)).

AI co-pilot/ AI persona: Virtual assistants trained in specific wisdom by AI to facilitate conversations and assist people in different tasks ([AISERA, accessed 2024](#)).

ANOVA test: Analysis of Variance, a statistical test to compare differences between two or more groups. It is used to test hypothesis. In this project, it will test influence of factors like age and income on survivors to find patterns ([Bevans, 2023](#)).

Chi-square test: Checks the accuracy by comparing hypothesis with results of categorization ([Turney, 2023](#)). It will be used to check the models' accuracy of categorization of survivor's experiences.

Executive Summary:

Misogyny is cultural, and finding new ways to fester. Online harassment, particularly against women, has increased in recent years. Women and gender minorities face backlash and violence in expressing themselves. Many have begun to withdraw from these spaces altogether.

The public remains critical of how these platforms address the issue, with 79% believing they are not doing enough. AI-driven interventions aim to moderate these platforms, but are trained in data from English-speaking cultures, i.e, the Global North. The words people use to depict misogyny varies greatly in India, but are not represented in most AI models, leading to a crucial gap in digital safety for this context. Many of these AI models are unilaterally adapted by resource-poor countries due to technical restraints and limited resources.

All this keeps the survivor facing Online Gender-Based Violence (OGBV) out of the equation of justice. They face lack of agency in deciding the type of harassment, the course for action, and are unable to reach out to stakeholders who have power and resources in this issue.

To stop the misogyny at the rate of millions of posts/second in this context, there is a need for datasets and AI models capable of recognising OGBV in India. Survivor knowledge and input is essential in building a multi-lingual corpus of data.

Our project, the Digital Support Ecosystem (DSE) is a thick-big data platform combining digital anthropology and data science for safe internet for women and gender minorities in India. This interdisciplinary framework, managed by Dignity in Difference, integrates survivor knowledge and cultural insights into the technological process and provides a practical path to realize it. The ecosystem allows survivors to report incidents of OGBV, recognize failures of the classification process, enhance their efficiency with cultural nuances, and lead the community co-creation and support mechanisms through collaboration with high-influence stakeholders like policymakers, journalists, lawyers, researchers, data scientists and law enforcement.

The outcome would build an early-warning, resistance mechanism against online misogyny and build deep insight into emerging gendered hate patterns. It would build a culturally-trained algorithm capable of understanding the landscape of OGBV leading to improved policy advocacy and interventions.

Foreword:

I'm a hijab-wearing Muslim woman journalist in India. These are some of the comments I received after I wrote a story on a redevelopment project by the Indian government during COVID-19 (Syed, 2020).



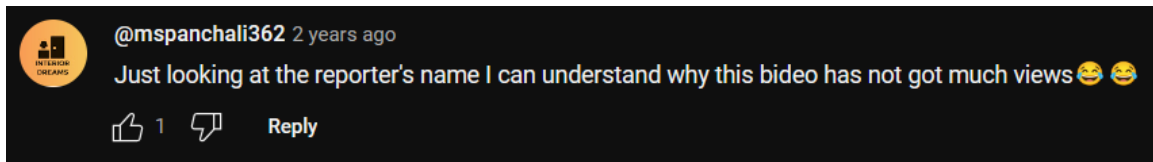
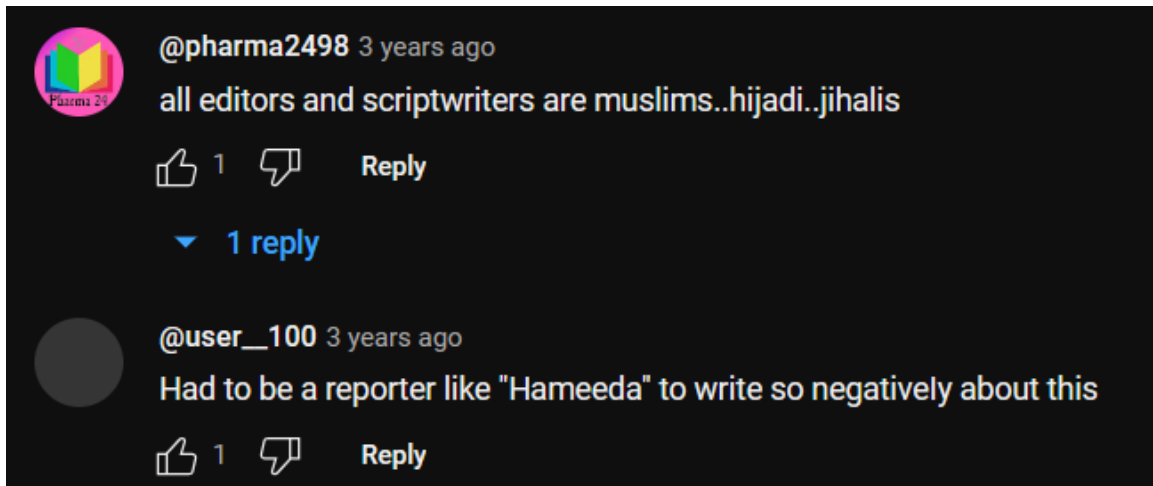


Figure 1: Screenshots of the comments I received on online platforms under my work (Source: [X \(formerly Twitter\)](#), 2020; [YouTube](#), 2020)

I criticized the government's choice to use 20,000 million for building a new Parliament instead of investing it in the failing Indian healthcare system during an ongoing pandemic. I was massively unprepared for what was coming. A flood of online comments tore away all parts of my identity from being a muslim woman to a journalist.

I was determined to do something about these hate comments. As scared and frustrated as I felt, I felt more helpless feeling that they could get away with this. I tried to report them, but the algorithm didn't find it violating community guidelines. I didn't know who to reach out to make the platforms accountable. When I shared the experience with friends, they had similar stories with advice to detach from the platforms. Ultimately, I felt the overwhelming pressure to just stay away from social media. I logged out of my account and didn't log in again. But what stayed with me was the feeling of fear and hurt. I admit that till today, I keep going back to see the comments from time to time. The hurt never goes away.

Since then, my perspective of 'safety' in the online space changed. I constantly stopped myself from sharing my opinions online. Outside of the offline space, if anyone looked at me for more than a couple of seconds, I felt they had realized I was the same journalist they hated so much. It was a distressing few months.

But, as more journalists around me signalled the same story of harassment, I realized it was a pattern. As a journalist, I felt narratives alone were not enough to curb the amount of online toxicity we were surviving. There needed to be structural interventions in place answering some of the key questions: Why didn't platform stop online hate in the first place? What did justice mean for a survivor of online hate and misogyny? What agency did I have in these choices?

At the heart of this project is the idea that reimagination of structures can aid in building kindness and empathy within conversations. We cannot just rely on ethical conduct to help us through hate, we must learn how to build an ecosystem of support that understands its source, and builds mechanisms of accountability centering the justice journey of survivors. I am grateful for the opportunity to share such a reimagination that has defined my worldview and choices and hope this document provides some tiny glimpses of what an ecosystem of care can look like.

Global Overview:



Figure 2: A quote on women's rights by Andrew Tate, a controversial online figure. (Source: [Scrollroll.com](https://scrollroll.com), 2023)

To 8.5 million of his followers, social media influencer, Andrew Tate has publicly stated his perceptions of men and women's place in society. It's not just him. Many around us think women and men are fundamentally unequal. We see it's examples in jokes (Ugh, women, am I right?), attitudes (Why are women so emotional?), and structures (Why should my woman colleague be paid equally?). While we may not realize or agree, it is prejudice, and is called misogyny.

Ten years ago, an independent developer made an online game inspired by their experience of depression ([Wikipedia](https://en.wikipedia.org/wiki/Gamergate), 2024). It got positive reviews. Their ex-boyfriend falsely alleged the reviews were due to sexual exchange favors between the developer and a reporter. Gamers were so enraged that they issued rape and death threats to the developer for months (Ibid, 2024). The developer eventually fled their house in fear. The year-long campaign was called "Gamergate" and gave rise to the alt-right movement online. It also gave a crucial clue: misogyny had found an online audience, and there were no mechanisms to protect the space ([Romano](https://www.romano.com), 2021).

Unfortunately, we tend to think of misogyny as an outcome of social media spaces, despite clear anthropological evidence ([Pew Research Center, 2022](#) , [Moloney et. al, 2018](#)). The train of thought also masks the influence of popular personalities like Andrew Tate worldwide in planting the prejudice in impressionable minds (Clea Skopeliti, 2024; [Shrabonti Bagchi, 2024](#)).

Learnings:

One of my first “aha” moments during the course was when I came across the concept of “masculinities”. Men are expected to perform individually and collectively as a gendered group. In our module, “Theories and Concepts of Gender”, Professors Thomas Brorsen Smidt and Tamara Shefer taught us the cost of those expectations. I thought back to Tate, US President Donald Trump and Russian President Vladimir Putin.

Anthropologists are more clear-headed in this. By speaking to the one experiencing misogyny, they show the connection between physical and digital space is cyclical. Offline attitudes are translated into online behaviors, which converts to real-life harm later (Citrol, 2016). Breaking the cycle means countering the behavior and preventing its reoccurrence in both spaces (Ibid 2016). But we rarely learn strategies from the survivor, even though they have the most stake and incentive to reduce it (Panday, 2023).

Class	Example Tweet
Call for action/violence	Oh @realamberheard You ignorant witch. We ALL already know you're the guilty one here. Johnny's innocence has been proven. You're just trying to buy time, before you (hopefully) have you sit your scronny ass in a jail cell. You speak nothing but venomous lies. #JohnnyDepp
Personal insult, denigration	Seriously, how fucking sick you have to be to pull a "prank" like this on someone ? What kind of gross bitch would think pooping in people's bed is funny ? Well, apparently @realamberheard does. #JusticeForJohnnyDepp
Gendered personal attack	Not a johnny Depp fan but @realamberheard claims have more holes than swiss cheese. I dont understand females who can't make their own money and want to pocket off someone elses. It's hard to find a victim that no one sides with in todays world but I think we all call bs on AH.
Weakness of character, intellectual inferiority	Look what headline just popped up on sky news! @realamberheard you dirty little Lier! #AmberHeardIsALiar #JusticeForJohnnyDepp
Cynicism, sarcasm	@realamberheard Yes, the excitement around #JusticeLeague was huge ... definitely nothing to do with you though. Imagine being in a 4 hour movie for 5 minutes and being the most insufferable part of it.
Skeptical attitude, distrust	I just noticed the 'actor/ activist' claims in your biog @realamberheard !! Well, you certainly are an actress for real!! Only trouble is that the majority of your acting seems to be done OFF stage!! And you have set 'activism' back decades dear!! Ugh, you are some piece of work!
Imputation	@realamberheard @realamberheard Put your hand down and stop exploiting Evan's story to sway the public perception back in your favor. Don't act like you didn't break bread and hang out with Marilyn Manson for years after his relationship with ERW/ your o
Refutation	Listen bitch, I just saw a video about you demanding Depp supporter info for some legal implications!!If you want any info about me just DM me and I'll be MORE than happy to bring you upto speed!! @realamberheard I am allowed my opinion and you are scum (&u better pay my airfare!)
Speculation, denying credibility	@realamberheard You do not represent women nor survivors. I stand with Johnny Depp, Kate James, Jennifer Howell, Lily-Rose Depp, Hilda Vargas, Samantha McMillen, Katherine Kendall, Trinity Esparza and ALL THE OTHER women and men who knows your true color
Demonstration of Power	Justice for Johnny Depp outside @wbpictures studio where @realamberheard is currently filming @aquamanmovie #JohnnyDepp #JusticeForJohnnyDepp #JOHNNY #AmberHeard
Taking a position	@realamberheard is not a victim, she is the perpetrator.

Figure 3: Layers of online misogyny, depicted in an Intimate Partner Violence (IPV) case between US celebrities Johnny Depp and Amber Heard. ([Source: Strathern et. al, 2022](#))

At the rate of million posts per second, one is more likely to come across Andrew Tate. But the sheer scale of his and other similar content makes it challenging to moderate ([Safer Scrolling, 2024](#)). Moderation means managing content in online spaces ([Brown, 2020](#)). Companies moderate content in two ways: through automated and human methods ([Gongane et. al, 2022](#)). Automated moderation uses machine learning (ML) systems to compare new content with a checklist of identified undesirable content. This creates a cycle of recognition based on existing patterns ([Ofcom, 2019](#)). This means the content flagged has to fit a particular gender harm standard to be removed (Ofcom, 2019).

Automated moderations has its limits. For example, it isn't effective for content falling outside the gender harm standard. But who sets the standard in the first place? Usually, top institutions and tech corporations in the United States, Germany, and Hong Kong (China) ([Mozilla Foundation, 2022](#)). This can challenge users experiencing harmful content outside of these cultures (Ibid, 2018).

This is where human moderation comes in. It's important in establishing the standards of learning for automated systems' use ([Thornham, 2020](#)). But they often work under tight time constraints, and need to make quick decisions on content removal, usually under a minute ([David et. al, 2023](#)). Sometimes, social media users contribute by marking content as inappropriate or harmful; this can inform the algorithms and help refine automated moderation practices (Ofcom, 2019). But the scale of content and the pace of flagging is not enough to perfectly identify and manage all toxic content ([Thornham, 2020](#)).

Experts also feel policing online content does little to address prejudice and overturn status quo ([Brooke, 2023](#)). Echoing their sentiments, acclaimed digital news platform WIRED, in 2016, published an open letter, "Dear Internet, It's Time to Fix the Mess You Made", on failure to stop online toxicity ([Wired, 2016](#)).

The sluggish change in gender harm standards also limit what new technologies, like artificial intelligence (AI) are capable of. AI broadly refers to machine learning (ML), natural language processing (NLP), computer vision, and others ([Mozilla Foundation, 2022](#)). Most AI models are trained in these gender harm standards. The standards from Global North countries tend to disproportionately represent English, American, white, and male perspectives (Gillespie, 2018).

But creating these standards needs expensive resources like engineers, data and computational power (Ibid, 2022). Private firms in resource-rich countries, like the US are able to invest the most in AI as a result ([Stanford University Human-Centered Artificial Intelligence Institute, 2022](#)).

Current gender harm standards mirror prejudices and power imbalances in the tech sector— a fact that is not recognized ([Lee et. al., 2019](#)). Women are unequal and face discrimination, limiting them from designing and consuming these systems and standards ([Ding, E Siapera, 2018](#)). There's a need for inclusive standards, which cannot come without inclusive representation in the sector(Mozilla Foundation, 2022).

Tidbits:

The Human Error Project in Switzerland found civil society actors in Europe concerned about racism and discrimination by AI in classifying human beings ([Phillip et. al, 2024](#)). It's recommendation? Bias is a symptom, rather than a cause of technological development.

Let's go back to Andrew Tate now. The same gender harm standards, feeding online moderation, fail to prevent his misogynistic content reaching young people and minorities ([Safer Scrolling, 2024](#)). Young boys have shifted from sympathizing with gender differences, to blaming women (Ibid 2024). In a psychologically tumultuous age, online misogynistic content weaponizes their insecurities ([Clea Skopeliti, 2024](#)).

Tidbits:

A [Vodafone](#) survey on internet practices of boys between 11-14 years of age revealed over 22% of parents have observed a gradual, degrading shift in the way their son speaks about women and girls over time.

AI and Global South:

Closer to my context, Global South countries have been quick to think of AI tools are answers to structural gaps ([Okolo, 2023](#)). Some examples include WhatsApp chatbots to spread sex education practices to rural Indian teenagers ([United Nations Population Fund, 2023](#)). These tools are innovative and democratic, and have led to inclusive development models for their cultures (Okolo, 2023).

Yet, these countries need technical infrastructure for research and development in AI tools. There is a heavy absence of tools, data, talent and capacity. This affects their readiness and feeds into the global AI divide ([Yu et. al, 2023](#)). Training AI models is expensive and unaffordable in the long-term (Ibid, 2023). When caught between AI innovation and looming political or civil crises, many countries would choose to funnel their resources to the second one.

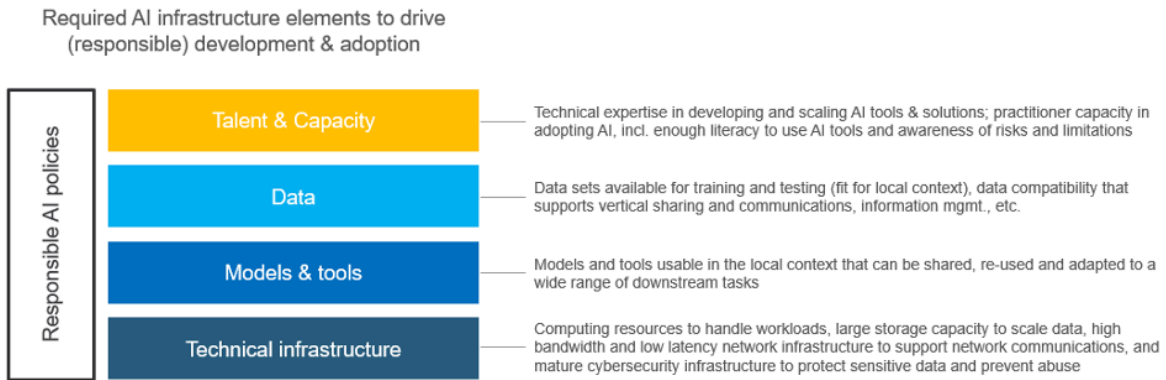


Figure 4: A visual of a country’s structural requirements for innovation in AI. Source: [World Economic Forum, 2023](#)

One way to address resource constraints is by re-using existing models. However, these models are (again) trained in English-speaking, Global North data (Mozilla Foundation, 2022). In the Global South context, these models are unfamiliar with detecting non-English language and cultures.

R FOR R*NDI
RUN DAY
🏃 DAY
🏃 🌞

Taken from VoiceValor data

Figure 5: The words mean “prostitute” in Hindi language. Automated moderation tools online could flag the first word more than the rest below. ([Source: Dignity in Difference, 2023](#))

Then there’s the issue of outsourcing, Global North tech companies have shifted data labeling labor to Global South humans, with India in special focus. Many human moderators are from here and exposed to graphic content without good pay or psychological safety ([Chandran et. al, 2023](#)). Labor and wage exploitation continue due to lax data protection, AI and labor policies ([Chan et. al, 2021](#)). Despite Google, IBM and Microsoft establishing research labs for advancement here, AI research centers have limited presence (Ibid, 2021).

For a survivor experiencing online misogyny, the journey of gender justice becomes unclear. They feel helpless by the lack of accountability or indepth understanding of gender

harm on online platforms; they may decide to limit their speech or their use of the space altogether ([UNESCO, 2021](#)). All this enables and amplifies online misogyny across digital systems and structures (Panday, 2023).

And so, real-life violence occurs. Close to 50% of women journalists said online violence converted into offline threats ([UN Women, 2022](#)). In India, where the ratio for every woman online is equivalent to three men, the trend is chilling ([DataReportal, 2024](#)). This has national consequences for the political and health fabric; women's voices are silenced, and their well-being deeply affected (UN Women, 2022).

Solutions:

Right To Be, a nonprofit combating street harassment and hate launched the "Hate and Hope" tracker, a webpage to document online and offline harassment. Anyone can submit their experiences of harassment, and request for support from the community. The website also offers tips of self-care.

The Indian Context:

India's population covers almost 18% of the entire world ([Worldometer, 2024](#)). It is incredibly diverse— with cultural, ethnic, linguistic, regional, economic, religious, caste and class groups ([Asia Society, accessed 2024](#)). It also has social hierarchies. Sometimes, these hierarchies create multiple intersections across different families, groups and regions. In such a country, what can online misogyny look like? And how can it be captured?

Social hierarchies:

Religion:

India has 78% Hindus, 14% Muslims and 2% Christians and other minorities ([Statista, 2023](#)). For the past ten years, increasing calls of “Hindus are in danger” (Hindi: *Hindu khatre mey hay*) have dominated right-wing discourse ([Somayajula, 2022](#)). The population paranoia goes back to pre-independence and is used by right-wing, nationalist Bharatiya Janata Party (BJP) politicians to incite fear and hate against the largest minority ([Appu et al, 2021](#); [Press Trust of India, 2015](#)).

In COVID-19, 70% of Hindus and Muslims felt it was important to stop women and men from marrying outside their religion ([Pew Research Center, 2021](#)). The party used these sentiments to legalize anti-conversion marriage laws, popularly known as ‘love jihad law’ in the country ([Ganguly, 2021](#)).

Tidbits:

The term ‘love jihad’ is a conspiracy theory coined by the right-wing groups in India. The word, “jihad”, comes from the Arabic word, meaning “a holy struggle”. By using this phrase, they allege that Muslim men are seducing Hindu women to compel them to convert to Islam, and increase the numbers of Muslims. (Source: [NPR, 2021](#)).

Tidbits:

For five centuries, Hindus and Muslims have been claiming ownership of a piece of land in the city of Ayodhya, Uttar Pradesh. Hindus think the land's mosque, built by a Muslim commander was constructed over the birthplace of a revered Hindu deity. The mosque

was demolished by Hindu nationalists in 1992, leading to widespread communal riots across India.

The BJP has been closely associated with the Ayodhya movement. The party included the construction of the Ram temple in its political and electoral strategy. The demolition played a key role in BJP's rise to national prominence.

In November 2019, the highest court in India ruled in favor of the construction of a Hindu temple in Ayodhya, and ordered an alternative piece of land to the Muslim community for mosque construction. This decision was seen as a victory for the BJP and its long-standing agenda, further solidifying its position in Indian politics.

(Source: [Al Jazeera, 2024](#))

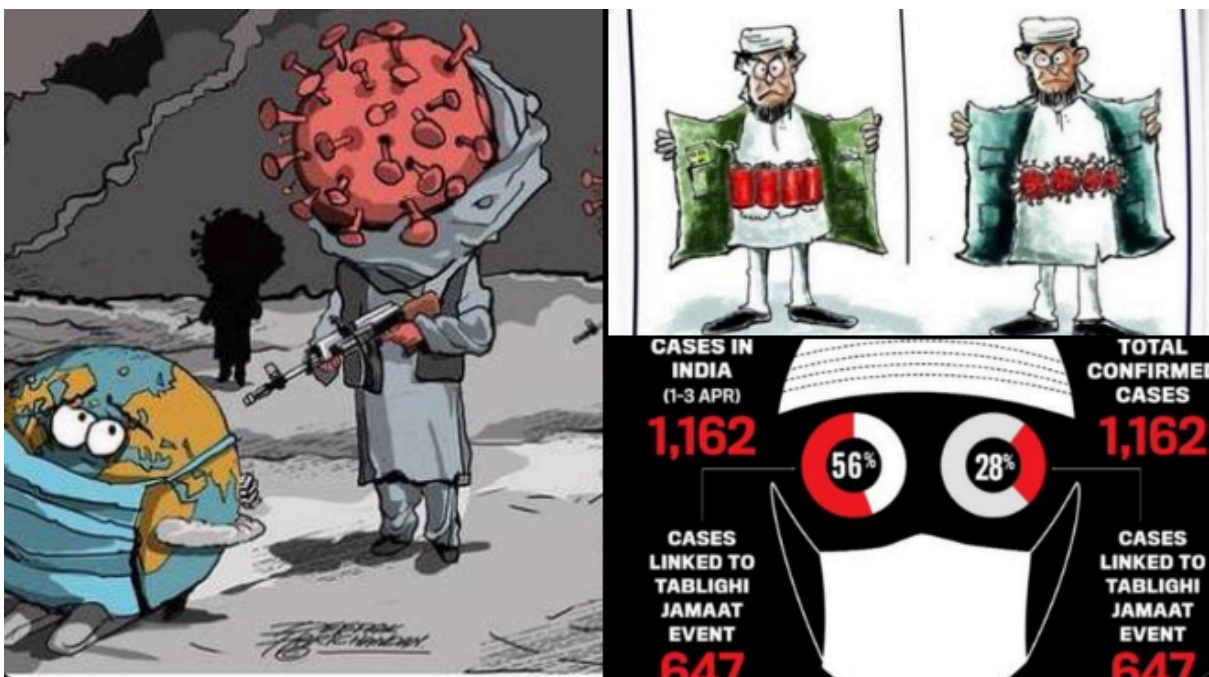


Figure 6: During COVID-19, cartoon depictions like above were published in national dailies after a Muslim organization organized a large gathering prior to the lockdown implementation. The national health ministry alleged the gathering being the single cause of the pandemic. (Source: [Patel, 2024](#); [Twitter, 2020](#))

The Muslim demographic has been kept 'in check' through COVID-19 disinformation and state laws (Pranay, 2022; [Jeffrey et al, 2019](#)). Among 668 anti-Muslim speech incidents, 75% of them took place in BJP-led states ([India Hate Lab, 2023](#)). Islamophobia has increased due to widespread misinformation, disinformation and fake news by politicians and right-wing during election season, and led to online violence ([Dutta et. al 2022](#); [Sharma, 2023](#)).

Caste

Present for 3,000 years, caste creates the standard of one's status, behavior, profession and personal relationships ([Wikipedia, accessed 2024](#)). Passed down through families, it was practiced originally by Hindus before becoming a broad social category (Ibid, 2024). There are many castes that came into existence through religion, traditional occupations, property ownership and region ([Jones, 2017](#)). One's caste draws the extent of one's economic and social capital, and limits the person from shedding the identities (BBC, 2019). Even today, low-caste individuals "suffer from landlessness, unemployment, and discriminatory practices" ([Asia Society, accessed 2024](#)).

Tidbits:

An ancient Hindu law text called "Manusmriti" said the caste system creates "order and regularity of society".

Hindus are four categories: Brahmins, the top caste who were mainly teachers and intellectuals, Kshatriyas, who were warriors and rulers, Vaishyas, who were merchants and the Shudras, who were considered 'untouchables' and the polluted caste.

Today, in the national constitution, lower castes are legally designated as "Scheduled Castes" and "Scheduled Tribes", to grant them reservations in recruitment, finance, education and housing to bring them at par with historically privileged groups. (Source: [BBC, 2019](#); [Dushkin, 1967](#))

Prejudice against castes exists in texts, customs and practices ([Kain et. al, 2021](#)). Low-caste groups face hate speech built on notions of "purity" (BBC, 2019). Major backlash and violence from upper castes maintains the status-quo (Ibid, 2019). For instance, 29% of Brahmin Indians are unwilling to live next to Scheduled Castes as neighbors ([Pew Research Center, 2021](#)).

Online, many castes own their identity to visibilise themselves. They build intersectional awareness, by calling themselves Dalit, Bahujan, Adivasi (DBA). These activists face severe humiliation, discrimination and violence for expressing their identities, rights and aspirations online ([Kain et. al, 2021](#)).



Figure 7: A snapshot of an Instagram post by transnational South Asian nonprofit, Equality Labs on reclaiming historically marginalized Dalit identities. (Source: [Equality Labs, 2024](#))

Gender:

On an ominous note, India holds the record for more girl deaths than boys ([UNICEF, 2021](#)). Societal values favor men more than women, with nine in ten Indians agreeing that a wife must always obey her husband ([Pew Research Center, 2022](#)). Most Indian adults believe sons should have greater rights and responsibilities than daughters (Ibid, 2022). The society's "honor" is tied to Indian women's mobility and growth in workplaces and otherwise— this is a challenging behavior to address while transforming gender relations ([Bhattacharya, 2018](#)).

In June 2016, the United Nations said access to Internet services is a fundamental human right ([United Nations General Assembly, 2016](#)). Yet India only has 26.5% of its women online ([DataReportal, 2023](#)). Women have restricted internet access in poor-resourced and socially underprivileged households, with their husbands usually holding power over resources ([GSMA, 2022](#)). With about 12% of Indians having regular access, most users are urban and from upper socio-economic strata and castes ([Kovacs et. al, 2013](#)).

The digital gap isn't just due to a lack of digital literacy, but also due to prejudice (GSMA, 2022). During the pandemic, women were pushed more out of digital spaces than before ([Saha, 2023](#)). Women from religious minorities, like me, struggle with stark power dynamics, both in and outside our religious community. Caught between the need for protection from violence (hostile sexism), and the desire to retain communal identities can make expressing ourselves feel riskier. We are uncertain of finding support and empathy outside of familiar spaces, and are influenced by strict caretaking roles that offer security

and respect (benevolent sexism), as the figure below demonstrates (Source: [Barretto et. al, 2022](#)). Beyond Muslim women, gender identity cuts across religious, ethnic, class and caste groups in India, leading to targeted toxicity online.

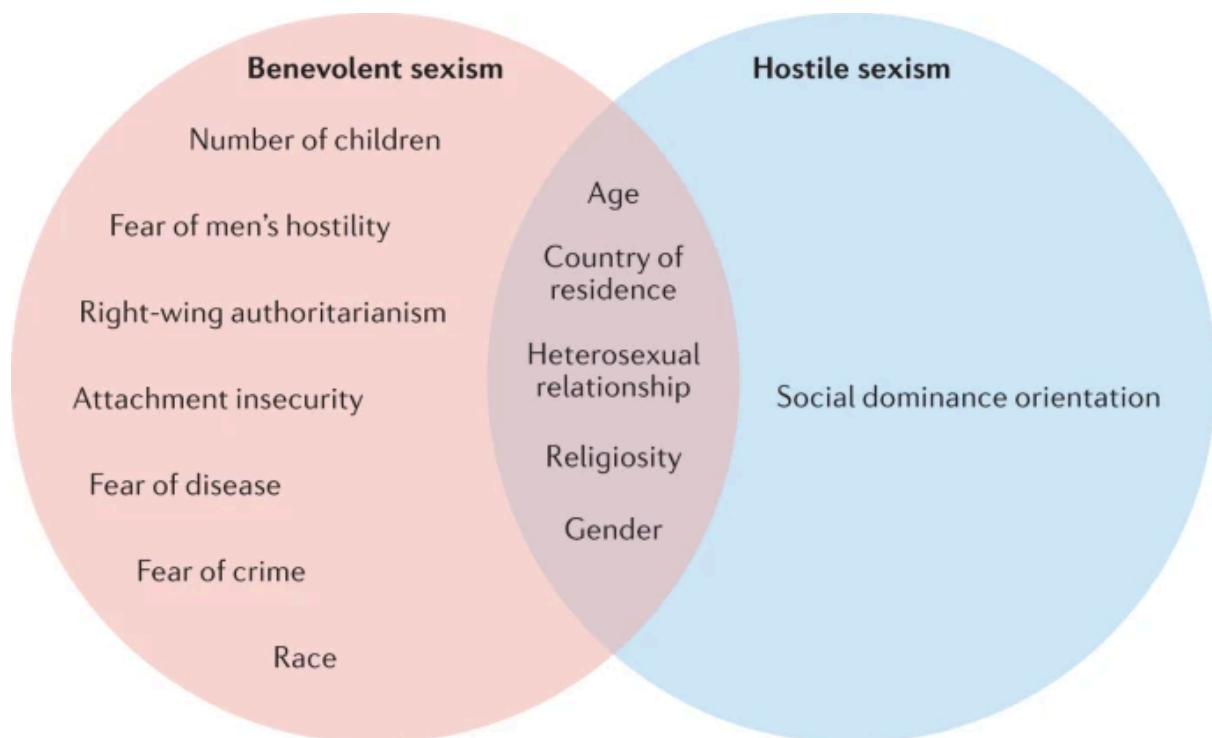


Figure 8: A Venn Diagram showing how different forms of prejudice interact with multiple cultural factors to build unique manifestations. (Source: [Barretto et. al, 2022](#))

Tidbits:

In 2019, BJP politicians stripped the autonomous rights of the legislation in Muslim-dominated regions of Kashmir, triggering online discourse around having the “right to marry Kashmiri women”. As a result, Kashmiri Muslim women faced targeted trolling and abuse. (Source: [Reuters, 2019](#))

Tidbits:

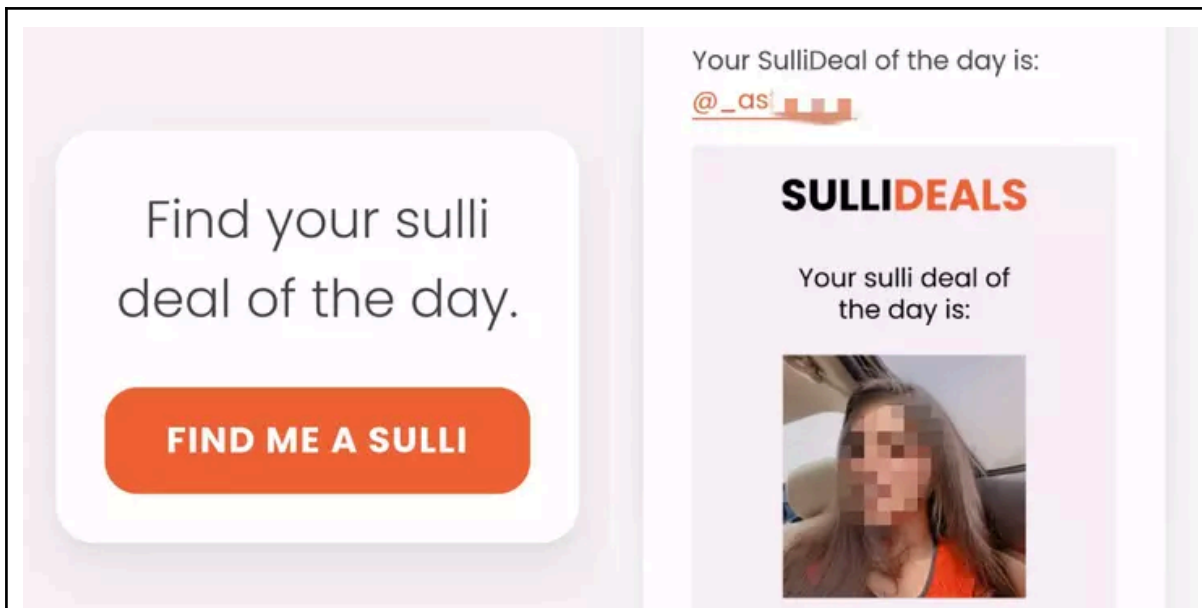


Figure 9: A snapshot of the Sull Deals app, auctioning the photos of 80-100 notable Muslim women personalities. ([Source: Rehbar et. al, 2021](#))

'Bulli/Sulli' are Islamophobic slurs referring to Muslim women. These are fashioned from the term 'Mulli', used by the right-wing groups to troll them. In 2022, two online apps were created to auction 80-100 Muslim women online. The photos on the apps were of well-known Muslim women activists and journalists critical of the BJP's policies. The apps were called "Sulli Deals" and "Bulli Bai". (Source: [Sengar, 2022](#); [Nabi, 2022](#); [Salim, 2022](#))

Religious and cultural minorities have used the internet and platforms like social media to voice their identities, experiences and inspire social change. The #MeToo movement, from a phrase used to empower victims and survivors of sexual abuse and harassment, is one global example in 2017. Survivors, mostly women, used Twitter and other online mediums to share stories of harassment by powerful media and business figures ([Nicholaou et. a, 2019](#)). In India, the movement sparked a year later with similar backgrounds of harassers ([Sharma, 2023](#)).

Women activists think #MeToo is a form of "fourth wave" of feminism, using digital infrastructure to advocate for rights and freedoms ([Women's History, 2021](#)). These "digital feminisms" still exist today, and have expanded more. For example, caste-based minorities use online spaces to voice lived experiences of discrimination ([Singh, 2018](#)).

Tidbits:

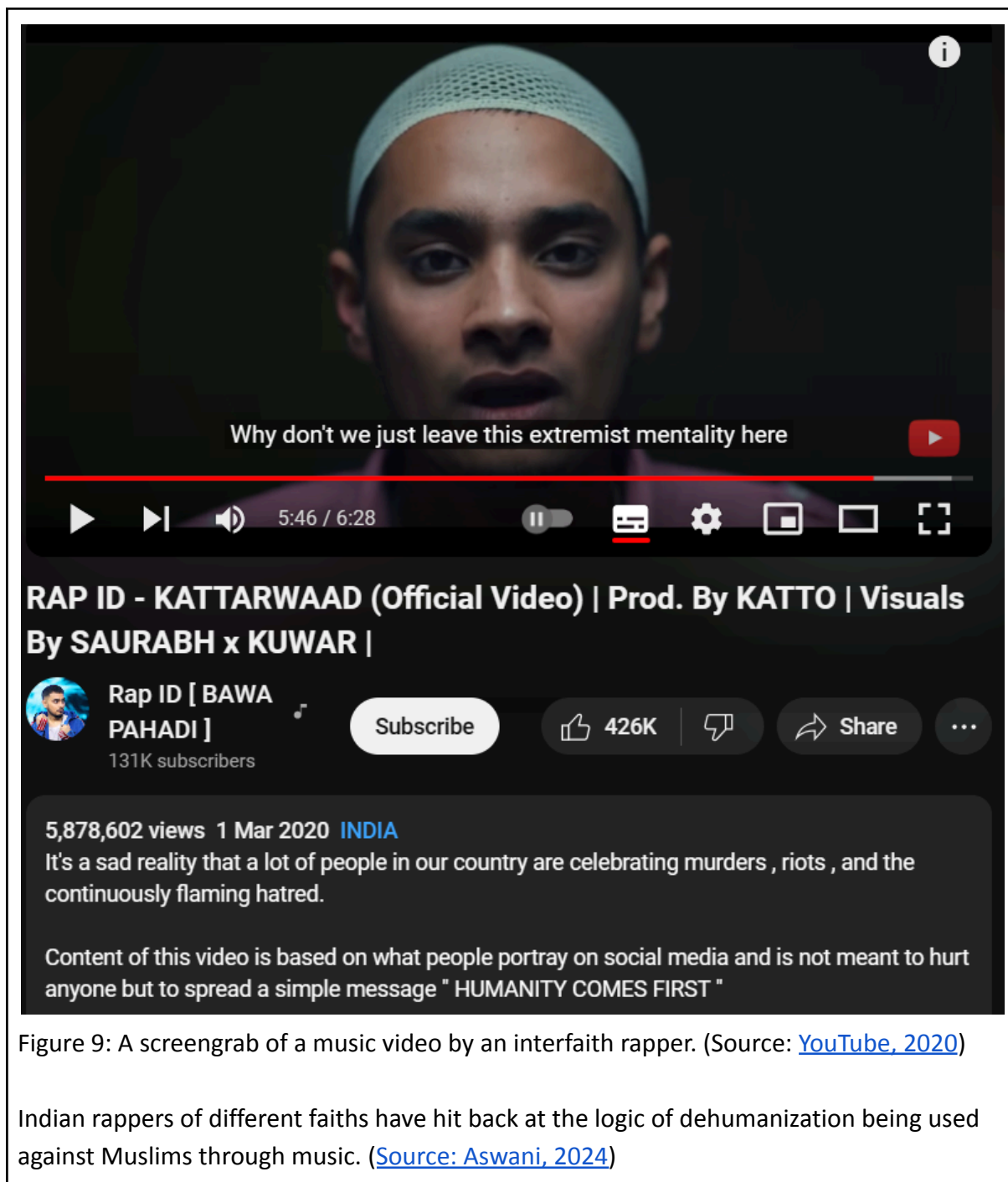


Figure 9: A screengrab of a music video by an interfaith rapper. (Source: [YouTube, 2020](#))

Indian rappers of different faiths have hit back at the logic of dehumanization being used against Muslims through music. (Source: [Aswani, 2024](#))

Top social hierarchies have resisted the spread of such expression through online violence ([Pillai et. al, 2022](#)). Here's one example: *OK Bhimer* mirrors the online catchphrase "OK boomer," used to disregard, ridicule, or invalidate the 'old-fashioned' views of the older generation ([International Dalit Solidarity Network, 2021](#)). The slang has been adapted by pro-caste supporters to 'cancel' the experiences of anti-caste activists. The word 'Bhimer' refers to the followers of the Indian Dalit leader and philosopher: Dr. Bhim Rao Ambedkar. ([International Dalit Solidarity Network, 2021](#)).

Law:

'Online violence' is known by many names, like gendered cyberhate, online abuse, and online violence against women (D.Ging, E. Siapera, 2018). Several authors and publications build evidence based on specific terms, like 'gender cyberbullying' and 'gendered hate speech' (Kristy et. al, 2020; European Union, 2018). The national policies show complex issues about defining harm and regulating free speech for the perpetrator and the survivor (D.Ging, E. Siapera, 2018).

I use the term misogyny to bring a cultural lens to the issue. The lens is important to establish the cyclical nature of online violence. Misogyny silences and intimidates survivors through cultural and structural practices. It signals the need to invest in debates and conflicts happening as a society. (Panday 2023; D.Ging, E. Siapera, 2018).

There are several laws within the Indian Penal Code (IPC) and the Information Technology Act 2021 (IT Act, 2021) for women to use to report online misogyny (SFLC.in et. al 2024). The IPC was, until last year, the official criminal code of the country.

These include:

	Application	Section	Penalty
Indian Penal Code	""DOXING, GENDER-BASED HATE SPEECH""	Section 499	""2 years imprisonment and/or fine.""
	""ONLINE SEXPLOITATION, SEXTORTION""	Section 354	""First time offence: 2 years imprisonment and/or fine.""
	""ONLINE SEXUAL HARASSMENT""	Section 354A	""Sexual harassment: Imprisonment for up to 3 years, and/or a fine. Sexual remarks: 1 year imprisonment and/or a fine.""
	""VOYEURISM""	Section 354C	""First time offence: 1 year up to 3 years imprisonment + fine. Repeat offence: increase up to 7 year imprisonment + fine.""
	""CYBER-STALKING""	Section 354 D	""First time offence: 3 years imprisonment + fine. Repeat offence: 5 years imprisonment + fine.""

Table 1.1: Laws related to Indian Penal Code on online misogyny. (Source: SFLC.in et. al, 2024)

	Application	Section	Penalty
Indian Penal Code	""ONLINE SEXPLOITATION, SEXTORTION, GENDER-BASED HATE SPEECH""	Section 503	""Criminal intimidation– imprisonment for a period of up to 2 years and/or a fine. If there was an intent to cause a serious injury or death, then the imprisonment could be for a term of up to 7 years and/or a fine.""
	""GENDER-BASED HATE SPEECH""	Section 507	""Threatening harm to someone’s person, reputation, or property: 2 year imprisonment and/or fine. Threatening to harm someone with the intention to cause a serious injury or death: up to 7 year imprisonment and/or fine.""
	""GENDER-BASED HATE SPEECH""	Section 509	""3 years imprisonment + fine.""

Table 1.2: Laws related to Indian Penal Code on online misogyny (Source: SFLC.in et al, 2024)

	Application	Section	Penalty
Information Technology Act	""IDENTITY THEFT""	Section 66C	""Upto 3 years imprisonment + fine up to INR 1 lakh.""
	""MORPHING, TRANSMOGRIFICATION""	Section 66D	""3 years imprisonment + fine up to INR 1 lakh.""
	""NON-CONSENSUAL DISSEMINATION OF INTIMATE PHOTOS AND VIDEOS, DOXING, ONLINE SEXPLOITATION, SEXTORTION""	Section 66E	""3 years imprisonment + fine up to INR 2 lakh.""
	""CYBER-FLASHING, VOYEURISM""	Section 67	""First time offence: 3 years imprisonment + fine upto INR 5 lakh. Repeat offence: 5 years imprisonment + fine upto INR 10 lakh.""

Table 1.3: Laws related to Information Technology Act on online misogyny (Source: SFLC.in et. al, 2024)

Application	Section	Penalty
<p>“CYBER-FLASHING, NON-CONSENSUAL DISSEMINATION OF PHOTOS AND VIDEOS, DOXING, MORPHING, TRANSMOGRIFICATION”</p>	Section 67A	<p>“First time offence: 5 years imprisonment + fine up to INR 10 lakh. Repeat offence: 7 years of imprisonment + a fine that could go upto INR 10 lakh.”</p>
<p>“OFFENCES SPECIFIC TO MINORS”</p>	Section 67B	<p>“First time offence: 5 years imprisonment + fine up to INR 10 lakh. Repeat offence: 7 years imprisonment + fine upto INR 7 lakh.”</p>

Table 1.4: Laws related to Information Technology Act and Code of Criminal Procedure on online misogyny and stakeholder accountability (Source: SFLC.in et. al, 2024)

Despite the law’s wide definition of violence, rampant under-reporting of offences is an issue. Survivors struggle to report due to “fear, ignorance of the law, the normalisation of the violence, and the lack of support mechanisms at the educational institution and especially, the home” (Radha, 2024). Survivors feel insecure, restrain from participation and eventually exit online spaces. Many times, police officials don’t realize the harasser is someone known to the survivor, making it difficult to report the crime (Ibid, 2024).

Many laws represent societal values which are prejudiced and employ a counter-harm approach to maintain the status quo of women (Bhat, 2021). Survivors harassed online do not receive quick protection and instant punishment (Alagh, 2023). The existing laws don't act fast enough to offer immediate safety for victims or penalize the harassers right away (Chetna, 2023). In the event of penalization, Indian courts tend to favor biases against women survivors and uphold conventional views (Malavika et. al, 2023).

The criminal legal system also employs a hierarchy while dealing with offences. Street sexual harassment is typically given more immediate priority and response, while online harassment is expected to be mitigated by the survivors themselves ([Gurumurthy et. al. 2018](#)). The importance over 'physical' injuries of the crime informs the approach of law and its implementers, thereby considering online misogyny as 'not real'. (Ibid, 2018).

Laws are ineffective without a thorough understanding of power dynamics and prejudice in India. Survivors struggle to access police stations and follow up on a laborious reporting process. After repeated sharing of their story, they find the judiciary complicit in maintaining the status quo. What agency does the survivor have in the entire process? Do they get to tell their story and build the gender harm benchmark themselves?

Praxis of action-research framework: Dignity in Difference

About the organization:

Dignity in Difference was born to bridge structural gaps in addressing online violence in South Asia. Our team consists of members who have faced online harassment either due to their gender, religious or political identity. Our experiences as survivors inform much of our approach towards making spaces safer.

Through the literature, we know that online misogyny is enabled and amplified by systems and structures across digital ecosystems. We build innovative tools, methods, and interventions to strengthen the resistance against online misogyny for and with survivors.

We acknowledge that misogyny is cultural and the preventive technology designed by Global North doesn't take cultural nuances into account. We believe that arbitrators of truth about misogyny are survivors and the failures of algorithms are an opportunity to deliver justice to them. Our mission is to build preventive and remedial interventions with and for survivors of misogyny, implemented with partners holding the highest power in the digital ecosystem to lead measurable change in online misogyny.

Our innovations, including Dignity At Scale, VoiceValor, and Bridge, are designed to empower technological capabilities, counter online toxicity, and promote respectful interactions. We also work with tech companies and policymakers to advocate for systematic reforms to prevent online misogyny.

We use a collaborative, interdisciplinary strategy to make digital environments safer and more inclusive, particularly for those most at risk of online violence.

Action Research:

Beginning my acceptance into GEST, my first choice was to understand the gaps in the ecosystem through a survivor-lens and imagine possible solutions. Through snowball sampling, I conducted semi-structure interviews with Muslim women in India.

Research Design:

The interviews for developing a deep, analytical framework of the project was done using the theoretical frameworks of action research, design justice and dialogic partnership. Action research implements actions while continuously generating knowledge through

thorough research ([Oosthuizen, 2002](#)). The interviews generated insights for current and future tools creation and heavily informed the design of this project.

Design Justice is an approach to design that challenges universal design principles, which structurally exclude marginalized communities ([Costanza-Chock, 2020](#)). The theory and practice explore building an inclusive world, through the idea of “Nothing about us without us”. The approach believes the digital world can be reimagined and made more inclusive through user-centred design practices and diversity in technology development.

With Design Justice as a foundational principle, dialogic partnership offers the methodology for co-creating digital tools and solutions while countering online misogyny. Through a multi-stakeholder collaboration with survivors, technologists and anthropologists, digital tools can be co-designed with equitable objectives (Panday, 2023).

Simply put, it is the co-creation and active participation of survivors in providing and incorporating their feedback in making design decisions with our organization (Panday, 2023; Costanza-Chock, 2018; Tacchi et al, 2009).

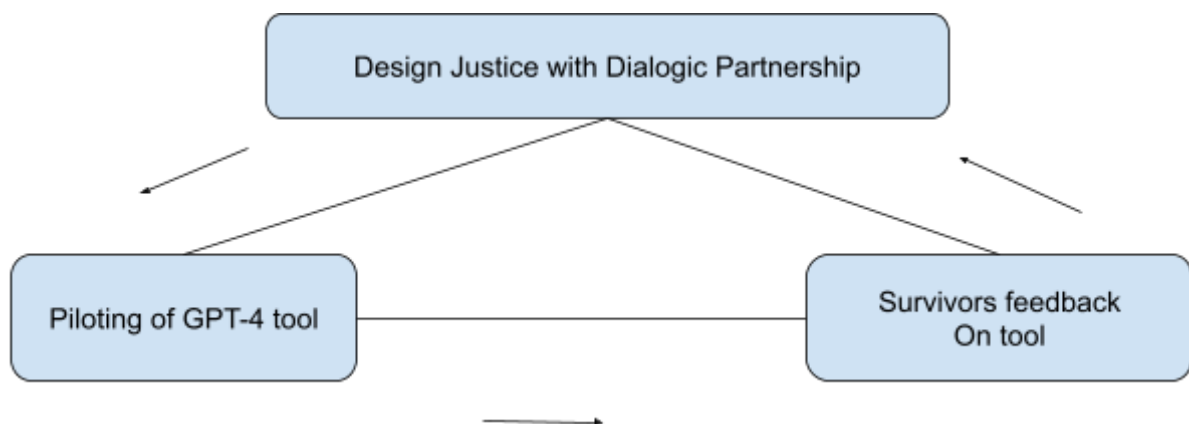


Figure 10.1: A visual framework for the theoretical framework for the action-research

Initially, my framework for the interviews was just to understand and document the survivors’ experiences. However, as I came to Iceland and began the program, I began to think very deeply about the concept of ‘situated knowledge’, shared by one of our academic coordinators, Mr. Thomas Borschen Schmidt, during one of our first classes and its similarities with dialogic partnership as a theory. Situated knowledge was a methodological theory which centered the knowledge arising from the person’s location and perspective. From the viewpoint of the survivors, their situated knowledge gave them an in-depth understanding of the gaps in the structural tool, Bridge, and opportunities to make it more cohesive.

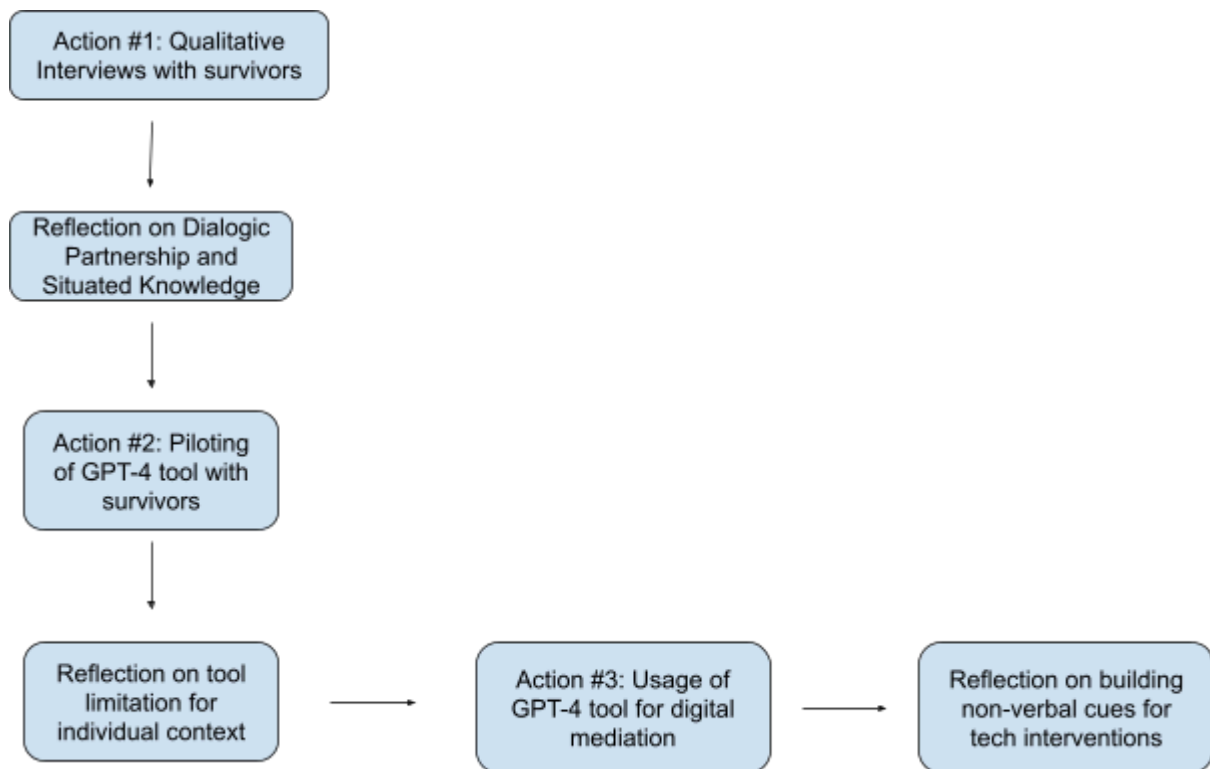


Figure 10.2: A visual framework for the action-research process throughout the GEST Fellowship

I began to reflect on how generative AI could create a **supportive ecosystem** for survivors of online misogyny.

Generative AI refers to technology that generates responses to prompts or questions ([UNESCO, 2023](#)). Known through ‘brand names’ like ChatGPT, and DALL-E, it is able to generate text, audios, images and videos through a “simple input” (Ibid, 2023). While it has been used to build creative outputs and services, it has equally been used as a tool in spreading disinformation and malicious content (Kelly, 2023).

From October 2023, I conducted action-research interviews with young Muslim women in India. I used one of the organization’s GPT-4 tools, called Bridge in the interviews. BridgeGPT supports users facing harassment, especially when targeted for their religious beliefs, by fostering dialogues rooted in empathy and kindness. I analyzed the tool’s role in mediating personal accounts of harassment and violence of the survivors.

For instance, if a person was being harassed in the name of religion through comments, and was feeling overwhelmed and exhausted to engage, they could ask Bridge to

craft a response by pasting the comment and its context. Bridge would respond in three parts:

1. By acknowledging the emotions caused by the comment
2. Pointing out the themes of discrimination in the comment and
3. Providing an editable reply, with the theories used for the reply.

In the process, the GPT would also help them understand how a response could be crafted differently to encourage kinder conversations. Below is an example of a response by Bridge (taken from an interview).

The survivor was referred to as a “penguin” by a comment, denoting the similarity in bodily and facial characteristics of the religious clothing (hijab) with the animal.

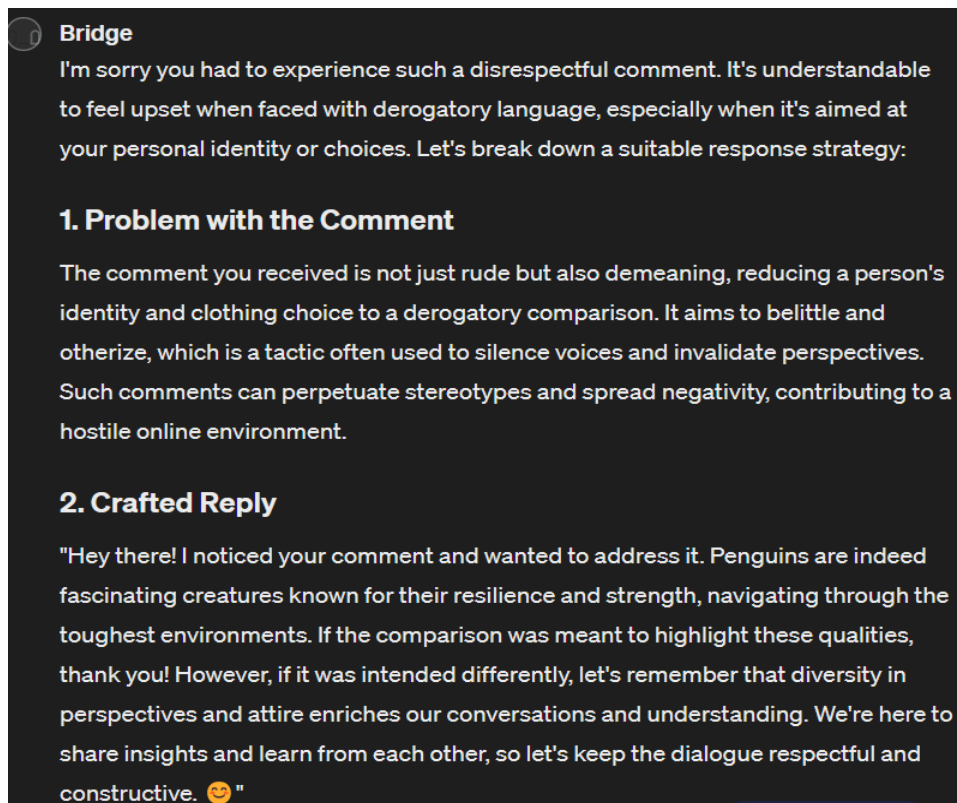


Figure 11.1: Part 1 of a snapshot of Bridge’s response towards a survivor facing online misogyny.

3. Theories Behind the Reply

- **Nonviolent Communication (NVC):** Developed by Marshall Rosenberg, NVC focuses on expressing our needs and feelings without blame or judgment. This approach aims to foster empathy and understanding, which is reflected in the reply by acknowledging the comment and redirecting the conversation positively.
- **Social Identity Theory:** This theory, proposed by Henri Tajfel and John Turner, suggests that our identities are partly shaped by our group memberships. The derogatory use of "penguins" attempts to create an "us vs. them" scenario. By addressing the comment without hostility, the response seeks to bridge divides and emphasize commonality.
- **Cognitive Dissonance Theory:** Leon Festinger's theory posits that people experience discomfort when holding conflicting beliefs or attitudes. By replying in a manner that contrasts the negative expectation of a confrontational response with a positive and inclusive one, it might prompt the commenter to reconsider their stance.

Combining these approaches, the reply aims to defuse negativity, encourage the commenter to view things from a different perspective, and maintain a respectful dialogue. It's crucial to stand up against derogatory language while fostering an environment where constructive conversations can flourish.

Figure 11.2: Part 2 of a snapshot of Bridge's response towards a survivor facing online misogyny.

Till February 2024, I spoke with nine survivors, all of them from the regions of Delhi and Kashmir, to gain an in-depth understanding of their experiences and their perception of the tool. I had initially received verbal confirmation for over 27 survivors. However, when fixing a time for the interview, many survivors showed hesitation in engagement. Some of them stopped responding to requests for interviews. As a survivor myself, I felt the hesitancy was justified, as it required calm dissection of a traumatic experience with little scope of immediate healing.

Disclaimer: My interview approach prioritised the safety and security of the survivors in a hostile political climate in India. Our organization took a deliberate step to only record the interview through written notes of the interviewer and refrained from using identifiable markers when referring to the survivors.

Discussion:

The action-research interviews centered around 3 key themes:

1. Experience of online harassment

2. Retrospection and response
3. Bridge as a possible intervention

The interviews were thematically analyzed to observe themes of different emotions when recounting the experience, the response to the moment of harassment, and the initial reaction to Bridge.

On experience of harassment:

Nine survivors aged 22-35 shared their experiences of online harassment in the interviews. The interview started with remembering a positive moment experienced that week and sharing the interview flow.

The survivors' experiences took place on WhatsApp, Twitter, Instagram, and Facebook. Six out of nine survivors didn't actively use social media. Four out of six began to use it less owing to single, few or multiple accounts of harassment. Each survivor's incident of harassment had a different intensity, and timeline of virality. Identity of those harassing the survivors ranged from Muslim men, Hindu men, and Hindu women. All survivors were able to share their experiences, and comments with complex detail, and show pictorial evidence of their words.

Survivor A spoke out about a class conflict on Twitter. They were targeted by a group of boys, who used their profile photo and posted it, saying they looked like a well-known pornstar.

Survivor A: "The guy posted my photo on Twitter [talking about how I looked like a pig and a pornstar] and had his allies posted comments saying, "Oh my god, go for it, you destroyed her"....They had a group chat in which they would make comments against girls.."

The second survivor was part of a Facebook group of Indian immigrants. One of the members shared discriminatory content around Muslims, which prompted the survivor to address it in the group. They were targeted by a different member, saying they should "go back to Pakistan", a phrase falsely classifying them to be non-Indian.

Survivor B: "There was this one lady who started posted and creating moments like "You should go back to Pakistan".. I wanted to diffuse the anger and not stoop down to their level. Some rubbish she kept saying, I don't remember. The thing that felt really bad "Pakistan jaana chahiye". That really felt bad to me. Fine, you have disagreements, but you don't have to say such things. What was really sad, like 200 something people didn't stand up. The admins didn't care to intervene or clarify the comment that it's not a conversation."

On emotions and regulation:

Survivor C: What did I do wrong? I wanted to understand: why me? Why is it that i say something and its blown out of proportion? Something very simple...there are women on the internet are expressing their opinions too. It took me awhile to understand that it was not about me...it was about them. People just want a bandwagon effect. Anybody who says something that doesn't agree with them, they were bothered. Sometimes you agree, but you don't speak their language. Its also easy on the internet for men to attack women on the internet. It was easy for them to talk about GMC, Bangladesh. They thought I was privileged- but they didn't know me...they didn't know where I come from. These people will go to any limits who threatens their opinion....they could take the twitter discussion and hurt me in real life. It scared me, and I felt I could not be a opinionated woman.

Survivors felt initial emotions of shock and anger. All of them became disturbed for a significant period, with a maximum of upto three months to rejuvenate from the incident(s) mentally. Only three out of nine survivors mentioned the incident to their family. None of the survivors reported the comments on the platform, citing reasons of distrust, feelings of alienation towards the process and uncertainty in the timeline of verdict towards the report.

Further, all survivors opted for 'ignoring' the comments after initial engagement. They agreed that their choice came from feeling futility towards the end goal of the harassment, and option to maintain their mental peace.

Last year, "We are Feminist Leaders" released a resource "Interrogating Silence and Speech", stating how 'silence', for people of colour and disempowered communities, could be perceived as a deliberate tool of protest ([We are Feminist Leaders, 2023](#)). The resource was used to inform my understanding of how all Muslim women survivors looked at 'silence' as a strategy and tool in responding to harassment. When asked to define why they chose it, some of the survivors said:

Survivor D: "Not responding to people keeps your power."

Survivor E: "So i can understand why people would do that...and i think that was also my response after a point because i want to prioritise my mental health and my sanity than engage a random person in my life...i was also telling myself that....But if you have someone who is informed about why they believe in (Indian Prime Minister) Modi, then you can have a really good conversation with people, and then you can also put perspectives and share and gain an understanding why people are such devout followers. But when somebody is coming from WhatsApp University, superficially, then I don't think there's any point in engaging with people either. After a point, I didn't have the mental

bandwidth... I'm not someone who's very aggressive...it also depends on who people are as well..."

Survivor F: "...That was one point where I thought yeah, these people wont stop. They make illogical assumptions and its like they don't even know what they're against. For you, it'll be a huge loop, so its better to save your reputation. At least you are coming with facts, research, truth and so much homework. You know you don't have to get into these petty conversations and its part of society."

The resource also raised important questions in the context of mediation, which is often built on the necessity to speak to conclude. How might one interrogate and build spaces for silence in BridgeGPT? For instance, one of the survivors posed this question on the use of BridgeGPT for closing conversations:

Survivor A: "So you know the answer validates the response. But also like, sometimes, the person in front of you is so ignorant that they don't understand the point you'll make. Even if I use the response, nobody will stop talking back. How do I keep dealing with the barrage of responses? How do you end this? There is no end- you can't convince them, and you don't want to end being like "listen I have nothing much to say". How do you close this on good terms?"

On BridgeGPT:

The third leg of the interview asked the survivors to recount one comment that they were unable to forget. The comment was used, under informed, conscious consent to generate a response from Bridge. The survivors' initial reactions were recorded through written word.

One of our survivors felt BridgeGPT was 'being too nice' in its response to a comment sexualising their identity. They felt BridgeGPT needed to set a particular tone and boundary, one that suggested legal, punitive measures. Another survivor believed Bridge's approach was 'bookish' yet 'genuine', but they wished it had 'eased into the conversation' better. As one survivor said,

Survivor G: "It's a huge deal to share my emotions like this, and I felt that Bridge could have validated me more."

Two survivors felt Bridge's response could have been crafted according to the intensity and scale of the harassment. One of them said,

Survivor H: "People face this kind of harassment everywhere- messenger, Instagram- women face harassment of many different kinds. Some random guy likes you and easy to stalk you on an online medium and there it is. To each kind of online harassment, you have

to respond accordingly. If there's a stalker online, they require a response. If you receive an inappropriate response, it's important to call out the other person, and take appropriate action. But trolling feeds on your response- they make an altogether different connotation- that part doesn't work well. But if they are just there for the sake of dragging it, you should understand that they only deserve your ignorance."

Another survivor felt that BridgeGPT's response could only be used in the middle of a conversation rather than at the start of it. They illustrated through their point:

Survivor 1: "Sometimes my first response to such a difficult comment is always a couple of words like, "why do you ask that?"- like inviting someone to share more thoughts. At times, it's difficult to share those expressions. It feels like I have to say something before this to set the tone of the conversation."

Recommendations:

Survivors felt BridgeGPT's responses could have been crafted according to the intensity and scale of the misogyny, ranging from trolling to identity theft with actionable steps. Bridge's utility was limited to ongoing conversations, lacking the capability to initiate or conclusively address discussions. The survivors recommended enhancements to BridgeGPT, including a broader spectrum of response options to misogyny, integration of mental health resources, and a deeper contextual understanding of misogyny scenarios. These improvements aim to bolster BridgeGPT's role in mediating conversations and greater efforts in integrating generative AI to address and mediate online toxicity in India.

Conclusion:

Every survivor had a different story, and different emotion to share about their experiences. Sometimes, their emotions interacted with their personalities and perspectives; other times, it was influenced by apathy towards online spaces. Yet, each interview made me realize how incredibly complex each story was. Even when the online space was the same, like Twitter, the way it unfolded was unique.

At the end of the interview, a question emerged. How could this individual, situated knowledge of 9 survivors be equally given justice? Was the concept of "justice" truly static? Who defined the ethics of these online spaces, and how could survivors be included in this practice?

Project Framework:

Project Rationale:

Online violence targets and disproportionately affects the gender and religion of a person (Raghvani et.al, 2021). Algorithmic frameworks built to identify and counter online violence often provide a skewed and prejudiced analysis due to unequal representation of poor-resource populations in the training datasets (Chandran, 2023). Calls for re-imagining such frameworks have been rooted in decolonising English language datasets and a conscious effort to move away from a Western-centric perspective of online communities (Google, 2021). Meanwhile, digital networks have witnessed organic efforts in countering discrimination and misogyny online through advocacy and 'digilante' strategies (Jane, 2016). There have also been several efforts across the globe to build digital tools to counter online misogyny. Some of them are:

Name	Topic	Type	Purpose	Focus Area	Link
Uli	Gender-based slurs	Chrome extension	Hide tweets containing gender-based violence words	India	https://uli.tattle.co.in/user-guide/
ShorrAI	Online gendered hate speech	AI-powered bot	Datasets for Indian vernacular languages	India	https://www.shhorai.com/
Sophia	Domestic violence	Chatbot	Gather evidence, learn about rights and options for action	Global	https://sophia.chat/about-2/
rAIrbow	Domestic violence	Chatbot	Conversational support towards abuse	South Africa	https://worldjusticeproject.org/world-justice-challenge-2021/r

					ainbow-chat bot-support- victims-dom estic-abuse
MySis	Domestic violence	Chatbot	Access to legal, psycho-social emergency support	Thailand	https://change-fusion.org/initiatives/11kdhvc0ebab7mgr9d85rviwj9axan
Saahas	Gender-based violence	Mobile and web application	Support directory on responding to gender-based violence according to region	Global	https://hundred.org/en/innovations/saahas

Table 2: A short mapping of existing tools to counter online gender-based violence in the Majority World (Source: Panday, 2023)

Despite these digital interventions, the scale of online misogyny remains too large to tackle. Isolated efforts focus on countering and immediate support. One nonprofit tech organization decided to remove their chatbot after a 3-year pilot, admitting tech blindspots in grasping the complexity and ambiguity of online experiences (Hussain, 2024).

India needs accurate gender harm indicators. The gaps highlight ethical challenges, from who gets to define online gender-based violence, and is it universal?

The Digital Support Ecosystem (DSE) provides an opportunity for counter-speech to be informed by not merely numbers but their lived experiences. The outcome would help stakeholders from multiple communities like researchers, journalists, counter-speech organisations, and advocacy campaigners in India. It would help in building an effective early resistance against emerging problems of online discrimination.

Project Goal:

The goal of the DSE is to operate as a one-stop feminist community platform for survivors of online misogyny and build data-informed interventions to reduce online violence in India. It empowers survivors by giving them an active role in countering discrimination and gives researchers, journalists, policymakers, and social media platforms with actionable, context-rich data. By building a more inclusive digital world, this solution stands to have a far-reaching, positive social impact.

Project Objectives:

- Design and implement a Digital Support Ecosystem (DSE) to enable survivors of online misogyny to actively engage in combating it.
- Create and maintain comprehensive, dynamic datasets on online misogyny that are multilingual and accurately labeled in the Indian context.
- Improve the detection of early trends in online violence, identify the actors involved, and understand the underlying drivers, facilitating more targeted interventions.
- Ensure that survivors do not feel isolated and have access to a supportive community throughout the process of reporting and recovery.
- Facilitate survivor-led community initiatives advocating against online misogyny to foster ownership and advocacy.
- Foster partnerships with technology companies to develop more effective reporting tools and algorithms that can detect and address online misogyny efficiently.

Target group:

The target group would be ethnic, religious, gender and sexual minorities from India, who own a smartphone, have access to the internet and online platforms, and are either moderate or frequent users of social media. Fluency in the English language is not essential, as our main focus is to build representation and awareness of what misogyny looks like and reads like in different regional languages.

Theory of Change:

The proposed solution is a Digital Support Ecosystem to counter online misogyny. We aim to create a dynamic, survivor-centric platform for reporting and documenting instances of online misogyny.

The platform allows survivors to report incidents, provide context, and contribute to a multi-lingual, time-series dataset of online misogyny. By labeling the misogynistic content and sharing their personal experiences, survivors are empowered and given agency in this process.

A dashboard, accessible to vetted stakeholders, displays real-time trends, topics, and targeted communities, acting as an early warning system for emerging patterns of misogyny. Stakeholders wishing to access the datasets must submit a plan and go through an ethical review process, ensuring respect for the survivors' privacy.

The impact of this solution is multi-faceted. It empowers survivors by giving them an active role in combating misogyny and provides researchers, journalists, policymakers, and social media platforms with actionable, context-rich data. By making strides towards a more inclusive digital world, this solution stands to have a far-reaching, positive social impact.

Theoretical Framework:

Digital Anthropology:

Digital anthropology examines the relationship between humans and digital technology, focusing on how digital tools and platforms influence human behavior, cultural practices, and worldviews ([Artz, accessed 2024](#)). Having emerged in the late 1990s, it observes how communities engage with digital technologies, its impact on community values and individual behaviors, and the role of technology in shaping our understanding of the world ([UNESCO et. al; 2023](#)). This branch of anthropology is into the complex interactions between technology and humanity, offering insights into the digital dimensions of human life. (Artz, 2024).

In this project, digital anthropology as the framework will allow a deeper assessment of the causes and structures that facilitate gender-based violence online in the context of societal values. As mentioned during 'Global Overview', cultural misogyny has found an online audience and reinforces existing stereotypes about gender and sexual minorities. By using digital anthropology in the DSE, we will examine patterns arising from these incidents signalling socio-cultural contexts and nuances. These nuances often lead to niche forms of misogyny in asymmetrical power relations in communities in India.

Citizen Science:

Citizen Science refers to amateur experts and volunteers engaging in building and analyzing data at a larger scale ([Gura, 2013](#)). It is a popular method in research requiring large-scale data while bypassing costs, time and labor (Gura, 2013). Through public

participation, citizen scientists with different expertise will generate new insights with more spread-out data ([Ullrich, 2024](#)). This can increase accessibility among the public and increase knowledge of the issue (Ullrich, 2024).

Through this project, survivors of online misogyny will engage in citizen science practices by sharing their stories, cultural context, and nuances of navigating online reporting mechanisms. From the start, survivors will have a platform to share incidents of online misogyny and choose how the platform uses the data. This acts as an ethical foundation for the theoretical framework of the project.

Data Science:

In simple terms, data science can provide a 'systems-level' perspective by collecting large-scale data from online communities, while lacking cultural context behind the data to derive insights ([Hiller, 2023](#)). Using data science allows for exploring the depth of online misogyny, without losing the magnitude. Data Science can aid in finding patterns by providing tools like algorithms to detect different forms of gender-based violence online.

In this project, Data Science will lend its expertise to Digital Anthropology by creating an algorithm that scrapes survivors' experiences into a multi-lingual, time-series dataset. This dataset can be converted into a dashboard to provide a 'big' picture of the scale of online misogyny in India. Our organization's innovations in digital anthropology combine thick and big data to understand how and why algorithms fail in preventing online misogyny and lead to actionable measures across the digital ecosystem.

Algorithmic Failure:

Algorithms are trained in data containing biases and misinterpretations. In the context of online misogyny, interactions usually have subtle, cultural contexts that seem ambiguous on a larger scale, and therefore, difficult to be picked up by algorithms. Such failures lead to survivors' experiences being unable to get labeled as harassment, leading to survivors feeling frustrated and invisible. Algorithmic failures also signal a need for cultural context in design to interpret data ethically. Considering how much online misogyny is a manifestation of societal prejudices, training algorithms should include cultural context to refrain from reinforcing existing inequalities.

The DSE will develop algorithmic classifiers that will be trained on the cultural inputs and contexts of the survivors. This will provide a more nuanced understanding of emerging trends, and target identities and build an effective early resistance against emerging problems of online misogyny.

Design Justice and Dialogic Partnership:

As explained during the action-research framework, Design Justice is an approach to design that challenges universal design principles, which structurally exclude marginalized communities ([Costanza-Chock, 2020](#)). The theory and practice explore building an inclusive world, through the idea of “Nothing about us without us”. The approach believes the digital world can be reimagined and made more inclusive through user-centred design practices and diversity in technology development.

With Design Justice as a foundational principle, dialogic partnership offers the methodology for co-creating digital tools and solutions while countering online misogyny. Through a multi-stakeholder collaboration with survivors, technologists and anthropologists, digital tools can be co-designed with equitable objectives (Panday, 2023).

The project’s foundation relies on the contribution of survivors and the power of community in shaping solutions towards online misogyny. The insights’ from the survivors experiences would shape the development of the platform’s design at different stages.

The conscious decision to involve survivors’ insights through a pilot testing of our tool also provides us with an understanding of further intentional design changes to help survivors feel comfortable during the time of harassment.

Contextual Integrity:

The theory of Contextual Integrity emphasizes collection and maintenance of data by respecting the norms of the context it has been taken from. Helen Nissenbaum (2010) shares this theory to build a larger, ethical framework that arises as a need out of democratisation of data collection and analysis. With the coming of technology and its regulation by private companies, privacy concerns need to be understood and evaluated according to the values of the context.

In this project, contextual integrity requires data collection and anonymization in ways through which the survivor is comfortable. In my data collection during interviews and pilot testing, I was careful not to audio-record any interviews as it could act as a deterrent in candid sharing of experiences. So, I wrote notes during the interview itself and kept the privacy of the survivors only limited to a manager in my organization for reflection and analysis purposes.

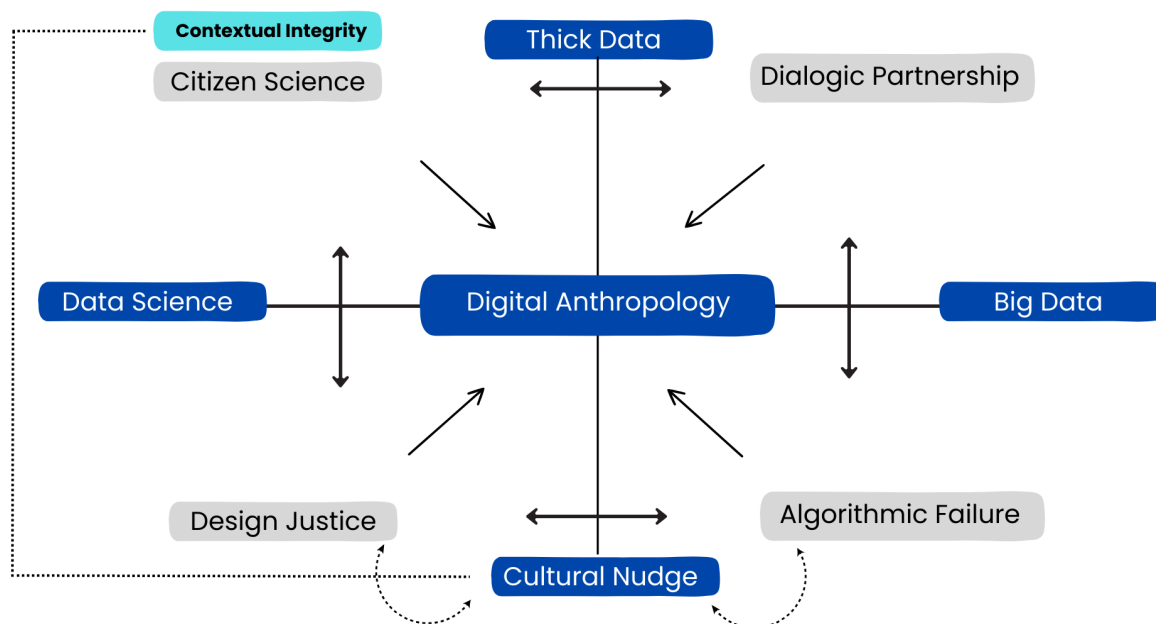


Figure 12: A visual Theoretical Framework used for the DSE.

Project Outcomes and Outputs:

Outcome 1: DSE algorithms identify, detect and take action against online misogyny in India

Output 1.1: Establishment of DSE for survivors of online misogyny in India on the internet.

Outcome 2: The DSE supports survivors with quality services.

Output 2.1: Platform builds a culturally-trained algorithm based on data collation.

Output 2.2: Survivor community builds initiatives countering online misogyny.

Output 2.3: Journalists, researchers, lawyers, and activists volunteer to provide their services to survivors for access to support.

Project Design and Principles:

The foundation of the project's design rests on building an algorithmic justice framework in India. The pathways of framework are partly adopted from a research study by Google on building an "AI Fairness research agenda" (Google, 2021). The framework outlines three pathways:

1. *Recontextualizing Data and Models:* In context of building multi-lingual datasets in India, emphasis is placed on the ethics of data collection (contextual integrity), of co-creation with the community (design justice), of conscious knowledge of power relations in social relationships defined by gender, caste, and class (algorithmic justice).

2. *Empowering Communities*: This pathway highly underlines participation of communities in identifying “problems, specifying fairness expectations, and designing systems.” (Google, 2021). The pathway merges with the theoretical framework of design justice, cultural inputs, dialogic partnership and citizen science. It further decolonizes epistemic traditions of knowledge through humility and collaboration.
3. *Enabling Fair (Just) Ecosystems*: Through joint and strategic interventions by grassroots stakeholders and thinkers, namely survivors, journalists, researchers, law enforcement and civil society actors, powerful stakeholders like policymakers and social media companies would be fundamentally accountable in building a justice ecosystem based on “critical transparency” (Google, 2021).

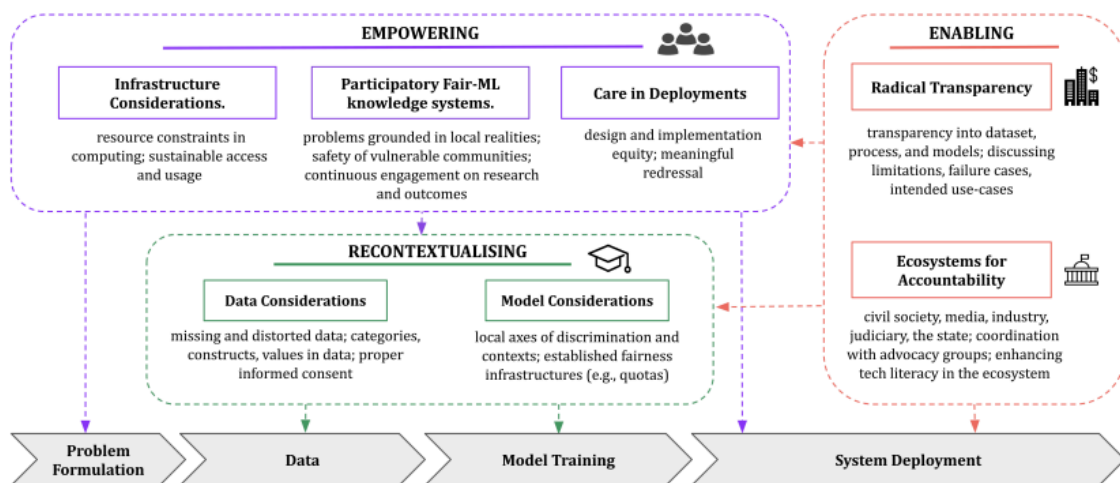


Figure 13: A visual Project Design building the foundation of the DSE, taken from [Google, 2021](#).

In the context of the theoretical framework and the design justice, the DSE would work as follows:

1. A survivor having experienced online misogyny will use the platform, operating very much like a website, to share their story with the link (citizen science, cultural inputs). They will be asked several optional questions and according to their needs, connected with a community of support, with mental health professionals or pursue legal action (design justice, dialogic partnership).
2. The platform will crawl data from the link, and based on the survivor’s inputs, thematically categorize the experience. The data would be converted into a dashboard (cultural inputs, algorithmic justice, digital anthropology).

- The inputs from the dashboard will help many stakeholders, like journalists, researchers and advocates against online misogyny to inform their approaches. It will also help in building a cohesive structure for the survivor to access (algorithmic justice, design justice, dialogic partnership).

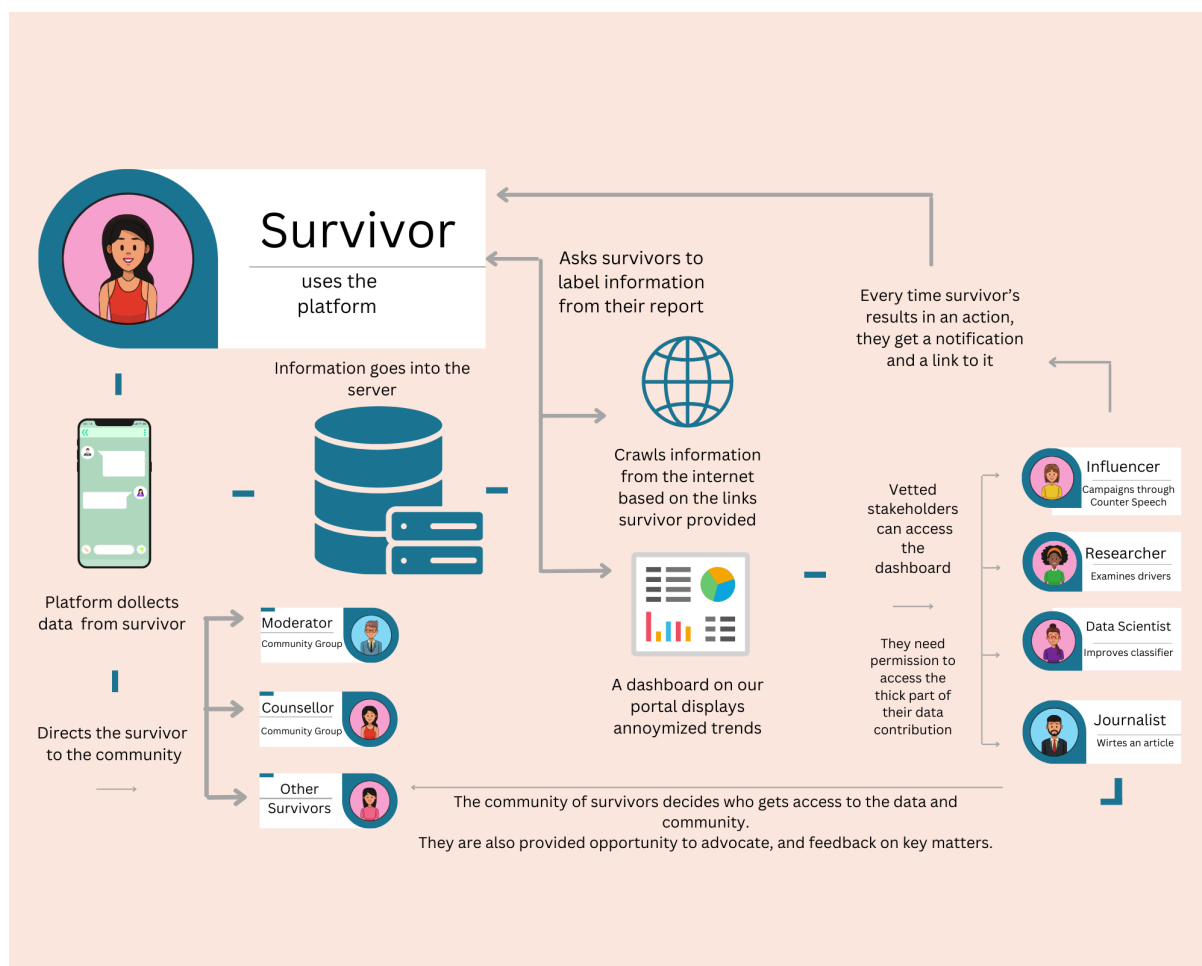


Figure 14: A visual mock-up of the DSE. (Dignity in Difference, 2024).

Based on this picture of the DSE, the assumption would be that survivors of online misogyny would be comfortable reporting incidents of online misogyny to the platform. Further, they would have access to digital tools and skills for navigating the internet and the platform. Another broad assumption would be that the survivors would be able to comfortably access quality services like mental health professional services, community of survivors and legal advisors.

Project Principles and SDGs:

The project is built on a set of foundational principles to ensure an intersectional and ethical approach to achieving safer digital spaces for women in South Asia. These principles are

aligned with the United Nations Sustainable Development Goals (SDGs). The SDGs are a global call to act on polycrises affecting the growth, sustainability and future of the world. The project aligns with:

1. SDG 5 - Gender Equality:

The fifth track aims to “achieve gender equality and empower all women and girls” ([United Nations, 2015](#)). Survivors using DSE to report an incident of online misogyny get actionable pathways by contributing their experience, data and get supported by a community of other survivors. DSE becomes a tool of empowerment and change to counter online misogyny and toxicity.

- a. Target: 5.1: “End all forms of discrimination against all women and girls everywhere.”
- b. Indicator: 5.1.1: “Whether or not legal frameworks are in place to promote, enforce and monitor equality and non-discrimination on the basis of sex.”
- c. Target 5.2: “Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation.”
- d. Indicator 5.2.1: “Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age.”
- e. Indicator 5.2.2: “Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence.”

2. SDG 9 - Industry, Innovation and Infrastructure:

The ninth track aims to “build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation” (UN, 2015). Through the DSE, survivors will see a dashboard showing different thematic categorizations of the incident they reported. The platform will tackle algorithmic failures through centering cultural inputs from the survivors. This builds comparative indicators of social media platforms’ response towards online misogyny, increases accuracy of context-specific classifiers around false positives and false negatives, and supports research on the boundary of digital and physical.

- a. Target 9.b: “Support domestic technology development, research and innovation in developing countries, including by ensuring a conducive policy

environment for, inter alia, industrial diversification and value addition to commodities.”

- b. Indicator: 9.b.1: “Proportion of medium and high-tech industry value added in total value added.”

3. SDG 10 - Reduced Inequalities:

The tenth track aims to “reduce inequality within and among countries” (UN, 2015). The DSE will help uncover physical and digital trends around gendered hate campaigns, and shape civil society interventions to reduce social, and political inequality and transform exclusive digital spaces.

- a. Target: 10.3: “Ensure equal opportunity and reduce inequalities of outcome, including by eliminating discriminatory laws, policies and practices and promoting appropriate legislation, policies and action in this regard.”
- b. Indicator: 10.3.1: “Proportion of population reporting having personally felt discriminated against or harassed within the previous 12 months on the basis of a ground of discrimination prohibited under international human rights law.”

4. SDG 16 - Peace, Justice and Strong Institutions:

The sixteenth track aims to “promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels”. (UN, 2015). Through the DSE, survivors experiential support, access to resources, agency in vetting stakeholders’ interests, and opportunities to advocate and influence online platforms and policymakers to build effective, safer digital ecosystems. Survivors co-create with Dignity in Difference to build inclusive and justice-centered design to reimagine accountability on online platforms.

- a. Target: 16.1: “Significantly reduce all forms of violence and related deaths everywhere.”
- b. Indicators: 16.1.4: “Proportion of population subjected to (a) physical violence, (b) psychological violence and (c) sexual violence in the previous 12 months.
- c. Target: 16.3: “Promote the rule of law at the national and international levels and ensure equal access to justice for all.”

- d. Indicators: 16.3.1: “Proportion of victims of violence in the previous 12 months who reported their victimization to competent authorities or other officially recognized conflict resolution mechanisms.”
- e. Target 16.6: Develop effective, accountable and transparent institutions at all levels.”
- f. Indicators: 16.6.2: “Proportion of population satisfied with their last experience of public services.”
- g. Target 16.7: “Ensure responsive, inclusive, participatory and representative decision-making at all levels.”
- h. Indicators: 16.7.2: “Proportion of population who believe decision-making is inclusive and responsive, by sex, age, disability and population group.”

Project Schedule:

The DSE prototype creation will take place in the latter half of 2024, immediately after the conclusion of the GEST Program. The assessment of the platform’s efficacy will be carried out in two parts: short-term and long-term to assess the stakeholders’ continuous evolution of inputs and insights and the long-term impact of the platform.

Timeline	Phase	Activity	Person Responsible	Oversight
June 2024	Protocol Design (India)	Vetting survivors and survivor-centric organizations, collecting inputs	Community Lead	Data Scientist
		Data Collation and Analysis	Community Lead + Data Scientist	Community Lead
		Platform design	Software Lead + Community Lead	Software Lead
July 2024		Running prototype tests in online communities	Data Scientist + Software lead	Community Lead
August 2024		Mid-Project Report	Community Lead + Software Lead	Software Lead
September 2024	Pilot Implementation	On-boarding stakeholders: lawyers, academicians, civil society	Community Lead	Data Scientist

October-November 2024		Running prototype with stakeholders through focus groups	Software Lead + Data Scientist	Community Lead
December 2024	Feedback Iteration	Stakeholders' assessment of project in contributing to data-informed interventions through surveys, likert-scale, and feedback mechanisms	Community Lead + Software Lead	Community Lead
		Final Project Report	Community Lead	Software Lead

Table 3: Project Timeline for the DSE from June-December 2024

Implementation Plan:

Capacity to implement:

The DSE will be created and managed by Dignity in Difference, which builds innovative tools, methods, and interventions to strengthen the resistance against online misogyny for and with survivors in South Asia. Our innovations in digital anthropology combine thick and big data to understand how and why algorithms fail in preventing online gender-based violence and lead to actionable measures across the digital ecosystem.

Our team consists of digital anthropologists, data analysts, advocacy specialists and community mobilizers. All of us come from a background in surviving violence online, which heavily informs our approach towards building any structural intervention. Our work has been recognized by United Nations Educational, Scientific and Cultural Organization (UNESCO), Liiv Center, UN Alliance of Civilizations (UNAOC), UN Office of Counter Terrorism (UNOCT).

We have previously built structural tools to inform key gaps in the digital ecosystem. Some of our tools include:

1. Dignity At Scale is an AI disinformation co-pilot for us to navigate complex tactics digital misinformation in anyone's locality and on the platforms they cherish. This low-cost, scalable solution gives deep insights into the online manipulation of truth. It could help women-led organisations and women leaders fight misinformation in India.
2. We empower young Global South feminist creators to identify and correct biases in toxicity classifiers, enhancing the effectiveness of online moderation through VoiceValor. It develops a practical and actionable framework of cultural insights, algorithmic accountability, and structural interventions to resist online misogyny, hate, and gender-based violence. It gives them access to online resources for support and 20 AI persons with domain-specific wisdom to create counter-narratives and receive coaching on strategies to navigate digital toxicity. For this tool, we received a confidence nomination by World Summit on the Information Society (WSIS) stakeholders, who evaluate how stakeholders are using technology to advance UN Sustainable Development Goals (SDGs).
3. We've trained 64 fellows and 26 research fellows from across South Asia to enable them with knowledge and tools to resist online misogyny and violence, using games

to teach complex issues. Ten of the members have went on build skills and research on online violence.

4. We've developed National Action Plans for Ministries in South Asia on countering violence through survivor-led co-creation on social impact issues.

Further, we are a young feminist organization that is value-aligned and encourages choices proposed by others on the team. We have used a consensus over voting approach for the rare conflict in decision-making.

Sustainability through stakeholder analysis:

Online misogyny is cultural, and requires a healthy ecosystem of actors supporting survivors. Much of the previous efforts have been limited to immediate support and counter-messaging, leaving a wide area of structural interventions away from survivor's influence. Survivors need opportunities to collaborate with stakeholders they cannot reach directly like journalists, policymakers, social media companies, law enforcement and researchers to determine:

1. Survivor knowledge and cultural inputs on online misogyny.
1. Demographic factors and correlation with accuracy or type of online misogyny.
2. Types of online misogyny and challenges in categorization for the platform.
3. The DSE's contribution to building data-informed interventions for OGBV.

Table 4 below gives an indepth understands of all the stakeholders involved.

Stakeholder	Interest	Influence	Engagement
Survivors of online misogyny (religious, cultural, gender, economic and sexual minorities)	Actionable-empowered pathways to achieve justice and safety	Advocate and influence high-power stakeholders	Contributing experience and data
Researchers and Data Scientists	Access to multi-lingual, labelled, time-series datasets with survivor-inputs	Creating new knowledge paradigms at the intersection of physical and digital	Use data to build comparative indicators of social media platforms' response towards online misogyny
Journalists	Pick early trends	Bring the impact	Interact with stories

Stakeholder	Interest	Influence	Engagement
	and survivor stories	of online misogyny in public discourse	of real incidents and people, with consent and anonymity to build counter-narratives.
Social Media Platforms	Trends from data to regulate online spaces	Remove offensive language, users, community networks and break the cycle of online misogyny	Training social media algorithms to identify specific online misogyny
Policymakers	Ability to create just and safe online spaces for gender minorities.	Shape public policies on online safety	Derive evidence-based insights through datasets and survivor narratives.
Law Enforcement	Ability to counter trends of online violence converting to offline harm and to ensure public safety	Use thick data to resist online misogyny in context of country law	Provide protection to survivors and stakeholders who do not choose anonymity against online and offline violence.

Table 4: Stakeholders analysis mapping of DSE

Partnerships:

To enhance more opportunities for the ecosystem to remain sustainable, Dignity in Difference will be working with a few partners on the project to manage cross-cutting and effective interventional needs. The needs include:

1. Data scraping and analysis
2. Legal and psychosocial support
3. Media advocacy

The organizations include:

Nippon Foundation:

Nippon Foundation is a grant-making organization based in Tokyo, Japan. They have been a knowledge partner of Dignity in Difference in conducting randomised control trials against online misogyny in South Asia through gamified interventions. They will support us in the data collection and analysis.

Counsel to Secure Justice:

It is an Indian non-profit that aims to provides legal, restorative justice support to survivors of sexual violence, including children. It will support the project by providing legal, psycho-social care to survivors who share their stories to the platform.

WISCOMP:

Women in Security, Conflict Management and Peace are a feminist pedagogical think-tank that work at the intersection of research, training, and practice. WISCOMP will support Dignity in Difference with building pedagogical awareness and practices on digital peace tools like the DSE.

Khabar Lahariya:

Khabar Lahariya is the first women-led, rural, nonprofit newsroom with a special focus on gender-based violence news in India. It will partner with DiD to share stories of survivors on its platform, while tying it to larger conversations around online safety and misogyny.

One Future Collective:

It is a nonprofit with heavy expertise on community, movement-building and building feminist structures. It will support DiD through leading community engagement initiatives, reviewing platform design and centering the voices of survivors in efforts to reduce online misogyny.

Budget:

Description	Unit	No of Units	Cost per Unit (€)	External funding (€)	Alumni fund contribution (€)	Total Cost (€)
-------------	------	-------------	-------------------	----------------------	------------------------------	----------------

Please insert a separate row for each cost adding more rows as needed	Describe the nature of the unit (e.g. rent per venue, printing costs per training material package, etc.)	(e.g. if one unit per beneficiary and 30 beneficiaries, No. of units=30)		Indicate project's external funding (not GEST funding)		
Personnel (Salaries, visiting expert, support personnel, trainer, consultants, etc.)						
Community Lead	Salary per month	6	600.00	3,600.00		3,600.00
Software Lead	Salary per month	6	600.00	3,600.00		3,600.00
Data Scientist	Salary per month	6	600.00	3,600.00		3,600.00
						-
						-
Sub-total				10,800.00	-	10,800.00
Equipment and supplies (procurement/or renting of small equipment and supplies needed for the project activities, incl. goods, material, etc.)						
Server hosting	per annum	1	3000		3,000.00	3,000.00
Computational costs	per annum	1	6000		6,000.00	6,000.00
Other digital services	per annum	1	1000.00		1,000.00	1,000.00
						-
						-
Sub-total				-	10,000.00	10,000.00
Community operations						
Regional internal and external meetings	2 per month	12	2000	2,000.00		24,000.00
						-
						-
						-
						-
Sub-total				2,000.00	-	24,000.00
Total				12,800.00	10,000.00	44,800.00

Table 5: Total costs for the DSE

Risk Management:

Risk management is a process to identify, ahead of time, potential factors that could affect the efficacy, implementation and sustainability of the project, in this case, the DSE.

Table 6 below provides an overall view of potential risks and mitigation strategies to ensure the future sustainability of the ecosystem.

Risk	Probability	Impact	Mitigation
Hesitancy of survivors in sharing their experiences	Low	High	Snowball sampling of survivors will ensure trust through established networks of connection. The network will encourage sharing of experiences and maintain transparent communication on purposes of data use.
Capturing the length and breadth of individual experiences in their contexts, while maintaining the scale	Low	High	Using the approach of digital anthropology with data science, algorithmic model will be trained with diverse, multilingual datasets
Handling sensitive data while maintaining privacy of survivors	Moderate	High	Data will be anonymized and encrypted with robust security protocols
Resistance in partnerships with tech companies	Moderate	High	Focus on corporate social responsibility and enhancing public image through cooperation
Shortfalls in funding	Moderate	High	Diversifying sources of funding through specialised grants and partnerships
Evolving online behaviors can be challenging to evaluate and measure	Moderate	High	Implementing adaptive monitoring and evaluation frameworks that use feedback loops to keep up to date with emerging trends.

Table 6: Risk analysis and mitigation for DSE

Evaluation and Impact Assessment:

The DSE will be monitored continuously through monthly feedback surveys with all of the stakeholders involved. The platform will be evaluated through quantitative and qualitative analysis. For quantitative analysis, statistical tests like ANOVA for demographic analysis and correlation studies, and Chi-square tests for comparing expected and observed categorization accuracy will be used. For qualitative evaluation, a thematic analysis for understanding user experiences and perceptions, particularly focusing on the cultural understanding of online misogyny will be used as a benchmark to understand the efficacy of the platform.

Table 7 below presents the indicators to be used to meet the targets set.

Monitoring and Evaluation	Measure	Method	Timeline	Responsible Personnel
Stakeholders: Survivors				
To what extent did survivors find DSE helpful in reporting cases of online misogyny	# of survivors accessing legal, psycho-social services	# of queries, pictorial evidence shared to the back-end, Pre/post assessment report of satisfaction, feedback surveys and likert-scale questions	Every month-end	Project Manager
Stakeholders: Researchers and Data Scientists				
To what extent did researchers and data scientists find DSE helpful in training algorithms to build multi-lingual datasets	# of researchers and data scientists using data from DSE to build comparative indicators of social media platforms'	Feedback surveys, likert-scale questions, focus group discussions, pre/post assessment report of learnings	Every month-end	Project Manager, Data Scientist

Monitoring and Evaluation	Measure	Method	Timeline	Responsible Personnel
Stakeholders: Survivors				
	response towards online misogyny			
Stakeholders: Journalists				
To what extent did journalists find DSE helpful as a structural interventions for survivors of online misogyny	# of journalists requesting access to survivors, referencing DSE in their articles with positive sentiments/statements	Feedback surveys, likert-scale questions, focus group discussions, pre/post assessment report of learnings	Every month-end	Project Manager
Stakeholders: Social Media Platforms				
To what extent did social media platforms find DSE helpful in providing insights on evolving trends of online misogyny	# of social media companies requesting permission to use dataset to train their algorithm	Feedback surveys, likert-scale questions, focus group discussions, pre/post assessment report of learnings	Every month-end	Project Manager
Stakeholders: Policymakers				
To what extent did policymakers find DSE helpful in providing key insights to shape public policy on online safety towards gender minorities	# of policymakers expressing interest for collaboration ,	Feedback surveys, likert-scale questions, focus group discussions, pre/post assessment report of learnings	Every month-end	Project Manager

Monitoring and Evaluation	Measure	Method	Timeline	Responsible Personnel
Stakeholders: Survivors				
	partnerships			
Stakeholders: Law enforcement				
To what extent did law enforcement find DSE helpful in providing actionable pathways in countering online misogyny	# of cyber police divisions expressing interest for collaboration , partnerships	Feedback surveys, likert-scale questions, focus group discussions, pre/post assessment report of learnings	Every month-end	Project Manager

Table 7: Monitoring and Evaluation system for DSE

Reporting, Communication and Advocacy:

The process of reporting will take place after sufficient use and feedback of the DSE has been determined internally. The expected timeline of the internal report for dissemination and discussion will be in December. Dignity in Difference will also disseminate an external report for stakeholders working at the intersection of tech and society nationally and globally, through forums, networks and social media to build awareness and future partnerships with similar stakeholders.

Table 8 gives an overview of the reporting process

Timeline	Type	Tool	Detail
December 2024-January 2025	Internal	Written Report	Report on process of DSE establishment and feedback from survivors
January 2025	External	Social Media communication	Short communication briefs for followers and

Timeline	Type	Tool	Detail
			actors in the ecosystem to take note
February 2025	External	Written Report, Video	Dissemination report on structural interventions of Dignity in Difference and multi-stakeholder recommendations
March 2025 - May 2025	External	Written Policy brief	National and international policy briefs for dissemination in multistakeholder forums

Table 8: Reporting process for the DSE by Dignity in Difference

Conclusion: 'S' is for survivor in shero and heroes

As a child, I loved superhero movies. Born and brought up in a monotheistic perspective of the world made it easier for me to imagine that one man could change it in a span of two hours, or even two months. The savior was especially a man— handsome, elegant, chivalrous and intelligent.

I grew up and carried that dream, sometimes imagining myself as a nonbinary superhero, or the female prophet who would crash the system and build a new one from its ashes. Much like a phoenix, my voice would be fierce, powerful and undeniable to everyone present. But as I began professional work, I was struck by my own limitations of knowledge. And god, did I have ego. It was hard to be curious, and even harder to accept feedback. But looking over my shoulder, if I hadn't, I wouldn't have applied for the GEST program.

We like to think complex issues need powerful, one-line solutions. Intelligent people in our lives tend to know exactly what to say and how to say it. But complexity is confusing in systems, and clumsy. It can redefine the stark boundaries in our minds, and make us question the limits of our imagination. It can mean accepting that our vice, is someone's virtue, and our hero is someone else's villain.

These confusions can make us scared to act, and go back to the one-line solutions that are easy to imagine. That's part of why right-wing politics is so powerful. It helps us silence the critic within us. But criticism, in constructive, strategic ways, is necessary and can help us get over our main-character syndrome.

Now that I think about it, the superheroes who taught me the most and that I still cling onto were Teenage Mutant Ninja Turtles— four mutant turtles adopted by a rat and taught kung-fu. Apart from my obvious affinity towards martial arts, I loved seeing them work together as a team, and build upon each other's strengths. I felt glimpses of that in the GEST program, when I was paired with my group for five months and given assignments. It helped me navigate the long-forgotten art of patience, empowerment and co-creation. This thesis, the project, and all the information here, is also, in some parts, a result of those co-learning, co-discussing, and co-realizing spaces.

Dignity in Difference will continue to build on the DSE, and expand it in other regions of the Global South, based on results of our prototype in India. Our aim is counter digital toxicity through different structural interventions, and maintain the open-access nature of our tools for other countries to adapt and use.

I hope the project, at least facilitated your imagination as a reader, and was an opportunity for you to re-imagine the system you're confused by. As one of our professors in the "Gender, Environment and Climate Change" module said, "Clumsy solutions aren't perfect, but they're a step to building better ones."

Thank you.

Bibliography:

1. Rogers, C. (2016). Of monsters and misogyny. Retrieved from https://nmu.edu/english/sites/english/files/d7files/WritingAwards/Barnard/Of_Monsters_and_Misogyny.pdf
2. Elliott, K., Gray, S., & Nussbaum, J. (2016). Online misogyny. *Feminist Media Studies*, 16(4), 634-646. <https://doi.org/10.1080/14680777.2016.1120490>
3. Silver, L., Johnson, C., & Jiang, J. (2022, December 6). Views of social media and its impacts on society in advanced economies - 2022. Pew Research Center. <https://www.pewresearch.org/global/2022/12/06/views-of-social-media-and-its-impacts-on-society-in-advanced-economies-2022/>
4. Data2X. (2021). Big data and the well-being of women and girls: Applications to the COVID-19 pandemic response. https://data2x.org/wp-content/uploads/2021/03/UCSD-Brief-3_BigDataGenderCOVID19SouthAsianMisogyny.pdf
5. Thomas, K. (2020). Misogyny and social media. *Sociology Compass*, 14(8), e12577. <https://doi.org/10.1111/soc4.12577>
6. Sharma, G. (2023, February 28). India manosphere: Men's rights influencers and gender divide. *Livemint*. <https://lifestyle.livemint.com/news/big-story/india-manosphere-mens-rights-influencers-gender-divide-111709889805465.html>
7. Association of School and College Leaders (ASCL). (n.d.). Safer scrolling: Improving mental health online. <https://www.ascl.org.uk/ASCL/media/ASCL/Help%20and%20advice/Inclusion/Safer-scrolling.pdf>
8. Marsh, S. (2024, March 1). There are some really extreme views: Young people face onslaught of misogyny online. *The Guardian*. <https://www.theguardian.com/society/2024/mar/01/there-are-some-really-extreme-views-young-people-face-onslaught-of-misogyny-online>
9. Global Action Plan. (n.d.). Desensitisation in social media. <https://www.globalactionplan.org.uk/online-climate/safer-socials/desensitisation>
10. Zhang, S., Lee, M. K., & Huang, Y. (2022). Analysis of online misogyny using data science approaches. *Social Network Analysis and Mining*, 12, 51. <https://doi.org/10.1007/s13278-022-00951-3>
11. Cambridge Consultants. (2023). AI content moderation technologies. Ofcom. https://www.ofcom.org.uk/data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf
12. Gillespie, T. (2020). Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. *New Media & Society*, 22(5), 874-889. <https://doi.org/10.1177/1461444820912540>

13. Chazan, G. (2022, March 3). AI bias and the challenge of ethical artificial intelligence. Financial Times.
<https://www.ft.com/content/afeb56f2-9ba5-4103-890d-91291aea4caa>
14. Miller, R. (2023, February 22). AI bias: The organised struggle against automated discrimination. The Conversation.
<https://theconversation.com/ai-bias-the-organised-struggle-against-automated-discrimination-223988>
15. Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.
https://www.researchgate.net/profile/Tarleton-Gillespie/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media/links/5dfcfa3a6fdcc2837318e10/Custodians-of-the-Internet-Platforms-Content-Moderation-and-the-Hidden-Decisions-That-Shape-Social-Media.pdf
16. Thompson, C. (2016, August 11). An open letter to the Internet. Wired.
<https://www.wired.com/2016/08/open-letter-to-the-internet/>
17. London School of Economics and Political Science. (n.d.). Perpetuating gendered inequality online.
<https://www.lse.ac.uk/research/research-for-the-world/society/perpetuating-gendered-inequality-online>
18. Jane, E. A. (2018). Misogyny online: A short (and brutish) history. Feminist Media Studies, 18(4), 482-499. <https://doi.org/10.1080/14680777.2018.1447345>
19. Mozilla Foundation. (2022). Internet health report 2022: Facts and figures.
<https://2022.internethealthreport.org/facts/>
20. Stanford University. (2022). AI Index 2022 report.
https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf
21. National University. (n.d.). Research datasets and resources.
<https://resources.nu.edu/researchprocess/datasets>
22. Van Dijck, J., & Poell, T. (2023). Platforms and algorithms: Governing digital societies. AI & Society, 38, 1-10. <https://doi.org/10.1007/s00146-023-01804-z>
23. West, D. M. (2022). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings Institution.
<https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>
24. UNESCO. (2020). The chilling: Global trends in online violence against women journalists. <https://en.unesco.org/sites/default/files/the-chilling.pdf>
25. UN Women. (2022). Accelerating efforts to tackle online and technology-facilitated violence against women and girls.
https://www.unwomen.org/sites/default/files/2022-10/Accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls-en_0.pdf

26. Kemp, S. (2024). Digital 2024: India. DataReportal.
<https://datareportal.com/reports/digital-2024-india>
27. Statista. (n.d.). Religion in India - Statistics & facts.
<https://www.statista.com/topics/4548/religion-in-india/#topicOverview>
28. Rao, A. (2022, June 5). Modi, India, and the U.S.: The far-right and the 'Great Replacement'. Jacobin.
<https://jacobin.com/2022/06/far-right-modi-india-us-great-replacement>
29. Al Jazeera. (2024, January 22). Babri Mosque to Ram Temple: A timeline from 1528 to 2024.
<https://www.aljazeera.com/news/2024/1/22/babri-mosque-to-ram-temple-a-timeline-from-1528-to-2024>
30. Pew Research Center. (2021, June 29). Religion in India: Tolerance and segregation.
<https://www.pewresearch.org/religion/2021/06/29/religion-in-india-tolerance-and-segregation/>
31. Roy, S. (2021, April 8). The problem with India's 'Love Jihad' laws. The Conversation.
<https://theconversation.com/the-problem-with-indias-love-jihad-laws-152675>
32. NPR. (2021, October 10). India's Hindu-Muslim interfaith weddings face religious conversion challenges.
<https://www.npr.org/2021/10/10/1041105988/india-muslim-hindu-interfaith-wedding-conversion>
33. Gettleman, J., & Raj, S. (2019, April 11). Modi's reelection in India. The New York Times. <https://www.nytimes.com/2019/04/11/world/asia/modi-india-elections.html>
34. India Hate Lab. (2024, February 25). Hate speech events in India: 2023 annual report.
<https://indiahatelab.com/2024/02/25/hate-speech-events-in-india-2023-annual-report/>
35. CARE Collective Against Racism. (2022, January 26). White Paper Issue
36. Rest of World. (2023). 3 minutes with Raqib Hameed Naik: Hindutva watch.
<https://restofworld.org/2023/3-minutes-with-raqib-hameed-naik-hindutva-watch/>
37. UNICEF. (n.d.). Gender equality in India.
<https://www.unicef.org/india/what-we-do/gender-equality>
38. Pew Research Center. (2022, March 2). How Indians view gender roles in families and society.
<https://www.pewresearch.org/religion/2022/03/02/how-indians-view-gender-roles-in-families-and-society/>
39. IDR. (2023). Closing the gender gap at work.
<https://idronline.org/close-the-gender-gap-at-work>
40. Article 19. (n.d.). Internet statement adopted.
https://www.article19.org/data/files/Internet_Statement_Adopted.pdf
41. Kemp, S. (2023). Digital 2023: India. DataReportal.
<https://datareportal.com/reports/digital-2023-india>

42. GSMA. (2022). The Mobile Economy: India.
<https://www.gsma.com/asia-pacific/wp-content/uploads/2022/09/India-report-FINAL-WEB.pdf>
43. Internet Democracy Project. (n.d.). Women and verbal online abuse in India.
<https://cdn.internetdemocracy.in/idp/assets/downloads/reports/women-and-verbal-online-abuse-in-india/Internet-Democracy-Project-Women-and-Online-Abuse.pdf>
44. Countering Backlash. (n.d.). Digital spaces must be safer for Muslim women in India.
<https://counteringbacklash.org/digital-spaces-must-be-safer-for-muslim-women-in-india/>
45. Firstpost. (2019, August 8). Surge in misogyny on social media as Indian men look for Kashmiri women for marriage, women's right activist decry objectification.
<https://www.firstpost.com/india/surge-in-misogyny-on-social-media-as-indian-men-look-for-kashmiri-women-for-marriage-womens-right-activist-decry-objectification-7135231.html>
46. Simmons-Duffin, S. (2022, February 21). Online auctions of Muslim women in India prompt calls for action. MIT Technology Review.
<https://www.technologyreview.com/2022/02/21/1046052/online-auctions-muslim-women-india/>
47. The Wire. (2022, January 5). Indian Muslim woman auction: 'Bulli Bai'.
<https://thewire.in/communalism/indian-muslim-woman-auction-bulli-bai>
48. Stewart, A. (2018, October 5). #MeToo movement history: Timeline & year after Weinstein. Refinery29.
<https://www.refinery29.com/en-us/2018/10/212801/me-too-movement-history-timeline-year-weinstein>
49. Outlook India. (2020, October 15). #MeToo movement in India: A timeline of key events.
<https://www.outlookindia.com/national/metoo-movement-in-india-a-timeline-of-key-events-news-276260>
50. Raghavan, M. (2018). DALIT GOES ONLINE: THE CONSTRUCTION OF IDENTITY AND SOCIAL SPACE.
https://www.researchgate.net/publication/328119338_DALIT_GOES_ONLINE_THE_CONSTRUCTION_OF_IDENTITY_AND_SOCIAL_SPACE
51. Amin, S. (2021, September 15). Love on the beat: How India's rappers are countering anti-Muslim hate. TRT World.
<https://www.trtworld.com/magazine/love-on-the-beat-how-indias-rappers-are-countering-anti-muslim-hate-16670462>
52. Rai, A., & Kumar, P. (2022). Navigating digital spaces for social justice. Palgrave Communications, 8(22). <https://www.nature.com/articles/s41599-022-01172-x>
53. International Dalit Solidarity Network (IDSN). (2021). Caste hate speech report.
<https://idsn.org/wp-content/uploads/2021/03/Caste-hate-speech-report-IDSN-2021.pdf>

54. European Parliament. (2018). Cyber violence and hate speech online against women. [https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU\(2018\)604979_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf)
55. Software Freedom Law Center. (2024). Online gender-based violence guide. https://sflc.in/wp-content/uploads/2024/01/OGBV-Guide-in-English_compressed.pdf
56. Radha, S. (2018). Electronic violence against women. IT for Change. https://projects.itforchange.net/e-vaw/wp-content/uploads/2018/01/Dr_Radha.pdf
57. Bhat, A. (n.d.). Rethinking legal and institutional approaches to sexist hate speech. IT for Change. https://itforchange.net/sites/default/files/1883/Aparna-Bhat-Rethinking-Legal-Institutional-Approaches-to-Sexist-Hate-Speech-ITfC-IT-for-Change_0.pdf
58. Gupta, J. (2023, May 22). Technology-facilitated gender-based violence against women journalists. Feminism in India. <https://feminisminindia.com/2023/05/22/technology-facilitated-gender-based-violence-women-journalists/>
59. IT for Change. (n.d.). The judiciary's tryst with online gender-based violence (Executive Summary). <https://itforchange.net/sites/default/files/2190/The%20Judiciary%27s%20Tryst%20with%20Online%20Gender-Based%20Violence%20%28Executive%20Summary%29.pdf>
60. D'Souza, A. (n.d.). Hidden figures: A look at technology-mediated violence against women in India. IT for Change. <https://itforchange.net/index.php/hidden-figures-a-look-at-technology-mediated-violence-against-women-india>
61. ScienceDirect. (n.d.). Topics on action research. <https://www.sciencedirect.com/topics/social-sciences/action-research>
62. Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. MIT Press. <https://direct.mit.edu/books/oa-monograph/4605/Design-JusticeCommunity-Led-Practices-to-Build-the>
63. Wexler, M. N. (2021). Social innovations in community engagement. Social Innovations Journal, 1(1), Article e593
64. Raghvani, V. (2021). User survey about exposure of hate speech among Instagram users in India. https://www.researchgate.net/publication/354009599_User_Survey_about_Exposure_of_Hate_Speech_among_Instagram_Users_in_India
65. Kumar, P. (2023). Racist, sexist, casteist: Is AI bad news for India? Context News. <https://www.context.news/digital-rights/racist-sexist-casteist-is-ai-bad-news-for-india>

66. Lee, H., & Horsley, J. S. (2016). Digital discourse: The impact of modern technology on communication. *Communication, Culture & Critique*, 9(2), 213-230.
<https://doi.org/10.1080/10304312.2016.1166560>
67. Artz, M. (n.d.). What is digital anthropology?
<https://www.mattartz.me/what-is-digital-anthropology/>
68. National Geographic. (Year). Citizen science and its implications.
<https://education.nationalgeographic.org/resource/citizen-science-article/>
69. United Nations. (n.d.). Sustainable Development Goal 5: Achieve gender equality and empower all women and girls. <https://sdgs.un.org/goals/goal5>

Appendices:

Appendix A: Interview Questions:

Table 9 shows the interview questions I made and used during interviewing Muslim women survivors on their experiences surrounding harassment.

Schedule	Time (60 en toto)	Activity	Notes/ Questions
Welcome and Introduction	5 minutes	Briefly introduce the session's purpose; outline structure	
Ice-breaker	5 minutes		Share one positive thing that happened to you this week.
Sharing Experiences	10 minutes	Encourage sharing an incident of online harassment	
Recounting experiences			
Emotional Impact	10 minutes	Explore feelings and emotional responses during and after the reporting process.	-How does it make you feel? -What emotional change did you experience? -How did you manage to move back into a 'regulated state' (a neutral state where one's emotions are under one's control)? Explain taking a meta-moment .
Emotional Regulation	10 minutes	Discuss various strategies for managing emotional	

Schedule	Time (60 en toto)	Activity	Notes/ Questions
		<p>responses.</p> <ul style="list-style-type: none"> - Explore actions taken in response to harassment, such as using silence 	
Pilot Intervention	5 minutes	<ul style="list-style-type: none"> - Participants select one toxic comment and input it into BridgeGPT. - Observe and discuss the tool's response. 	
Closing	15 minutes	<p>Summarise key insights and ask for feedback. Regulate any feelings of pain together.</p>	

Table 9: Interview questions used during action-research phase for DSE

Appendix B: Project Logframe:

Table 10 gives a more detailed overview of the Project Logframe for DSE.

Objective	Indicators	Means of measurement	Assumptions
Goal: Algorithmically just structural interventions for survivors of online misogyny in India using thick-big data	1. Rate of algorithmic success in reducing impact of online misogyny for survivors after six months	1. Feedback surveys and interviews with survivors to assess satisfaction	1. Multilingual datasets with cultural nuances will aid in efforts to prevent online harassment
	2. Average satisfaction level of survivors after six months	2. Using data analytics tools to assess pre and post-interventions scenarios	2. DSE is capable of complex data processing and adaptation to new contexts.
	3. Measurement fairness and equity in interventions every 2 months	3. Fairness audits to ensure bias reduction against groups	3. Engagement of survivors and other stakeholders in designing the DSE.
	4. Degree of understanding of algorithmic decisions	4. Maintaining monthly transparency logs on decisions taken for algorithmic fairness	4. Interviews and collected data maintain privacy and ethics of collation.
Outcome: DSE algorithms identify and detect against online misogyny in India	1. Number of instances when the algorithm correctly labeled scenarios of online harassment per month	1. Automated moderation with human oversight to assess accuracy of algorithm	1. Survivors have access to digital equipment to access the platform.
	2. Average response time of the algorithm in detecting online content per run time	2. Reviewing system logs to measure accuracy and response time	2. Algorithms can adapt to new and evolving knowledge.
	3. Proportion of true positives to false positives	3. Collecting and analyzing feedback surveys from survivors to assess efficacy.	3. Survivors accurately report the cases.

Output 1.1: Establishment of DSE for survivors of online misogyny in India on the internet	1. DSE is accessible to survivors every day	1. Tracking usage patterns, engagement levels and demographics through web analytics tools	1. Survivors have access to Internet and technical equipment to access the platform.
	2. Average frequency of interactions with platform per month	2. Collecting and analyzing feedback surveys from survivors to assess usability.	2. Survivors have adequate digital literacy skills to understand and navigate the DSE.
Outcome 2: The DSE supports survivors with quality services.	1. Average level of satisfaction reported by the survivor	1. Impact assessments to understand the benefits like safety and empowerment every three months	1. DSE protects survivors' information from being accessed without authorization.
	2. Percentage of queries resolved with survivor satisfaction	2. Backend monitoring of analytics to see response rate	2. Adequate resources to maintain quality of financial, technical and social services.
	3. Rate of returning survivors to use DSE	3. Focus group discussions and audit reviews every 3 months	3. Survivors have easy access to effective feedback mechanisms.
Output 2.1: Platform builds a culturally-trained algorithm based on data collation.	1. Accuracy of algorithm in identifying cultural nuances	1. Standard metrics to quantify algorithm's bias and accuracy	1. Availability of high-quality, culturally rich data.
	2. Number of broad range of cultural contexts the algorithm covers	2. Researchers analyze datasets to include culturally rich data	2. Stakeholders and experts guide the process of determining absence or presence of cultural nuances
	3. Measures of bias in algorithm	3. Conducting regular audits per month	3. Algorithm is able to continuously learn
Output 2.2: Survivor community builds initiatives countering online misogyny.	Total number of initiatives	Record maintained of all initiatives	Adequate support form survivors