



# **Human Factors**

SAVA -Society against Violence and Abuse

**UoM-17VCS012**

VCSB

# IDEA 1

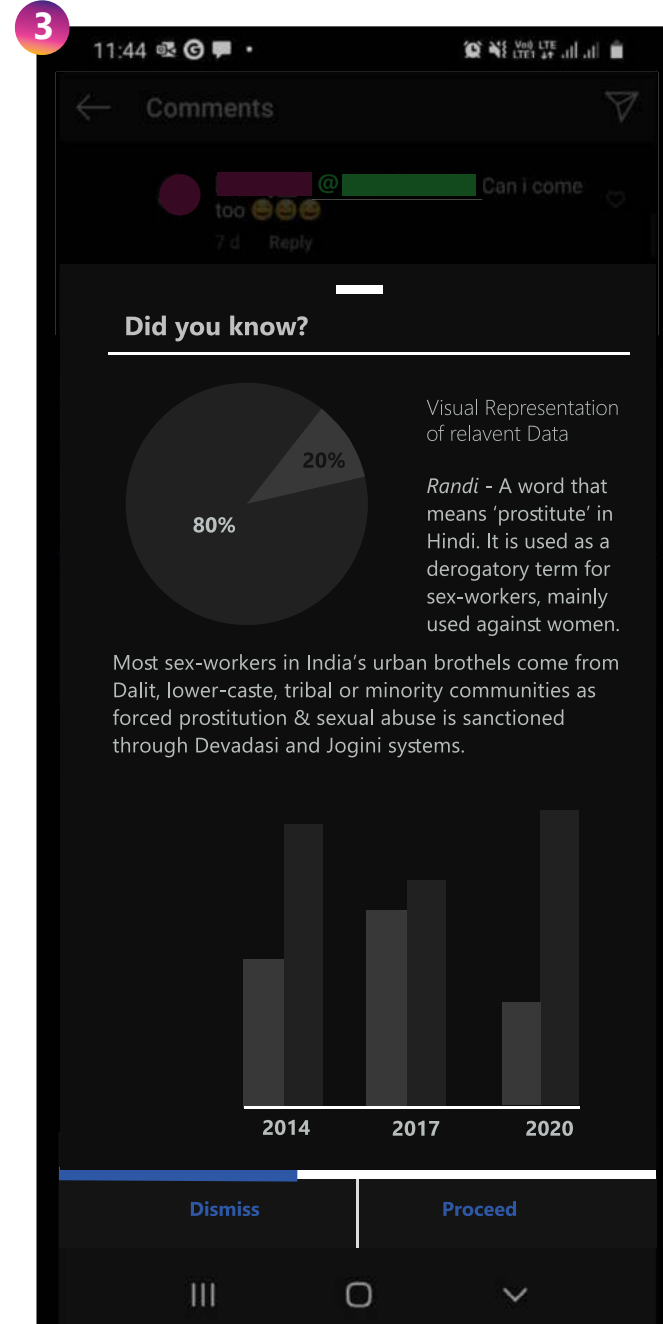
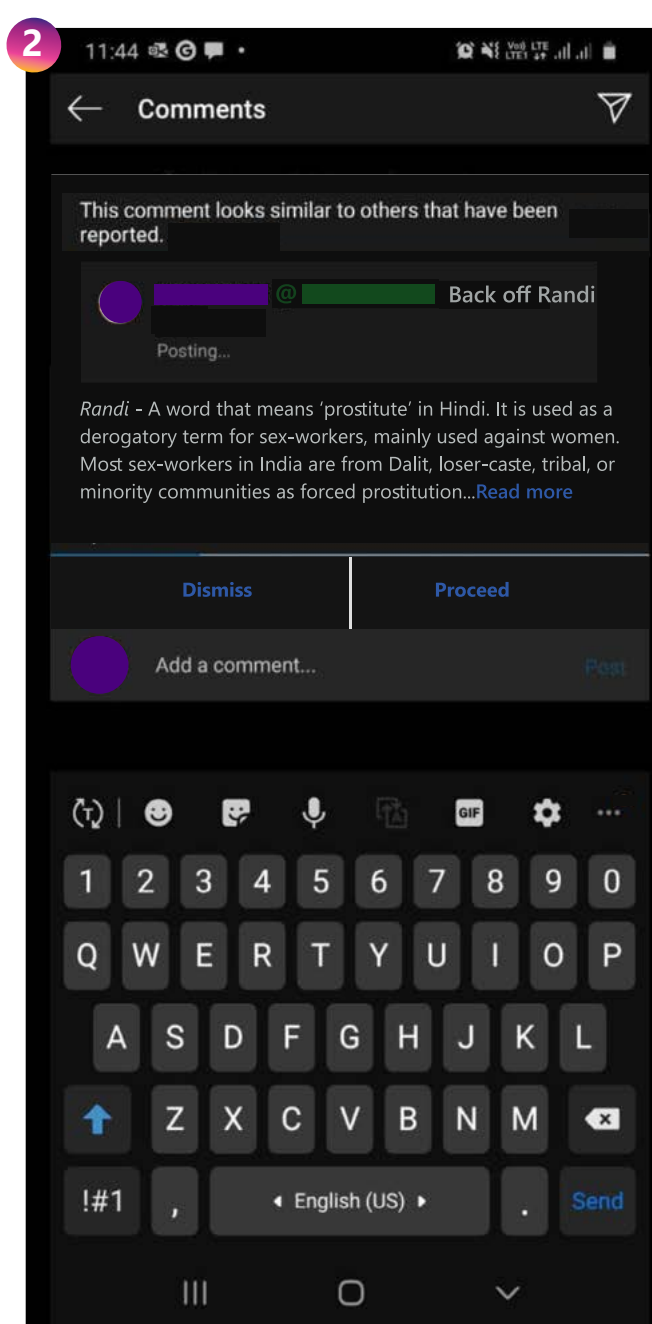
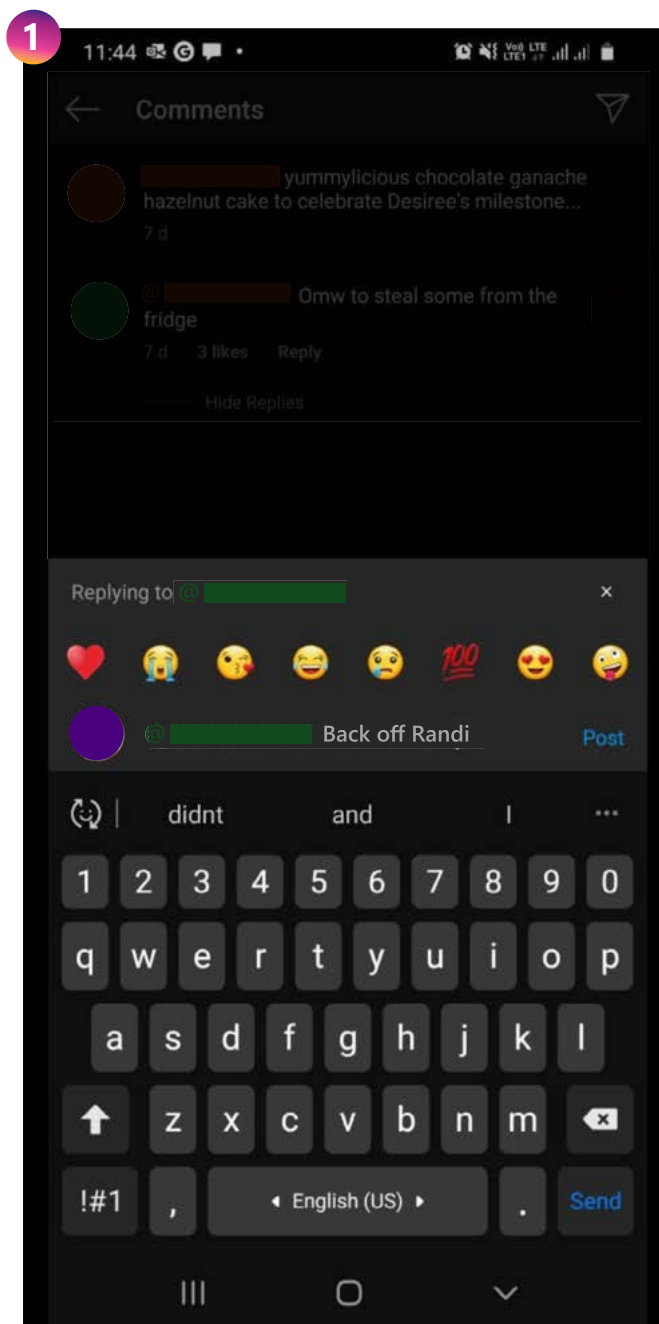
Reverse reporting system by putting abuser through similar harrowing steps that usually abused person needs to go through.

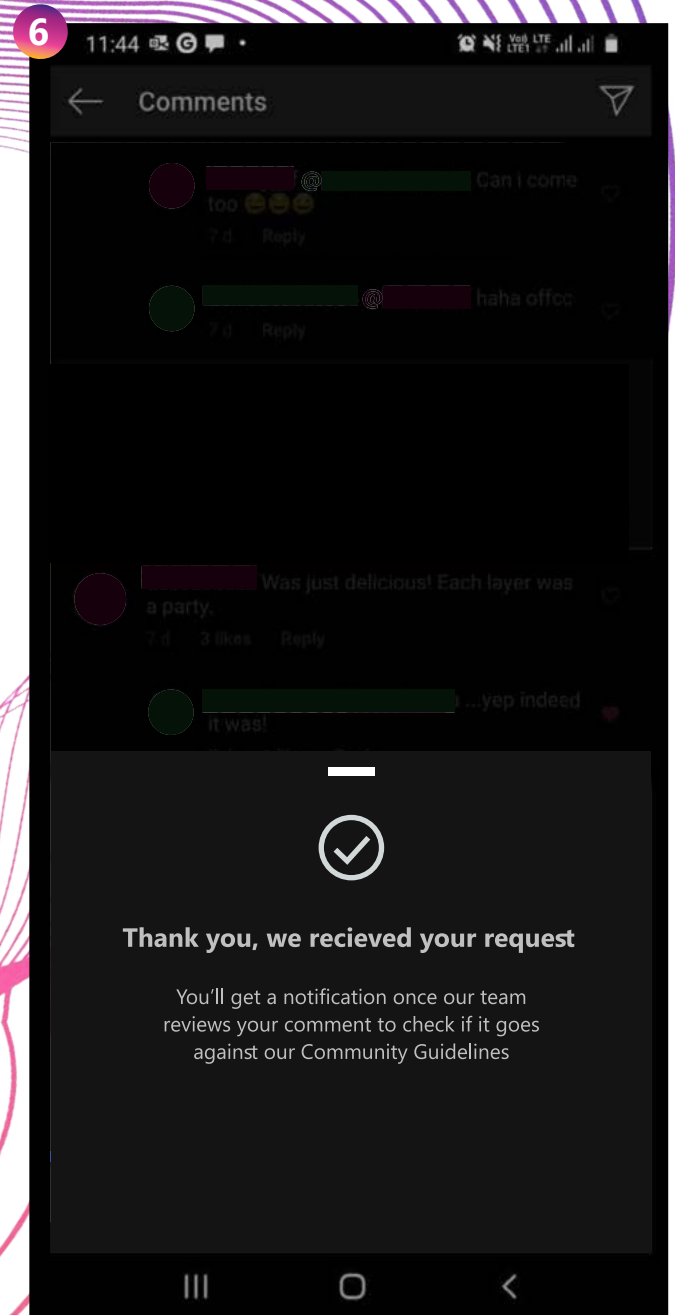
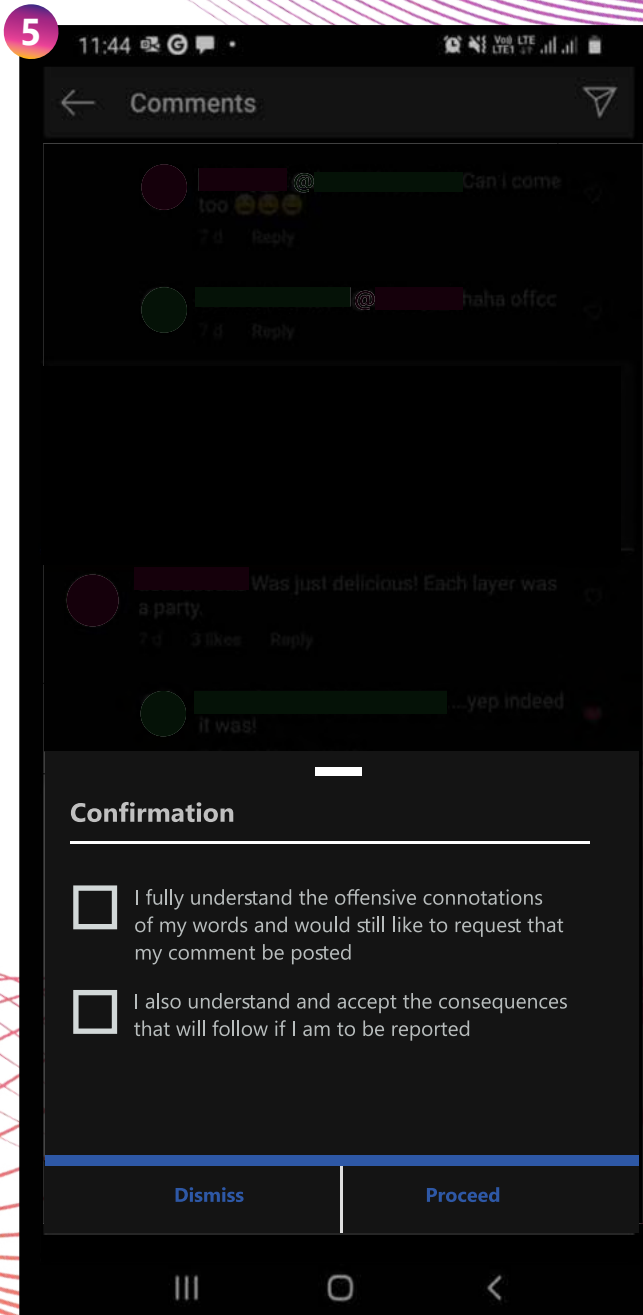
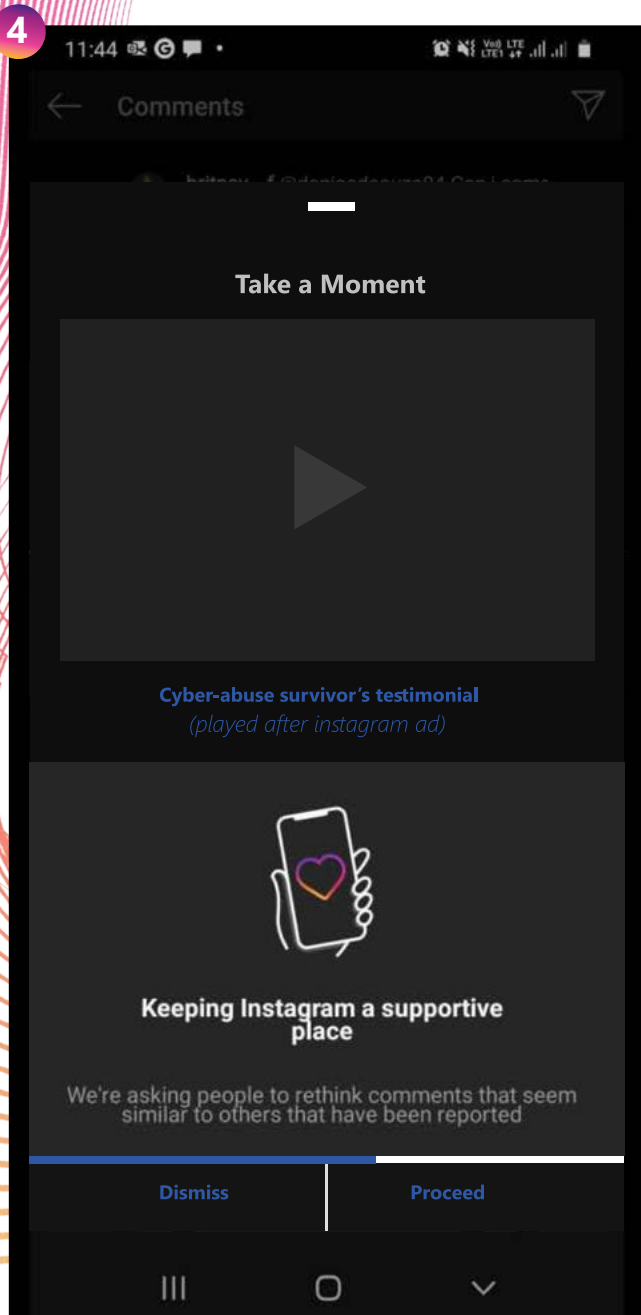
#self-surveillance  
#archeologyofknowledge  
#habitus #communicativeaction

The number of steps to post an abusive comment is three while the number to report one is six. In adopting survivor-centric preventative measures in response to GBV, it is unfair that the onus to seek justice for the abused individuals/communities lies only with themselves. It is possible that survivors of online abuse have already been traumatized and have had this affect their mental health negatively. To ask them to concentrate their energy in a process that firstly, is very objective in nature and secondly, is one that doesn't guarantee action against their abuser doesn't only seem redundant but also reproduces the toxic idea that victims are responsible for the abuse they face. This proposed new feature for posting abusive comments aims at eliminating and minimizing the need to report by attempting to nip the abuse in the bud while also educating the abuser.

1. The desired outcome of self-surveillance is attempted through three ways in this process.
- The nature of posting abusive comments are usually time-bound and hence delaying the posting of the comment might make the abuser reconsider their words and surveil themselves.
  - the content designed for this process aims to appeal to the more humane side of the abuser. It involves asking the abuser to compulsorily watch a testimonial by an abuse survivor and also by asking them to confirm that they are willing to accept consequences that will follow if they are to be reported. The first attempts to invoke self-surveillance through empathy and the latter through some level of fear.

- Lastly, many users have lost confidence in Instagram's current reporting process as their efforts of reporting abusive content have been met with uncertain responses and lack of assurance of any kind of action. Hence it becomes easier to ignore and follow the path of least resistance. Reversing this system that otherwise invokes self-surveillance in the abused, the process will end with providing the same kind of uncertainty to the abuser leaving them in suspense with very little information about if their comment will be posted or not.
2. Casteist, racist, sexist, ableist slurs are used casually and often mindlessly. Providing information as reminders attempts to make the abusers question and become aware of their own habitus which makes it acceptable for them to use abusive language
  3. By reminding the abuser of information about abusive language that they are using, this idea uses the concept of archeology of knowledge by reminding them of how words over the years have become abusive because of its socio-cultural associations.
  4. Additionally this aims at communicative action as equipping the abuser with information of how their comments can be negatively impactful might serve as a learning process and give them the opportunity to do better. This process is also beneficial for Instagram as a platform as an advertisement that is deemed useful to the abuser could be detected by Instagram's algorithm and played right before the testimonial of the cyber-abuse survivor







## IDEA 2

If someone has 'sensitive content' warnings on, the post could have more specific descriptors about the content or the profile it originated.

#self-surveillance  
#cognitiveinterests

This idea looks at posts labeled as 'sensitive content' on Instagram, where posts that can trigger users with explicit imagery are blurred. However, this becomes counterproductive since the blurriness of the image only make users more curious to tap on it and end up getting scandalized with the graphic images/video, and possibly reporting it too. In this case, having more context through keywords about what the 'sensitive content' is really about, gives users more information, making them aware and allowing them to make an informed decision.

Instagram could also benefit from this since keywords can be used to improve image recognition, for it to describe the content without the need of a person. Taking a step further, we're proposing a "Hide" feature under each post, where users can opt to not see posts of a certain kind. We're also extending one of the ideas from the account creation process here, where users decide what kind of content they don't wish to see by customizing hashtags. Since hashtags make up a certain meaning about a post, they can enter the hashtags that they don't wish to see. This feature would be available in their profile page.

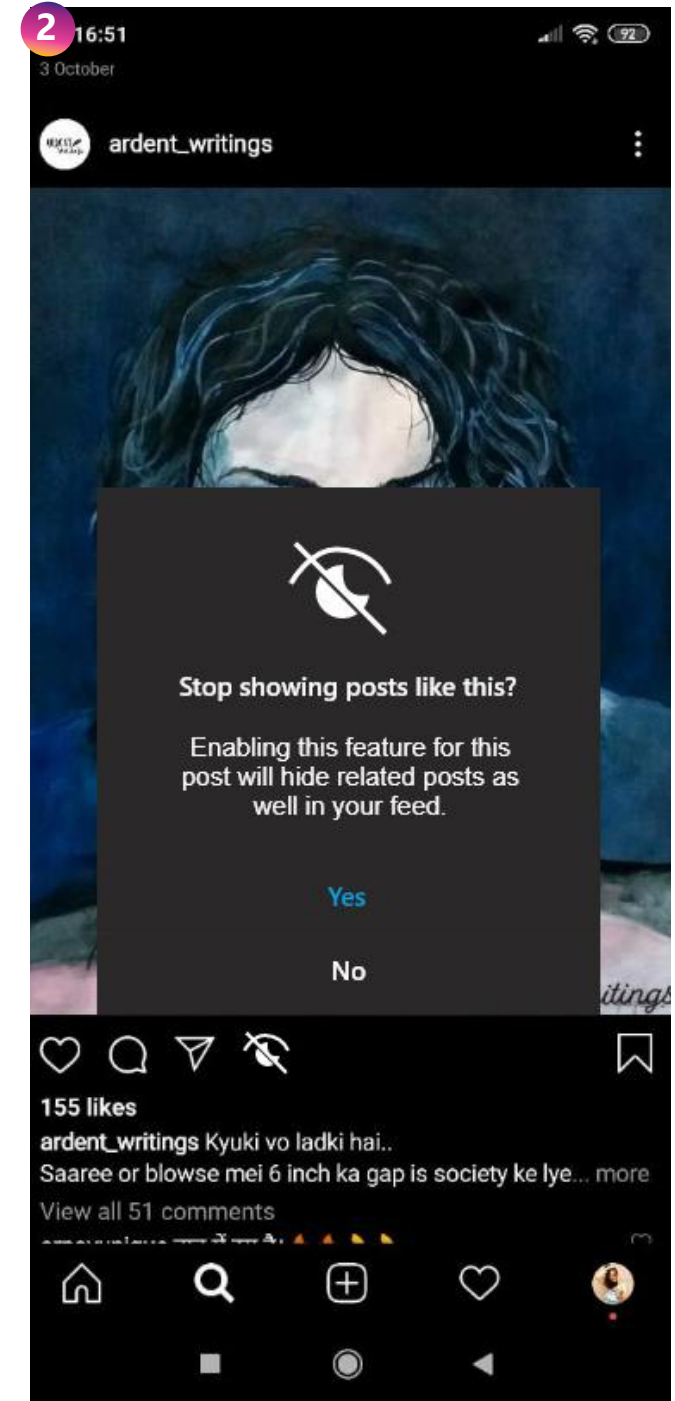
This feature could help reduce the number of reports made, making it easier for IG. The idea for this feature comes from Habermas' theory of cognitive interests. Cognitive interests are strategies that aim to guide people on how they acquire and use knowledge to pursue their interests. This specifically looks at emancipatory interest, where users overcome distortions through self reflection. This is an attempt

for users to achieve a sense of emancipation by curating the kind of content that they wish to see.

We're also proposing Instagram that they caution the user when self breach of content occurs on the "Discover" feature where they try looking up the kind of posts that they specified not to see in the first place through customizing hashtags. The idea of cautioning the user comes from Foucault's concept of self awareness and self surveillance. This is being imposed by the user alone and not anyone else. Instagram would just help in reminding the user about what kind of content they didn't want to view, in an attempt to prevent them from feeling uncomfortable or scandalized.

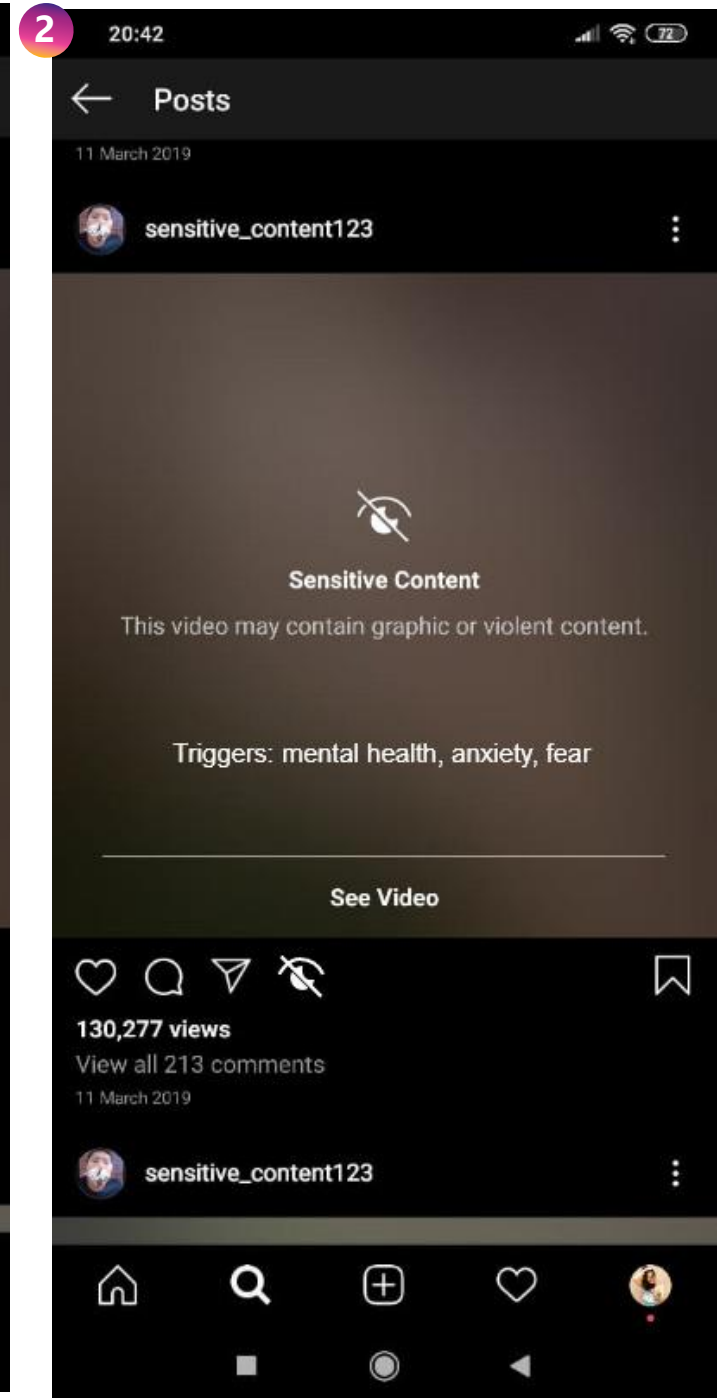
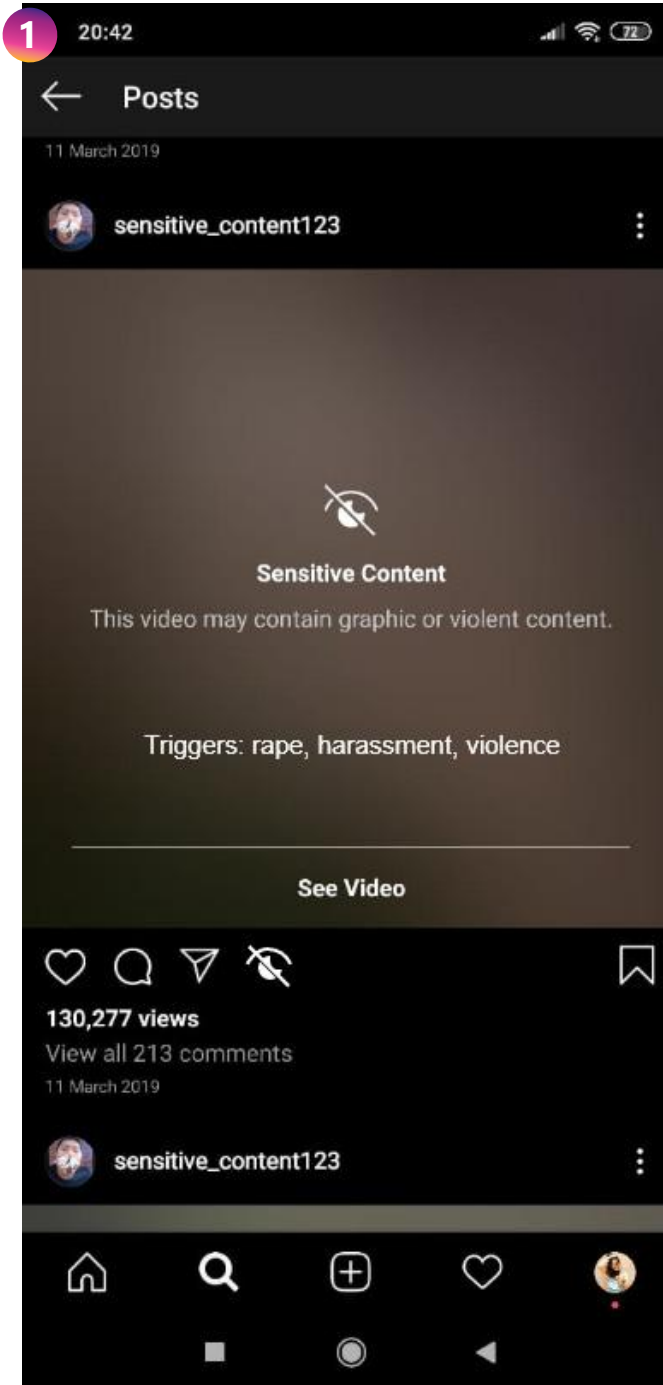
## Feature 1

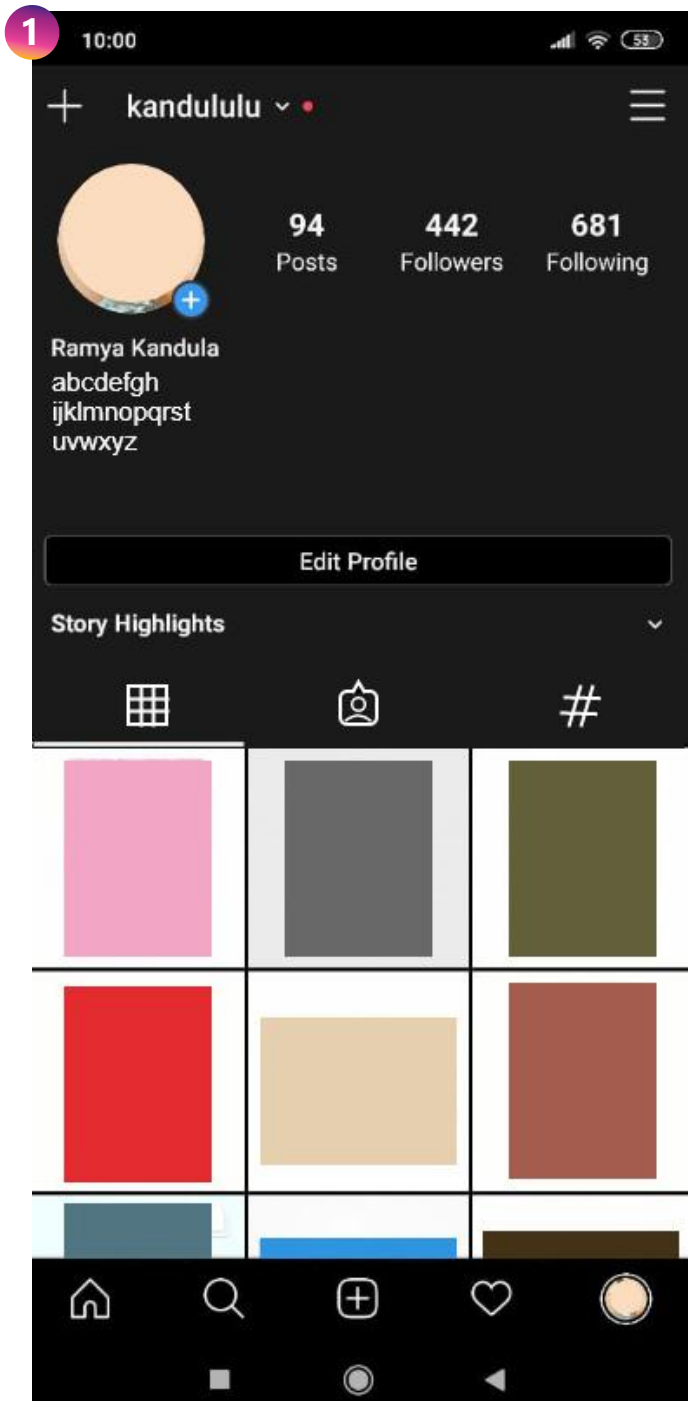
Option for users to choose the kind of posts they wish to see



## Feature 2

Keywords on sensitive content to give user context about the post





### Feature 3

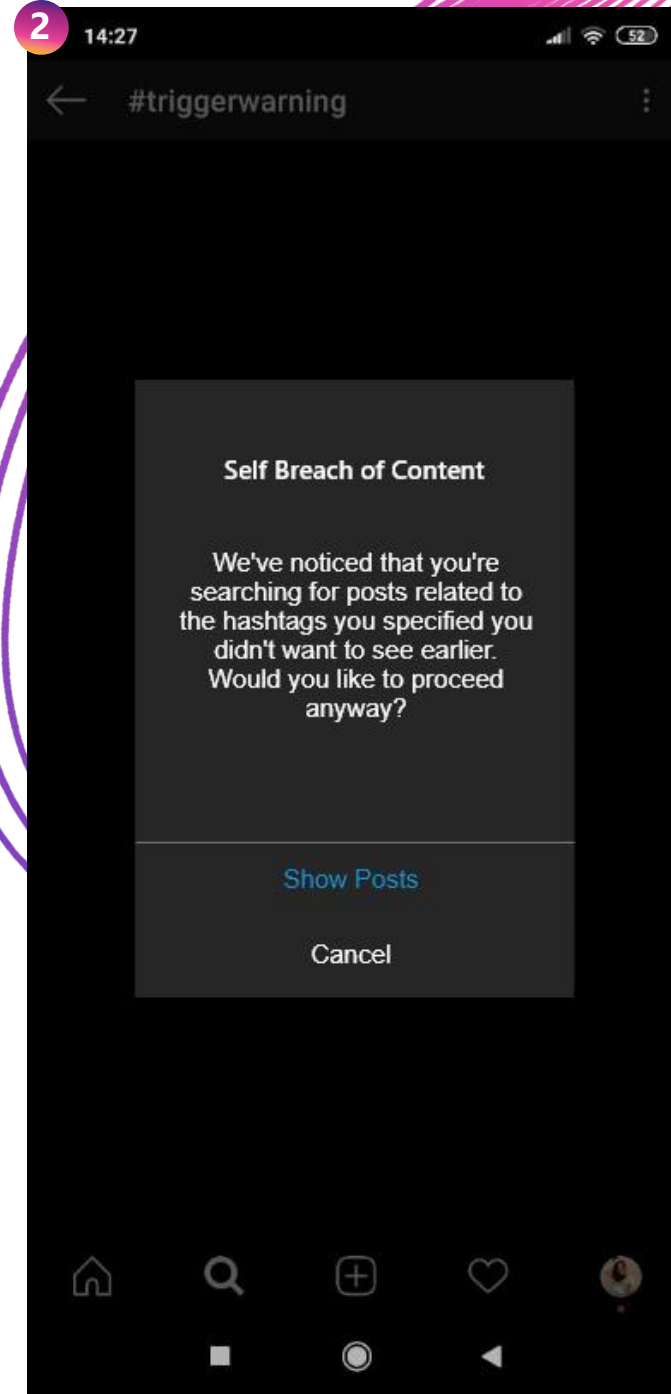
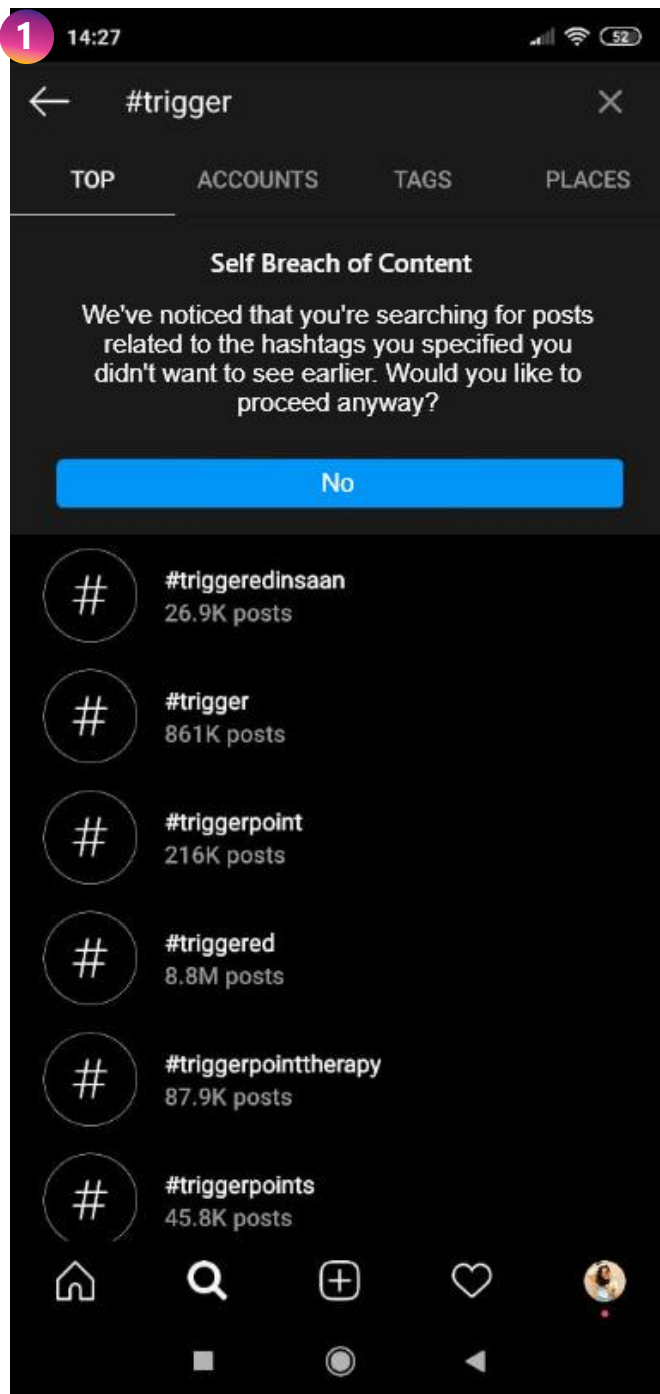
Customize content through hashtags





## Feature 3

Cautioning user  
when self breach  
of content occurs





# IDEA 3

Modifying the account creation process, which would make users take the policies and code of conduct seriously

#discourse

#self-surveillance

Instagram policies and code of conduct exist, but aren't taken seriously because of the loophole in the account creation process. Instagram provides complete freedom by allowing users to create multiple accounts irrespective of whether or not they have been blocked/reported before. When users understand that their behavior and actions on social media does not have any serious consequences, users tend to take it lightly. Hence verifying account history and linking accounts to a person's biometrics helps Instagram keep track of their users and debars them from creating multiple fake accounts. This can be applied in a situation where they have been blocked or reported in another alias or if they are under evaluation. The idea was based on the panopticon idea of self surveillance. When users know their actions have severe consequences such as them not being allowed to create another account, it imbibes fear and regulates discipline.

Instagram is just like any other social media platform, that provides freedom to its users to post whatever they feel like. Sometimes people might post offensive content and Instagram's AI detects these posts that are a violation of Instagram's policies and take necessary steps to make sure it doesn't happen again. Sometimes users end up seeing certain content, which does not necessarily violate Instagram's policies but might make them feel uncomfortable. Customization during account creation process is a new add on which allows users to choose but only from a given set of hashtags to avoid social reproduction. Posts on Instagram are categorized and labeled based on hashtags.

Selecting them under the customization section, Instagram would make sure to not show any posts related to those hashtags, thus, providing users a comfortable user experience.

Though Instagram has policies and code of conduct, most users are not aware of them, since the platform does not explicitly inform its users about it. Not knowing about the policies, terms and conditions, and the consequences of breaking them, makes the user question Instagram when they are reported for something. A clear explanation of the policies and code of conduct at the beginning (during the account creation process) is necessary. This idea was based on the discourse theory by Foucault.



1

2:14

VZWIR 78%

## Add Phone or Email

PHONE

EMAIL

IN +91

Phone

You may receive SMS updates from Instagram and can opt out at any time.

Next

1

2

3

-

4

5

6

⌋

7

8

9

⌫

,

0

.

→

v

2

## Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.



Place your left thumb



3

## Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.



Success



4

## Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.



Place your right thumb

5

## Input Biometrics

Biometrics data is necessary to maximise safety and to ensure users abide to the terms and policies of Instagram.

Press and make your finger cover the fingerprint identification area as much as possible, then lift after you feel a vibration.



Success

Next

6

## < Location/Region

Allowing Instagram to access your location, would allow us to customize your posts more relevant to your location.



Select your country

United States  
United Kingdom  
Australia  
**India**  
South Africa  
Spain  
Germany

I allow Instagram to access my device location.

Accept

Skip

7

## Personalization

Instagram cares a lot about user experience. If there are certain things that are triggering for you that you do not wish to see, pick from the hashtags



Posts on Instagram are categorized and labeled based on #tags. select #tags and we will make sure to hide posts associated with it.

#flashinglights

#Dickpics

#Abuse

#DomesticViolence

#Pornvideos

#Violence

#Terrorism

#Murder

#Rape

#AnimalCreulty

#Sex

Next

Skip

8

## < Personalization

Instagram cares a lot about user experience. If there are certain things that are triggering for you that you do not wish to see, pick from the hashtags



Posts on Instagram are categorized and labeled based on #tags. select #tags and we will make sure to hide posts associated with it.

#flashinglights

#Dickpics

#Abuse

#DomesticViolence

#Pornvideos

#Violence

#Terrorism

#Murder

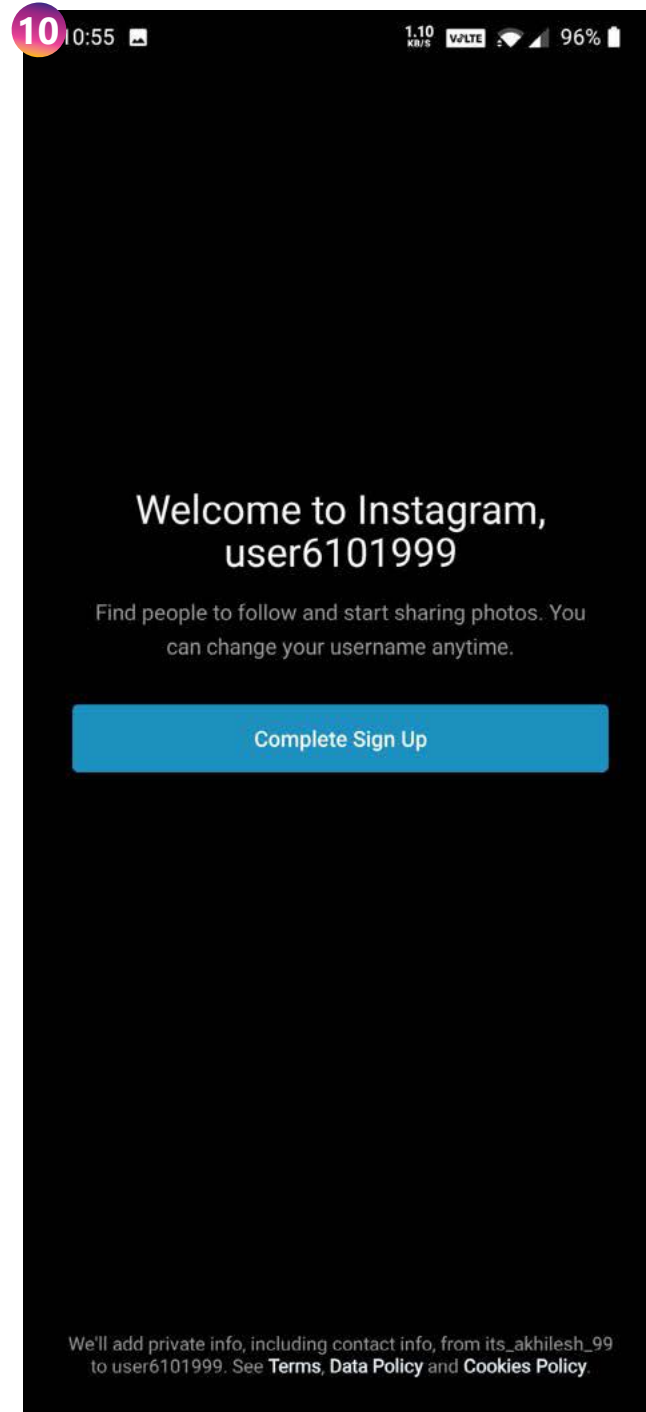
#Rape

#AnimalCreulty

#Sex

Next

Skip



# IDEA 4

Substituting reporting options with examples of abuse for the user to understand which category to choose

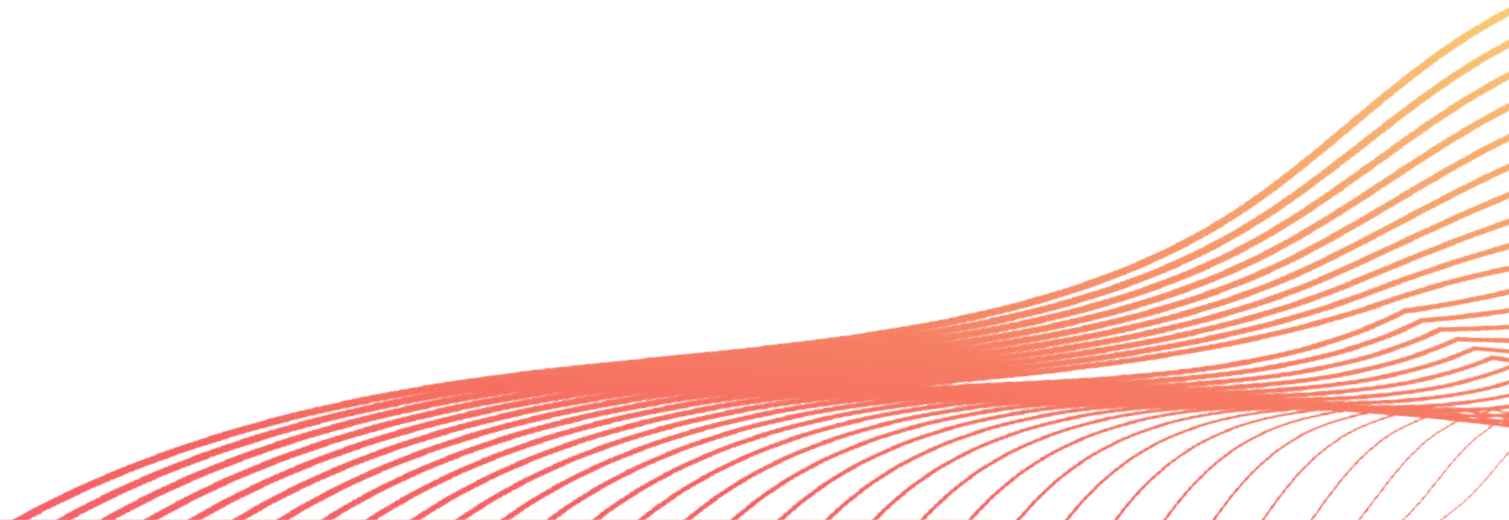
#socialcapital  
#culturalcapital

The purpose of this idea is to simplify and humanize the reporting process for the user by replacing overwhelming terminology with more accessible language as well as examples of abusive content to help the user categorize the post better. The current reporting options on Instagram display 'Spam' and 'Inappropriate' as two of the initial options which then take the user to a long list of vague topics to categorize the post or comment into. These options are more in relation to what goes against Instagram's policies or codes of conduct (such as nudity, misinformation, scam or fraud) rather than focusing on how the content affected the user's experience.

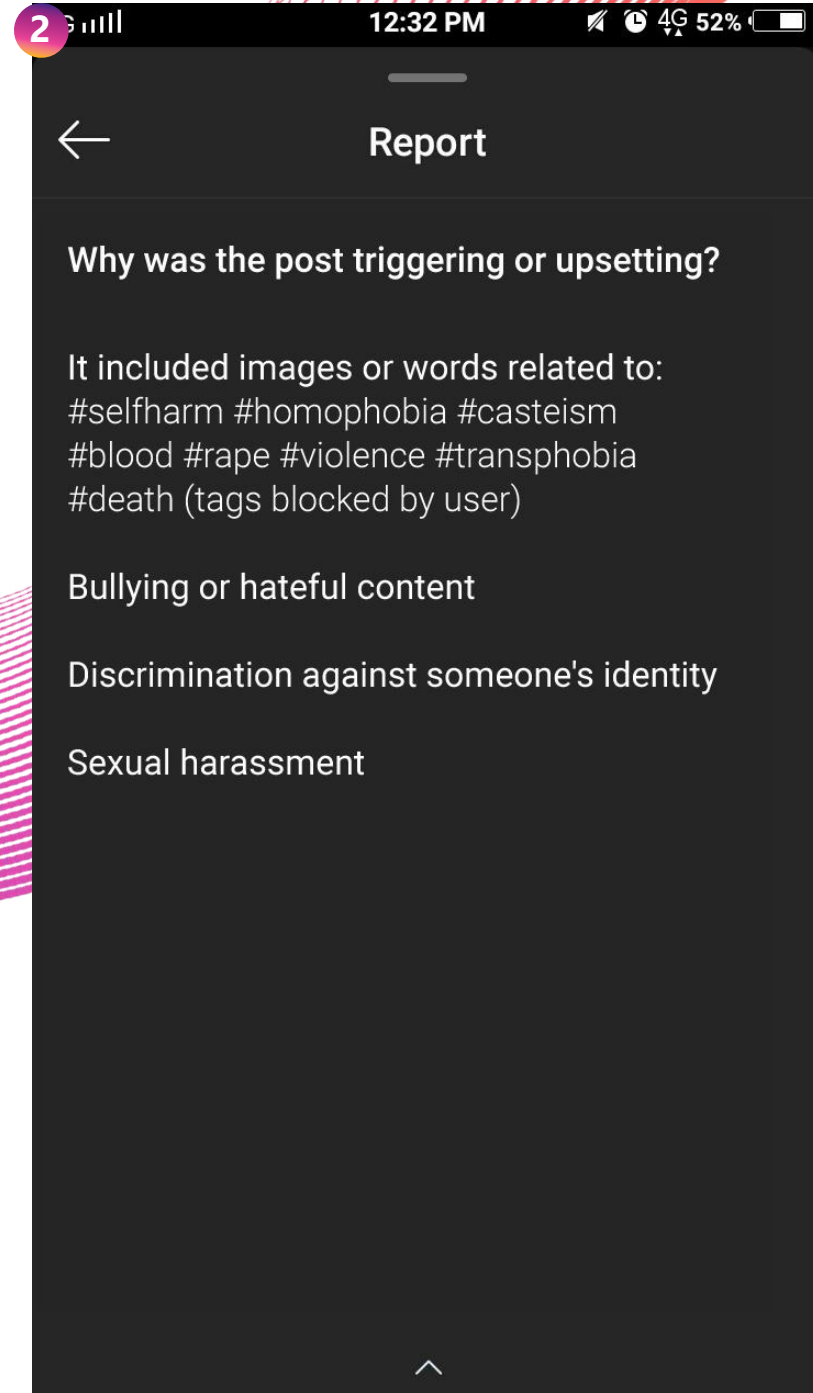
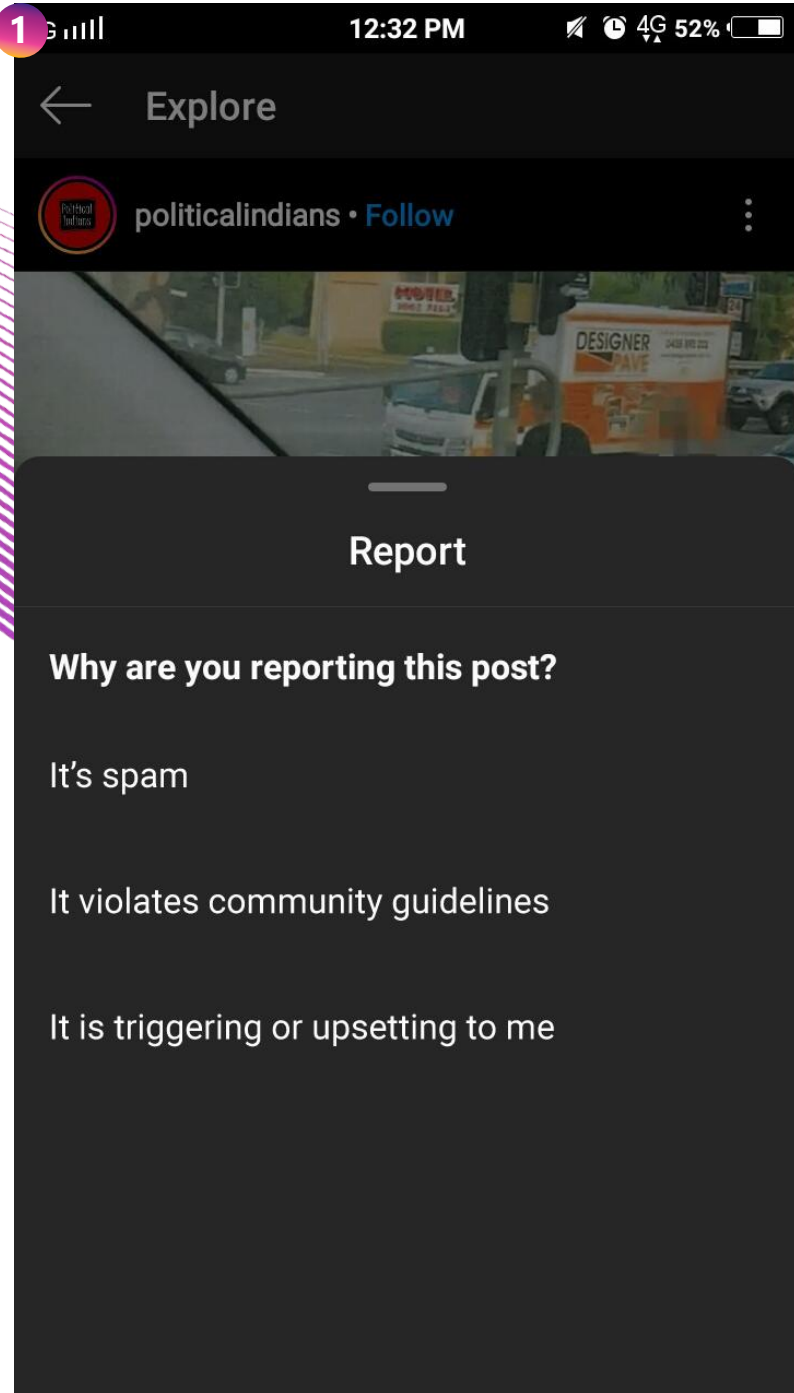
The idea makes use of Bourdieu's critical theory concepts of social and cultural capital by making the reporting options more understandable and relatable to the user's social capital. The list of types of abuse or

harassment with examples would take into account the 'blocked tags' from the user's account customization and the type of content that they have chosen not to see or have reported in the past. The options could also be region specific and display categories and examples related to the user's geographic location and local language.

At the end of the report, along with a message that notifies the user of their report being sent and an option to block the reported user, there would be a link to a feature that allows them to keep track of the progress of consequences on the accounts that they reported. By having more emphasis on emotion and making the procedure less robotic, there might be more usage of Instagram's reporting feature, diminishing the amount of offensive content and benefiting the platform through positive engagement.







3

12:34 PM

4G 52%

←

Report

The post discriminated against someone's:

Race or ethnicity

☐

Gender or sexual orientation

☒

Religion

☐

Caste

☐

Disability

☐

Proceed

4

12:34 PM

4G 52%

←

Report

The content of the post was:

Objectifying or vulgar

☐

Sexist or misogynistic

☒

Making jokes about sexual assault

☐

Threatening violence

☐

Proceed


5

12:34 PM4G 52%

←



Report


Was the post similar to any of the following?



So it's just a dude then ?



1d996 likesReply






You can't switch genders it's impossible



1d1149 likesReply






LGBT and Demokkrats 2020 🇺🇸🇺🇸🇬🇪🇬🇪🙏🙏



14h19 likesReply





If he can't even figure out what gender he is then how is he supposed to make political decisions

1d152 likesReply



Submit Report


6

12:34 PM4G 52%

←



Report


Was the post similar to any of the following?



if you don't cover up, you're asking for it so badly



1d26 likesReply






#nofatbitches2020



2hReply






She needs to cover up is what she needs to do 🤔🤔



12h44 likesReply



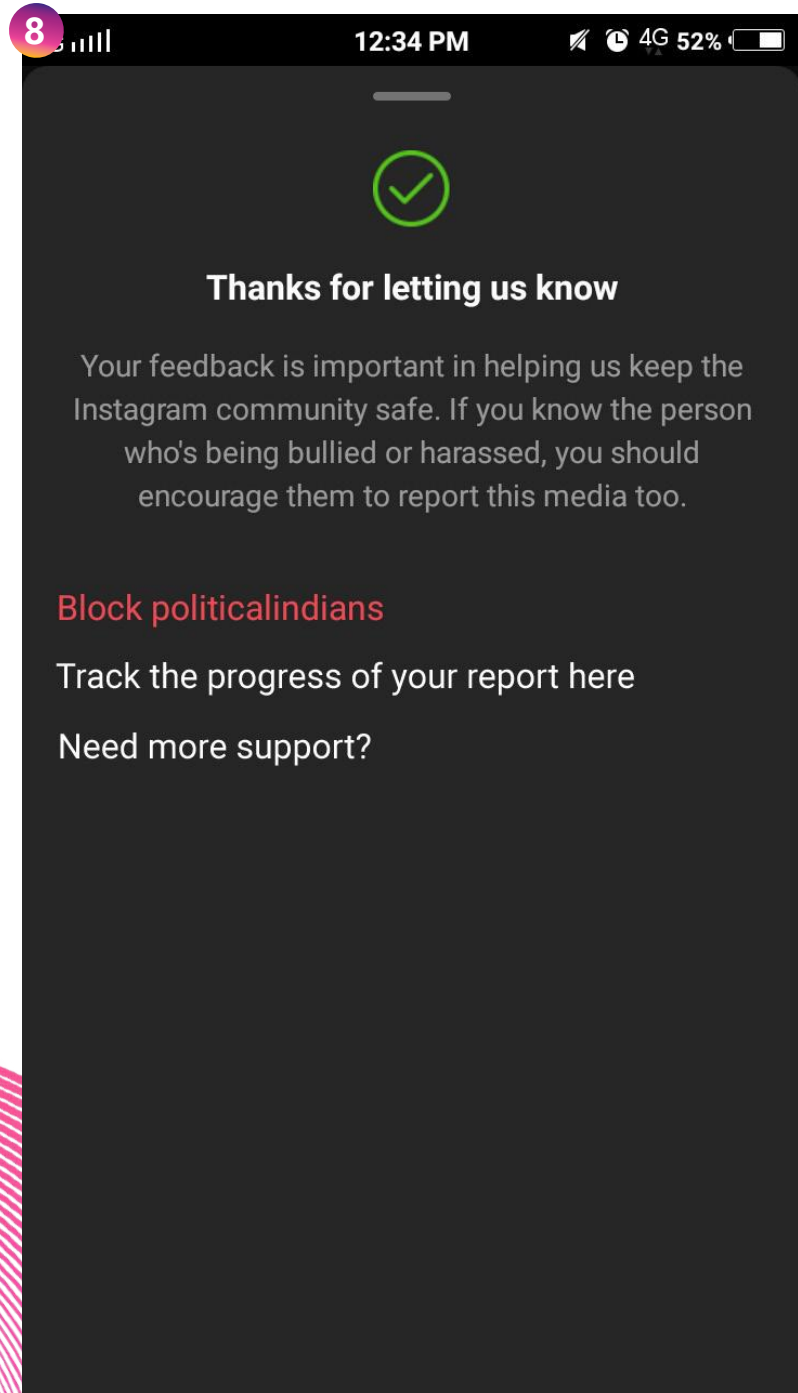
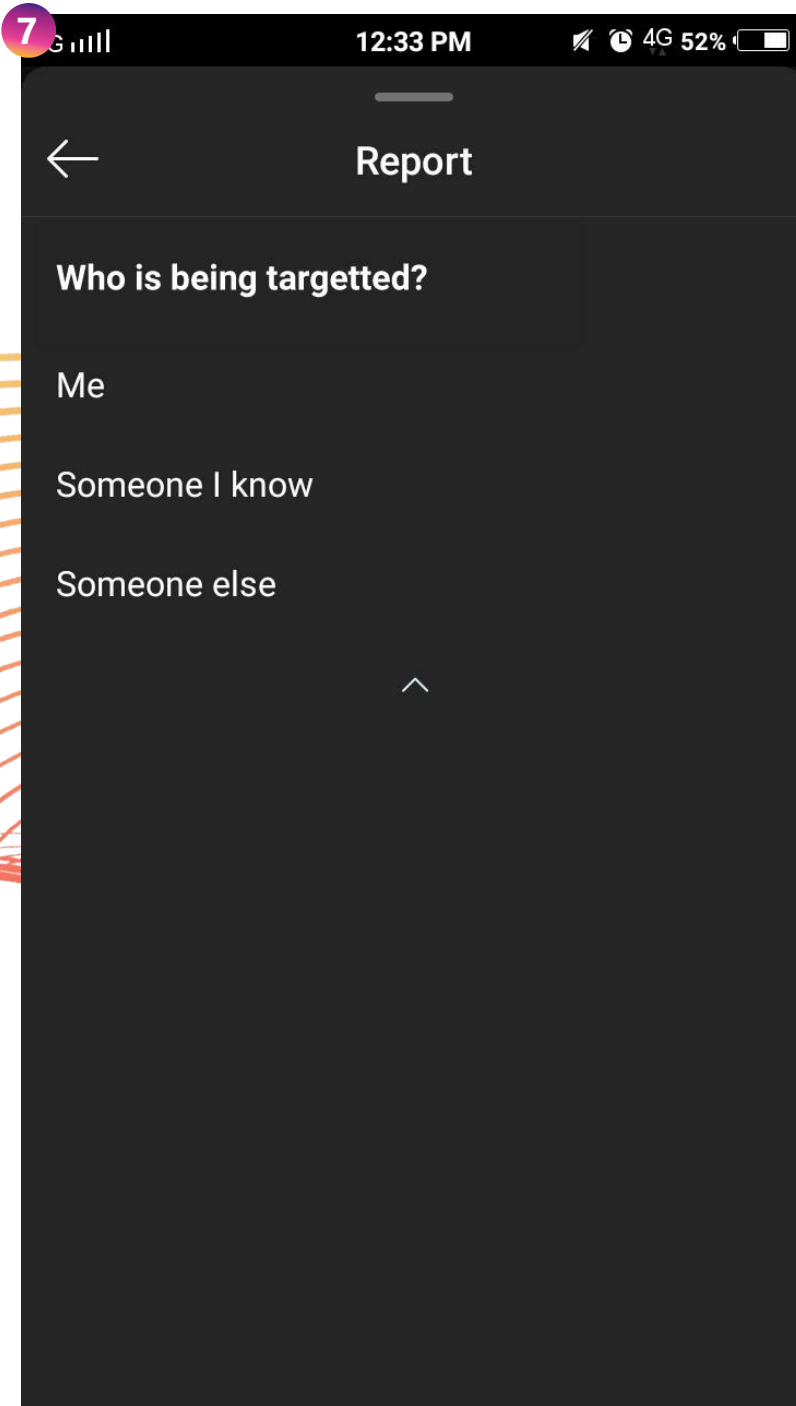


Killing babies isn't cool #FeminismIsCancer

3d644 likesReply



Submit Report





# IDEA 5

History of a user's abusive or offensive behavior can be transferred to other platforms or employers.

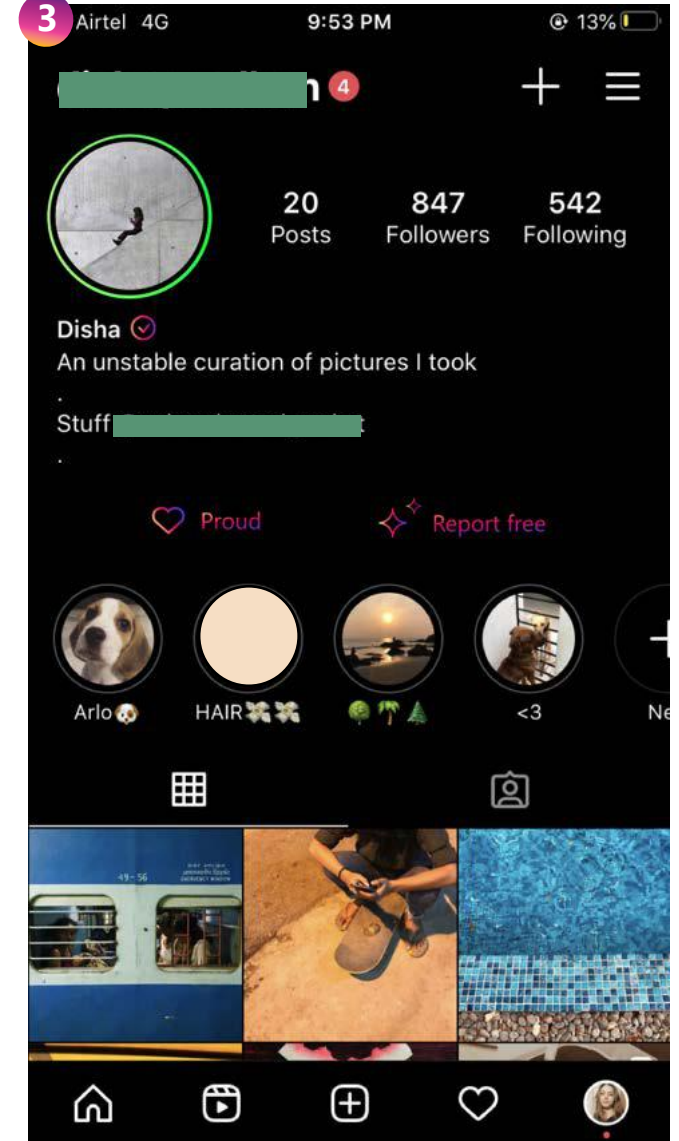
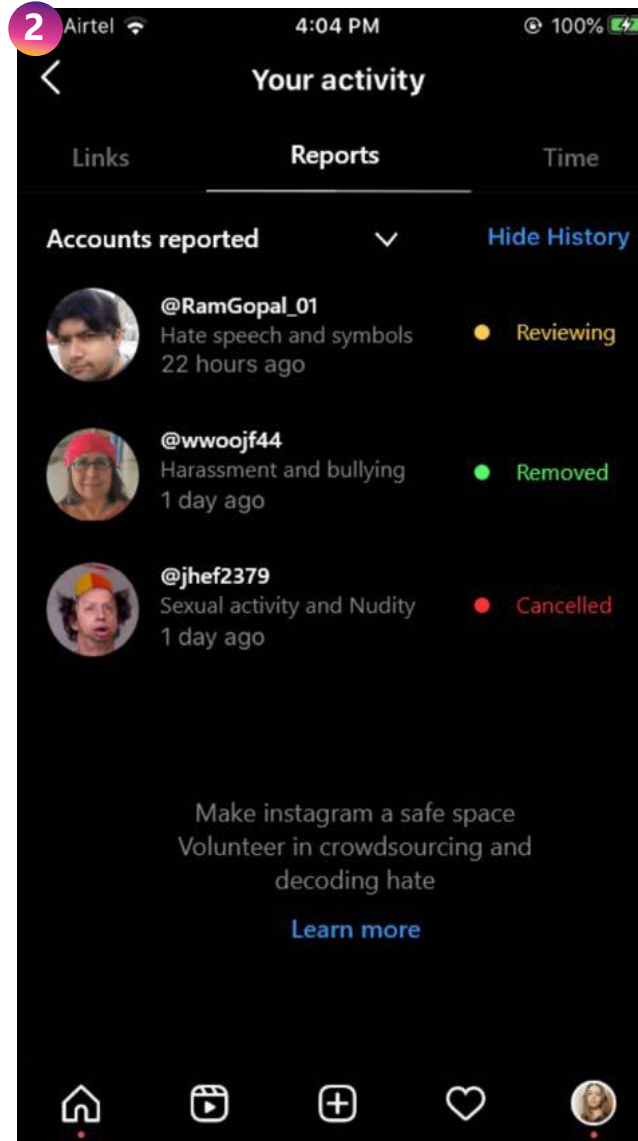
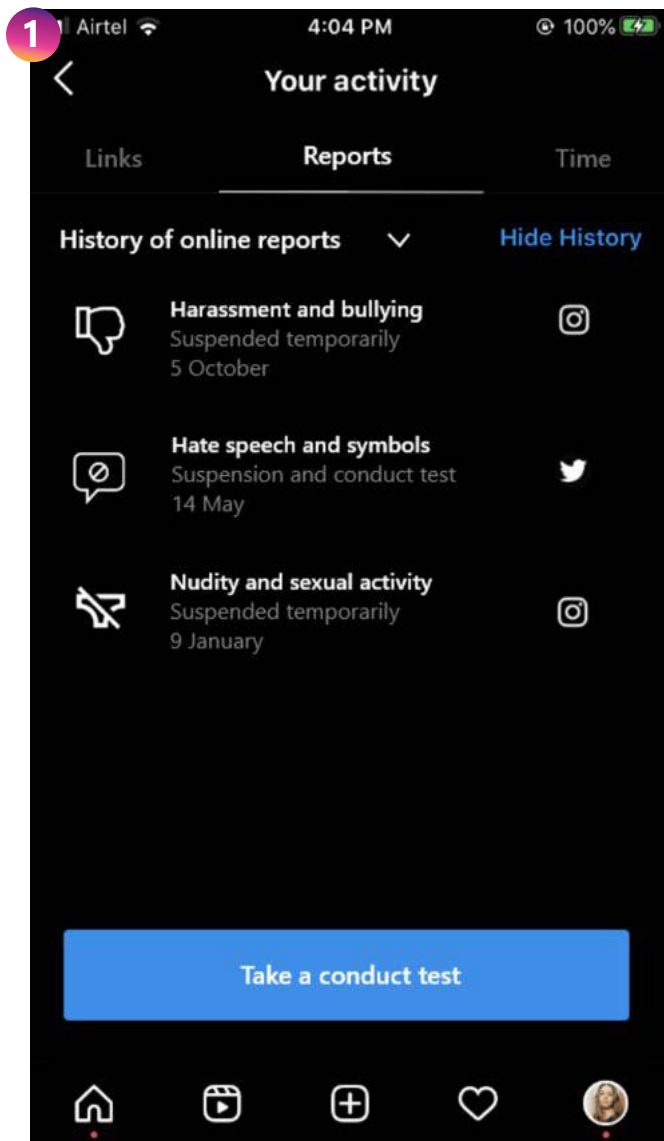
#capital  
#self-surveillance

This feature came into being, to create an archive of reports. People generally tend to forget about what happens to the account or content they report, in turn leading to an uncertain response to the same from Instagram. This allows the users to keep track of their reports and hold Instagram accountable in case the reported content doesn't check mark their guidelines, and have a constant scope for improvement. The other part of this archive is the history of the user's own reports. The times they have been reported on different social media platforms are collated in a singular space.

This space was created for an ethical background check for employees of a workspace. While this creates a set of real world consequences to their online behavior, it also tends to be a reality check. This archive is a space for self surveillance of the user's behavior online, to make them more conscious about their words and actions. The internet's culture is free expression and making users conscious about their actions contradicts it. But making the users conscious about the right things is important. This imbibes a positive change in the user's cultural capital that follows self surveillance. Rewarding positive changes and conscious behavior online is equally important. Which is why curated highlights on a user's profile is a small but meaningful addition - like being "report free".

Real life crimes get imprinted on our papers, become a part of us for life and offer no way to redeem the action. Such need not be the case for online reports. While the upper hand here is the platform and there is no involvement of the government, there is much scope for doing

things differently. The users could go through the sexual and verbal conduct test to redeem the label of being reported, and gives a chance to the user to do better the next time. Discipline and power structures on social media platforms, are created by the company themselves based on the priority balance between the stakeholders like advertisers and users. The company can create a joint database with other platforms as well, create guidelines for reporting content and have the power to remove hate online, while also improving their Ai in recognizing hate from all social media platforms.



-----XXX-----